REVIEWS, REFINEMENTS AND NEW IDEAS IN FACE RECOGNITION

Edited by **Peter M. Corcoran**

INTECHWEB.ORG

Reviews, Refinements and New Ideas in Face Recognition
Edited by Peter M. Corcoran

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Statements and opinions expressed in the chapters are these of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published articles. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

Publishing Process Manager Mirna Cvijic
Technical Editor Teodora Smiljanic
Cover Designer Jan Hyrat
Image Copyright hfng, 2010. Used under license from Shutterstock.com

First published July, 2011
Printed in Croatia

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Reviews, Refinements and New Ideas in Face Recognition, Edited by Peter M. Corcoran
p. cm.
ISBN 978-953-307-368-2

Contents

Preface

As a baby one of our earliest stimuli is that of human faces. We rapidly learn to identify, characterize and eventually distinguish those who are near and dear to us. This skill stays with us throughout our lives.

As humans, face recognition is an ability we accept as commonplace. It is only when we attempt to duplicate this skill in a computing system that we begin to realize the complexity of the underlying problem. Understandably, there are a multitude of differing approaches to solving this complex problem. And while much progress has been made many challenges remain.

This book is arranged around a number of clustered themes covering different aspects of face recognition. The first section on Statistical Face Models and Classifiers presents some reviews and refinements of well-known statistical models. The second section presents two articles exploring the use of Infrared imaging techniques to refine and even replace conventional imaging. After this follows the section with a few articles devoted to refinements of classical methods. Articles that examine new approaches to improve the robustness of several face analysis techniques are followed by two articles dealing with the challenges of real-time analysis for facial recognition in video sequences. A final article explores human perceptual issues of face recognition.

I hope that you find these articles interesting, and that you learn from them and perhaps even adopts some of these methods for use in your own research activities.

Sincerely,

Peter M. Corcoran
Vice-Dean,
College of Engineering & Informatics,
National University of Ireland Galway (NUIG),
Galway, Ireland

Part 1

Statistical Face Models & Classifiers

1

A Review of Hidden Markov Models in Face Recognition

Claudia Iancu and Peter M. Corcoran
College of Engineering & Informatics
National University of Ireland Galway
Ireland

1. Introduction

Hidden Markov Models (HMMs) are a set of statistical models used to characterize the statistical properties of a signal. An HMM is a doubly stochastic process with an underlying stochastic process that is not observable, but can be observed through another set of stochastic processes that produce a sequence of observed symbols. An HMM has a finite set of states, each of which is associated with a multidimensional probability distribution; transitions between these states are governed by a set of probabilities. Hidden Markov Models are especially known for their application in 1D pattern recognition such as speech recognition, musical score analysis, and sequencing problems in bioinformatics. More recently they have been applied to more complex 2D problems and this review focuses on their use in the field of *automatic face recognition*, tracking the evolution of the use of HMMs from the early-1990's to the present day.

Our goal is to enable the interested reader to quickly review and understand the state-of-art for HMM models applied to face recognition problems and to adopt and apply these techniques in their own work.

2. Historical overview and Introduction to HMM

The underlying mathematical theory of Hidden Markov Models (HMMs) was originally described in a series of papers during the 1960's and early 1970's [Baum & Petrie, 1966; Baum et al., 1970; Baum, 1972]. This technique was subsequently applied in practical pattern recognition applications, more specifically in speech recognition problems [Jelinek et al., 1975]. However, widespread understanding and practical application of HMMs only began a decade later, in the mid-1980s. At this time several tutorials were written [Levinson et al., 1983; Juang, 1984; Rabiner & Juang, 1986; Rabiner, 1989]. The most comprehensive of these was the last, [Rabiner, 1989], and provided sufficient detail for researchers to apply HMMs to solve a broad range of practical problems in speech processing and recognition. The broad adoption of HMMs in automatic speech recognition represented a significant milestone in continuous speech recognition problems [Juang & Rabiner, 2005].

The mathematical sophistication of HMMs combined with their successful application to a wide range of speech processing problems has prompted researchers in pattern recognition to consider their use in other areas, such as character recognition, keyword spotting, lipreading, gesture and action recognition, bioinformatics and genomics. In this chapter we present a review of the most important variants of HMMs found in the *automatic face recognition literature*. We begin by presenting the initial 1D HMM structures adapted for use in face recognition problems in section 3. Then a number of papers on hybrid approaches used to improve the performance of HMMs for face recognition are discussed in section 4. In section 5 the various 2D variants of HMM are described and evaluated in terms of the recognition rates achieved from each. Finally section 6 includes some recent refinements in the application of HMM techniques to face recognition problems.

3. HMM in face recognition - initial 1D HMM structures

As mentioned in the previous section, HMMs have been used extensively in speech processing, where signal data is naturally one-dimensional. Nevertheless, HMM techniques remain mathematically complex even in the one-dimensional form. The extension of HMM to two-dimensional model structures is exponentially more complex [Park & Lee, 1998]. This consideration has led to a much later adoption of HMM in applications involving two-dimensional pattern processing in general and face recognition in particular.

3.1 Initial research on ergodic and top-to-bottom 1D HMM

In 1993, a new approach to the problem of automatic face recognition based on 1D HMMs was proposed by [Samaria & Fallside, 1993]. In this paper faces are treated as two-dimensional objects and the HMM model automatically extracts statistical facial features. For the automatic extraction of features, a 1D observation sequence is obtained from each face image by sampling it using a sliding window. Each element of the observation sequence is a vector of pixel intensities (or greyscale levels).

Two simple 1D HMMs were trained by these authors in order to test the applicability of HMMs in face recognition problems. A test database was used comprising images of 20 individuals with a minimum of 10 images per person. Images were acquired under homogeneous lighting against a constant background, and with very small changes in head pose and facial expressions. For a first set of tests an ergodic HMM was used. The images were sampled using a rectangular window, size 64 × 64, moving left-to-right horizontally with a 25% overlap (16 pixels), then vertically with 16 pixels overlap and starting again horizontally right-to-left. Using the observation sequence thus extracted, an 8-state ergodic HMM was built to approximately match the 8 distinct regions that seem to appear in the face image (eyes, mouth, forehead, hair, background, shoulders and two extra states for boundary regions). Figure 1 taken from [Samaria & Fallside, 1993] shows the training data used for one subject and the mean vectors for the 8 states found by HMM for that particular subject.

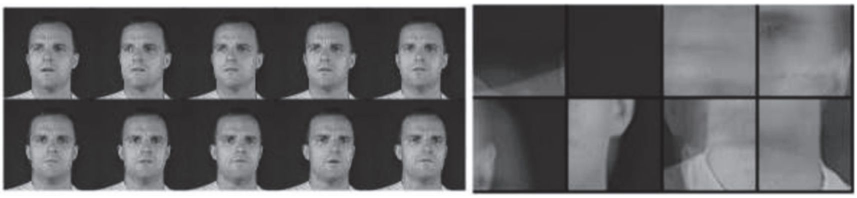


Fig. 1. Training data and states for ergodic HMM [Samaria & Fallside, 1993]

In the second set of tests, a left-to-right (top-to-bottom) HMM was used. Each image was sampled using a horizontal stripe 16 pixels high and as wide as the image, moving top-to-bottom with 12 lines overlap. The resulting observation sequence was used to train a 5-state left-to-right HMM where only transitions between adjacent states are allowed. The training images and the mean vectors for the 5 states found by HMM are presented in Figure 2.

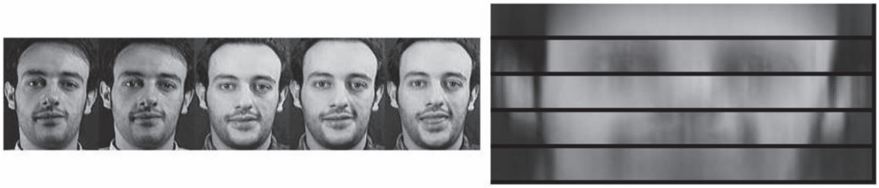


Fig. 2. Examples of training data and states for top-to-bottom HMM from [Samaria & Fallside, 1993]

In both of these models the statistical determination of model features, yields some states of the HMM which can be directly identified with physical facial features. Training and testing were performed using the HTK toolkit[1]. According to these authors, successful recognition results were obtained when test images were extracted from the same video sequence as the training images, proving that the proposed approach can cope with variations in facial features due to small orientation changes, provided the lighting and background are constant. Unfortunately these authors did not provide any explicit recognition rates so it is not possible to compare their methods with later research. It is reasonable, however, to surmise that their experimental results were marginal and are improved upon by the later refinements of [Samaria & Harter, 1994].

3.2 Refinement of the top-to-bottom 1D HMM

In a later paper [Samaria & Harter, 1994] refined the work begun in [Samaria & Fallside, 1993] on a top-to-bottom HMM. These new experiments demonstrate how face recognition rates using a top-to-bottom HMM vary with different model parameters. They also indicate the most sensible choice of parameters for this class of HMM. Up until this point, the parameterization of the model had been based on subjective intuition.

For such a 1D top-to-bottom HMM there are three main parameters that affect the performance of the model: the height of the horizontal strip used to extract the observation sequence, L (in pixels), the overlap used, M (in pixels) and the number of states N of the HMM. The height of the strip, L, determines the size of the features and the length of the observation sequence, thus influencing the number of states. The overlap, M, determines how likely feature alignment is and also the length of the observation sequence. A model with no overlap would imply rigid partitioning of the faces with the risk of cutting across potentially discriminating features. The number of states, N, determines the number of features used to characterize the face, and also the computational complexity of the system.

These experiments were performed using the Olivetti Research Lab (ORL) database, containing frontal facial images with limited side movements and head tilt. The database was comprised of 40 subjects with 10 pictures per subject. The experiments used 5 images

[1] http://htk.eng.cam.ac.uk/

per person for training and the remaining 5 images for testing. The results were reported as error rates, calculated as the proportion of incorrectly classified images. Three sets of tests were done, varying the values of each of the three parameters as follows: 2 ≤ N ≤ 10, 1 ≤ L ≤ 10 and 0 ≤ M ≤ L−1. For M varied, the number of states was fixed at N = 5 and window height L was varied between 2 and 10. According to the tests, the error rates drop as the overlap increases, approximately from 28% to 15%. However a greater overlap implies a bigger computational effort. When L was varied, N was fixed to 5 and the overlaps considered were 0, 1 and L-1. In this case if there is little or no overlap, the smaller the strip height the lower the error rate is, with values between 13% for L = 1 up to 28% for L = 10. However, for sufficiently large overlap the strip height has marginal effect on the recognition performance, the error rate remaining almost constant around 14%. In the third set of tests N was varied, with L = 1 and 0 overlap and L = 8 and maximum overlap (M=L-1). The performance is fairly uniform for values of N between 4 and 10, with an increase in error for values smaller than three.

The conclusions of this paper are: (i) a large overlap in the sampling phase (the extraction of observation sequences) yields better recognition rates; the error rate varies from up to 30% for minimum overlap down to 15% for maximum overlap; (ii) for large overlaps the height of the sampling strip has limited effect. The error rate remains almost constant at 15% for maximum overlap, regardless of the value of L, and (iii) best results are obtained with a HMM with 4 or more states. Error rate drops from around 25% for 1-2 states to 15% from 4 states onward. We remark that these early models were relatively unsophisticated and were limited to fully frontal faces with images taken under controlled background and illuminations conditions.

3.3 1D HMM with 2D-DCT features for face recognition

In [Nefian & Hayes, May 1998], Samaria's version of 1D HMM, is upgraded using 2D-DCT feature vectors instead of pixel intensities. The face image is divided into 5 significant regions, *viz*: hair, forehead, eyes, nose, and mouth. These regions appear in a natural order, each region being assigned to a state in a top-to-bottom 1D continuous HMM. The state structure of the face model and the non-zero transition probabilities are shown in Figure 3.

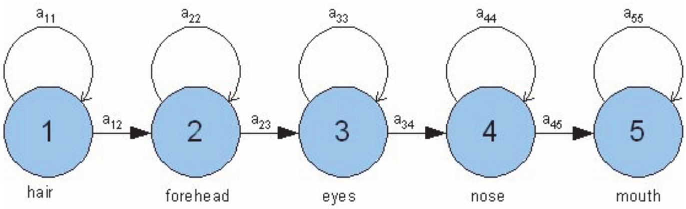


Fig. 3. Sequential HMM for face recognition

The feature vectors were extracted using the same technique as in [Samaria & Harter, 1994]. Each face image of height H and width W is divided into overlapping strips of height L and width W, the amount of overlap between consecutive strips being P, see Figure 4. The number of strips extracted from each face image determines the number of observation vectors.

The 2D-DCT transform is applied on each face strip and the observation vectors are determined, comprising the first 39 2D-DCT coefficients. The system is tested on ORL[2] database containing 400 images of 40 individuals, 10 images per individual, image size 92 × 112, with small variations in facial expressions, pose, hair style and eye wear. Half of the database is used for training and the other half is used for testing. The recognition rate achieved for L=10 and P=9 is 84%. Results are compared with recognition rates obtained using other face recognition methods on the same database: recognition rate for the eigenfaces method is 73%, and for the 1D HMM used by Samaria is also 84%, but the processing time for DCT based HMM is an order of magnitude faster - 2.5 seconds in contrast to 25 seconds required by the pixel intensity method of [Samaria & Harter, 1994].

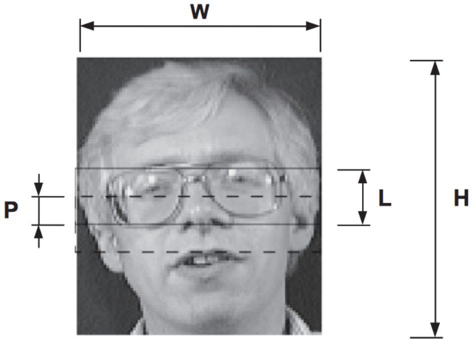


Fig. 4. Face image parameterization and blocks extraction [Nefian & Hayes, May 1998].

3.4 1D HMM with KLT features for face detection and recognition

In a second paper [Nefian & Hayes, October 1998] introduce an alternative 1D HMM approach, which performs the face detection function in addition to that of face recognition. This employs the same topology and structure as in the previous work of these authors, described above, but uses different image features. In contrast with the previous paper, the observation vectors used here are the coefficients of Karhunen-Loeve Transform. The KLT compression properties as well as its decorrelation properties make it an attractive technique for the extraction of the observation vectors. Block extraction from the image is achieved in the same way as in the previous paper. The eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the extracted vectors form the KLT basis set. If μ is the mean of the vectors used to compute the covariance matrix, a set of vectors is obtained by subtracting this mean from each of the vectors corresponding to a block in the image. The resulting set of vectors is then projected onto the eigenvectors of the covariance matrix and the resulting coefficients form the observation vectors.

The system is used both for face detection and recognition by the authors. For face detection, the system is first trained with a set of frontal faces of different people taken under different illumination conditions, in order to build a face model. Then, given a test image, face detection begins by scanning the image with horizontally and vertically overlapping rectangular windows, extracting the observation vectors and computing the probability of

[2] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

data inside each window given the face model, using Viterbi algorithm. The windows that have face model likelihood higher than a threshold are selected as possible face locations. The face detection system was tested on MIT database with 48 images of 16 people with background and with different illuminations and head orientations. Manually segmented faces from 9 images were used for training and the remaining images for testing, with a face detection rate of 90%.

For face recognition this system was applied to the ORL database containing 400 images of 40 individuals, 10 images per individual, at a resolution of 92 × 112 pixels, with small variations in facial expressions, pose, hairstyle and eye wear. The system was trained with half of the database and tested with the other half. The accuracy of the system presented in this paper is increased slightly over earlier work to 86% while the recognition time decreases due to use of the KLT features.

3.5 Refinements to 1D HMM with 2D-DCT features

Following on the work of [Samaria, 1994] and [Nefian, 1999], Kohir & Desai wrote a series of three papers using the 1D HMM for face recognition problems. In a first paper, [Kohir & Desai, 1998], these authors present a face recognition system based on 1D HMM coupled with 2D-DCT coefficients using a different approach for feature extraction than that employed by [Nefian & Hayes, May 1998 & October 1998]. The extracted features are obtained by sliding square windows in a raster scan fashion over the face image, from left to right and with a predefined overlap. At every position of the window over the image (called sub-image) 2D DCT are computed, and only the first few DCT coefficients are retained by scanning the sub-image in a zigzag fashion. The zigzag scanned DCT coefficients form an observation vector. The sliding procedure and the zigzag scanning are illustrated in Figure 5 [Kohir & Desai, 1998].

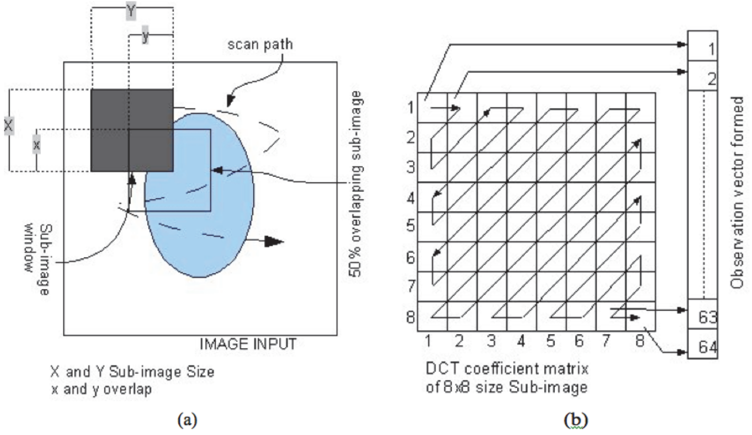


Fig. 5. (a) Raster scan of face image with sliding window. (b) Construction of 1D observation vector from zigzag scanning of the sliding window [Kohir & Desai, 1998].

The performance of this system is tested using the ORL database. Half of the images were used in the training phase and the other half for testing (5 faces for training and the remaining 5 for testing), sampling windows of 8 × 8 and 16 × 16, were used with 50% and

75% overlaps, and 10, 15 and 21 DCT coefficients were extracted. The number of states in the HMM was fixed at 5 as per the earlier work of [Nefian & Hayes, May 1998]. The recognition rates vary from 74.5% for a 16 × 16 window, with a 50% overlap and 21 DCT coefficients to 99.5% for 16 × 16 window, 75% overlap and 10 DCT coefficients.

In a second paper [Kohir & Desai, 1999] these authors further refined their research contribution. To evaluate the recognition performances of the system, 2 new experiments are performed:

- In a first experiment the proposed method is tested with different numbers of training and testing faces per subject. The tests were performed on the ORL database, and the number of training faces was increased from 1 to 6, while the remaining faces were used in the testing phase. A sampling window of 16 × 16 with 75% overlap was used with 10 DCT coefficients as these had provided optimal recognition rates in their earlier work. The recognition rates achieved are from 78.33% for a single training image and 9 testing images up to 99.5% which is the rate obtained when 5 or 6 training images and 5 or 4 testing images are used. It is worth noting that the ORL database comprises frontal face images in uniform lighting conditions and that recognition rates close to 100% are often achieved when using such datasets.
- In a second experiment the system was tested while increasing the number of states in the HMM. Again the ORL database is used, with 5 images for training and 5 for testing. The recognition rates vary as follows: 92% for a 2-states HMM, increasing to 99.5% for a 5-states HMM and stabilizing around 97%-98% when using up to 17 states. The system was also tested with the SPANN database[3] containing 249 persons, each with 7 pictures, with variations in pose, 3 pictures were used for training and the remaining 4 for testing, and the recognition rate achieved was 98.75%.
- A third paper, [Kohir & Desai, 2000] describes the same 1D HMM with DCT features, with a variation in the training phase. In this paper, first a *mean image* is constructed from all the training images, and then each training image is subtracted from the *mean image* to obtain a *mean subtracted image*. The observation vectors are extracted from these *mean subtracted images* using the same window sliding method. The observation vector sequences are then clustered using the K-means technique, and thus an initial state segmentation is obtained. Subsequently, the conventional training steps are followed. In the recognition phase, each test image is first subtracted from the *mean image* obtained during the training phase and recognition is performed on the resulting *mean subtracted image*.

The experiments for face recognition were performed on the same two databases, ORL and SPANN. For ORL database 5 pictures were used for training and the remaining 5 for testing, and the recognition rate obtained is 100%, compared to 88% when the eigenfaces method is used. For SPANN database 3 pictures were used for training and the remaining 4 for testing, the obtained recognition rate was 90%, compared again with the eigenfaces method where a 77% recognition rate was achieved. For the ORL database different resolutions were also tested, the highest recognition rate, 100% being obtained for 96 × 112.

Also, 'new subject rejection' for authentication applications was tested on the ORL database. The database was segmented into 2 sets: 20 subjects corresponding to an 'authorized' subject class - 5 pictures used in training phase and the rest in the testing phase. The

[3] http://www.khayal.ee.iitb.ernet.in/usr/SPANN_DATA_BASE/2D_Signals/Face/faces

remaining 20 subjects are assigned to an ‘unauthorized’ class - all 10 pictures are used in the testing phase. For each ‘authorized’ subject a HMM model is built. Also a separate ‘common HMM’ model is built using all *mean subtracted training images* of all the ‘authorized’ subjects. For each test face, if the probability of the ‘common HMM’ is the highest, the input face image is rejected as ′unauthorized′, otherwise the input face image is treated as ′authorized′. The results are: 100% rejection of any new subjects and 17% rejection of known subjects (false negatives).

3.6 Refinement of 1D HMM with sequential prunning

As proved by [Samaria & Harter, 1994], the number of states used in a 1D HMM can have a strong influence on recognition rates. The problem of the optimal selection of the structure for an HMM is considered in [Bicego at al., 2003a]. The first part of this paper presents a method of improving the determination of the optimal number of states for an HMM. These authors then proceed to prove the equivalence between (i) a 1D HMM whose observation vectors are modelled with *multiple Gaussians per state* and (ii) a 1D HMM with *one Gaussian per state* but employing a larger number of states. According to the authors, there are several possible methods for solving the first problem, e.g. cross-validation, Bayesian inference criterion (BIC), minimum description length (MDL). These are based on training models with different structures and then choosing the one that optimizes a certain selection criterion. However, these methods involve a considerable computational burden plus they are sensitive to the local-greedy behaviour of the HMM training algorithm, i.e. the successful training of the model is influenced by the initial estimates selected.

The approach proposed by [Bicego et al., 2003a] addresses both the computational burden of model selection, and the initialization phase. The key idea is the use of a decreasing learning strategy, starting each training session from a ‘nearly good’ situation derived from the previous training session by pruning the ‘least probable’ state. More specifically, the authors proposed starting the model training with a large number of states. They next run the estimation algorithm and, on convergence, evaluate the model selection criterion. The ′least probable′ state is then pruned, and the resulting configuration of the model with one less state is used as a starting point for the next sequence of iterations. In this way, each training session is started from a ′nearly good′ estimate. The key observation supporting this approach is that, when the number of states is extremely large, the dependency of the model behaviour on the initial estimates is much weaker. An additional benefit is that using ′nearly good′ initializations drastically reduces the number of iterations required by the learning algorithm at each step in this process. Thus the number of model states can be rapidly reduced at low computational cost.

In order to assess the performance of their proposed method, these authors tested the pruning approach and the standard approach (consisting in training one HMM for varying number of states) with BIC criterion and MMDL (mixture minimum description length) [Figueiredo et al., 1999] criterion. These two strategies are compared in terms of: (i) accuracy of the model size estimation, (ii) total computational cost involved in the training phase, and (iii) classification accuracy. In all the HMMs considered in this paper the emission probability density for each state is a single Gaussian. For the accuracy of the model size estimation, synthetically generated test sets of 3 known HMMs were used. The authors set the number of states allowed from 2 to 10. The selection accuracy ranged from 54% to 100% for standard BIC and MMDL, and from 98% to 100% for pruning BIC and MMDL, with up to 50% less iteration required for the latter.

Classification accuracy was tested on both synthetic and real data. For the synthetic data, the test sets used previously to estimate the accuracy of the model size estimation were used, obtaining 92% to 100% accuracy for standard BIC and MMDL compared to 98% to 100% accuracy for pruning BIC and MMDL, with 35% less iterations for pruning. For classification accuracy on real data, two experiments were conducted. The first involves a 2D shape recognition problem, and uses a data set with four classes each with 12 different shapes. The results obtained are 92.5% for standard BIC, 94.37% for standard MMDL, and 95.21% for pruning BIC and MMDL. The second experiment was conducted on the ORL database, using the method proposed by [Kohir & Desai 1998]. The results are 97.5% for standard BIC and MMDL and 97.63% for pruning BIC and MMDL. The classification accuracies are similar, but the pruning method reduces substantially the number of iterations required.

3.7 A 1D HMM with 2D-DCT features and Haar wavelets

In a following paper [Bicego et al., 2003b], a comparison between DCT coding and wavelet coding is undertaken. The aim is to evaluate the effectiveness of HMMs in modelling faces using these two different forms of image features. Each compresses the relevant image data, but employing different underlying techniques. Also, the suitability of HMM to deal with the JPEG 2000 image compression standard is considered by these authors. They adopt the 1D HMM approach introduced by [Kohir & Desai, 1998]. However, the optimum number of states for the model is selected using the sequential pruning strategy presented in [Bicego et al., 2003a] and described in the preceding section. The same feature extraction used by [Kohir & Desai, 1998] is employed, and both 2D DCT and Haar wavelet coefficients are computed.

These experiments have been conducted on the ORL database, consisting of 40 subjects with 10 sample images of each. The first 5 images are used for training the HMM while the remaining 5 are used in the testing phase. The number of states for each HMM is estimated using the pruning strategy. For feature extraction, a 16 × 16 pixel sliding window is used, with 50% and 75% overlaps being tested, and in each case the first 4, 8 and 12 DCT or Haar coefficients are retained. The recognition rate scores for 50% overlap are between 97.4% for 4 coefficients to 100% for 12 coefficients, and for 75% overlap between 95.4% for 4 coefficients to 99.6% for 12 coefficients. Slightly better results were obtained for DCT coefficients throughout the experiments. It is worth noting that unlike [Samaria & Harter, 1994] and [Nefian & Hayes, 1998] in the case of [Kohir & Desai, 1998] the method of extracting observation vectors results in better performance for a 50% overlap than for 75% overlap.

A second experiment was performed to prove the effectiveness of HMM in solving the face recognition problem regardless of the coefficients used, by replacing in the proposed system the wavelet coding with a trivial coding represented by the mean of the square window. The results obtained are 84.9% for 50% overlap and 77.8% for 75% overlap.

4. Hybrid approaches based on 1D HMM

From the discussions of the preceding section it can be seen that 1D HMM can perform successfully in face recognition applications. However, the vast majority of early experiments were performed on the ORL database. The images in this dataset only exhibit very small variations in head pose, facial expressions, facial occlusions such as facial hair and glasses, and almost no variations in illumination. For practical applications a face recognition system must be able to handle significant variations in facial appearance in a robust manner. Thus in this next section more challenging face recognition applications are described and further HMM approaches are considered from the literature. Specifically, in this section we consider hybrid approaches based on HMMs used successfully in more challenging applications of face recognition.

There are several core problems that a face recognition system has to solve, specifically those of variations in illumination, variations in facial expressions or partial occlusions of the face, and variations in head pose. Firstly an attempt at solving recognition problems caused by facial occlusions is considered [Martinez, 1999]. The solution adopted by this author was to explore the use of *principle component analysis* (PCA) features to characterize 6 different regions of the face and use 1D HMM to model the relationships between these regions. A second group of researchers [Wallhoff et al., 2001] have tackled the challenging task of recognizing side-profile faces in datasets where only frontal faces were used in the training stage. These authors have used a combination of *artificial neural network* (ANN) techniques combined with 1D HMM to solve this challenging problem.

4.1 Using 1D HMM with PCA derived features

A face recognition system is introduced [Martinez, 1999] for indexing images and videos from a database of faces. The system has to tackle three key problems, identifying frontal faces acquired, (i) under differing illumination conditions, (ii) with varying facial expressions and (iii) with different parts of the face occluded by sunglasses/scarves. Martinez's idea was to divide the face into N different regions analyzing each using PCA techniques and model the relationships between these regions using 1D HMMs.

The problem of different lighting conditions is solved in this paper by training the system with a broad range of illumination variations. To handle facial expressions and occlusions, the face is divided into 6 distinct local areas and local features are matched. This dependence on local rather than global features should minimize the effect of facial expressions and occlusions, which affect only a portion of the overall facial region. Each of these local areas obtained from all the images in the database is projected into a primary eigenspace. Each area is represented in vector form. Figure 6 [Martinez, 1999] shows the local feature extraction process.

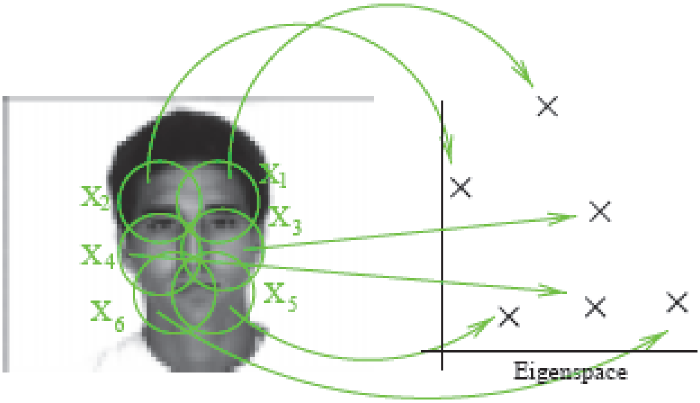


Fig. 6. Projection of the 6 different local areas into a global eigenspace Martinez, 1999].

Note that face localization is performed manually in this research and thus cannot be precise enough to guarantee that the extracted local information will always be projected accurately into the eigenspace. Thus information from pixels within and around the selected local area is also extracted, using a rectangular window. By considering these six local areas as hidden states, a 1D HMM was built for each image in the database. However, a more desirable case is to have a single HMM for each person in the database, as opposed to a HMM for each image. To achieve this, all HMMs of the same person were merged together into a single 1D HMM, where the transition probability from one state to another is *1/number of HMMs per person*. In the recognition phase, instead of using the forward-backward algorithm, the authors used the Viterbi algorithm [Rabiner, 1989] to compute the probability of an observation sequence given a model.

Two sets of tests were performed, using pictures and video sequences. The image database[4] was created by Aleix Martinez and Robert Benavente. It contains over 4,000 colour facial images corresponding to 126 people - 70 men and 56 women. There are 12 images per person, the first 6 frontal view faces with different facial expressions and illumination conditions and the second 6 faces with occlusions (sun-glasses and scarf) and different illumination conditions. These pictures were taken under strictly controlled conditions. No restrictions on appearance including clothing, accessories such as glasses, make-up or hairstyle were imposed on participants. Each person participated in two sessions, separated by 14 days. The same pictures were taken in both sessions. In addition, 30 video sequences were processed consisting of 25 images almost all of them containing a frontal face. Five different tests were run, using 50 people (25 males and 25 females) randomly selected from the database, converted to greyscale images and sampled at half their size, and also using 30 corresponding video sequences. In a first test, all 12 images per person were used in training, and the system was tested with every image by replacing each one of the local features with random noise with mean 0. The recognition rate obtained was 96.83%. For a second test training was with the first six images and testing with the last six images, featuring occlusions. A recognition rate of 98.5% was achieved. In a third test the last six images were used for training and the first six for testing and the resulting recognition rate was 97.1%. A fourth test consisted of training with only two non-occluded images and testing with all the remaining images. A lower recognition rate of 72% was obtained. Finally, the system was trained with all 12 images for each person, and tested with the video sequences, achieving a 93.5% recognition rate.

4.2 Artificial Neural Networks (ANN) in conjunction with 1D HMM

[Wallhoff et al., 2001] approached the challenging task of recognizing profile views with previous knowledge from only frontal views, which may prove a challenging task even for humans. The authors use two approaches based on a combination of Artificial Neural Networks (ANN) and a modelling technique based on 1D HMMs: a first approach uses a synthesized profile view, while a second employs a joint parameter-estimation technique. This paper is of particular interest because of its focus on non-frontal faces. In fact these authors are one of the first to address the concept of training the recognition system with conventional frontal faces, but extending the recognition to include faces with only a side-profile view.

[4] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

The experiments are performed on the MUGSHOT[5] database containing the images of 1573 cases, where most individuals are typically represented by only two photographs: one showing the frontal view of the person's face and the other showing the person's right hand profile. The database contains pairs of mostly male subjects at several ages and representatives of several ethnic groups, subjects with and without glasses or beards and a wide range of hairstyles. The lighting conditions and the background of the photographs also change. The pictures in the database are stored as 8-bit greyscale images. Prior to applying the main techniques of [Wallhoff et al., 2001] a pre-processing of each image is conducted. Photographs with unusually high distortions, perturbations or underexposure are discarded; all images are manually labelled so that all faces appear in the centre of the image and with a moderate amount of background, and resized to 64x × 64 pixels. Then two sets are defined: a first set consisting of 600 facial image pairs, *frontal* and right-hand *profile*, are used for training the neural network. A second set with 100 facial image pairs is used for testing. The features used for experiments are pixel intensities. In order to obtain the observation vectors, each image which was resized to 64 × 64 pixels is divided into 64 columns. So from each image 64 observation vectors are extracted. The dimension of the vectors is the number of rows in the image, which is also 64, and these vectors consist of pixel intensities. In the training phase an appropriate neural network is used, estimated by applying the following intuitions: (i) a point in the frontal view will be found in approximately the same row as in the profile view, (ii) considering the right half of the face to be almost bilaterally symmetrical with the left half, only the first 40 columns of the image are used in the input layer to the ANN. Figure 7 taken from [Wallhoff et al., 2001] shows how a frontal view of the face is used to generate the profile view. In the testing phase, a 1D left to right first order HMM is used, allowing self transitions and transitions to the next state only. The models consist of 24 states, plus two non-emitting start and end states.

In the first hybrid approach for face profile recognition there are two training stages. Firstly, a neural network is trained using the first set of 600 images, the frontal image of each individual representing the input and the profile view the output. In this way the neural network is trained to synthesize profiles from the frontal image. In figure 8 [Wallhoff et al., 2001] an example of synthesized profile is shown. In the second training stage, the 100 frontal images are introduced in the neural network and their corresponding profiles are synthesized. Using these profiles, an average profile HMM model is obtained. Then for each testing profile, an HMM model is built using for initialization the average profile model. The Baum-Welch estimation procedure is used for training the HMM.

In a second approach only one training stage is performed, the computation speed being vastly improved as a result. This proceeds as follows: the NN is trained using the frontal images as input; the target outputs are in this case the mean values of each Gaussian mixture used for describing the observations of the corresponding profile image. First, an average profile HMM model is obtained using the 600 training profile images. Using this average model, the mean values for each individual in the training set are computed and used as the target values for the NN to be trained. In the recognition phase, for each frontal face the mean value for profile is returned by the NN. Using this mean and the average profile model, the corresponding HMM is built, then the probability of the test profile image given the HMM model is computed. The recognition rates achieved for the

[5] http://www.nist.gov/srd/nistsd18.cfm

systems proposed in this paper are around 60% for the first approach and up to 49% for the second approach, compared to 70%-80% when humans perform the same recognition task. The approach presented by the authors is very interesting in the context of a mugshot database, where only the two instances, one *frontal* and one *profile* of a face are present. Also the results are quite impressive compared to the human recognition rates reported. However, both ANN and HMM are computationally complex, and using pixel intensities as features also contributes to making this approach very greedy in terms of computing resources.

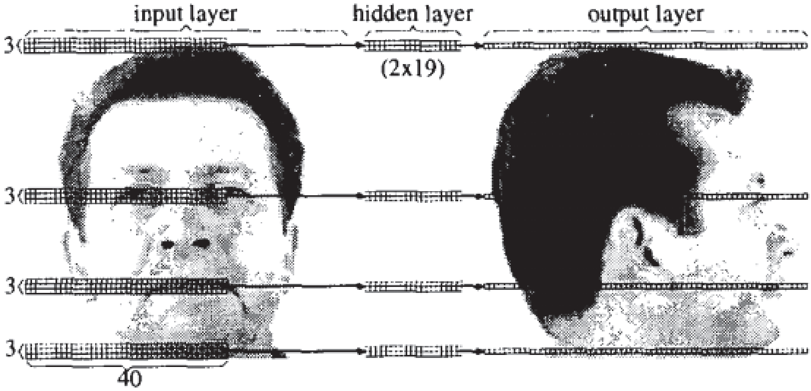


Fig. 7. Generation of a profile view from a frontal view [Wallhoff et al., 2001].

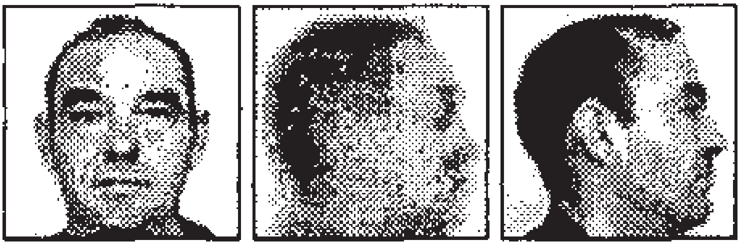


Fig. 8. Example of frontal view, generated and real profile [Wallhoff et al., 2001].

5. 2D HMM approaches

In section 3 and section 4 we showed how 1D HMMs might be adapted for use in face recognition applications. But face images are fundamentally 2D signals and it seems intuitive that they would be more effectively processed with a 2D recognition algorithm. Note however that a fully connected 2D extension of HMM exhibits a significant increase in computational complexity making it inefficient and unsuitable for practical face recognition applications [Levin & Pieraccini, 1992]. As a consequence of this complexity of the full 2D HMM approach a number of simpler structures were developed and are discussed in detail in the following sections.

5.1 A first application of pseudo 2D HMM to Facial Recognition

In his PhD thesis, [Samaria, 1994] was the first researcher to use *pseudo-2D* HMMs in face recognition, with pixel intensities as features. In order to obtain a P2D HMM, a one-dimensional HMM is generalized, to give the appearance of a two-dimensional structure, by allowing each state in a one-dimensional global HMM to be a HMM in its own right. In this way, the HMM consists of a top-level set of super states, each of which contains a set of embedded states. The super states may then be used to model the two-dimensional data in one direction, with the embedded HMMs modelling the data along the other direction. This model is appropriate for face images as it exploits the 2D physical structures of a face, namely that a face preserves the same structure of states from top to bottom – forehead, eyes, nose, mouth, chin, and also the same left-to-right structure of states inside each of these super states. An example of state structure for the face model and the non-zero transition probabilities of the P2D HMM are shown in figure 9. Each state in the overall top-to-bottom HMM is assigned to a left-to-right HMM.

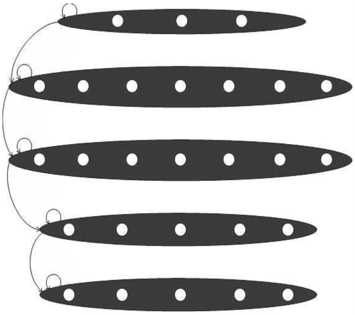


Fig. 9. Structure of a P2D HMM.

In order to simplify the implementation of P2D-HMM, the author used an equivalent 1D HMM to replace the P2D-HMM as shown in figure 10. In this case, the shaded states in the 1D HMM represent end-of-line states with two possible transitions: one to the same row of states - *superstate self-transition* - and one to the next row of states - *superstate to superstate transition*. For feature extraction a square window is used sliding from left-to-right and top-to-bottom. Each observation vector contains the intensity level values of the pixels contained by the window, arranged in a column-vector. In order to accommodate the extra end-of-line state, a white frame is added at the end of each line of sampling. Each state is modelled by one Gaussian with mean and standard deviation set, initialized at the beginning of training, to mid-intensity values for normal states and to white with near zero standard deviation for the end-of-line states. The parameters of the model are then iteratively re-estimated using the Baum-Welch algorithm.

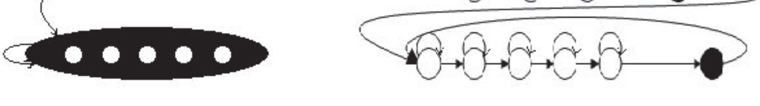


Fig. 10. P2D HMM and its equivalent 1D HMM.

Samaria's experiments were carried out on the ORL database. Different topologies and sampling parameters were used for the P2D-HMM: from 4 to 5 superstates and from 2 to 8 embedded states within each superstate. In addition these experiments considered different sizes of sampling windows with different overlaps ranging from 2 × 2 pixels with 1 × 1 overlap up to 24 × 22 (horizontal × vertical) pixels with 20 × 13 pixels overlap. The highest error rate of 18% was obtained for a 3-5-5-3 P2D-HMM, using a 10 × 8 scanning window with an 8 × 6 overlap, while the smallest error rate of 5.5% was obtained for 3-6-6-6-3 P2D-HMM, with 10 × 8 (and 12 × 8) window and 8 × 6 (and 9 × 6 respectively) overlap. In the same thesis Samaria also tested the standard *unconstrained* P2D HMM, which does not have an end-of-line state. In this case no attempt is made to enforce the fact that the last frame of a line of observations should be generated by the last state of the superstate. The recognition results for the *unconstrained* P2D HMM are similar to those obtained with constrained P2D-HMM, the error rates ranging from 18% to 6%. We remark that Samaria also obtained a 2% error rate for a 3-7-7-5-3 P2D HMM with 12×8 sampling window and 4 × 6 overlap, but considering that for only slightly different overlaps (8 × 6 and 4 × 4) the error rates were 6% and 8.5% respectively, this particular result appears to be a statistical anomaly. It does serve to remind that these models are based on underlying statistical probabilities and that occasional aberrations can occur.

5.2 Refining pseudo 2D HMM with DCT features

In [Nefian & Hayes, 1999] the authors adapted the P2D-HMM developed by [Kuo & Agazzi, 1994] for optical character recognition analysis, showing how it represented a valid approach for facial recognition and detection. These authors renamed this technique as *embedded* HMM. In order to obtain the observation vectors, a set of overlapping blocks are extracted from the image from left to right and top to bottom as shown in figure 11, the observation vector finally consisting of the 6 lower-frequency 2D-DCT coefficients extracted form each image block. Each state in the embedded HMMs is modelled using a single Gaussian.

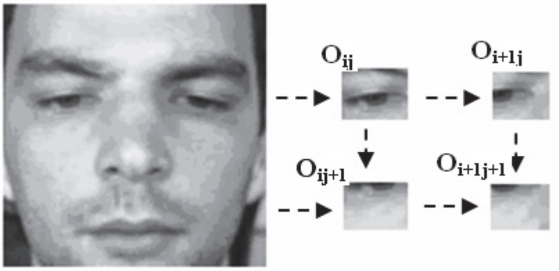


Fig. 11. Face image parameterization and blocks extraction.

For face recognition the ORL database was used. The system was trained with half of the database and tested with the other half. The recognition performance of the method presented in this paper is 98%, improving by more than 10% compared with the best results obtained in using 1D HMM in earlier work [Nefian & Hayes, May 1998, October 1998].

This research also considered the problem of face detection. In the testing phase for detection, 288 images of the MIT database were used, representing 16 subjects with different illuminations and head orientations. A set of 40 images representing frontal views of 40 different individuals from the ORL database is used to train one face model. The testing is performed using a doubly embedded Viterbi algorithm described by [Kuo & Agazzi, 1994]. The detection rate of the system described in this paper is 86%. While this version of HMM appears to be relatively efficient in face detection, it is however computationally very complex and slow, particularly when compared with state of art algorithms [Viola & Jones, 2001].

5.3 Improved initialization of pseudo 2D HMM

Also employing a P2D HMM, [Eickler et al., 2000] describe an advanced face recognition system based on the use of standard P2D HMM employing 2D DCT features is presented. The performance of the system is enhanced using improved initialization techniques and mirror images. It is very important to use a good initial model therefore the authors used all faces in the database to build a 'common initial model'. Then for each person in the database a P2D HMM model is refined using this 'common model'. Feature extraction is based on DCT. The image is scanned with a sliding window of size 8 × 8 from left-to-right and top-to-bottom with an overlap of 6 pixels (75%). The first 15 DCT coefficients are extracted. The use of DCT coefficients allows the system to work directly on images compressed with JPEG standard without a need to decompress these images. The size of the sampling window was chosen as 8 × 8 because the DCT portion of JPEG image compression is based on this window size.

Tests are performed on the ORL database, described previously, with the first 5 images per person used for training and the remaining 5 for testing. Three sets of experiments are performed in this paper. *First Experiment Set:* the system is tested on different quadratic P2D HMM model topologies (4×4 states to 8×8 states) with 1 to 3 Gaussian mixtures to model the probability density functions. The recognition rates achieved range from 81.5% for 4×4 states with 1 Gaussian to 100% for 8 × 8 states with 2 and 3 Gaussians. *Second Experiment Set*: the effect of overlap on recognition rates is tested. An overlap of 75% is used for all training while for testing overlaps between 75% and 0% were used, with a 7×7 HMM and from 1 to 3 Gaussian mixtures. The overall result of this experiment is that recognition rates decrease slightly when the overlap is reduced, however, very good recognition rates of 94.5%-99.5% were still obtained even for 0% overlap, compared with 98.5%-100% for 75% overlap. Thus wide variations in overlap have relatively minor effects on overlap for a sophisticated 7x7 HMM model.

Third Experiment Set: comprises an evaluation of the effect of compression artefacts on the recognition rate. Recognition was performed on JPEG compressed images across a range of quality settings ranging from 100 for the best quality to 1 for the highest compression ratio as shown in figure 12 [Eickler et al., 2000]. The results are as follows: for compression ratios of up to 7.5 to 1, the recognition rates remain constant around 99.5%±0.5%. For compression ratios over 12.5 to 1, the recognition rates drop below 90%, down to approximately 5% for 19.5 to 1 compression ratio.

There are some additional conclusions we can draw from the work of [Eickler et al., 2000]. Firstly, building an initial HMM model using all faces in the database is an improvement over the intuitive initialization used by [Samaria, 1994] or [Nefian & Hayes, 1999], however this may lead to the dependency of the initial model on the composition of the database. Secondly, these authors obtain excellent results when using JPEG compressed images in the testing phase (overlap 0%), speeding up the recognition process significantly. Note however, in the training stage they use uncompressed images scanned with a 75% overlap and as they have used a very complex HMM model with 49 states the training stage of their approach is resource and time intensive offsetting the benefits of faster recognition speeds.

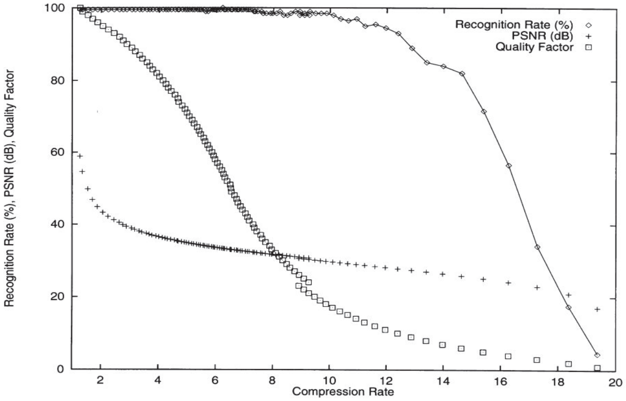


Fig. 12. Recognition rates versus compression rates [Eickler et al., 2000].

5.4 Discrete vs continuous modelling of observation vectors for P2D HMM

In another paper on the subject of face recognition using HMM, [Wallhoff et al., 2001] consider if there is a major difference in recognition performance between HMMs where the observation vectors are modelled as continuous or discrete processes. In the continuous case, the observation probability is expressed as a density probability function approximated by a weighted sum of Gaussian mixtures. In the case of a discrete output probability, a discrete set of observation probabilities is available for each state, and input vector. This discrete set is stored as a set of codebook entries. The codebook is typically obtained by k-means clustering of all available training data feature vectors.

The authors used for their experiments 321 subjects selected from the FERET database[6]. For testing the system, two galleries of images were used: f_a gallery, containing a regular frontal image for each subject, and f_b gallery, containing an alternative frontal image, taken seconds after the corresponding f_a image. First the images are pre-processed, using a semi-automated feature extraction that starts with the manual labelling of the eye and mouth centre-coordinates. The next step is the automatic rotation of the original images so that a line through the eyes is horizontal. After this the face is divided vertically and processing continues on a half-face image. The images are re-sized to the smallest image among the resulting images being 64 × 96 pixels.

For feature extraction, the image is scanned using a rectangular window, with an overlap of 75%. After the DCT coefficients for each block are calculated, a triangular shaped mask is applied and the first 10 coefficients are retained, representing the observation vector. Two

[6] http://www.frvt.org/feret/default.htm

sets of experiments were performed, for continuous and discrete outputs. For the case of *continuous output*, the experiments used 8 × 8 and 16 × 16 scanning windows, and 4 × 4 to 7 × 7 state structures for the P2D HMM. Initially only one Gaussian per state was used. The best recognition rate in this case was 95.95%, for 8 × 8 block size and 7 × 7 states for HMM. When the number of Gaussians was increased form 1 to 3, the recognition rate dropped, maybe due to the fact that only one image per person was used in the training phase. In the case of *discrete output values*, identical scanning windows and HMM were used, and two codebook sizes of 300 and 1000 values were used to generate the observation vectors. The highest recognition rate obtained was 98.13%, for 8 × 8 pixels block size, 7 × 7 states HMM, and a codebook size of 1000. In both cases, continuous and discrete, better results were obtained for the smaller size of scanning window.

5.5 Face retrieval on large databases

After using the combination of 2D DCT and P2D HMM for face recognition on small databases, a new HMM-based measure to rank images within a larger database is next presented, [Eickeler, 2002]. The relation of the method presented to confidence measures is pointed out and five different approximations of the confidence measure for the task of database retrieval are evaluated. These experiments were carried out on the C-VIS database, containing the extracted faces of three days of television broadcast resulting in 25000 unlabeled face images. Normal HMM-based face recognition for database retrieval entails building a model for each person in the database. However, in the case of a very large and unlabeled database, that would imply building a model λ_j for each image O_j in the database, which is not only computationally expensive, but results in poor modelling, considering that a robust model for one person requires multiple training images of that person. In this case, calculating the probability of a query image for each built model $P(O_{query} \mid \lambda_j)$ is simply not practical.

A more feasible method for database retrieval is to train a query HMM λ_{query} using the query images O_{query} of the person searched for ω_{query}, but noting that the probability derived by the Forward-Backward algorithm, $P(O_j \mid \lambda_{query})$ cannot be used as ranking measure for the images in the database because inaccuracies in the modelling of the face images have a big influence on the probability. In order to fix this problem, the ranking of the images uses the query model λ_{query} as a representation of the person being searched for and a set of cohort models Λ_{cohort} representing people not being searched for. An easy way to form the cohort is by using former queries or by taking some images form the database. So instead of calculating $P(O_{query} \mid \lambda_j)$, the probability of an image O_j given the person being searched is used:

$$P(O_j \mid \omega_{query}) \propto P(\lambda_{query} \mid O_j) = \frac{P(O_j \mid \lambda_{Query})}{P(O_j \mid \Lambda_{cohort})} \tag{1}$$

In this research five different confidence measures were used for database retrieval based on this formula. For the confidence measure using normalization, the denominator is replaced:

$$P(O_j \mid \Lambda_{cohort}) = \sum_{\lambda_k \in \Lambda_{cohort}} P(O_j \mid \lambda_k) \tag{2}$$

Another confidence measure uses one filler (common) model instead of a cohort of HMMs for a group of people. The filler model can be trained on all people of the cohort group. If the denominator is set to a fixed probability, it can be dropped from the formula, in which case the confidence measure will be $P(O_j \mid \lambda_{query})$. The fourth confidence measure is based on the sum of ranking differences between the ranking of the cohort models on the query image and the ranking of the cohort models on each of the database images. Finally, the Levenshtein Distance (the Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other) is considered as an alternative measure for the comparison of the rankings of the cohort models for the query image and the database images.

For the experimental part 14 people with 8 to 16 face images each were used as query images, and also as cohort set. A NN-based face detector was used to detect the inner facial rectangles in the video broadcast and the rectangle of each image is scaled to 66 × 86 pixels. In order to remove the background an ellipsoid mask is applied. A P2D HMM with 5 × 5 states is used. The results of the query are evaluated using precision and recall: precision is the proportion of relevant images among the retrieved images while recall is the proportion of relevant images in the database that are part of the retrieval result. In a first experiment a database retrieval for each person of the query set using the normalization is performed and only the precision is calculated considering the database is unlabeled hence an exact number for each person in unknown. For 12 out of 14 people the precision is constant at 100% for around 40 retrieved images (the number of images per person varies between 20 and 300). In a second experiment all five measures were tested for one person. The results are almost perfect for normalization, a little worse but much faster for the filler model. The 'sum of ranking differences' and Levenshtein Distance measures return relatively good results but are inferior to normalization, while the use of a fixed probability gives significantly worse results than all other measures.

5.6 A low-complexity simplification of the Full-2D-HMM

An alternative approach to 2D HMM was proposed by [Othman & Aboulnasr, 2000]. These authors propose a *low-complexity* 2D HMM (LC2D HMM) system for face recognition. The aim of this research is to build a full 2D HMM but with reduced complexity. The challenge is to take advantage of a full 2D HMM structure, but without the full complexity implied by an unconstrained 2D model. Their model is implemented in the 2D DCT compressed domain with 8×8 pixel non-overlapping blocks to maintain compatibility with standard JPEG images. The authors claim a computational complexity reduction from N^4 for a fully connected 2D HMM to $2N^2$ for the LC2D HMM, where N is the number of states. Although the accuracy of the system is not better than other approaches, these authors claim that the computational complexity involved is somewhat less than that required for a 1D HMM and significantly less than that of P2D HMM.

The LC2D HMM is based on 2 key assumptions: (i) the active state at the observation block $B_{k,l}$ is dependant only on immediate vertical and horizontal neighbours, $B_{k-1,l}$ and $B_{k,l-1}$;[7]; (ii) the active states at the 2 observation blocks in anti-diagonal neighbourhood locations, $B_{k-1,l}$ and $B_{k,l-1}$ are statistically independent given the current state. This assumption allows

[7] From a mathematical perspective this assumption is equivalent to a second-order Markov Model, requiring a 3D transition matrix.

separating the 3D state transition matrix into two distinct 2D transition matrices, for horizontal and vertical transitions. This decreases the complexity of the model quite significantly. This *low-complexity* model topology and image scanning are illustrated in figure 13.

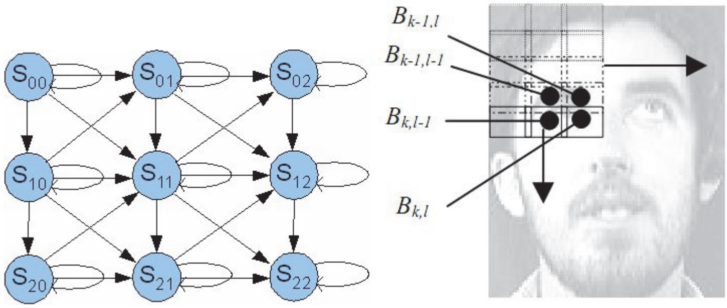


Fig. 13. (a) Image scanning b) Model topology [Othman & Aboulnasr, 2000]

The authors state that the two assumptions are acceptable for non-overlapped feature blocks, but have less validity for very small sized feature blocks or as the allowable overlap increases. The tests were performed on the ORL database. The model for each person was trained with 9 images, and the remaining image was used in the testing phase. Image scanning is performed in a two dimensional manner, with block size set to 8×8. Only the first 9 DCT coefficients per block were used. Different block overlap values were used to investigate the system performance and the validity of the design assumptions. The recognition rates are around 70% for 0 or 1 pixel overlap, decreasing dramatically down to only 10% for a 6-pixel overlap. This is explained because the assumptions of statistical independence, which are the underlying basis of this model, lose their validity as the overlap increases.

5.7 Refinements of the low-complexity approach

In a subsequent publication by the same authors, [Othman & Aboulnasr, 2001], a hybrid HMM for face recognition is introduced. The proposed system comprises of a LC2D HMM, as described in their earlier work used in combination with a 1D HMM. The LC2D HMM carries out a complete search in the compressed JPEG domain, and a 1D HMM is then applied that searches only in the candidate list provided by the first module.

In the experiments presented in this paper, a 6×2 states model was used for the LC2D HMM, and 4 and 5 state top-to-bottom models were used for the 1D HMM. For the 1D HMM, DCT feature extraction is performed on a horizontal 10 × 92 scanning window. For the 2D HMM, a 8×8 block size is used for scanning the image, and the first 9 DCT coefficients are retained from each block. No overlap is allowed for the sliding windows. Tests are performed on the ORL database. In a first series of tests the effects of training data size on the model robustness were studied. The accuracy of the system ranges from 48%-58% when trained with only 2 images per person, to almost 95%-100% if trained with 9 images per person. A second series of experiments provides a detailed analysis of the trade-off between recognition accuracy and computational complexity and determines an optimal operating point for this hybrid approach. This appears to be the first research in this field to consider such trade-offs in a detailed study and this methodology should provide a useful approach for other researchers in the future.

In a third paper, [Othman & Aboulnasr, 2003], these authors propose a 2D HMM face recognition system that limits the independence assumptions described in their original work to conditional independence among adjacent observation blocks. In this new model, the active states of the two anti-diagonal observation blocks are statistically independent given the current state and knowledge of the past observations. This translates into a more flexible model, allowing state transitions in the transverse direction as shown in figure 14, taken from, [Othman & Aboulnasr, 2003].

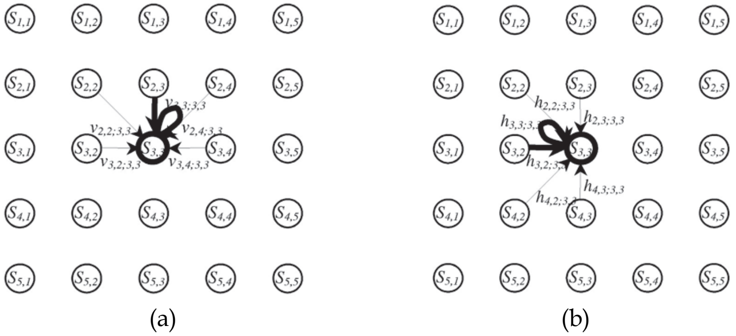


Fig. 14. Modified LC2D HMM [Othman & Aboulnasr, 2003]. (a) Vertical transitions to state S3,3 for 5×5 state model (b) Horizontal transitions to state S3,3 also for 5×5 state model.

This modified LC2D HMM face recognition system is examined for different values of the structural parameters, namely number of states per model and number of Gaussian mixtures per state. These tests are again conducted on the ORL database. The images are scanned using 8×8 blocks and the first 9 2D DCT coefficients comprise the observation vector. The HMMs were trained using 9 images per person, and tested using the 10th image. The test is repeated 5 times with different test images and the results are averaged over a total of 200 test images for 40 persons. Test images are not members of the training data set at any time. The results vary from a very low 4% recognition rate for a 7 × 3 HMM with 64 Gaussian mixtures per state, up to 100% for a 7 × 3 HMM with 4 Gaussian mixtures per state. Best results are obtained for 4 and 8 Gaussians per state. The reason for the poor performance for a higher number of Gaussian mixtures is that the model becomes too discriminating and cannot recognize data with any flexibility, outside the original training set. Finally, the reader's attention is drawn to detailed comments by [Yu & Wu, 2007] on the key assumption of conditional independence in the relationship between adjacent blocks. In this communication, [Yu & Wu, 2007] it is shown that this key assumption is entirely unnecessary.

6. More recent research on HMM in face recognition

While there have been more recent research which applies HMM techniques to face recognition, most of this work has not refined the underlying methods, but has instead combined known HMM techniques with other face analysis techniques. Some work is worth mentioning, such as that of [Le & Li, 2004] who combined a one-dimensional discrete hidden Markov model (1D-DHMM) with new way of extracting observations and using observation sequences. All subjects in the system share only one HMM that is used as a means to weigh a pair of observations. The Haar wavelet transform is applied to face images to reduce the dimensionality of the observation vectors. Experiments on the AR face database[8] and the CMU PIE face database[9] show that the proposed method outperforms PCA, LDA, LFA based approaches tested on the same databases.

Also worth mentioning is the work of [Yujian, 2006]. In this paper, several new analytic formulae for solving the three basic problems of 2-D HMM are provided. Although the complexity of computing these is exponential in the size of data, it is almost the same as that of a 1D HMM for cases where the numbers of rows or columns are a small constant. While this author did not apply these results specifically to facial recognition problem they appear to offer some promise in simplifying the application of a full 2D HMM to the face recognition problem.

Another notable contribution is the work of [Chien & Liao, 2008] which explores a new discriminative training criterion to assure model compactness combined with ability for accurate to discrimination between subjects. Hypothesis testing is employed to maximize the confidence level during model training leading to a *maximum-confidence* model (MC-HMM) for face recognition. From experiments on the FERET[10] database and GTFD[11], the proposed method obtains robust segmentation in the presence of different facial expressions, orientations, and so forth. In comparison with the maximum likelihood and minimum classification error HMMs, the proposed MC-HMM achieves higher recognition accuracies with lower feature dimensions. Notably this work uses more challenging databases than the ORL database.

Finally we conclude this chapter referring to our own recent work in face recognition using EHMM, presented in [Iancu, 2010; Corcoran & Iancu 2011]. This work can be divided in three parts according to our objectives. The tests were performed on a combined database (BioID, Achermann, UMIST) and on the FERET database. The first objective was to build a recognition system applicable on handheld devices with very low computational power. For this we tested the EHMM-based face recognizer for different sizes of the model, different number of Gaussians, picture size, features, and number of pictures per person used for training. The results obtained for very small picture size (32 × 32), with 1 Gaussian per state and on a simplified EHMM are only 58% recognition for only 1 image per person used for training, when we use 5 pictures per person for training the recognition rates go up to 82% [Corcoran & Iancu 2011]. A second objective was to limit the effect of illumination variations on recognition rates. For this three illumination normalization techniques were used and various combinations of these were tested: histogram equalization (HE), contrast limited adaptive histogram equalization (CLAHE) and DCT in logarithm domain (logDCT). The best recognition rates were obtained for a combination of CLAHE and HE (95.71%) and the worst for logDCT (77.86%) on the combined database [Corcoran & Iancu 2011].

[8] http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[9] http://www.ri.cmu.edu/research_project_detail.html?project_id=418&menu_id=261

[10] http://www.frvt.org/feret/default.htm

[11] http://www.anefian.com/research/face_reco.htm

A third objective was to build a system robust to head pose variations. For this we tested the face recognition system using frontal, semi-profile and profile views of the subjects. The first set of tests was performed on the combined database. Here the maximum head pose angle is around 30°. We compared recognition rates obtained when building one EHMM model per person versus one EHMM model per picture. The second set of tests was performed on FERET database which has a much bigger variety of head poses. In this case we used one frontal, 2 semi-profiles and 2 profiles for each subject in the training stage and all pictures of each subject in the testing stage. We compared the recognition rates when building 1 model per person versus 2 models per person versus 3 models per person. We obtained better recognition rates for one model per person for the first set of tests where the database has little head pose variation but better recognition rates for 2 models per person for the second set of tests where the database has a very high head pose variation [Iancu, 2010].

7. Review and concluding remarks

The focus of this chapter is on the use of HMM techniques for face recognition. For this review we have presented a concise yet comprehensive description and review of the most interesting and widely used techniques to apply HMM models in face recognition applications. Although additional papers treating specific aspects of this field can be found in the literature, these are invariably based on one or another of the key techniques presented and reviewed here.

Our goal has been to quickly enable the interested reader to review and understand the state-of-art for HMM models applied to face recognition problems. It is clear that different techniques balance certain trade-offs between computational complexity, speed and accuracy of recognition and overall practicality and ease-of-use. Our hope is that this article will make it easier for new researchers to understand and adopt HMM for face analysis and recognition applications and continue to improve and refine the underlying techniques.

8. References

Baum, L. E. & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state Markov chains. *Annals of Mathematical Statistics*, vol. 37, 1966.

Baum, L. E.; Petrie, T.; Soules, G. & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics*, vol. 41, 1970.

Baum, L. E. (1972). An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes, Inequalities, vol. 3, pp. 1-8, 1972.

Bicego, M.; Castellani, U. & Murino, V. (2003b). Using hidden markov models and wavelets for face recognition. *Proceedings of Image Analysis and Processing 2003. 12th International Conference on*, pp. 52–56, 2003.

Bicego, M.; Murino, V. & Figueiredo, M. (2003a). A sequential pruning strategy for the selection of the number of states in hidden markov models. *Pattern Recognition Letters*, Vol. 24, pp. 1395–1407, 2003.

Chien, J-T. & Liao, C-P. (2008). Maximum Confidence Hidden Markov Modeling for Face Recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , Vol. 30, No 4, pp. 606-616, April 2008

Corcoran, P.M. & Iancu, C. (2011). Automatic Face Recognition System for Hidden Markov Model Techniques, *Face Recognition Volume 2, Intech Publishing*, 2011.

Eickeler, S. (2002). Face database retrieval using pseudo 2d hidden markov models. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition, Proceedings*, pp. 58–63, May 2002.

Eickeler, S.; Muller, S. & Rigoll, G. (2000). Recognition of jpeg compressed face images based on statistical methods. *Image and Vision Computing Jour- nal, Special Issue on Facial Image Analysis*, Vol. 18, No 4, pp. 279–287, March 2000.

Figueiredo, M.; Leitao, J. & Jain, A. (1999). On fitting mixture models. *Proceedings of the Second International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer-Verlag*, pp. 54-69, 1999.

Iancu, C. (2010). Face recognition using statistical methods. *PhD thesis*, NUI Galway, 2010.

Jelinek, F.; Bahl, L.R. & Mercer, R.L. (1975). Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Transactions on Information Theory*, Vol. 21, No 3, pp. 250 – 256, 1975.

Juang, B.H. (1984). On the hidden markov model and dynamic time warping for speech recognition-a unified view. *AT&T Technical Journal*, Vol. 63, No 7, pp. 1213–1243, September 1984.

Juang, B.H. & Rabiner, L.R. (2005). Automatic speech recognition - a brief history of the technology development. *Elsevier Encyclopedia of Language and Linguistics*, Second Edition, 2005.

Kohir, V.V. & Desai, U.B. (1998). Face recognition using a dct-hmm approach. *Applications of Computer Vision, WACV '98, Proceedings, Fourth IEEE Workshop on*, pp. 226–231, October 1998.

Kohir, V.V. & Desai, U.B. (1999). A transform domain face recognition approach. *TENCON 99, Proceedings of the IEEE Region 10 Conference*, Vol. 1, pp. 104–107, September 1999.

Kohir, V.V. & Desai, U.B. (2000). Face recognition. *IEEE International Symposium on Circuits and Systems*, Geneva, Switzerland, May 2000.

Kuo, S. & Agazzi, O. (1994). Keyword spotting in poorly printed documents using pseudo 2-d hidden markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, pp. 842–848, August 1994.

Le, H.S. & Li, H. (2004). Face identification system using single hidden markov model and single sample image per person. *IEEE International Joint Conference on Neural Networks*, Vol. 1, 2004.

Levin, E. & Pieraccini, R. (1992). Dynamic planar warping for optical character recognition. *Proceedings ICASSP 1992,* San Francisco, Vol. 3, pp. 149–152, March 1992.

Levinson, S.E.; Rabiner, L.R. & Sondhi, M.M. (1983). An introduction to the application of the theory of probabilistic functions of a markov process to automatic speech recognition. *Bell System Technical Journal*, Vol. 62, No 4, pp. 1035–1074, April 1983.

Martinez, A. (1999). Face image retrieval using hmms. *IEEE Workshop on Content-Based Access of Image and Video Libraries, (CBAIVL '99) Proceedings*, pp. 35–39, June 1999.

Nefian, A.V. (1999). A hidden markov model based approach for face detection and recognition. *PhD Thesis*, 1999.

Nefian, A.V. & Hayes III, M.H. (Oct. 1998). Face detection and recognition using hidden markov models. *Image Processing, ICIP 98, Proceedings. 1998 International Conference on*, Vol. 1, pp. 141–145, October 1998.

Nefian, A.V. & Hayes III, M.H. (May 1998). Hidden markov models for face recognition. *Acoustics, Speech, and Signal Processing ICASSP '98. Proceedings of the 1998 IEEE International Conference on*, Vol. 5, pp. 2721–2724, May 1998.

Nefian, A.V. & Hayes III, M.H. (1999). An embedded hmm-based approach for face detection and recognition. *Acoustics, Speech, and Signal Processing, ICASSP '99. Proceedings,* IEEE International Conference, 6:3553–3556, March 1999.

Othman, H. & Aboulnasr, T. (2000). Hybrid hidden markov model for face recognition. *4th IEEE Southwest Symposium on Image Analysis and In- terpretation*, pp. 34–40, April 2000.

Othman, H. & Aboulnasr, T. (2001). A simplified second-order hmm with ap- plication to face recognition. *ISCAS 2001 IEEE International Symposium on Circuits and Systems*, Vol. 2, pp. 161–164, May 2001.

Othman, H. & Aboulnasr, T. (2003). A separable low complexity 2d hmm with application to face recognition. *Pattern Analysis and Machine Intelli- gence, IEEE Transactions on*, 2003.

Park, H.S. & Lee, S.W. (1998). A Truly 2D Hidden Markov Model For Off-Line Handwritten Character Recognition. *Pattern Recognition*, Vol. 31, No 12, pp. 1849-1864, December 1998.

Rabiner, L.R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of IEEE*, Vol. 77, No 2, pp. 257–286, February 1989.

Rabiner, L.R. & Juang, B.H. (1986). An introduction to hidden markov models. *IEEE ASSP Magazine*, Vol. 3, No 1, pp. 4–16, 1986.

Samaria, F. (1994). Face recognition using hidden markov models. *Ph.D. thesis*, Department of Engineering, Cambridge University, UK, 1994.

Samaria, F. & Fallside, F. (1993). Face identification and feature extraction using hidden markov models. *Image Processing: Theory and Applications, Elsevier*, pp. 295–298, 1993.

Samaria, F. & Harter, A.C. (1994). Parameterization of a stochastic model for human face identification. *Applications of Computer Vision, 1994., Pro- ceedings of the Second IEEE Workshop on*, Vol. 77, pp. 138–142, December 1994.

Viola, P. & Jones, M. (2001). Robust real-time object detection, *Technical report 2001/01*, Compaq CRL, 2001.

Wallhoff, F.; Eickeler, S. & Rigoll, G. (2001). A comparison of discrete and continuous output modeling techniques for a pseudo-2d hidden markov model face recognition system. *International Conference on Image Processing, Proceedings*, Vol. 2, pp. 685–688, October 2001.

Wallhoff, F., Müller, S. & Rigoll, G. (2001). Hybrid face recognition system for profile views using the mugshot database. *IEEE ICCV Workshop on Recognition, Analysis and*

Tracking of Faces and Gestures in Real-Time Systems, Proceedings, pp. 149–156, July 2001.

Yu, L. & Wu, L. (2007). Comments on ’a separable low complexity 2d hmm with application to face recognition’. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 29, No 2, pp. 368–368, February 2007.

Yujian, L. (2007). An analytic solution for estimating two-dimensional hidden Markov models. *Applied Mathematics and Computation,* Vol. 185, No 2, pp. 810-822, February 2007.

2

GMM vs SVM for Face Recognition and Face Verification

Jesus Olivares-Mercado, Gualberto Aguilar-Torres, Karina Toscano-Medina,
Mariko Nakano-Miyatake and Hector Perez-Meana
National Polytechnic Institute
Mexico

1. Introduction

The security is a theme of active research in which the identification and verification identity of persons is one of the most fundamental aspects nowadays. Face recognition is emerging as one of the most suitable solutions to the demands of recognition of people. Face verification is a task of active research with many applications from the 80's. It is perhaps the biometric method easier to understand and non-invasive system because for us the face is the most direct way to identify people and because the data acquisition method consist basically on to take a picture. Doing this recognition method be very popular among most of the biometric systems users. Several face recognition algorithms have been proposed, which achieve recognition rates higher than 90% under desirable's condition (Chellapa et al., 2010; Hazem & Mastorakis, 2009; Jain et al., 2004; Zhao et al., 2003).

The recognition is a very complex task for the human brain without a concrete explanation. We can recognize thousands of faces learned throughout our lives and identify familiar faces at first sight even after several years of separation. For this reason, the Face Recognition is an active field of research which has different applications. There are several reasons for the recent increased interest in face recognition, including rising public concern for security, the need for identity verification in the digital world and the need for face analysis and modeling techniques in multimedia data management and computer entertainment. Recent advances in automated face analysis, pattern recognition, and machine learning have made it possible to develop automatic face recognition systems to address these applications (Duda et al., 2001).

This chapter presents a performance evaluation of two widely used classifiers such as Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) for classification task in a face recognition system, but before beginning to explain about the classification stage it is necessary to explain with detail the different stages that make up a face recognition system in general, to understand the background before using the classifier, because the stages that precede it are very important for the proper operation of any type of classifier.

1.1 Face recognition system

To illustrate the general steps of a face recognition system consider the system shown in Fig. 1, which consists of 4 stages:

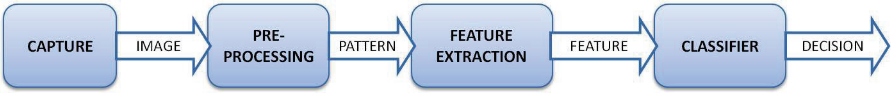


Fig. 1. General Structure of a face recognition system.

1.1.1 Capture

This stage is simple because it only needs a camera to take the face image to be procesed. Due to this is not necessary to have a camera with special features, currently cell phones have a camera with high resolution which would serve or a conventional camera would be more than enough because the image can be pre-processed prior to extract the image features. Obviously, if the camera has a better resolution can be obtained clearer images for processing.

1.1.2 Pre-processing

In this stage basically apply some kind of cutting, filtering, or some method of image processing such as normalization, histogram equalization or histogram specification, among others. This is to get a better image for processing by eliminating information that is not useful in the case of cutting or improving the quality of the image as equalization. The pre-processing of the image is very important because with this is intended to improve the quality of the images making the system more robust for different scenarios such as lighting changes, possibly noise caused by background, among others.

1.1.3 Feature extraction

The feature extraction stage is one of the most important stages in the recognition systems because at this stage are extracted facial features in correct shape and size to give a good representation of the characteristic information of the person, that will serve to have a good training of the classification models.

Today exists great diversity of feature extraction algorithms, the following are listed some of them:

- Fisherfaces (Alvarado et al., 2006).
- Eigenfaces(Alvarado et al., 2006).
- Discrete Walsh Transform (Yoshida et al., 2003).
- Gabor Filters (Olivares et al., 2007).
- Discrete Wavelet Transform (Bai-Ling et al., 2004).
- Eigenphases (Savvides et al., 2004).

1.1.4 Classifiers

The goal of a classifier is to assign a name to a set of data for a particular object or entity. It defines a set of training as a set of elements, each being formed by a sequence of data for a specific object. A classifier is an algorithm to define a model for each class (object specific), so that the class to which it belongs an element can be calculated from the data values that define the object. Therefore, more practical goal for a classifier is to assign of most accurate form to new elements not previously studied a class. Usually also considered a test set that allows measure the accuracy of the model. The class of each set of test is known and is used to validate the model. Currently there are different ways of learning for classifiers among which are the supervised and unsupervised.

In supervised learning, a teacher provides a category label or cost for each pattern in a training set, and seeks to reduce the sum of the costs for these patterns. While in unsupervised learning or clustering there is no explicit teacher, and the system forms clusters or "natural groupings "of the input patterns. "Natural"is always defined explicitly or implicitly in the clustering system itself; and given a particular set of patterns or cost function, different clustering algorithms lead to different clusters.

Also, is necessary to clarify the concept of identification and verification. In identification the system does not know who is the person that has captured the characteristics (the human face in this case) by which the system has to say who owns the data just processed. In verification the person tells the system which is their identity either by presenting an identification card or write a password key, the system captures the characteristic of the person (the human face in this case), and processes to create an electronic representation called live model. Finally, the classifier assumes an approximation of the live model with the reference model of the person who claimed to be. If the live model exceeds a threshold verifying is successful. If not, the verification is unsuccessful.

1.1.4.1 Classifiers types.

Exist different types of classifiers that can be used for a recognition system in order to choose one of these classifiers depends on the application for to will be used, it is very important to take in mind the selection of the classifier because this will depend the results of the system. The following describes some of the different types of classifiers exist.

Nearest neighbor. In the nearest-neighbor classification a local decision rule is constructed using the k data points nearest the estimation point. The k-nearest-neighbors decision rule classifies an object based on the class of the k data points nearest to the estimation point x_0. The output is given by the class with the most representative within the k nearest neighbors. Nearness is most commonly measured using the Euclidean distance metric in x-space (Davies E. R., 1997; Vladimir & Filip, 1998).

Bayes' decision. Bayesian decision theory is a fundamental statistical approach to the problem of pattern recognition. This approach is based on quantifying the tradeoffs between various classification decisions using probability and the costs that accompany such decisions. It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known (Duda et al., 2001).

Neural Networks. Artificial neural networks are an attempt at modeling the information processing capabilities of nervous systems. Some parameters modify the capabilities of the network and it is our task to find the best combination for the solution of a given problem. The adjustment of the parameters will be done through a learning algorithm, i.e., not through explicit programming but through an automatic adaptive method (Rojas R., 1996).

Gaussian Mixture Model. A Gaussian Mixture Model (GMM) is a parametric probability density function represented as a weighted sum of Gaussian component densities. GMMs are commonly used as a parametric model of the probability distribution of continuous measurements or features in a biometric system, such as vocal-tract related spectral features in a speaker recognition system. GMM parameters are estimated from training data using the iterative Expectation-Maximization (EM) algorithm or Maximum A Posteriori MAP) estimation from a well-trained prior model (Reynolds D. A., 2008).

Support Vector Machine. The Support Vector Machine (SVM) is a universal constructive learning procedure based on the statistical learning theory. The term "universal"means that the SVM can be used to learn a variety of representations, such as neural net (with the usual sigmoid activation), radial basis function, splines, and polynomial estimators. In more general sense the SVM provides a new form of parameterization of functions, and hence it can be applied outside predictive learning as well (Vladimir & Filip, 1998).

In This chapter presents only two classifiers, the Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) as it classifiers are two of the most frequently used on different pattern recognition systems, and then a detailed explanation and evaluation of the operation of these classifier is required.

2. Gaussian Mixture Model

2.1 Introduction.

Gaussian Mixture Models can be used to represent probability density functions complex, from the marginalization of joint distribution between observed variables and hidden variables. Gaussian mixture model is based on the fact that a significant number of probability distributions can be approximated by a weighted sum of Gaussian functions as shown in Fig. 2. Use of this classifier has excelled in the speaker's recognition with very good results (Reynolds & Rose, 1995; Reynolds D. A., 2008).

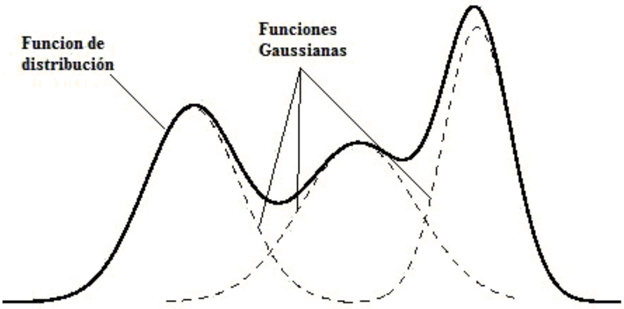


Fig. 2. Approximation of a probability distribution function by a weighted sum of Gaussian functions.

To carry out the development of Gaussian Mixture Model must consider 3 very important points:

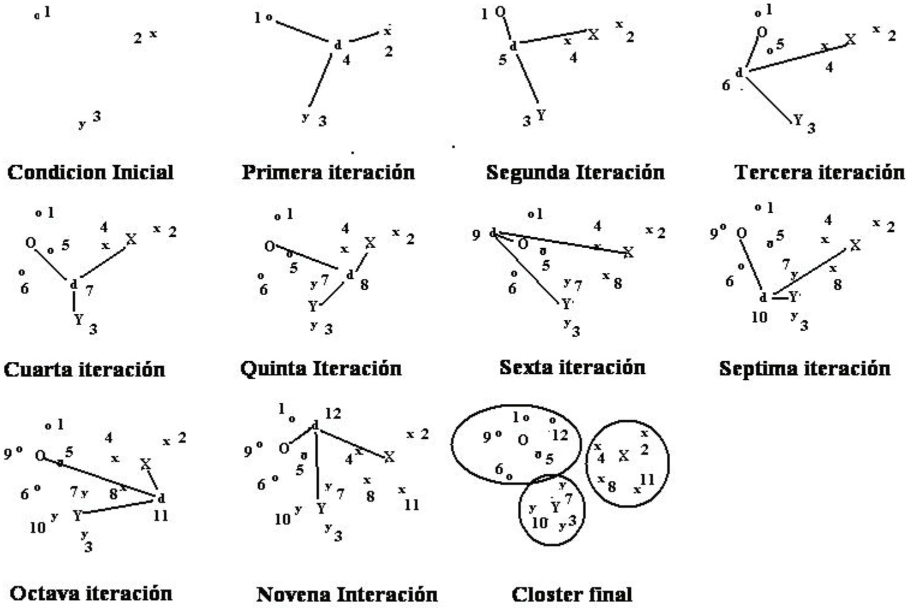
- Model initialization.
- Model development.
- Model evaluation.

2.1.1 Model initialization

Gaussian mixture models allow grouping data. The K-means algorithm is an algorithm that corresponds to a non-probabilistic limit, particular of the maximum likelihood estimation applied to Gaussian mixtures.

The problem is to identify data groups in a multidimensional space. It involves a set $x1..., x_N$ of a random variable of D-dimensions in a Euclidean space. A group can be thought of as a data set whose distance between them is small compared to the distance to points outside the group.
Introducing a set of D-dimensional vectors μ_k, with $k = 1, 2, \ldots, K$ where μ_k is the prototype associated with the k-th group. The goal is to find an assignment of the observed data to the groups, as well as a set of vectors μ_k as to minimize the sum of the squares of the distances between each point to its nearest vector μ_k.
For example, initially select the first M feature vectors as the initial centers, as shown in Figure 3, ie:

$$\mu_i = X_i \tag{1}$$



Condicion Inicial

Primera iteración

Segunda Iteración

Tercera iteración

Cuarta iteración

Quinta Iteración

Sexta iteración

Septima iteración

Octava iteración

Novena Interación

Closter final

Fig. 3. Illustration of K-Means algorithm for M3.

Then is added a vector more and get the distance between the new vector and M centers, determining that the new vector belongs to the center with which the distance is the lowest. Subsequently the new center is calculated by averaging the items belonging to the center. Thus denoting by $X_{i,j}$ the characteristic vectors belonging to the center $\mu - k$, the new center is given by:

$$\mu_k = \frac{1}{N} \sum_{j=1}^{N} X_{k,j} \tag{2}$$

This process is repeated until the distance between the $k - th$ center on the iteration n and $n + 1$ is less than a given constant.
Figure 3 shows that the first three vectors are used as initial centers. Then insert the fourth vector which has the shortest distance from the center x. Subsequently the new center is

calculated by averaging the two vectors belonging to the center X. Then is analyzed the vector 5, which has a shorter distance from the center O, which is modified by using the vectors 1 and 5, as shown in the second iteration in Figure 3. Then is analyzed the vector 6, which has a minimum distance to the center O. Here the new center O vectors are modified using $1, 5$ and 6. The process continues until ninth iteration, where the center O is calculated using the vectors $1, 5, 6, 9, 12$; the center X is calculated using the vector $2, 4, 8, 11$, while the Y is obtained $3, 7, 10$ from the vectors.
After obtaining the centers, the variance of each center is obtained using the relationship:

$$\sigma_k = \frac{1}{N} \sum_{j=1}^{N_k} \left(\mu_k - X_{k,j} \right)^2 \tag{3}$$

2.1.2 Model development

Gaussian Mixture Models(GMM) are statistical modeling methods while a model is defined as a mixture of a certain numbers of Gaussian functions for the feature vectors (Jin et al., 2004). A Gaussian mixture density is a weighted sum of M component densities , this is shown in Figure 4 and obtained by the following equation:

$$p(x|\lambda) = \sum_{i=1}^{M} p_i b_i(x) \tag{4}$$

Where x is a N-dimensional vector, $b_i(\overrightarrow{x})$,$i = 1, 2, \ldots, M$, are the components of density and p_i, $i = 1, 2, \ldots, M$, are weights of the mixtures. Each component density is a D-Gaussian variation of the form:

$$b_i(\overrightarrow{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\sigma_i|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}(x - \mu_i)' \sigma_i^{-1} (x - \mu_i) \right\} \tag{5}$$

Where $()'$ denotes the transposed vector, $\mu - i$ denotes the average value of N dimensions and σ_i covariance matrix which is diagonal, and p_i the distribution of weights which satisfy the relationship:

$$\sum_{i=1}^{M} p_i = 1 \tag{6}$$

So the distribution model is determined by the mean vector, covariance matrix and the weights of the distribution with which the model is represented as:

$$\lambda = p_i, \mu_i, \sigma_i, \qquad i = 1, 2, \ldots, M \tag{7}$$

The estimation of system parameters using the ML algorithm (Maximum Likeklihood) seeks to find the parameters to approximate the best possible distribution of the characteristics of the face under analysis and will seek to find the parameters of λ to maximize distribution. For a sequence of T training vectors $X = x_1, \ldots, x_T$, the GMM likelihood can be written as:

$$p(X|\lambda) = \prod_{t=1}^{T} p(X_t|\lambda) \tag{8}$$

Unfortunately, Equation 8 is nonlinear in relation to the parameters of λ, so to is possible to maximize directly, so it must use an iterative algorithm called Baum-Welch. Baum-Welch

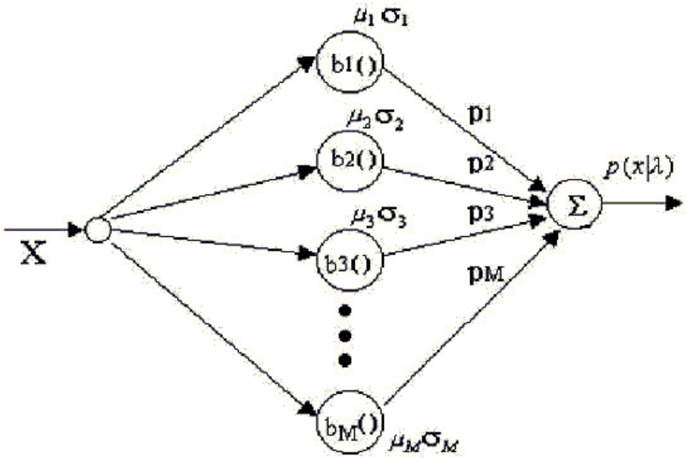


Fig. 4. Gaussian Mixture Model, GMM

algorithm is used by HMM algorithm to estimate its parameters and has the same basic principle of the algorithm of Expectation Maximization (EM Expectation-Maximization), which is part of an initial set of parameters $\lambda(r-1)$ and a new model is estimated $\lambda(r)$, where r denotes the $r-th$ iteration, so to:

$$p(X|\lambda(r)) \geq P(X|\lambda(r-1)) \tag{9}$$

Thus, this new model ($\overrightarrow{\lambda}$), becomes the initial model for the next iteration. Each T elements must update the model parameters as follows:

Pesos de la mezcla

$$p_i = \frac{1}{T}\sum_{t=1}^{T} p(i|X_{t+k}, \lambda) \tag{10}$$

Media

$$\mu_i = \frac{\sum_{t=1}^{T} p(i|X_{t+k}, \lambda) X_{t+k}}{\sum_{t=1}^{T} p(i|X_{t+k}, \lambda)} \tag{11}$$

Covarianza

$$\sigma_i = \frac{\sum_{t=1}^{T} p(i|X_{t+k}, \lambda)(X_{t+k} - \sigma_i)^2}{\sum_{t=1}^{T} p(i|X_{t+k}, \lambda)} \tag{12}$$

To calculate the posterior probability is obtained by:

$$p(i|X_{t+k}, \lambda) = \frac{p_i b_i(X_{t+k})}{\sum_{j=1}^{M} p_j b_j(X_{t+k})} \tag{13}$$

2.1.3 Model evaluation

In order to carry out the evaluation of the model considers that the system will be used to identify R people, which are represented by models $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_R$. The aim is then to find the model with maximum posterior probability for a given observation sequence. Formally, the person identified is one that satisfies the relation:

$$\widehat{R} = \arg\max Pr(\lambda_k|X), \qquad k = 1, 2, \ldots, R \tag{14}$$

Using Bayes theorem, equation 14 can be expressed as:

$$\widehat{R} = \arg\max \frac{p(X|\lambda_k)Pr(\lambda_k)}{p(X)}, \qquad k = 1, 2, \ldots, R \tag{15}$$

Assuming to the probability of each person is equally likely, then $Pr(\lambda_k) = \frac{1}{R}$ and taking into account to $P(X)$ is the same for all models of speakers, equation 15 simplifies to:

$$\widehat{R} = \arg\max p(X|\lambda_k), \qquad k = 1, 2, \ldots, R \tag{16}$$

Replacing $p(X|\lambda_k)$,

$$p(X|\lambda) = \prod_{t=1}^{T} p(X_t|\lambda_k) \tag{17}$$

in equation 16 yields:

$$\widehat{R} = \arg\max \prod_{t=1}^{T} p(X_t|\lambda_k), \qquad k = 1, 2, \ldots, R \tag{18}$$

Finally using logarithms have:

$$\widehat{R} = \arg\max \sum_{t=1}^{T} \log_{10}(p(X_t|\lambda_k)), \qquad k = 1, 2, \ldots, R \tag{19}$$

where $p(X_t|\lambda_k)$ is given by the equation 4, that is by the output of the system shown in Figure 4.

3. Support Vector Machine

The Support Vector Machine (SVM) (Vladimir & Filip, 1998) is a universal constructive learning procedure based on the statistical learning theory. Unlike conventional statistical and neural network methods, the SVM approach does not attempt to control model complexity by keeping the number of features small. Instead, with SVM the dimensionality of z-space can be very large because the model complexity is controlled independently of its dimensionality. The SVM overcomes two problems in its design: The *conceptual problem* is how to control the complexity of the set of approximating functions in a high-dimensional space in order to provide good generalization ability. This problem is solved by using penalized linear estimators with a large number of basis functions. The *computational problem* is how to perform numerical optimization in a high-dimensional space. This problem is solved by taking advantage of the dual kernel representation of linear functions.
The SVM combines four distinct concepts:

1. *New implementation of the SRM inductive principle.* The SVM use a special structure that keeps the value of the empirical risk fixed for all approximating functions but minimizes the confidence interval.

2. *Input samples mapped onto a very high-dimensional space using a set of nonlinear basis functions defined a priori.* It is common in pattern recognition applications to map the input vectors into a set of new variables which are selected according to a priori assumptions about the

learning problem. For the support vector machine, complexity is controlled independently of the dimensionality of the feature space (z-space).

3. *Linear functions with constraints on complexity used to approximate or discriminate the input samples in the high-dimensional space.* The support vector machine uses a linear estimators to perform approximation. As it, nonlinear estimators potentially can provide a more compact representation of the approximation function; however, they suffer from two serious drawbacks: lack of complexity measures and lack of optimization approaches which provide a globally optimal solution.
4. *duality theory of optimization used to make estimation of model parameters in a high-dimensional feature space computationally tractable.* For the SVM, a quadratic optimization problem must be solved to determine the parameters of a linear basis function expansion. For high-dimensional feature spaces, the large number of parameters makes this problem intractable. However, in its dual form this problem is practical to solve, since it scales in size with the number of training samples. The linear approximating function corresponding to the solution of the dual is given in the kernel representation rather than in the typical basis function representation. The solution in the kernel representation is written as a weighted sum of the support vectors.

3.1 Optimal separating hyperplane

A separating hiperplane is a linear fuction that is capable of separating the training data without error (see Fig. 5). Suppose that the training data consisting of n samples $(x_1, y_1), \ldots, (x_n, y_n), x \in \Re^d$, y $\in$ $+1, -1$ can be separated by the hypoerplane decision function

$$D(x) = (w \cdot x) + w_0 \tag{20}$$

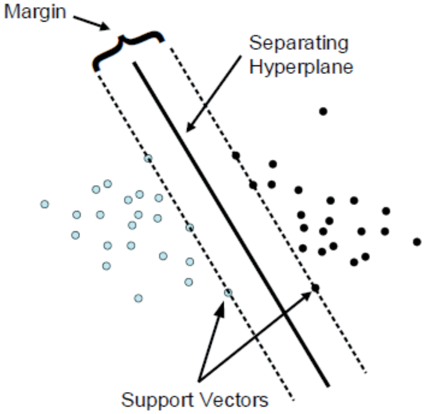


Fig. 5. Classification (linear separable case)

with appropriate coefficients w and w_0. A separating hyperplane satisfies the constraints that define the separation of the data samples:

$$\begin{aligned} (w \cdot x) + x_0 &\geq +1 \qquad \text{if } y_i = +1 \\ (w \cdot x) + x_0 &\leq -1 \qquad \text{if } y_i = -1, i = 1, \ldots, n \end{aligned} \tag{21}$$

For a given training data set, all possible separating huperplanes can be represented in the form of equation 21.

The minimal distance from the separating hyperplane to the closest data point is called the *margin* and will denoted by τ. A separating hyperplane is called *optimal* if the margin is the maximum size. The distance between the separating hyperplane and a sample x' is $|D(x')|/||w||$, assuming that a margin τ exists, all training patterns obey the inequality:

$$\frac{y_k D(x_k)}{||w||} \geq \tau, \qquad k = 1, \ldots, n \tag{22}$$

where $y_k \in -1, 1$.

The problem of finding the optimal hyperplane is that of finding the w that maximizes the margin τ. Note that there are an infinite number of solutions that differ onlu in scaling of w. To limit solutions, fix the scale on the product of τ and norm of w,

$$\tau ||w|| = 1 \tag{23}$$

Thus maximizing the margin τ is equivalent to minimizing the norm of w. An optimal separating hyperplane is one that satisfies condition (21 above and additionally minimizes

$$\eta(w) = ||w||^2 \tag{24}$$

with respect to both w and w_0. The margin relates directly to the generalization ability of the separating hyperplane. The data points that exist at margin are called the *support vectors* (Fig. 5). Since the support vectors are data points closest to the decision surface, conceptually they are the samples that are the most difficult to classify and therefore define the location of the decision surface.

The generalization ability of the optimal separating hyperplane can be directly related to the number of support vectors.

$$E_n[Errorrate] \leq \frac{E_n[Numberofsupportvectors]}{n} \tag{25}$$

The operator E_n denotes expectation over all training sets of size n. This bound is independent of the dimensionality of the space. Since the hyperplane will be employed to develop the support vector machine, its VC-dimension must be determined in order to build a nested structure of approximating functions.

For the hyperplane functions (21) satisfying the constraint $||w||^2 \leq c$, the VC-dimension is bounded by

$$h \leq \min(r^2 c, d) + 1 \tag{26}$$

where r is the radius of the smallest sphere that contains the training input vectors $(x_1, \ldots, x_n)$. The factor r provides a scale in terms of the training data for c. With this measure of the VC-dimension, it is now possible to construct a structure on the set of hyperplanes according to increasing complexity by controlling the norm of the weights $||w||^2$:

$$S_k = (w \cdot x) + w_0 : ||w||^2 \leq c_k, c_1 < c_2 < c_3 \ldots \tag{27}$$

The structural risk minimization principle prescribes that the function that minimizes the guaranteed risk should be selected in order to provide good generalization ability.

By definition, the separating hyperplane always has zero empirical risk, so the guaranteed risk is minimized by *minimizing the confidence interval*. The confidence interval is minimized

by minimizing the VC-dimension h, which according to (26) corresponds to minimizing the norm of the weights $||w||^2$. Finding an optimal hyperplane for the separable case is a quadratic optimization problem with linear constraints, as formally stated next.
Determine the w and w_0 that minimize the functional

$$\eta(w) = \frac{1}{2}||w||^2 \tag{28}$$

subject to the constraints

$$y_i[(w \cdot x) + w_0] \geq 1, \qquad i = 1, \ldots, n \tag{29}$$

given the training data $(x_i, y_i), i = 1, \ldots, n, x \in \Re^d$. The solution to this problem consists of $d + 1$ parameters. For data of moderate dimension d, this problem can be solved using quadratic programming.
For training data that cannot be separated without error, it would be desirable to separate the data with a minimal number or errors. In the hyperplane formulation, a data point is nonseparable when it does not satisfy equation (21). This corresponds to a data point that falls within the margin or on the wrong side of the decision boundary. Positive slack variables $\xi_i, i = 1, \ldots, n$, can be introdiced to quantify the nonseparable data in the defining condition of the hyperplane:

$$y_i[(w \cdot x) + w_0] \geq 1 - \xi_i \tag{30}$$

For a training sample x_i, the slack variable ξ_i is the deviation from the margin border corresponding to the class of y_i see Fig. 6. According to our definition, slack variables greater than zero correspond to misclassified samples. Therefore the number of nonseparable samples is

$$Q(w) = \sum_{i=1}^{n} I(\xi_i > 0) \tag{31}$$

Numerically minimizing this functional is a difficult combinatorial optimization problem because of the nonlinear indicator function. However, minimizing (31) is equivalent to minimizing the functional

$$Q(\xi) = \sum_{i=1}^{n} \xi_i^p \tag{32}$$

where p is a small positive constant. In general, this minimization problem is NP-complete. To make the problem tractable, p will be set to one.

3.2 Inner product kernel

The inner product kernel (H) is known a priori and used to form a set of approximating functions, this is determined by the sum

$$H(x, x') = \sum_{j=1}^{m} g_j(x) g_j(x') \tag{33}$$

where m may be infinite.
Notice that in the form (33), the evaluation of the inner products between the feature vectors in a high-dimensional feature space is done indirectly via the evaluation of the kernel H between support vectors and vectors in the input space. The selection of the type of kernel function

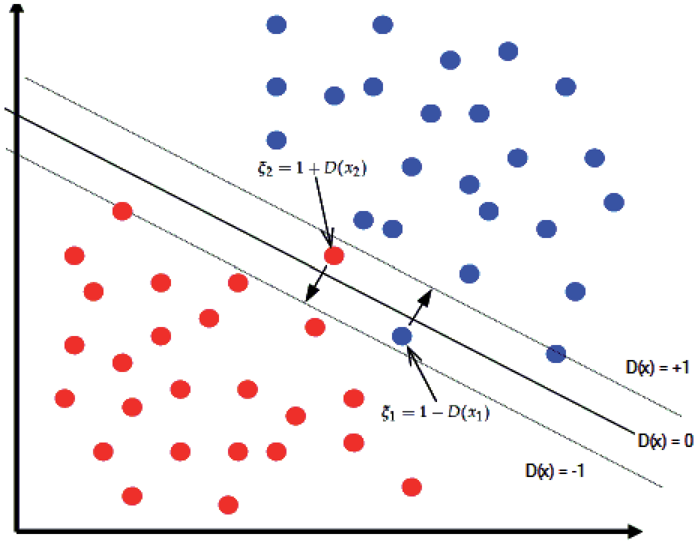


Fig. 6. Nonseparable case

corresponds to the selection of the class of functions used for feature construction. The general expression for an inner product in Hilbert space is

$$(z \cdot z') = H(x, x') \tag{34}$$

where the vectors z and z' are the image in the m-dimensional feature space and vectors x and x' are in the input space.
Below are several common classes of multivariate approximating functions and their inner product kernels:

Polynomials of degree q have inner product kernel

$$H(x, x') = [(x \cdot x') + 1]^q \tag{35}$$

Radial basis functions of the form

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i \exp\left\{-\frac{|x - x_i|^2}{\sigma^2}\right\}\right) \tag{36}$$

where σ defines the width have the inner product kernel

$$H(x, x') = \exp\left\{-\frac{|x - x'|^2}{\sigma^2}\right\} \tag{37}$$

Fourier expansion

$$f(x) = v_o + \sum_{j=1}^{q} (v_j \cos(jx) + w_j \sin(jx)) \tag{38}$$

has a kernel

$$H(x, x') = \frac{\sin(q + \frac{1}{2})(x - x')}{\sin(x - x')/2} \tag{39}$$

4. Evaluation results

Here are some results with both classifiers, GMM and SVM combined with some feature extraction methods mentioned above, like Gabor, Wavelet and Eigenphases. The results shown were performed using the database "The AR Face Database"(Martinez, 1998) is used, which has a total of 9, 360 face images of 120 people (65 men and 55 women) that includes face images with several different illuminations, facial expression and partial occluded face images with sunglasses and scarf. Two different training set are used, the first one consists on images without occlusion, in which only illumination and expressions variations are included. On the other hand the second image set consist of images with and without occlusions, as well as illumination and expressions variations. Here the occlusions are a result of using sunglasses and scarf. These images sets and the remaining images of the AR face database are used for testing.

Tables 1 and 2 shows the recognition performance using the GMM as a classifier. The recognition performance obtained using the Gabor filters-based, the wavelet transform-based and eigenphases features extraction methods are shown form comparison. Table 1 shows that when the training set 1 is used for training, with a GMM as classifier, the identification performance decrease in comparison with the performance obtained using the training set 2. This is because the training set 1 consists only of images without occlusion and then system cannot identify several images with occlusion due to the lack of information about the occlusion effects. However when the training set 2 is used the performance of all of them increase, because the identification system already have information about the occlusion effects.

	Image set 1	Image set 2
Gabor	71.43 %	91.53 %
Wavelet	71.30 %	92.51 %
Eigenphases	60.63 %	87.24 %

Table 1. Recognition using GMM

	Image set 1		Image set 2	
Average	False acceptance	False reject	False acceptance	False reject
Gabor	4.74 %	7.26 %	1.98 %	4.13 %
Wavelet	6.69 %	6.64 %	1.92 %	5.25 %
Eigenphases	37.70 %	14.83 %	21.39 %	21.46 %

Table 2. Verification using GMM

Tables 3 and 4 show the obtained results with Gabor filters, Wavelet and Eigenphases as features extractors methods in combination with SVM for identification and verification task. Also shows the same characteristics like GMM when the training set 1 and training set 2 are used for training.

Figs. 7 and 8 shows the ranking performance evaluation of Gabor, Wavelets and eigenphases feature extractions methods, using GMM for identitification and Figs. 9 and 10 shows the ranking performance evaluation of Gabor, Wavelet and Eigenphases with the Support Vector Machine for identification.

In Figs. 11-13 shows the evaluation of the GMM as verifier using different thresholds for acceptance in these graphs shows the performance of both the false acceptance and the false rejection. Showing the moment when both have the same percentage, depending on the needs

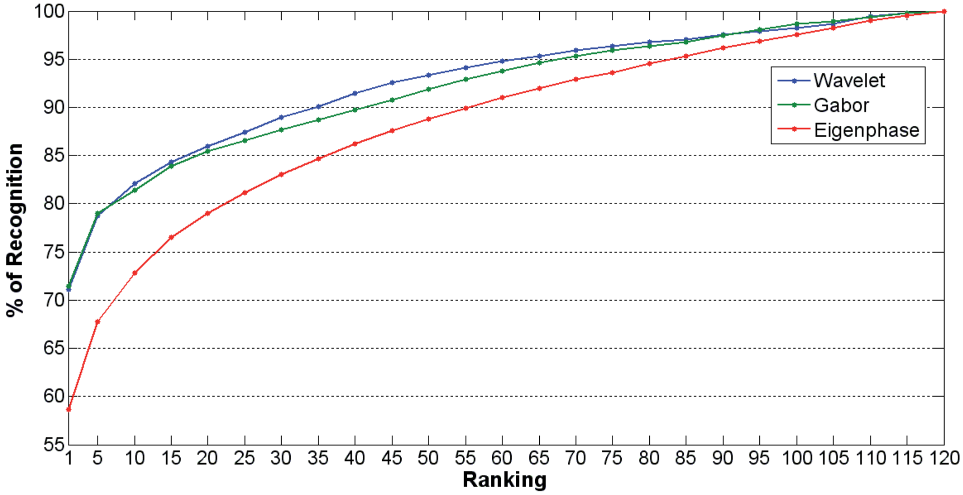


Fig. 7. Ranking performance evaluation using GMM and Training set 1.

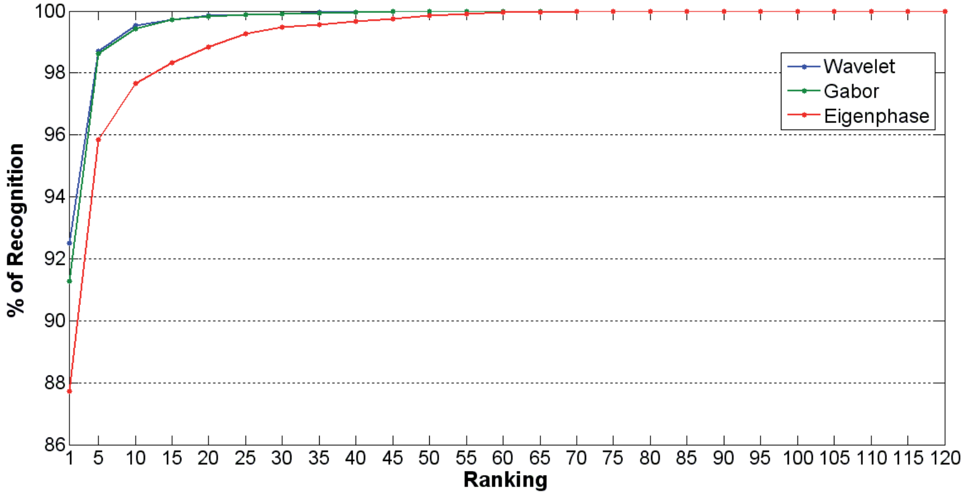


Fig. 8. Ranking performance evaluation using GMM and Training set 2.

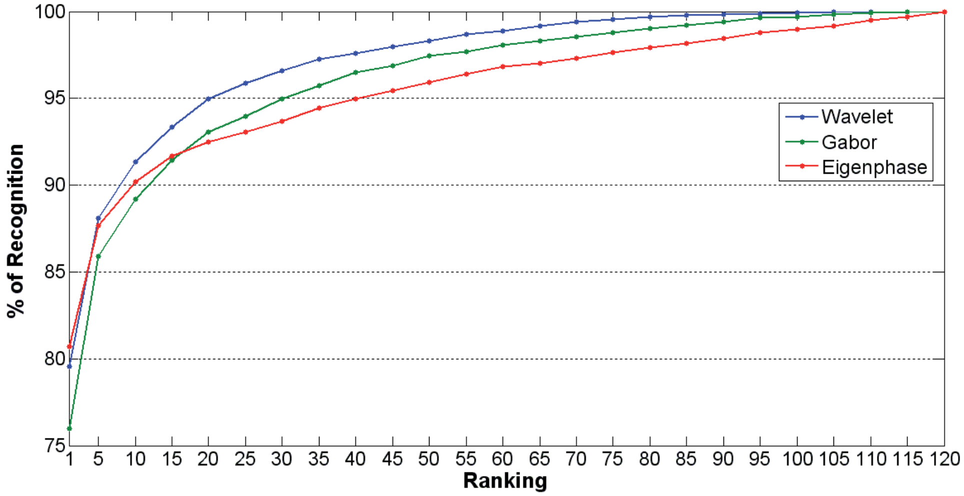


Fig. 9. Ranking performance evaluation using SVM and Training set 1.

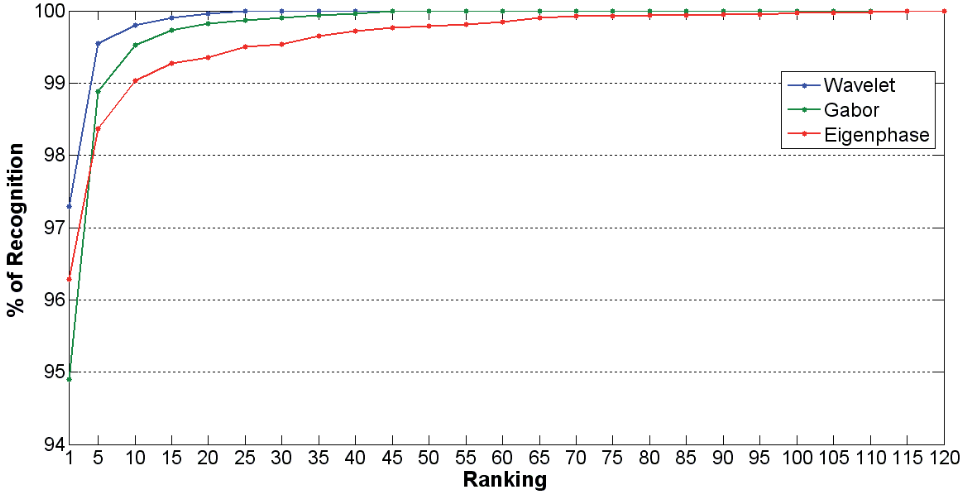


Fig. 10. Ranking performance evaluation using SVM and Training set 2.

	Image set 1	Image set 2
Gabor	75.98 %	94.90 %
Wavelet	79.54 %	97.29 %
Local 3	84.44 %	97.67 %
Local 6	81.05 %	97.29 %
Local fourier 3	85.92 %	97.92 %
Local fourier 6	85.59 %	97.85 %
Eigenphases	80.63 %	96.28 %

Table 3. Recognition using SVM

	Image set 1		Image set 2	
Average	False acceptance	False reject	False acceptance	False reject
Gabor	0.43 %	22.65 %	0.12 %	8.38 %
Wavelet	0.17 %	22.27 %	0.04 %	4.64 %
Eigenphases	0.001 %	34.74 %	0.002 %	16.04 %

Table 4. Verification using SVM

will have to choose a threshold. In Figs. 14-16 shows the evaluation of the SVM as verifier using also different thresholds.

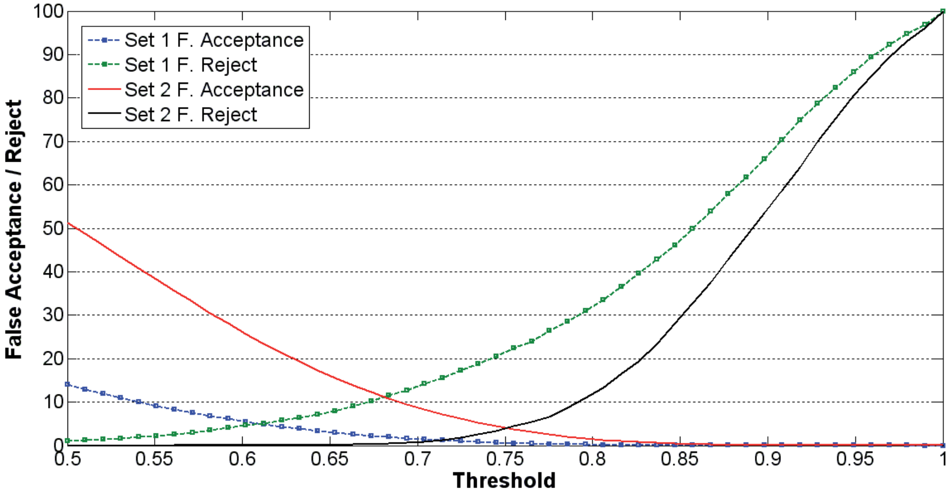


Fig. 11. Verification performance of Gabor-based feature extraction method, for several threshold values using GMM.

5. Conclusion

In this chapter presented two classifiers that can be used for face recognition, and shown some evaluation results where the GMM and SVM are used for identification and verification tasks. Two different image sets were used for training. One contains images with occlusion and the

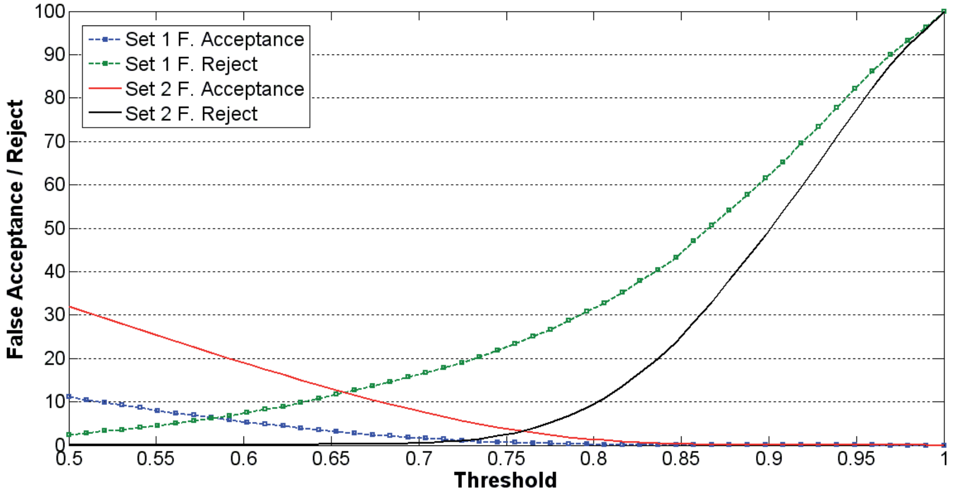


Fig. 12. Verification performance of Wavelet-based feature extraction method, for several threshold values using GMM.

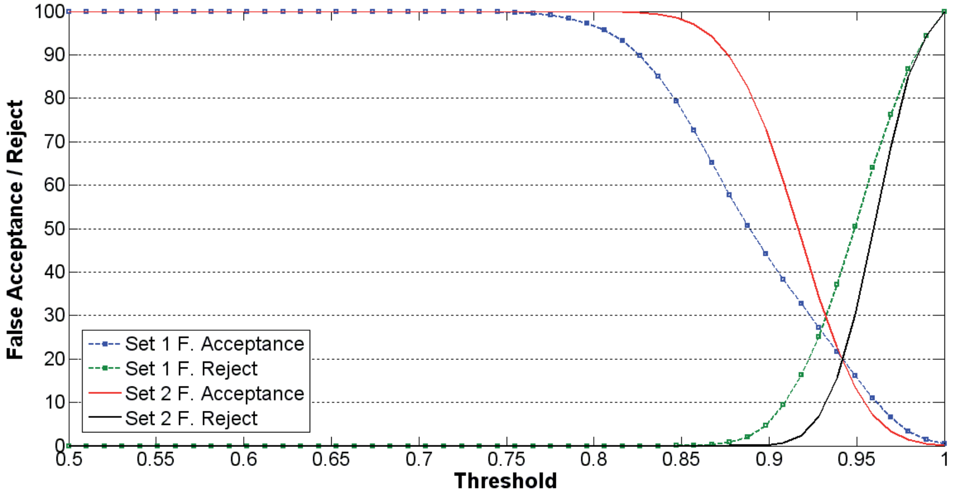


Fig. 13. Verification performance of Eigenphases feature extraction method, for several threshold values using GMM.

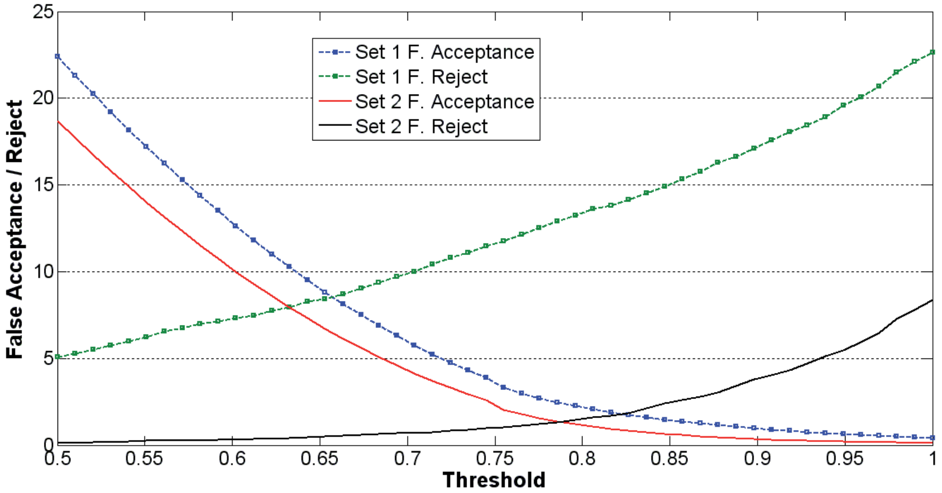


Fig. 14. Verification performance of Gabor-based feature extraction method, for several threshold values using SVM.

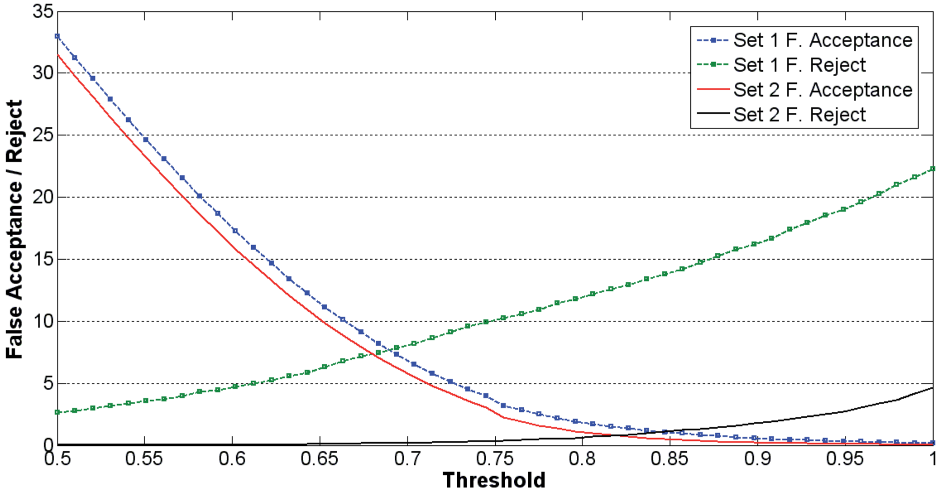


Fig. 15. Verification performance of Wavelet-based feature extraction method, for several threshold values using SVM.

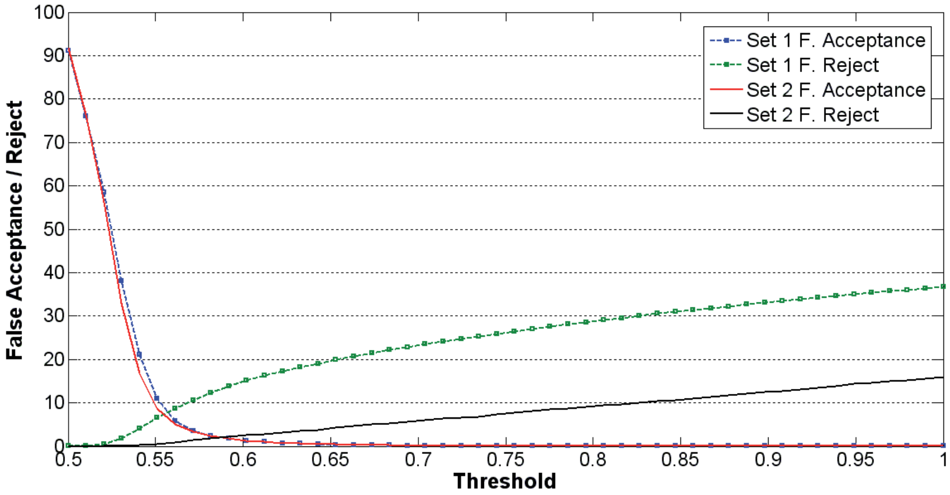


Fig. 16. Verification performance of Eigenphases feature extraction method, for several threshold values using SVM.

other one contains images without occlusions. The performance of this classifiers are shown using the Gabor-based, Wavelet-based and Eigenphases method for feature extraction.

It is important to mention, at the verification task it is very important to keep the false acceptation rate as low as possible, without much increase of the false rejection rate. To find a compromise between both errors, evaluation results of both errors with different thresholds are provided. To evaluate the performance of proposed schemes when they are required to perform an identification task the rank(N) evaluation was also estimated.

Evaluation results show that, in general, the SVM performs better that the GMM, specially, when the training set is relatively small. This is because the SVM uses a supervised training algorithm and then it requires less training patterns to estimate a good model of the person under analysis. However it requires to jointly estimating the models of all persons in the database and then when a new person is added, the all previously estimated models must be computed again. This fact may be an important limitation when the database changes with the time, as well as, when huge databases must be used, as in banking applications. On the other hand, because the GMM is uses a non-supervised training algorithm, it requires a larger number of training patterns to achieve a good estimation of the person under analysis and then its convergence is slower that those of SVM, however the GMM estimated the model of each person independently of that of the other persons in the database. It is a very important feature when large number of persons must be identified and the number of persons grows with the time because, using the GMM, when a new person is added, only the model of the new person must be added, remaining unchanged the previously estimated ones. Thus the GMM is suitable for applications when large databases must be handed and they change with the time, as in banking operations. Thus in summary, the SVM is more suitable when the size of databases under analysis is almost constant and it is not so large, while the GMM is more suitable for applications in which the databases size is large and it changes with the time.

6. References

Alvarado G.; Pedrycz W.; Reformat M. & Kwak K. C. (2006). Deterioration of visual information in face classification using Eigenfaces and Fisherfaces. *Machine Vision and Applications*, Vol. 17, No. 1, April 2006, pp. 68-82, ISSN: 0932-8092.

Bai-Ling Z.; Haihong Z. & Shuzhi S. G. (2004). Face recognition by applying wavelet subband representation and kernel associative memory. *Neural Networks, IEEE Transactions on*, Vol. 15, Issue:1, January. 2004, pp. 166 - 177, ISSN: 1045-9227.

Chellapa R.; Sinha P. & Phillips P. J.(2010), Face recognition by computers and humans. *Computer Magazine*, Vol. 43, Issue 2, february 2010, pp. 46-55, ISSN: 0018-9162.

Davies E. R. (1997). *Machine Vision: Theory, Algorithms, Practicalities*, Academic Press, ISBN: 0-12-206092-X, San Diego, California, USA.

Duda O. R; Hart E. P. & Stork G. D. (2001). *Pattern Classification*, Wiley-Interscience, ISBN: 0-471-05669-3, United States of America.

Hazem. M. El-Bakry & Mastorakis N. (2009). Personal identification through biometric technology, *AIC'09 Proceedings of the 9th WSEAS international conference on Applied informatics and communications*, pp. 325-340, ISBN: 978-960-474-107-6, World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA.

Jain, A.K.; Ross, A. & Prabhakar, S. (2004). An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, Jan. 2004, pp. 4-20, ISSN: 1051-8215.

Jin Y. K.; Dae Y. K. & Seung Y. N. (2004), Implementation and enhancement of GMM face recognition systems using flatness measure. *IEEE Robot and Human Interactive Communication*, Sept. 2004, pp. 247 - 251, ISBN: 0-7803-8570-5.

Martinez A. M. & Benavente R. (1998). The AR Face Database. CVC Technical Report No. 24, June 1998.

Olivares M. J.; Sanchez P. G.; Nakano M. M. & Perez M. H. (2007). Feature Extraction and Face Verification Using Gabor and Gaussian Mixture Models. *MICAI 2007: Advances in Artificial Intelligence*, Gelbukh A. & Kuri M. A., pp. 769-778, Springer Berlin / Heidelberg.

Reynolds D.A. & Rose R.C. (1995). Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE Transactions on Speech and Audio Processing*, Vol. 3, Issue 1, Jan. 1995, pp. 72-83, ISSN: 1063-6676.

Reynolds D.A. (2008). Gaussian Mixture Models, *Encyclopedia of Biometric Recognition, Springer*, Feb 2008, , ISBN 978-0-387-73002-8.

Rojas R. (1995). *Neural networks: a systematic introduction*, Springer-Verlag, ISBN: 3-540-60505-3, New York.

Savvides, M.; Kumar, B.V.K.V. & Khosla, P.K. (2004). Eigenphases vs eigenfaces. *Pro- ceedings of the 17th International Conference on Pattern Recognition*, Vol. 3, Aug. 2004, pp. 810-813, ISSN: 1051-4651.

Vladimir C. & Filip M. (1998). *Learning From Data: Concepts, Theory, and Methods*, Wiley Inter-Science, ISBN: 0-471-15493-8, USA.

Yoshida M.; Kamio T. & Asai H. (2003). Face Image Recognition by 2-Dimensional Discrete Walsh Transform and Multi-Layer Neural Network, *IEICE Transactions on Fundamentals of Electronics, Communications and Computer.*, Vol. E86-A, No. 10, October 2003, pp.2623-2627, ISSN: 0916-8508.

Zhao W.; Chellappa R.; Phillips P. J. & RosenfeldA. (2003). Face recognition: A literature survey. *ACM Computing Surveys (CSUR)*, Vol. 35, Issue 4, December 2003, pp. 399-459.

3

New Principles in Algorithm Design for Problems of Face Recognition

Vitaliy Tayanov
Vyacheslav Chornovil State Institute of Modern Technologies and Management of Lviv
Ukraine

1. Introduction

This chapter is devoted to two main problems in pattern recognition. First problem concerns the methodology of classification quality and stability estimation that is also known as classification reliability estimation. We consider general problem statement in classification algorithms design, classification reliability estimation and the modern methods solution of the problem. On the other hand we propose our methods for solution of such kind of a problem. In general this could be made by using of different kind of indicators of classification (classifier) quality and stability. All this should summarize everything made before with our latest results of solution of the problem. Second part of the chapter is devoted to new approach that gives the possibility to solve the problem of classifier design that is not sensitive to the learning set but belongs to some kind of learning algorithms.

Let us consider recognition process, using the next main algorithms, as some parts of the some complicated recognition system: algorithm for feature generating, algorithms for feature selection and classification algorithms realizing procedure of decision making. It is really important to have good algorithms for feature generating and selection. Methods for feature generating and selection are developed a lot for many objects to be recognized. For facial images the most popular algorithms use 3D graph models, morphological models, the selection some geometrical features using special nets, etc. From other hand very popular algorithms for feature generating and selection are those which use Principle Component Analysis (PCA) or Independent Component Analysis (ICA). PCA and ICA are good enough for a number of practical cases. However there are a lot of different deficiencies in classifier building. For classifiers using learning the most essential gap is that all such classifiers can work pretty well on learning stage and really bad for some test cases. Also they can work well enough if classes are linearly separated e.g. Support Vector Machines (SVM) in linear case. For non-linear case they have a number of disadvantages. That is why it is important to develop some approaches for algorithms building that are not sensitive for the kind of sample or complexity in classification task.

All classification algorithms built for the present could be divided almost into 5 groups: algorithms built on statistical principles, algorithms, built on the basis of potential functions and non-parametrical estimation, algorithms, using similarity functions (all kinds of metrical classifiers like 1NN and kNN classifiers) algorithms, using logical principles like decision lists, trees, etc, hierarchical and combined algorithms. A lot of recognition systems use a number of algorithms or algorithm compositions. For their optimization and tuning one uses special algorithms like boosting. These compositions can be linear or non-linear as well.

To build any effective classifier we need to use some algorithms that allow us to measure the reliability of classification. Having such algorithms we can find estimates of optimal values of classifier parameters. Accuracy of these estimates allows us to build reliable and effective classification algorithms. They perform the role of indicators of measurement of different characteristics and parameters of the classifier.

We propose the new approach for object classification that is independent of the learning set but belongs to some kind of learning algorithms. Methods, using for new classification approach concerns the field of results combination for the classification algorithms design. One of the most progressive directions in this area assumes using of the consensus in recognition results, produced by different classifiers.

Idea of usage of consensus approach is the following. One divides all objects to be recognized into three groups: objects, located near separation hyperplane (ambiguity objects), objects, located deeply inside the second class and belong to the first one (misclassified objects) and objects that are recognized correctly with large enough index of reliability (easy objects). The group of ambiguity objects is the largest one that can cause errors during recognition due to their instability. Because of that it is extremely important to detect such kind of objects. Next step will be detecting of the true class for every of ambiguity objects. For this it is planned to use apparatus of cellular automata and Markov models. It is important to mark that such an approach allows us to reduce the effect of overestimation for different recognition tasks. This is one of the most principle reasons for using such kind of algorithms.

The practical application of consensus approach for the task of face recognition could be realized by the following way. If we use one of the following classifiers e.g. 1NN, kNN, classifier, built on the basis of potential functions, SVM, etc. we can have some fiducial interval of the most likely candidates. Fiducial interval in general is the list of the candidates that are the most similar to the target object (person). If we use decision making support system controlled by operator result of the system work could be one or two candidates, given to the operator for the final decision or expertise. If we use an autonomous system the final decision should be made by the system that has to select one of the most likely candidates using some special verification algorithms that analyse the dynamics of behaviour of the object in the fiducial interval under the verifying external conditions and parameters of the algorithm.

2. Estimations building in pattern recognition

2.1 Some important tasks of machine learning

The modern theory of machine learning has two vital problems: to obtain precise upper bound estimates of the overtraining (overfitting) and ways of it's overcoming. Now the most precise familiar estimates are still very overrated. So the problem is open for now. It is experimentally determined the main reasons of the overestimation. By the influence reducing they are as follows [Vorontsov, 2004]:

1. The neglect of the stratification effect or the effect of localization of the algorithms composition. The problem is conditioned by the fact that really works not all the composition but only part of it subject to the task. The overestimation coefficient is from several tens to hundreds of thousands;

2. The neglect of the algorithms similarity. The overestimation coefficient for this factor is from several hundreds to tens of thousands. This factor is always essential and less dependent from the task than first one;
3. The exponential approximation of the distribution tail area. In this case the overestimation coefficient can be several tens;
4. The upper bound estimation of the variety profile has been presented by the one scalar variety coefficient. The overestimation coefficient is often can be taken as one but sometimes it can be several tens.

The reason of overtraining effect has been conditioned by the usage of an algorithm with minimal number of errors on the training set. This means that we realize the one-sided algorithms tuning. The more algorithms are going to be used the more overtraining will be. It is true for the algorithms, taken from the distribution randomly and independently. In case of algorithm dependence (as rule in reality they are dependent) it is suggested that the overtraining will be reduced. The overtraining can be in situation if we use only one algorithm from composition of two algorithms. Stratification of the algorithms by the error number and their similarity increasing reduces the overtraining probability.

Let us consider a duplet algorithm-set. Every algorithm can cover a definite number of the objects from the training set. If one uses internal criteria [Kapustii et al., 2007; Kapustii et al., 2008] (for example in case of metrical classifiers) there is the possibility to estimate the stability of such coverage. Also we can reduce the number of covered objects according to the stability level. To cover more objects we need more algorithms. These algorithms should be similar and have different error rate.

There is also interesting task of redundant information decrease. For this task it is important to find the average class size guaranteeing the minimal error rate. The reason in such procedure conditioned also by the class size decrease for the objects interfering of the recognition on the training phase.

The estimation of the training set reduction gives the possibility to define the data structure (the relationship between etalon objects and objects that are the spikes or non-informative ones). Also the less class size the less time needed for the decision making procedure. But the most important of such approach consists in possibility to learn precisely and to understand much deeper the algorithms overtraining phenomenon.

In this paper we are going to consider the metrical classifiers. Among all metrical classifiers the most applied and simple are the *k*NN classifiers. These classifiers have been used to build practical target recognition systems in different areas of human's activity and the results of such classification can be easily interpreted. One of the most appropriate applications of metrical classifiers (or classifiers using the distance function) concerns the biometrical recognition systems and face recognition systems as well.

2.2 Probabilistic approach to parametrical optimization of the *k*NN classifiers

The most advanced methods for optimization composition algorithm, informative training set selection and feature selection are bagging, boosting and random space method (RSM). These methods try to use the information containing in the learning sample as much as they can. Let us consider the metrical classifier optimization in feature space, using different metrics. The most general presentation of the measure between feature vectors $\mathbf{x}$ and $\mathbf{y}$ has been realized through Manhatten measure as the simple linear measure with weighted coefficients a_i [Moon & Stirling, 2000]:

$$d(x,y) = \sum_{i=1}^{n} a_i \mid x_i - y_i \mid , \tag{1}$$

where $d(x,y) = \sum_{i=1}^{n} a_i \mid x_i - y_i \mid$ is the arbitrary measure between vectors $\mathbf{x}$ and $\mathbf{y}$.

Minkovski measure as the most generalized measure in pattern recognition theory can be presented in form of

$$d(x,y) = (\sum_{i=1}^{n} \mid x_i - y_i \mid^p)^{\frac{1}{p}} = (\sum_{i=1}^{n} a_i \mid x_i - y_i \mid)^{\frac{1}{p}} = C(p) \sum_{i=1}^{n} a_i \mid x_i - y_i \mid , \tag{2}$$

where parametrical multiplier $C(p)$ have been presented in form of

$$C(p) = (\sum_{i=1}^{n} a_i \mid x_i - y_i \mid)^{\frac{1-p}{p}} ; a_i = \mid x_i - y_i \mid^{p-1} ; p > 0 \tag{3}$$

One can make the following conclusions. An arbitrary measure is the filter in feature space. It determines the weights on features. The weight must be proportional to the increase of one of indexes when it has been added to general feature set used for class discrimination procedure. Such indexes are: correct recognition probability, average class size, divergence between classes, Fisher discriminant [Bishop, 2006]. One can use another indexes, but the way of their usage should be similar. If one of the features does not provide the index increase (or worsen it) the value of such feature weight should be taken as zero. So by force of supplementary decrease of feature number one can accelerate the recognition process retaining the qualitative characteristics. The feature optimization problem and measure selection has been solved uniquely. This procedure has been realized using weighted features and linear measure with weighted coefficients. Feature selection task at the same time has been solved partially. First the feature subset from general set is determined. Such set has been determined by some algorithm (for example by the number of orthogonal transforms). Such algorithm should satisfy the definite conditions like follows: class entropy minimization or divergence maximization between different classes. These conditions have been provided by the Principle Component Analysis [Moon & Stirling, 2000]. The last parameter using in the model is the decision function or decision rule. Number of decision functions can be divided into functions working in feature space and the functions based on distance calculation. For example the Bayes classifier, linear Fisher discriminant, support vector machine etc. work in feature space. The decision making procedure is rather complex in multidimensional feature space when one uses such decision rules. Such circumstance is especially harmful for continuous recognition process with pattern series that have been recognized. Thus realizing the recognition system with large databases in practice one uses classifiers based on distance function. The simplest classifier is 1NN. But this classifier has been characterized by the smallest probability indexes. Therefore one should use *k*NN one. So the task consists in selection of *k* value that is optimal for decision making procedure in bounds of fiducial interval. This interval corresponds to the list of possible candidates. Unlike the classical approach *k* value has upper bound by class size. In classical approach the nearest neighbor value should be taken rather large, approximating Bayes classifier.

Let us consider RS with training. The calculation and analysis of the parameters of such systems is carried out on the basis of learning set. Let there exists the feature distribution in linear multidimensional space or unidimensional distribution of distances. We are going to analyse the type of such distribution. The recognition error probability for $\mu = 0$ could be presented as $\int_{|x| \geq \theta} p(x)dx$, where θ is the threshold. According to the Chebyshev inequality [Moon & Stirling, 2000] we obtain $\int_{|x| \geq \theta} p(x)dx \leq \frac{\sigma^2}{\theta^2}$.

Let us consider the case of mean and variance equality of $p(x)$ distribution. The upper bound for single mode distributions with mode $\mu = 0$ with help of Gauss inequality [Weinstein, 2011] is:

$$P(|x - \mu| \geq \lambda\tau) \leq \frac{4}{9\lambda^2} \tag{4}$$

where $\tau^2 \equiv \sigma^2 + (\mu - \mu_0)^2$.

Let $\mu = \mu_0 = 0$ and $\tau \equiv \sigma$. Then the threshold θ is $\theta = \lambda\tau = \lambda\sigma$ and $\lambda = \frac{\theta}{\sigma}$. Thus the Gauss inequality for the threshold θ could be presented in form of:

$$\int_{|x| \geq \theta} p(x)dx \leq \frac{4\sigma^2}{9\theta^2}. \tag{5}$$

As seen from (5), the Gauss upper bound estimate for the single-mode distribution is better in 2.25 times then for the arbitrary distribution. So the influence of the distribution type on the error probability is significant. The normal distribution has equal values of mode, mean and median. Also this distribution is the most popular in practice. On the other hand the normal distribution has been characterized by the maximum entropy value for the equal values of variance. This means that we obtain the minimal value of classification error probability for the normally distributed classes. For the algorithm optimization one should realize the following steps:

- to calculate the distance vector between objects for the given metric;
- to carry out the non-parametrical estimation of the distance distribution in this vector by the Parzen window method or by the support vector machines;
- to estimate the mean and variance of the distribution;
- on the basis of estimated values to carry out the standardization of the distribution ($\mu = 0$, $\sigma = 1$);
- to build the distributions both for the theoretical case and estimated one by the non-parametrical methods;
- to calculate the mean square deviation between the distributions;
- to find out the parameter space, when deviation between the distributions less then given δ level.

2.2.1 Probability estimation for some types of probability density functions

Let us consider some probability density functions (pdfs) that have a certain type of the form (presence of the extremum, right or left symmetry). If pdf have not one of such types of structure one can use the non-parametrical estimation. As the result of such estimation we get the uninterrupted curve describing pdf. This function can be differentiated and integrated by the definition. Because the Gaussians have been characterized by the minimal error of the classification for the given threshold θ and does not exceed $\frac{4\sigma^2}{9\theta^2}$ (see eq.5) for the unimodal and symmetric pdf or pdf with right asymmetry, the double-sided inequality for the given value of recognition error can be presented in form of :

$$0.5(1 - erf(\frac{\theta}{\sigma})) \le \varepsilon \le \frac{4\sigma^2}{9\theta^2} . \tag{6}$$

where $\mu = 0$.

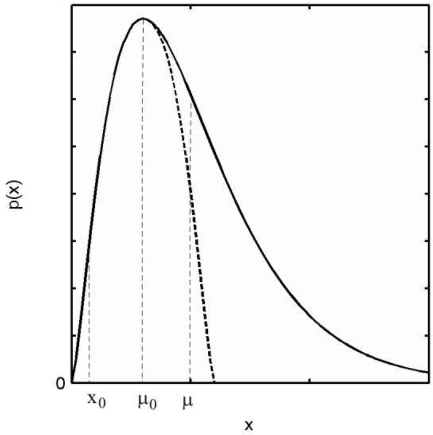


Fig. 1. Right asymmetry of pdf

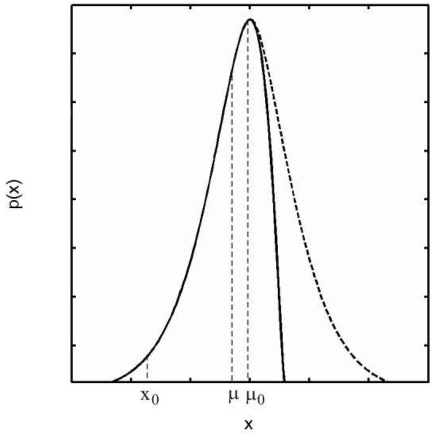


Fig. 2. Left asymmetry of pdf

Let us analyse the form of potentially generated pdfs of distances between objects. All of the distributions will have extremum. This will be conditioned by following facts. All of the pdfs have been determined on the interval $[0,\infty)$ and the density near zero and for the large distances is not high because these values are mostly unlikely. The right asymmetry is much more likely because pdf of distances is limited by zero and from the other side it has no strictly determined limitations.

Let's consider a widespread problem of classification in the conditions of two classes. We will denote the size of classes as s_1 and s_2 correspondingly. Then if the probability of replacement of object of a class having size s_1 within a fiducial interval is equal ε_1 the probability of no replacement of objects from the same class by objects from a class s_2 in this interval is equal to $(1-\varepsilon_1)^{s_2}$ under the condition of independence of objects [Kapustii et al, 2008; Kyrgyzov, 2008; Tayanov & Lutsyk, 2009]. For other class at corresponding changes this probability is equal to $(1-\varepsilon_2)^{s_1}$. If now one selects some virtual class and admits that replacement of any object of this class by objects from the mentioned two classes is authentic event it is possible to write down a following equation:

$$\gamma((1-\varepsilon_1)^{s_2}+(1-\varepsilon_2)^{s_1})=1, \tag{7}$$

where the proportionality multiplier is calculated trivially.

Sometimes there are situations when distances between objects are equal to 0. Thus non-parametrically estimated distribution of one of the classes can have a maximum in a point corresponding to zero distance. Let density of distributions are equal $p_1(0)$ and $p_2(0)$ in a zero point. The estimation of relation between probabilities can be set in a form of $\frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$ or $\ln\frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$. Thus it is necessary to make boundary transition from cumulative density function (cdf) to pdf as they are connected among themselves by differentiation operation. The relation $\ln\frac{p_1(0)^{s_2}}{p_2(0)^{s_1}}$ or generally ($\ln\frac{p_2(0)^{s_1}}{p_1(0)^{s_2}}$) can be used for construction of the following classifier

$$\begin{array}{ccc} \ln\dfrac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}} > \gamma_1; & & \ln\dfrac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}} > \gamma_2; \\ & \text{or} & \\ \ln\dfrac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}} < \gamma_1, & & \ln\dfrac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}} < \gamma_2, \end{array} \tag{8}$$

where values $\ln\frac{p_1(\theta)^{s_2}}{p_2(\theta)^{s_1}}=0$ or $\ln\frac{p_2(\theta)^{s_1}}{p_1(\theta)^{s_2}}=0$ have no influence on classification results and the decision can be accepted for benefit of any class. In case of non-parametric estimation the probability of such value is almost equal to 0. This approach is especially useful for the recognition tasks with similar objects i.e. objects that are week separated in the feature space. It should be noted that such type of algorithms have been oriented on the tasks with high level of class overlapping. Face recognition belongs to the tasks that have sufficiently a lot of objects that could not be separated so easy.

2.3 Combinatorial approach

Let us present the recognition results for *k*NN classifier in form of binary sequence:

$$\underbrace{\underbrace{1111111111}_{l_1}\underbrace{000}_{m_1}\underbrace{111111111}_{l_2}\underbrace{00000}_{m_2}}_{I_t}\underbrace{111}_{l_3}\underbrace{00}_{m_3}\ldots$$

Fig. 1. The recognition results in form of binary sequence for *k*NN classifier

Using *k*NN classifier it is important that among *k* nearest neighbours we have the related positive objects majority or the absolute one. Let us consider the simpler case meaning the related majority. The *k*NN classifier correct work consists in fact that for *k* nearest neighbours it has to be executed the condition

$$\left|\bigcup_i \tilde{l}_i\right| > \left|\bigcup_i \tilde{m}_i\right|,\ i = 1,2,3\ldots, \tag{9}$$

where $\tilde{l}_i$, $\tilde{m}_i$ are the groups that appear after class size decrease. Under the group one understands the homogeneous sequence of elements. In such sequence (see Fig.1) there exist patterns of all classes. In general case there is no direct conformity between the group number and the class number although.

Let us consider the case of non-pair *k* value in *k*NN classifier only. This means that we have the case of synonymous classification. Such univocacy could disappear in case of pair *k* value and votes equality for different classes.

Let us estimate the effect of class size reduction in case of *k*NN classifier. Note that reduced class sizes are equal to each other and equal s^*. Let us consider the *k*NN classifier correct work condition: $ENT\left(\frac{k}{2}\right) + 1 \le s^*$. In contradistinction to 1NN classifier there is no such an importance of the first nearest patterns of the true class. Thus all such sequences one could denote as l_i. Let us determine the probabilities that it will be selected s^* patterns from the true class by the combinatorial approach. These probabilities have fiducial sense. This means that for the given part of positive objects there will be no selections among the patterns of the false classes by the correspondent combinatorial way. The multiplication of pointed two probabilities determines the probability of *k*NN classifier correct work. Let assign q_j as the recognition error probability for the corresponding m_i groups:

$$\begin{aligned}
&q_1 = P\left(\inf\left(\left|\bigcup_i m_i\right|\right) \ge ENT\left(\frac{k}{2}\right)+1\right);\\
&q_2 = P\left(\inf\left(\left|\bigcup_i m_i\right|\right) + \left|m_{i+1}\right| \ge ENT\left(\frac{k}{2}\right)+1\right);\\
&q_3 = P\left(\begin{array}{l}\inf\left(\left|\bigcup_i m_i\right|\right) + \left|m_{i+1}\right| + \left|m_{i+2}\right| \ge \\ \ge ENT\left(\frac{k}{2}\right)+1\end{array}\right);\ldots\\
&q_j = P\left(\begin{array}{l}\inf\left(\left|\bigcup_i m_i\right|\right) + \left|\bigcup_j m_{i+j-1}\right| \ge \\ \ge ENT\left(\frac{k}{2}\right)+1\end{array}\right);\ldots
\end{aligned} \tag{10}$$

The combinatorial expression for q_j probability could be written in form of:

$$q_j = \frac{\sum\limits_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^{j}_{\left|\bigcup_{i,j} m_{i+j-1}\right|} C^{s^*-j}_{s-\left|\bigcup_{i,j} m_{i+j-1}\right|}}{C_s^{s^*}}, \left|\bigcup_{i,j} m_{i+j-1}\right| \ge ENT\left(\frac{k}{2}\right)+1 . \tag{11}$$

The fiducial probability for arbitrary true pattern sequence is equal:

$$P_{q_j} = \frac{\sum\limits_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^{j}_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{s-\left|\bigcup_i l_i\right|}}{C_s^{s^*}} . \tag{12}$$

Thus the correct recognition probability for *k*NN classifier has been determined by probability (12) and addition to probability (11):

$$\begin{aligned}
P_j = P_{q_j}(1-q_j) = {}& \frac{\sum\limits_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^{j}_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{s-\left|\bigcup_i l_i\right|}}{C_s^{s^*}} - \\
& \frac{\left(\sum\limits_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^{j}_{\left|\bigcup_i l_i\right|} C^{s^*-j}_{s-\left|\bigcup_i l_i\right|}\right)\left(\sum\limits_{j=ENT\left(\frac{k}{2}\right)+1}^{s^*} C^{j}_{\left|\bigcup_{i,j} m_{i+j-1}\right|} C^{s^*-j}_{s-\left|\bigcup_{i,j} m_{i+j-1}\right|}\right)}{\left(C_s^{s^*}\right)^2} .
\end{aligned} \tag{13}$$

It is modelled the recognition process with different sequences of patterns of true and false classes for the 1NN and *k*NN classifiers in case of absolute majority. For modelling the face recognition system has been taken. The class size (training set) has been taken as 18 according to the database that it was made. On the Fig.1 the results of modelling of the training set decrease influence on the recognition results for the 1NN classifier have been presented. On the Fig.2 the similar results for *k*NN classifier under condition $ENT\left(\frac{k}{2}\right)+1=s^{*}$ have been presented.

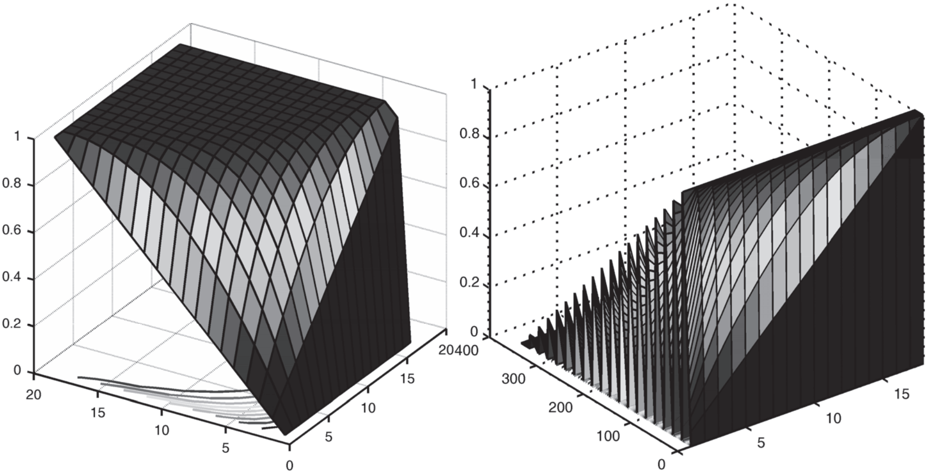


Fig. 2. The probability of correct recognition as function of training set (x axis) and number of true/false objects in the target sequence (y axis) for the 1NN classifier

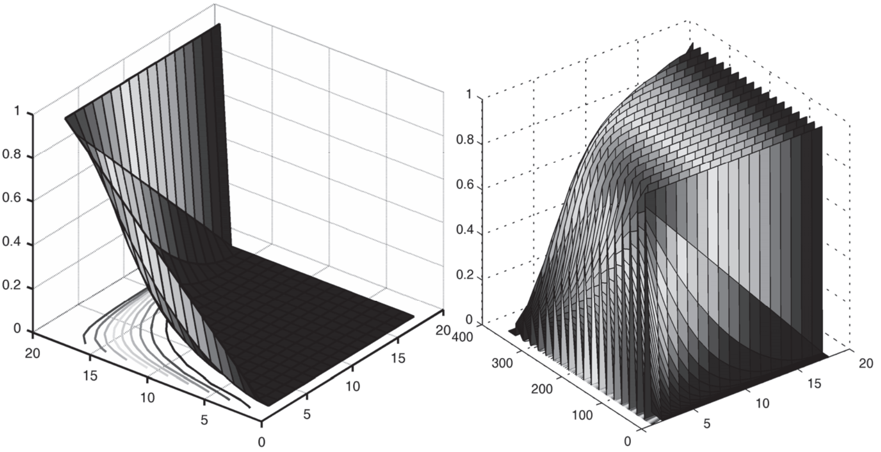


Fig. 3. The probability of correct recognition as function of training set (x axis) and number of true/false objects in the target sequence(y axis) for the *k*NN classifier

On the Fig.1,2 x axis means the size of the training set and the y axis means the size of the true patterns sequence (left picture) and sequence of both true and false patterns (right picture). The y axis has been formed by the following way. We organized 2 cycles where we changed the number of true and false patterns. For every combination of these patterns and different class sizes we calculate the probability of correct recognition.

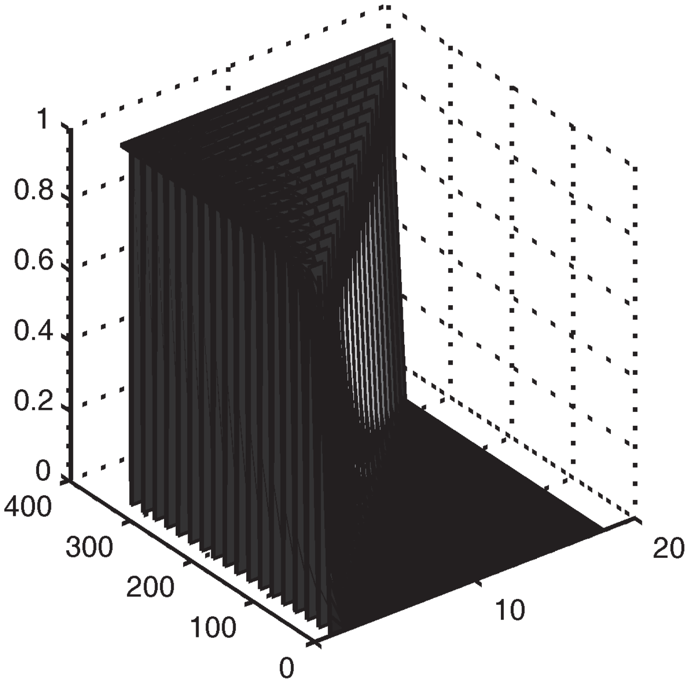


Fig. 4. The probability of correct recognition as function of training set (x axis) and $ENT\left(\frac{k}{2}\right)+1$ value (y axis)

On the Fig.4,5 the results of *k*NN classifier modelling have been presented. Here it has been satisfied the following condition: $ENT\left(\frac{k}{2}\right)+1 \le s^*$. On the Fig.5 the fiducial probability as function of training set size (x axis) and $ENT\left(\frac{k}{2}\right)+1$ value (y axis).

The probability part of proposed approach is based in following idea. Despite of combinatorial approach, where the recognition results were determined precisely, we define only the probability of the initial sequence existence. Due to low probability of arbitrary sequence existing (especially for the large sequences) it has been determined the probability of homogeneous sequences existing of the type $\{0\}$ or $\{1\}$. This probability has been determined on the basis of the last object in given sequence as probability of replacing this object (the object from the true class $\{1\}$ by the others objects of the false classes from the database. This means that the size of homogeneous sequence has been determined by the most "week" object in the homogeneous pattern sequence. The probability of existing of the non-homogeneous sequences is inversely proportional to the $2^{|l+m|}$ value, where $|l+m|$ is the sequence size. This procedure could be realized using distribution function (fatigue function) of the distances between the objects. This approach has been developed for metrical classifiers and classifiers on the basis of distance function in [Kapustii et al., 2008; Tayanov & Lutsyk, 2009]. Thus we need to calculate the probability of sequence with true patterns existing that has definite size or for the given probability rate we need to calculate the maximal size of the sequence that satisfies this probability. For the binary sequence the sum of the weights of the lower order bits is always less than the next most significant bit.

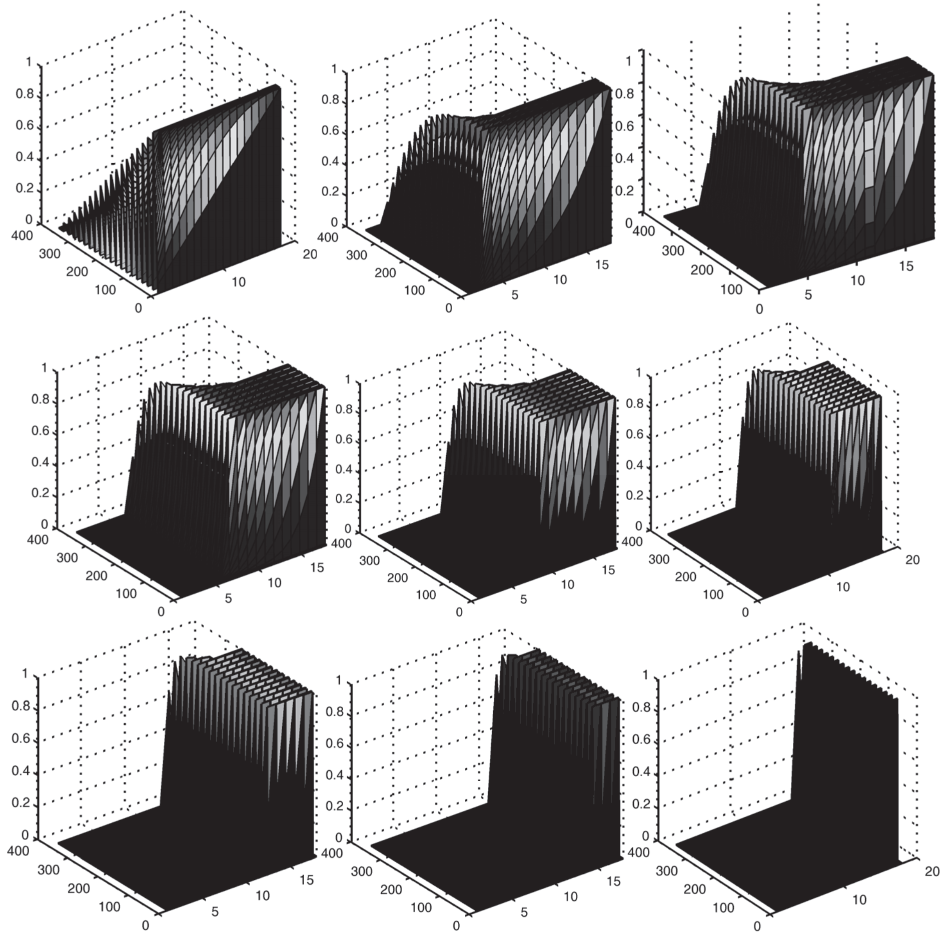


Fig. 5. The probability of correct recognition as function of $ENT\left(\dfrac{k}{2}\right)+1$ (x axis) and number of true/false objects in the target sequence (y axis)

The difference is equal to 1. This means that arbitrary pattern replacement of the true class in the fiducial interval is equivalent to the alternate replacement of the previous ones. The minimal whole order of the scale of notation that has such peculiarity is equal to 2. Thus we need to calculate the weights of the true patterns position and compare them with binary digit. Such representation of the model allows us to simplify the probability calculation of the patterns replacement from the true sequences by the patterns of false classes. On the other side the arbitrary weights can be expressed through the exponent of number 2 that also simplifies the presentation and calculation of these probabilities. So the probability of the homogeneous sequence of the true patterns existence has been calculated on the basis of distance distribution function and is the function of the algorithm parameters. We should select the sequence of the size that has been provided by the corresponding probability. We after apply the combinatorial approach that allows us to calculate the influence effect of the class size decrease on the recognition probability rate. Thus the probabilistic part of the given approach has been determined by the recognition algorithm parameters. So the integration of both probabilistic and combinatorial parts allows us to define more precise the influence of the effect of the training set reduction.

Let us consider step by step the example of fast computing of replacement of true pattern probability from the sequence where relation between weights of the objects is whole exponent of number 2. Thus for example the weights can be presented by the following way: $w = \{2^9,\ 2^6,\ 2^4, 2^3, 2^2, 2^1, 2^0\}$. As known the probability of replacement of true object from the sequence by the false one when it is known that replacement is true event is inversely proportional to the weights of these objects. Let define the probability of replacement of the object having the 2^9 weight comparatively to the object with 2^6 weight. As far as we do not know what object has been replaced the total weight of the fact that there will not be replaced the objects with 2^6 weight and lower is equal: $w = \{2^9,\ 2^6,\ 2^4, 2^3, 2^2, 2^1, 2^0\}$. This weight can be expressed trough 2^6 weight accurate within 1 by following way: $2^6(1+0.5) = 1.5 * 2^6$. In case of large sequences this one has week influence on the accuracy. The relation between 2^9 and 2^6 is equal to 8. In case of divisible group of events we obtain the $8\lambda + 1.5\lambda = 1$ equation, where the proportional coefficient λ approximately equal to 0.11. So the probability of non-replacement of the object with 2^9 weight is equal $8 * 0.11 = 0.88$. The object with 2^6 weight has the corresponding probability equal to $1 - 0.88 = 0.12$. Since we know exactly that replacement is the true event and the last object has weight equal to 1 the accuracy correction that equal to 1 makes the appropriate correction of probability calculation.

3. Classification on the basis of division of objects into functional groups

Algorithms of decision making are used in such tasks of pattern recognition as supervised pattern recognition and unsupervised pattern recognition. Clustering tasks belong to unsupervised pattern recognition. They are related to the problems of cluster analysis. Tasks where one provides the operator intervention in the recognition process belong to the learning theory or machine learning theory. The wide direction in the theory of machine learning has the name of statistical machine learning. It was founded by V.Vapnik and Ja. Chervonenkis in the sixties-seventies of the last century and continued in nineties of the same century and has the name of Vapnik-Chervonenkis theory (VC theory) [Vapnik, 2000].

It should be noted that classification algorithms built on the basis of training sets are mostly unstable because learning set is not regular (in general). That is why it has been appeared the idea of development of algorithms that partially use statistical machine learning but have essentially less sensitivity to irregularity of the training sets.

This chapter focuses on tasks that partially use learning or machine learning. According to the general concept of machine learning a set is divided into general training and test (control) subsets. For the training subset one assumes that the class labels are known for every object. Using test subsets one verifies the reliability of the classification. The reliability of algorithms has been tested by methods of cross-validation [Kohavi, 1995; Mullin, 2000].

Depending on the complexity of the classification all objects can be divided into three groups: items that are stable and are classified with high reliability ("easy" objects), objects belonging to the borderline area between classes ("ambiguous" objects) and objects belonging to one class, and deeply immersed inside another one ("misclassified" objects). Among those objects that may cause an error the largest part consists of terminal facilities. Therefore it is important to develop an algorithm that allows one to determine the largest number of frontier facilities. The principal idea of this approach consists in preclassification of objects by dividing them into three functional groups. Because of this it is possible to achieve much more reliable results of classification. This could be done by applying the appropriate algorithms for every of obtained groups of objects.

3.1 The most stable objects determination

The idea of the model building is as follows. The general object set that have to be classified is divided on three functional groups. To the first group of objects the algorithm selects the objects with high level of classification reliability. The high level of reliability means that objects are classified correctly under the strong (maximal) deviations of the parameters from optimal ones. From the point of view of classification complexity these objects belongs to the group of so called "easy" objects. The second group includes objects, on which there is no consensus. If one selects two algorithms in a composition of algorithms, they should be as dissimilar as possible and they should not be a consensus. If one uses larger number of algorithms, the object belongs to the second group if there is no consensus in all algorithms. If consensus building uses intermediate algorithms, parameters of which are within the intervals between the parameters of two the most dissimilar algorithms, this makes it impossible to allocate a larger number of objects, on which there is no consensus. Dissimilarity between algorithms is determined on the basis of the Hamming distance between results of two algorithms defined as binary sequences [Kyrgyzov, 2008; Vorontsov, 2008]. In practice this also means that in general it will not be detected the new objects, if one uses composition of more than two algorithms, on which one builds the consensus. The third group consists of those objects, on which both algorithms have errors, while they are in consensus. The error caused by these objects can not be reduced at all. Thus the error can not be less than the value determined by the relative amount of objects from the third group. The next step will be the reclassification of the second group of objects. This special procedure allows us to determine the true class, to which a particular object belongs to. Reclassifying the second group of objects we can also have some level of error. This error together with the error caused by the third group will give the total error of all proposed algorithms.

The research carried out in this paper concerns the analysis of statistical characteristics of the results of a consensus generating by two algorithms. The objective of the task analysis is a statistical regularity of characteristics of various subsets taken by division of the general set into blocks of different size. Probability distribution by the consensus for three groups of objects has been carried out by nonparametric estimation using Parzen window with Gauss kernels.

3.1.1 Experimental results

Figs 6-11 show the parametrically estimated pdfs for the probability of a correct consensus, the probability of incorrect consensus and the probability that consensus will not be reached.

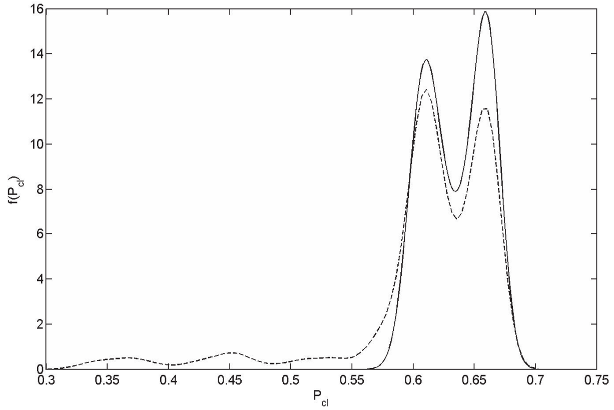


Fig. 6. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of the correct consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly) .

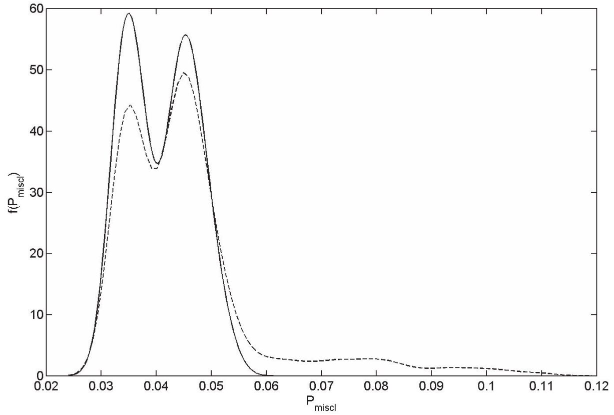


Fig. 7. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of the incorrect consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

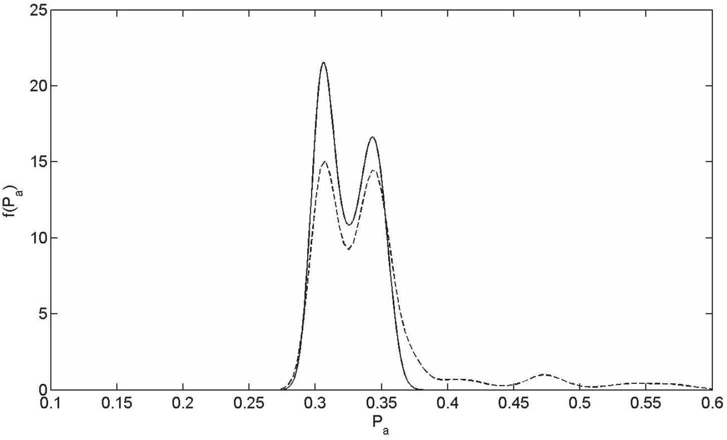


Fig. 8. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of no consensus between two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

As can be seen from the figures these distributions can be represented using one-component, two-component or multicomponent Gauss mixture models (GMM). In multicomponent GMM weights determined according to their impact factors. Distribution parameters (mean and variance) and weights of impact in the model are estimated using EM algorithm. Estimation of corresponding probability values was carried out by blocks with a minimal size of $Q = 30$ and $Q = 200$ elements. The size of these blocks has been driven by a small sample size which according to various criteria ranges from 30 to 200 items. According to the standard definition of a small sample it is assumed that sample is small when it is characterized by irregular statistical characteristics.

As seen from all figures the estimates obtained by blocks with a minimal size of 30 elements and some more are irregular. This means that for these tasks the sub-sample size of 30 items and some more is small. This has been indicated by long tails in the corresponding probability distributions. The maximum in zero point for two-component model is characterized by a large number of zero probabilities. This can be possible if there are no mistakes in the consensus of two algorithms. Estimates of probabilities on the basis of average values and the corresponding maximum probability distributions (for maximum likelihood estimation (MLE)) are not much different, which gives an additional guarantee for the corresponding probability estimates. Significance of obtained consensus estimates of probabilities of correct consensus, incorrect consensus and probability that consensus will not be achieved, provides a classification complexity estimate. Problems and algorithms for the complexity estimation of classification task is discussed in [Basu, 2006]. For example, tasks "pima" and "bupa" are about the same level of complexity because values of three probabilities are approximately equal. Tab. 1 shows that all algorithms excepting proposed new one have large enough sensitivity to the equal by the classification complexity tasks they work with. Mathematical analysis of composition building of algorithms has been considered in details in [Zhuravlev, 1978].

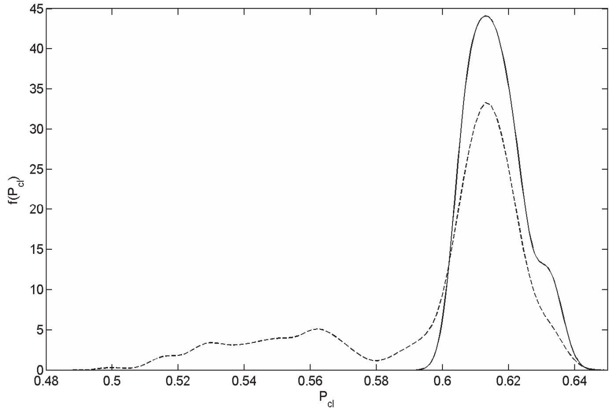


Fig. 9. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of the correct consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

Figs 6-11 show graphic dependencies of consensus results for problems taken from repository UCI. This repository is formed at the Irvin's University of California. The data structure of the test tasks from this repository is as follows. Each task is written as a text file where columns are attributes of the object and rows consist of a number of different attributes for every object. Thus the number of rows corresponds to the number of objects and the number of columns corresponds to the number of attributes for each object. A separate column consists of labels of classes, which mark each object. A lot of data within this repository has been related to biology and medicine. Also all these tasks could be divided according to the classification complexity. In the data base of repository there exists a number of tasks with strongly overlapped classes. Some of them will be used for research.

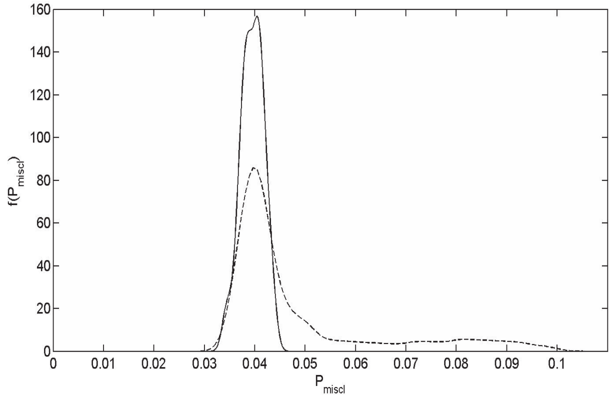


Fig. 10. Task "bupa" from UCI repository: non-parametrical estimation of the pdf of the incorrect consensus that consists of two algorithms (solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

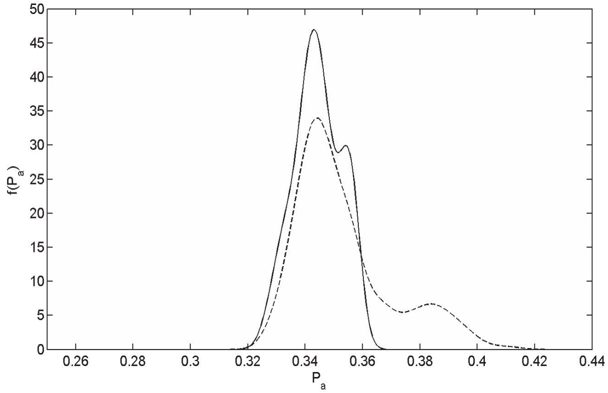


Fig. 11. Task ″bupa″ from UCI repository: non-parametrical estimation of the pdf of no consensus between two algorithms(solid line is used for the set of 200 objects and dot line is used for set of 30 objects correspondingly)

In Tab. 1 one gives the probabilities of errors obtained on the test data for different classifiers or classifier compositions. All these algorithms were verified on two tasks that are difficult enough from the classification point of view. For the proposed algorithm it has been given the minimal and maximal errors that can be obtained on given tested data.

In Tab. 1 the value of minimal error is equal to consensus error for the proposed algorithm. The value of maximal error has been calculated as sum of minimal error and the half of the related amount of objects, on which there is no consensus (fifty-fifty principle). As seen from the table the value of maximal error is much less than the least value of error of all given algorithms for two tasks from UCI repository. In comparison with some algorithms given in the table the value of minimal error is approximately 10 times less for the proposed algorithm then the error of some other algorithms from the table. The proposed algorithms are characterized by much more stability of the classification error in comparison with other algorithms. It can be seen from corresponding error comparison for two tasks from the UCI repository.

Algorithm	Task	bupa	pima
Monotone (SVM)		0.313	0.236
Monotone (Parzen)		0.327	0.302
AdaBoost (SVM)		0.307	0.227
AdaBoost (Parzen)		0.33	0.290
SVM		0.422	0.230
Parzen		0.338	0.307
RVM		0.333	-
Proposed algorithm (min/max)		0.040/0.212	0.041/0.203

Table 1. Error of classification for different algorithms

	Q=200		Q=30	
	μ	σ	μ	σ
P_c	0.635	0.024	0.611	0.064
P_e	0.041	0.006	0.046	0.013
$P_{\bar{c}}$	0.324	0.019	0.344	0.052

Table 2. Task "pima" from UCI repository

	Q=200		Q=30	
	μ	σ	μ	σ
P_c	0.635	0.024	0.611	0.064
P_e	0.041	0.006	0.046	0.013
$P_{\bar{c}}$	0.324	0.019	0.344	0.052

Table 3. Task "bupa" from UCI repository

In tabs 2-3 the estimates of probability of belonging of every object from the task of repository UCI to every of three functional groups of objects have been given. In this case the objects, on which consensus of the most dissimilar algorithms exists (P_c), belong to the class of so called "easy" objects. Then objects, on which both of algorithms that are in consensus make errors (P_e), belong to the class of objects that cause uncorrected error and this error can not be reduced at all. The last class of objects consists of objects, on which there is no consensus of the most dissimilar algorithms ($P_{\bar{c}}$). This group of objects also belongs to the class of border objects. In the tables one gives variances of corresponding probabilities too. Minimal size of the blocks, on which one builds estimates using algorithms of cross-validation changes from 30 to 200.

3.1.2 Case of three classifiers in the consensus composition

In the previous case we analysed the classifier composition that consists of two the most dissimilar algorithms. Now we are going to build the classifier composition that consists of three algorithms. The third algorithms we choose considering the following requirements. These algorithms have to be exactly in the middle of two the most dissimilar algorithms. This means that the Hamming distance between the third algorithm and one of the most dissimilar algorithms is equal to the distance between the "middle" algorithm and the second algorithm in the consensus composition of two the most dissimilar algorithms. In Tabs 4 and 5 the results of comparison of two consensus compositions have been given. The first composition consists of two algorithms and the second one consists of three algorithms correspondingly. As in the previous case we used "pima" and "bupa" testing tasks from UCI repository.

	consensus of two classifiers		consensus of tree classifiers	
	μ	σ	μ	σ
P_c	0.635	0.024	0.607	0.021
P_e	0.041	0.006	0.0347	0.006
$P_{\bar{c}}$	0.324	0.019	0.358	0.017

Table 4. Task ”pima” from UCI repository

As seen from the both tabs there is no big difference between two cases. Consensus of two algorithms can detect a bit lager quantity of correctly classified objects that means a bit more reliable detection of correctly classified objects. Consensus of three algorithms can detect a bit larger quantity of objects on which we have no consensus (the third group of objects). But if we will use the “fifty-fifty” principle for detection objects from the third group the general error of classification will be the same. We can also note that the variances of two consensuses compositions have no large differences between each other.

	consensus of two classifiers		consensus of tree classifiers	
	μ	σ	μ	σ
P_c	0.616	0.008	0.586	0.012
P_e	0.040	0.002	0.037	0.002
$P_{\bar{c}}$	0.344	0.008	0.377	0.013

Table 5. Task ”bupa” from UCI repository

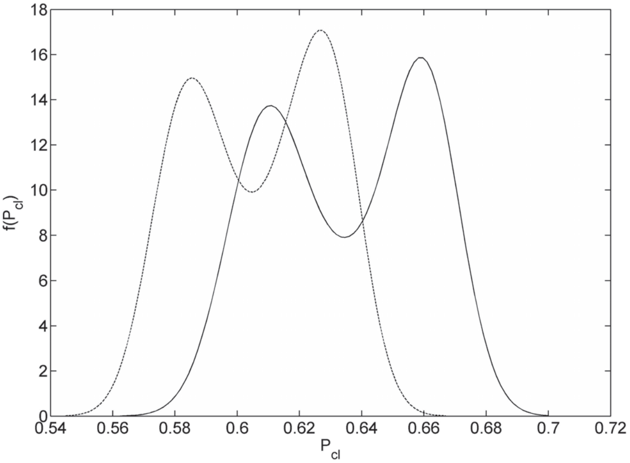


Fig. 12. Task ”bupa” from UCI repository: non-parametrical estimation of the pdf of correct consensus between two (solid line) and tree (dot-line) algorithms

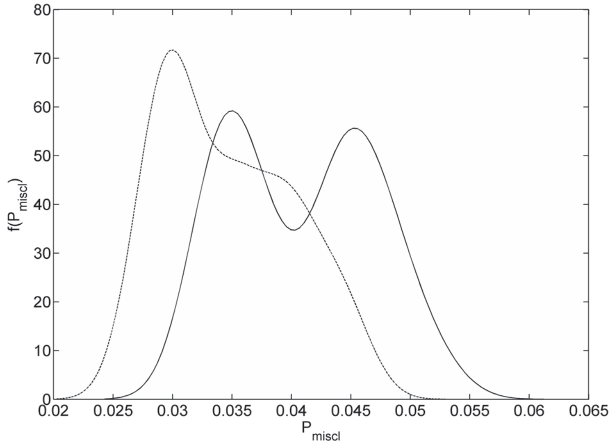


Fig. 13. Task ″bupa″ from UCI repository: non-parametrical estimation of the pdf of incorrect consensus between too (solid line) and tree (dot-line) algorithms

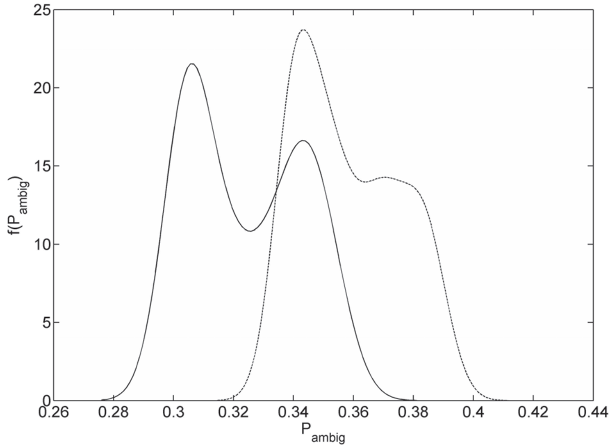


Fig. 14. Task ″bupa″ from UCI repository: non-parametrical estimation of the pdf of no consensus between too (solid line) and tree (dot-line) algorithms

On figs 12-17 the results of consensus building for three algorithms have been given. Here we also use two tasks from the UCI repository as in the case of two algorithms. According to figs corresponding to the task of "bupa" we can make the following conclusions. In comparison to the case of two algorithms we can see that for the number of preclassified groups of objects we have just shifts between the corresponding pdfs and the form of curves is approximately the same. We can also note that relative value of the shift is rather small (about 5% for the pdf of correct probability). This shift is almost conditioned by the statistical error of determining of the most different algorithms.

According to figs corresponding to the task of "pima" we can mark that differences in forms of pdfs are more essential than in previous task. This circumstance could be used for comparison of the task complexity using the value of overtraining as stability to learning. Using such approach it is possible to obtain much more precise and informative estimations of the complexity from the learning process point of view.

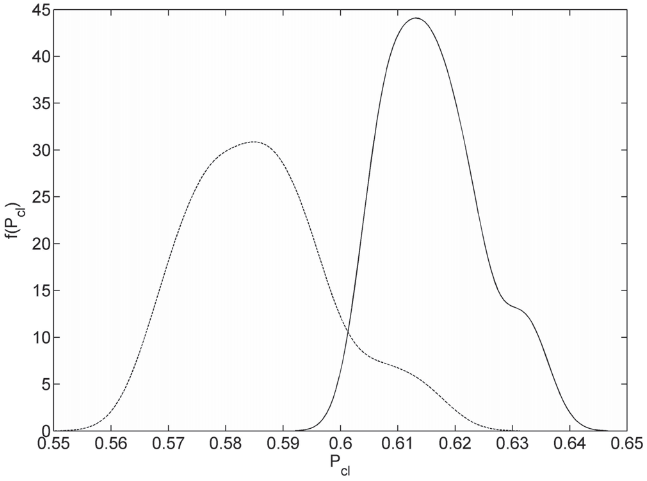


Fig. 15. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of correct consensus between too (solid line) and tree (dot-line) algorithms

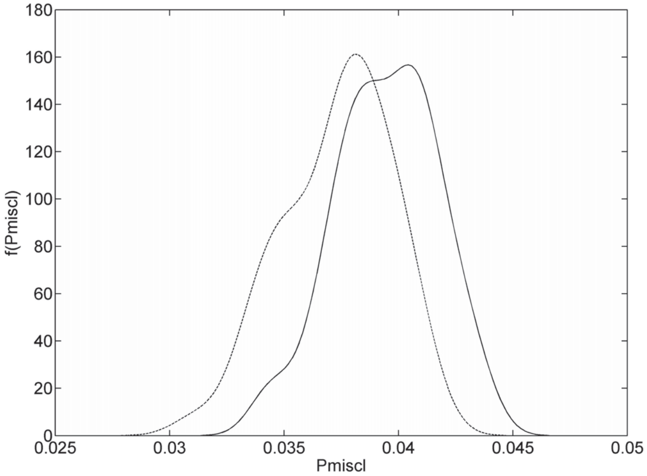


Fig. 16. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of incorrect consensus between too (solid line) and tree (dot-line) algorithms

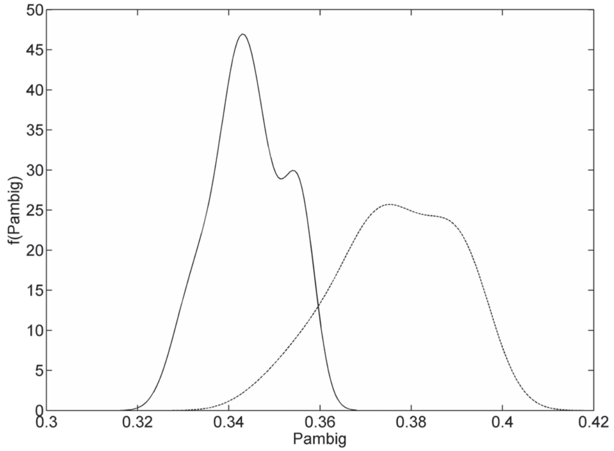


Fig. 17. Task ″pima″ from UCI repository: non-parametrical estimation of the pdf of no consensus between too (solid line) and tree (dot-line) algorithms

4. Specific of usage of the proposed approach for problems of face recognition

The problem of face recognitions is one of the principle tasks of the large project connected with determining of human behaviour and psychoanalysis of the human, based on the face expression and body movement. Such type of systems belongs to the class of no contact systems. Unlike to the human recognition systems based on fingerprints or images of iris these systems do not require human to keep finger of eyes near (or on) the scanner. This is also very important from law point of view. It is impossible to force a human to put the finger on the scanner if he does not want to do this and if this is not a criminal case. The same fact is concerned the case of iris recognition systems. To take a picture of somebody is not forbidden and this or that person could not be familiar with the fact that somebody took already picture of the face of such person. This is really important when creating the training and test databases. Face recognitions systems can be joined with hidden video cameras installed in shops, supermarkets, banks and other public places. Here it is important to hide the fact of video surveillance. This could be done with help of no contact recognition systems only. On other hand the facial information and mimicry could be used for the human behaviour determination and psychophysical state of the human. This is important to avoid and predict of acts of terrorism. Here it is very important information about dynamics of face expression and movement of the separate parts of the face.

In spite of the fact that face recognition systems has larger value of error of both of the types than finger print recognition systems, iris recognition systems and others, they find a lot of different applications because of their flexibility of installation, training and testing. In this situation it is very important to make research in the field of recognition probability estimation, overtraining estimation, model parameters estimation, etc. to find the most optimal parameters of the face recognition systems. To build very reliable recognition systems it is important to use proposed approach that allows us to build hierarchical recognition on the basis of objects division into functional groups and due to this to use the effect of preclassification.

For the procedure of decision making one proposes to use the notation of fiducial interval. By the fiducial interval one understands the list of possible candidates for the classification. Usage of the fiducial interval is very useful for the decision making support systems with presence of an operator. The result of the system work is the list of candidates that are the most similar to the object to be recognized. In this case the final decision about the object will be made by operator. The system can work as completely autonomous one with using of the fiducial interval for the decision making. In fiducial interval there exist several group of objects that belong to their own class. Our task is to find the group of objects that corresponds to the object to be recognised or to make decision that there is no corresponding objects in fiducial interval. The idea of fiducial interval consists in following concepts. The size of the fiducial interval (the number of possible candidates) has to be enough to be sure that if the corresponding objects are in the database of the recognition system they will drop into this interval. The size of the fiducial interval corresponds to the fiducial probability. The larger is fiducial interval the larger is fiducial probability. That is why it is convenient to use the notation of fiducial interval for the probability of the fact that corresponding objects will drop in the list of possible candidates. The second paragraph of this chapter has been devoted to the problems of forming of some types of fiducial intervals.

5. Discussion and future work

In this chapter we shortly considered some approaches for solution of such important problems as recognition reliability estimation and advanced classification on the basis of division of objects into three functional groups. In domain of reliability estimation there exist two principal problems. First problem concerns the tasks of statistical estimation of the probability of correct recognition especially for small training sets. This is very important when we can not achieve additional objects so fast and make our training set more representative. That could be in situations when we work with data slowly changing in time.

Another important problem concerns the effect of overestimation in pattern classification. The value of overestimation could be found as difference between the recognition results on training and test sets. In the beginning of the chapter one mentioned the main problems of the statistical learning theory and overestimation as one of the most principal problems. One did not pay attention to this problem in this chapter but it is planned to do in future research. The attention has been payed to the problems of recognition reliability estimation. In this chapter the results of both combinatorial and probabilistic approach to recognition reliability estimation have been presented. As seen from the figures there was realized the advanced analysis and estimation of the recognition results when the training set is decreased. So we can make the prognosis of the recognition probability for reduced training sets using combinatorial approach. The reliability of such approach can be provided on the basis of probabilistic approach.

It was considered some methods of the reliability estimation for some types of classifiers. Such of the classifiers belongs to the group of so called metrical classifiers or classifiers on the basis of dissimilarity functions or distance functions. It will be interesting to consider the proposed methods in case of other types of classifiers e.g. classifiers using separating hyperplane, classifiers built on logic functions and others. It will be interesting to consider the idea of how to express one classifier through another or to build relations between the different types of classifiers. All this could give us the possibility to use one approach to reliability estimation for any type of the classifiers.

In the second part of the chapter the probability of belonging of every object to each of the three groups of objects: a group of "easy" objects, on which it is reached the correct consensus of two algorithms, a group of objects, on which two the most dissimilar algorithms have an incorrect consensus and a group of objects, on which one does not achieve consensus have been considered. The analysis shows that there are probability distributions of data that can be presented as a multicomponent models including GMM. All this makes it possible to analyze the proposed algorithms by means of mathematical statistics and probability theory. From the figures and tables one can see that the probability estimations using methods of cross-validation with averaged blocks of 30 and 200 elements minimum differ a little among themselves, which makes it possible to conclude that this method of consensus building, where consensus consists in the most dissimilar algorithms, is quite regular and does not have such sensitivity to the samples as other algorithms that use training. As seen from the corresponding tables, the minimum classification error is almost less by order of magnitude than error for the best of existing algorithms. The maximal error is less of 1.5 to 2 times in comparison with other algorithms. Also, the corresponding errors are much more stable both relatively to the task, on which one tests the algorithm and the series of given algorithms where the error value has significantly large variance. Moreover, since the minimal value of error is quite small and stable, it guaranties the stability of receipt of correct classification results on objects, on which consensus is reached by the most dissimilar algorithms. Relatively to other algorithms such a confidence can not be achieved. Indeed, the error value at $30-40\%$ (as compared to 4%) gives no confidence in results of classification. The fact that the number of ambiguous objects selected by two the most different algorithms is less than the number of objects selected by three algorithms conditioned by the overtraining of two the most dissimilar algorithms. So the future research in this domain should be devoted to the problem of overtraining of the ensemble of two the most dissimilar algorithms. This means that it should be reduced the overtraining of the preclassification that allows us to reduce the error of classification gradually and due to this to satisfy much more reliable classification.

6. References

Basu, M. & Ho, T. (2006). *Data complexity in pattern recognition*, Springer-Verlag, ISBN 1-84628-171-7, London

Bishop, C. (2006). *Pattern recognition and machine learning*, Springer-Verlag, ISBN 0-387-31073-8, New York

Kapustii, B.; Rusyn, B. & Tayanov, V. (2007) Mathematical model of recognition systems with smalldatabases. *Journal of Automation and Information Sciences,* vol. 39, No. 10, pp. 70–80, ISSN 1064-2315

Kapustii, B.; Rusyn, B. & Tayanov, V. (2008). Features in the Design of Optimal Recognition Systems . *Automatic control and computer sciences*, vol. 42, No. 2, pp. 64–70, ISSN 0146-4116

Kapustii, B.; Rusyn, B. & Tayanov, V. (2008). Estimation of the Influence of Information Class Coverage on Generalized Ability of the k-Nearest-Neighbors Classifier. *Automatic control and computer sciences*, vol. 42, No. 6, pp. 283–287, ISSN 0146-4116

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceedings of 14th International Joint Conference on Artificial Intelligence*, pp. 1137-1145, Palais de Congres Montreal, Quebec, Canada, 2010

Kyrgyzov, I. (2008). *Recherche dans les bases de donnes satellitaires des paysages et application au milieu urban :clustering, consensus et categorisation: Ph.D. thesis*, l'ecole Nationale Superiere des Telcommunications, Paris

Moon, T., Stirling, S. (2000). *Mathematical methods and algorithms for signal processing*, Prentice-Hall, ISBN 0-201-36186-8, New Jersey.

Mullin, M., Sukthankar, R. (2000) Complete cross-validation for nearest neighbour classifiers, *Proceedings of International Conference on Machine Learning*, pp. 639-646, 2000

Tayanov, V. & Lutsyk, O. (2009). Classifier Quality Definition on the Basis of the Estimation Calculation Approach. *Computers & Simulations in Modern Science*, Mathematical and Computers in Science and Engineering, A series of Reference Books and Textbooks, pp. 166-171, ISSN 1790-2769

Vapnik, V. (2000). *The Nature of Statistical Learning Theory* (2), Springer-Verlag, ISBN 0-387-98780-0, New York

Vorontsov, K. (2010). Exact combibatorial bounds on the probability of overfitting for empirical risk minimization. *Pattern Recognition and Image Analysis,* vol.20, No. 3 pp. 269-285 ISSN 1054-6618

Vorontsov, K. (2008) On the influence of similarity of classifiers on the probability of overfitting pattern recognition and image analysis: new information technologies, *Proceedings of International Conference on Pattern Recognition and Image Analysis: new information technologies (PRIA-9)*, Volume 2., Nizhni Novgorod, Russian Federation, pp. 303-306, 2000

Weinstein, E. (March 2011). *Gauss's Inequality*, In: *A Wolfram Web Resource*, 16.03.2011, Available from http://mathworld.wolfram.com

Zhuravlev, J. (1978). An Algebraic Approach to Recognition and Classification Problems. *Problems of cybernetics,* vol.33, pp.5–68 (in Russian)

4

A MANOVA of LBP Features for Face Recognition

Yuchun Fang, Jie Luo, Gong Cheng, Ying Tan and Wang Dai
School of Computer Engineer and Science College, Shanghai University
China

1. Introduction

Face recognition is one of the most broadly researched subjects in pattern recognition. Feature extraction is a key step in face recognition. As an effective texture description operator, Local Binary Pattern (LBP) feature is firstly introduced by Ahonen et al into face recognition. Because of the advantages of simplicity and efficiency, LBP feature is widely applied and later on becomes one of the bench mark feature for face recognition. The basic idea of LBP feature is to calculate the binary relation between the central pixel and its local neighborhood. The images are described with a multi-regional histogram sequence of the LBP coded pixels. Since most of the LBP pattern of the images are uniform patterns, Ojala et al, 2002 proposed Uniform Local Binary Pattern (ULBP). Through discarding the direction information of the LBP feature, they proposed the Rotation Invariant Uniform Local Binary Pattern (RIU-LBP) feature. The Uniform LBP feature partly reduces the dimension and retains most of the image information. RIU-LBP greatly reduces the dimension of the feature, but its performance in face recognition decreases drastically. This chapter mainly discusses the major factors of the ULBP and RIU-LBP features and introduces an improved RIU-LBP feature based on the factor analysis.

Many previous works also endeavored to modify the LBP features. Zhang and Shan et al, 2006 proposed Histogram Sequence of Local Gabor Binary Pattern (HSLGBP), whose basic idea was to perform LBP coding to the image in multi-resolution and multi-scale of the images, thereby enhancing the robustness to the variation of expression and illumination; Jin et al, 2004 handled the center pixel value as the last bin of the binary sequence, the formation of the new LBP operator could effectively describe the local shape of face and its texture information; Zhang and Liao et al, 2007a, 2007b proposed multi-block LBP algorithm (MB-LBP), the mean of pixels in the center block and the mean of pixels in the neighborhood block were compared; Zhao & Gao, 2008 proposed an algorithm for multi-directional binary mode (MBP) to perform LBP coding from four different directions; Yan et al, 2007 improved the robustness of the algorithm by fusing the mult-radius LBP feature; He et al, 2005 believed that every sub-block contained different information, and proposed an enhanced LBP feature. The original image was decomposed into four spectral images to calculate the Uniform LBP codes, and then the waterfall model was used to combine them as the final feature. In order to effectively extract the global and local features of face images, Wang Wei et al 2009 proposed LBP pyramid algorithm. Through multi-scale analysis, the algorithm first constructed the pyramid of face images, and then the histogram sequence in a hierarchical way to form the final features.

No matter how the ULBP features are modified, the blocking number, the sampling density, the sampling radius and the image resolution dominantly control the performance of the algorithms. They affect the memory consumption and computation efficiency of the final feature drastically. However, the values of these factors have to be pre-selected and in most previous work, their values are decided with some experience, which obviously ignores the influence degree of each factor and the experience values are hard to be generalized to other databases. In order to seek a general conclusion, in this chapter, we use statistical method of multivariate analysis of variance (MANOVA) to discuss the contribution of four factors for face recognition based on both ULBP and RIU-LBP features. Besides, we research the correlation of the factors and explore which factors play a key role in face recognition. We also analyze the characteristics of the factors; discuss the change of influence of factors for different LBP features. Based on the factor analysis, we propose a modified RIU-LBP feature.

The chapter is organized as follows. In Section 2, we introduce the LBP operators, the LBP features and the four major factors. In Section 3, we illustrate how the MANOVA is applied in exploring the importance of four factors and the results obtained for the two types of LBP features. Based on the above analysis results, an improved RIU-LBP algorithm is introduced in Section 4, which is a fusion of multi-directional RIU-LBP features. We summarize the chapter with several key conclusions in Section 5.

2. LBP features and factors

LBP feature is a sequence of histograms of blocked sub-images of face images coded with LBP operator. The image is divided into rectangle regions and histograms of the LBP codes are calculated over each of them. Finally, the histograms of each region are concatenated into a single one that represents the face image.

2.1 Three LBP operators

With the variation of the LBP operator, the obtained LBP features are of different computation complexity. Three types of LBP operators are compared in this chapter.

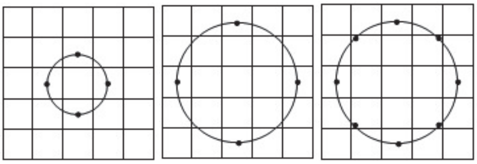
The basic LBP operator is formed by thresholding the neighborhood pixels into binary code 0 or 1 in comparison with the gray value of the center pixel. Then the central pixel is coded with these sequential binary values. Such coding denoted as $LBP_{(P,R)}$ is determined by the radius of neighborhoods R and the sampling density P. With various values of R and P, the general LBP operator could adapt to different scales of texture features, as shown in Figure 1. The order of binary code reserves the direction information of texture around each pixel with 2^P variations.

When there exist at most 2 times of 0 to 1 or 1 to 0 variation, the binary pattern is called a uniform pattern. The Uniform LBP operator $LBP_{(P,R)}^{u2}$ codes the pixel with uniform patterns and denotes all un-uniform patterns with the same value. Its coding complexity is $P^2 - P + 2$.

Rotation Invariant Uniform (RIU) LBP operator $LBP_{(P,R)}^{Riu2}$ is another very popular texture operator. It neglects the order of binary coding and the center pixel of RIU-LBP is denoted by simply counting the number of 1s in the neighborhood as denoted in Equation (1)

$$LBP_{(P,R)}^{Riu2} = \begin{cases} \sum_{p=0}^{P} s(g(p) - g(c)), & \text{for uniform patterns} \\ \quad P+1, & \text{otherwise.} \end{cases} \tag{1}$$

where c is the center pixel, $g(\cdot)$ denotes gray level of pixel and $s(\cdot)$ is the sign function. The coding complexity of the RIU-LBP operator is $P+2$.



$LBP_{(1,4)}$ $LBP_{2,4}$ $LBP_{(2,8)}$

Fig. 1. The General LBP operator

2.2 Three LBP features

After the original face image is transformed into an LBP image with the LBP operators, the LBP image is blocked into M-by-N squares (See examples in Figure 2) to reserve the space structure of face, and then the LBP histogram is calculated for each square to statistically reflect the edge sharpness, flatness of region, existence of special points and other attributes of a local region. The LBP feature is the concatenated serial of all M-by-N LBP histograms. Hence, LBP feature is intrinsically a statistic texture description of the image containing a sequence of histograms of blocked sub-images. The blocking number and the sampling density determine the feature dimensions. For the three LBP operators introduced in Section 2.1, the corresponding LBP feature is denoted as $LBP_{(P,R)}(M,N)$, $LBP_{(P,R)}^{u2}(M,N)$ and $LBP_{(P,R)}^{Riu2}(M,N)$ respectively by taking into consideration the blocking parameters.

Due to different coding complexity, the above three LBP features $LBP_{(P,R)}(M,N)$, $LBP_{(P,R)}^{u2}(M,N)$ and $LBP_{(P,R)}^{Riu2}(M,N)$ are of various dimensions as shown in Equation (2) to (4) respectively. For the former two types of LBP feature, increasing the sampling density will result in explosion of dimension. Examples of dimension comparison are listed in Table I. The blocking number and sampling density are two major factors affecting the dimensions of the LBP feature.

$$D = (M \times N) \times 2^P \tag{2}$$

$$D = (M \times N) \times [P^2 - P + 2 + 1] \tag{3}$$

$$D = (M \times N) \times (P + 2) \tag{4}$$

$M \times N / P$	$7 \times 8 / 8$	$7 \times 8 / 16$	$14 \times 16 / 8$
The general LBP $M \times N \times 2^P$	14336	28672	57344
Uniform LBP $M \times N \times (P^2 - P + 2 + 1)$	3304	13608	13216
RIU-LBP $M \times N \times (P+2)$	560	1008	2240

Table 1. Dimension comparison of three LBP features ($M \times N$ denotes the blocking number, P denotes the sampling density)

2.3 The four factors of LBP feature

The blocking number, the sampling density, the sampling radius and the image resolution are four factors that determine the LBP features.

The blocking number and sampling density are two important initial parameters affecting the dimensions and arouse more attentions in previous research. In addition, the blocking number and the image resolution determine the number of pixels of each sub-image. It means how much local information of face contains in each sub-image. If the image resolution is $H \times W$, the blocking number is $M \times N$, and then each sub-image contains $\left[\frac{H}{M}\right] \times \left[\frac{W}{N}\right]$ pixels, [ ] is rounding. For example, when the image resolution is 140 * 160, the blocking numbers are respectively 3×4, 7×8, 14×16, 21×24 and the sub-images contain respectively 1880, 400, 100, 49 pixel as shown in Figure 2. The position of the neighbour points of the LBP operator is decided by the size of the sampling radius, so the vale of radius also directly affects the LBP features.

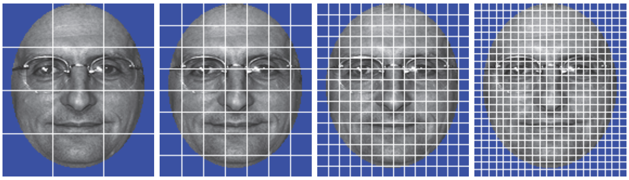


Fig. 2. Comparison of the blocking number of sub-image

Among the four factors, the blocking number and the sampling density are two factors deciding the dimension of the LBP features. For an example shown in Table 1, in the case of different sampling density, the dimension of $LBP_{(8,1)}^{u2}(8,7)$ feature is $D = 3304$. When P doubles, $D = 13608$ for $LBP_{(16,1)}^{u2}(8,7)$. Figure 2 shows that the more the blocking number is, the higher the dimension of features is. Such feature will inevitably cost huge amount of memory and lowers the speed in computation. Does it deserve to spend so much memory to prompt precision of merely a few percents? We do some preliminary experiments to find the answer.

For different values of the four factors, we compare the face recognition rate on a face database containing 2398 face images (1199 persons, each of 2 images) selected from the FERET database. Experimental results are evaluated with the curve of rank with respect to CMS (Cumulative Matching Score), meaning the rate of correct matching lower than a certain rank. The closer this curve towards the line $CMS = 1$, the better is the performance of the corresponding algorithms.

Figure 3 is the comparison of recognition rate for three LBP features. It can be observed that the Uniform LBP has very close performance with the basic LBP but of much lower dimension. While the recognition rate of the RIU-LBP feature decreases significantly due to the loss of direction information. The results indicate that it is good enough to adopt ULBP instead of the basic LBP feature and the direction information has major impact to the recognition rate.

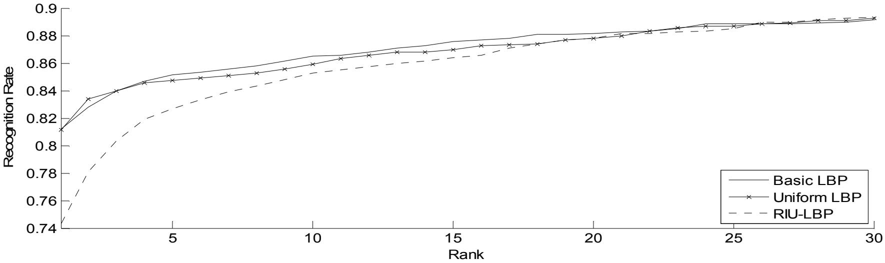


Fig. 3. Comparison of recognition rate for three LBP features (With blocking number 7 x 8, sampling density 8, sampling radius 2 and image resolution 70*80)

Figure 4 compares the performance of the same kind of LBP feature with various values of the blocking numbers and sampling density. The comparison shows that higher sampling density and more blocking numbers could result in better performance. It demonstrates that the two factors affect not only the feature dimensions, but also the face recognition rate. Besides, the sampling radius determines the sampling neighborhood of sub-blocks. The image resolution and the blocking number determine the number of pixels of each sub-block.

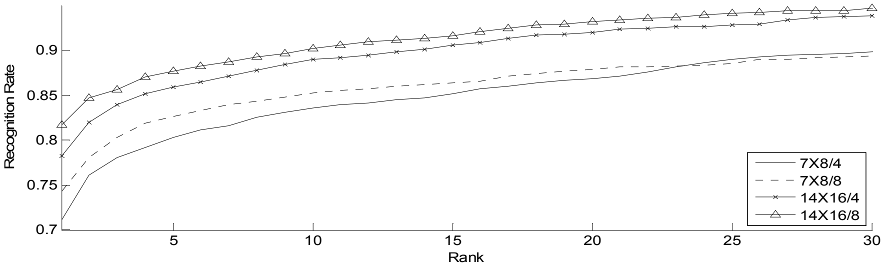


Fig. 4. Comparison of recognition rate for different blocking number and sampling density for RIU-LBP (With sampling radius 2, image resolution 70*80 and 7 x 8 /4 denotes blocking number 7 x 8 and sampling density 4)

Figure 5 compares the performance of the same kind of LBP feature with various values of the sampling radius and image resolution. The results show that higher resolution and larger sampling density are better for face recognition. Though, these two factors do not affect the feature dimensions, they have unneglectable impact on the recognition rate.

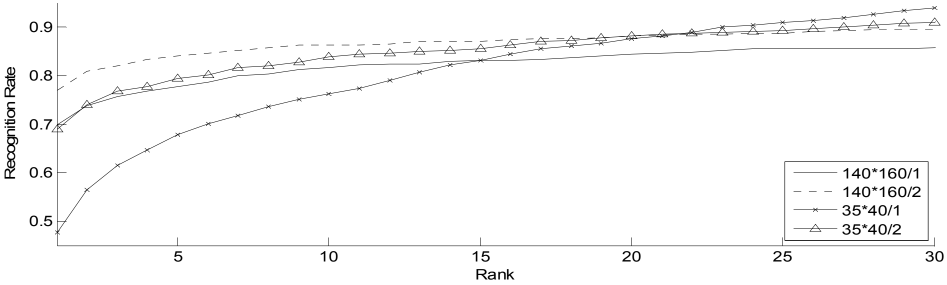


Fig. 5. Comparison of recognition rate for different sampling radius and image resolution for RIU-LBP (With blocking number 7 x 8, sampling density 8 and 140*160/1 denotes image resolution 140*160 and sampling radius 1)

3. The importance of four factors

As described in the above, when we use LBP feature in face recognition, the blocking number, the sampling density, the sampling radius, the image resolution would affect the recognition rate in varying degrees.

In most present research, the values of these factors are selected according to experience in experiments. Ahonen , et al, 2004 compared different levels of the blocking number, the sampling density and the sampling radius based on the Uniform LBP feature. Under the image resolution of 130 * 150, they selected blocking number 7×7 , sampling density 8 and radius 2 as a set of best value which could balance the recognition rate and the feature dimensions. Moreover, they referred to that if the sampling density dropped from 16 to 8, it could substantially reduce the feature dimension and the recognition rate only lowered by 3.1%. Later on, Ahonen, et al, 2006 also analyzed the effect of blocking number for the recognition rate by several experiments. The conclusion is that the blocking number 6×6 are better than blocking number 4×4 in case of less noise, and vice versa. Chen, 2008 added fusion of decision-making in LBP feature extraction method. They selected the sampling density 8, radius 2 blocking number 4×4 and image resolution 128 * 128. Xie et al, 2009 proposed LLGP algorithm, and they selected image resolution 80* 88 and blocking number 8×11 as the initial parameters. Zhang et al 2006 proposed HSLGBP algorithm and also discussed the size of sub-images and its relationship with recognition rates. Wang et al, 2008 used multi-scale LBP features to describe the face image. On the basis, they discussed the relationship between the blocking number and recognition rate, and summarized that too large or too small size of blocks would affect recognition rate. In some other papers, the researchers fixed the size of sub-image determined by the blocking number and the resolution of images.

However, more open problems could not be explained with the experienced values. How is the degree of impact of the four factors? Are they contributes the same in efficiency? Are there interactions between pairs of factors? How will the parameters affect the recognition? How to compare the performance of different LBP features? In order to solve these problems, we endeavor to compare four major factors for the two most typical LBP features, i.e. the ULBP and the RIU-LBP feature with MANOVA. Our purpose is to explore the contribution of the four factors for recognition and the correlation among them. The results of the studies to these problems provide important merits for the improvement of LBP features.

3.1 MANOVA

MANOVA is an extensively applied tool for multivariate analysis of variance in statistics. For our problem, the four variables waiting to be explored are the resolution of images, the blocking numbers, the sampling density and the radius of LBP operators for face recognition tasks. With MANOVA, we could identify whether the independent variables have notable effect and whether there exist notable interactions among the independent variables [12].

By denoting the four factors as follows:

I - Resolution of images

B - Blocking numbers

P - Sampling density of the LBP operator

R - Sampling radius of the LBP operator

And taking the face recognition rate as the dependent variable, the total sum of squares deviations S_T is denoted in Equation (5):

$$\begin{aligned} S_T = S_B + S_P + S_R + S_I + S_{B\times P} + S_{B\times R} \\ + S_{B\times I} + S_{P\times R} + S_{P\times I} + S_{R\times I} + S_E \end{aligned} \quad (5)$$

where $S_{B\times P} + S_{B\times R} + S_{B\times I} + S_{P\times R} + S_{P\times I} + S_{R\times I}$ is the sum of interaction, and S_E is the sum of squares of the errors.

MANOVA belongs to the F-test, in which the larger F value and the smaller P value correspond to independent variables that are more significant. Hence, the significance of the factors is evaluated through checking and comparing the F value and the P value. If the P value is less than a given threshold, the factor has dominant effect, or there exist notable interactions between two factors. If the F value of one factor is the largest, its effect is the most important.

3.2 Experiment design of factors

We use the same face database as described in Section 2.3 in MANOVA. As analyzed in Section 2.3, the general LBP feature has much higher dimensions than the ULBP feature but the performance is close, so we conduct the experiments for ULBP and RIULBP features.

We set each factor three or four different levels as shown in Table 2. Under the RIU-LBP features, 108 sets of experimental data were obtained. Under the ULBP features, 81 sets of experimental data were obtained (level of blocking number 21×24 is missed due to too high computation complexity).

	B	*P*	*R*	*I*
Level1	3×4	4	1	35*40
Level2	7×8	8	2	70*80
Level3	14×16	16	3	140*160
Level4	21×24	—	—	—

Table 2. Different levels of four factors in experiment

3.3 Analysis of the factors of RIU-LBP

3.3.1 The significance and interaction

We firstly analyze the independent influence of four factors and significance of interaction. Table 3 shows the results based on RIU-LBP feature. The row of Table 3 is in descending order by *F* value.

The first part of Table 3 shows the independent effect of the four factors. The value of *P* were less than 0.05, it means all four factors have significant effects for recognition rate. The impact from the largest to the smallest, respectively, is the blocking number, the sampling radius, the image resolution and the sampling density. Specially, the *F* value of blocking number is greater than the other three factors, which reflects the importance of the blocking number in face recognition. The *F* value of the sampling density is much smaller than the other three factors, meaning the weakest degree of influence.

	Df	Sum Sq	Mean Sq	*F* value	*P* value
B	3	1.509	0.503	**900.683**	<0.001
R	2	0.464	0.232	415.919	<0.001
I	2	0.381	0.190	341.202	<0.001
P	2	0.048	0.024	42.772	<0.001
B * *R*	6	0.067	0.011	20.058	<0.001
R * *I*	4	0.043	0.010	19.239	<0.001
B * *I*	6	0.021	0.004	6.300	<0.001
P * *I*	6	0.011	0.003	4.729	0.002
R * *P*	6	0.007	0.002	2.923	0.027
B * *P*	6	0.006	0.001	1.700	0.134
Residuals	68	0.038			
Total	107	2.594			

Table 3. MANOVA results based on RIU-LBP (Df is freedom, Sum Sq is sum of squares, Mean Sq is mean square and * is interaction)

3.3.2 Analysis of levels of each single factor

For each single factor, we also design the MANOVA for each pair of levels. The *P* values are recorded for all four factors shown in Table 4-7 respectively. The second column lists the average recognition rate for each level in each of these tables.

Table 4 is the analysis result based on four levels of blocking numbers. From the second column it can be seen that the more the blocking number is, the higher the mean of recognition rate. For blocking numbers, the *P* value between level 14*16 and level 21*24 is 0.241. So there are no notable difference between them while for other pairs of levels, the interactions are notable.

Level	Mean	3×4	7×8	14×16	21×24
3×4	0.477	1.000	<0.001	<0.001	<0.001
7×8	0.679	<0.001	1.000	0.016	0.002
14×16	0.754	<0.001	0.016	1.000	0.410
21×24	0.777	<0.001	0.002	0.410	1.000

Table 4. Level comparison for the blocking number based on RIU-LBP

Through the first part of the analysis, we know that the sampling density is of the least affect among four factors. The pair-wise results for sampling density in Table 5 show that there is no significant difference for various levels since the *P* values are all larger than 0.05. Besides, the second column shows that the highest recognition rate is 0.689, corresponding to the sampling density is 8. It gives us an important information that the sampling density is not the bigger the better. So we should not blindly pursue high sampling density.

Level	Mean	4	8	16
4	0.642	1.000	0.61	0.61
8	0.689	0.61	1.000	0.89
16	0.684	0.61	0.89	1.000

Table 5. Level comparison for sampling density based on RIU-LBP

Table 6 and Table 7, respectively are the analysis result based on the three levels of sampling radius and image resolution. Although these two factors would not affect the feature dimension, but from previous discussion we already know that they also affect recognition rate. If we select the appropriate parameter values of these factors, it could help to improve the recognition rate.

For the resolution of image, there exists significant difference between 35*40 and 140*160. The significance between 35*40 and 70*80 is larger than that between 70*80 and 140*160 as shown in Table 6. Because the clearer the images are, the more useful information could be extracted. But at the same time, we should also consider that the higher the resolution, the larger the storage space is required. For the massive database, high-resolution images would increase the storage difficulties and need to spend more time to load image information. So according to the requirement of applications, we could select the lower resolution images to reduce storage space with acceptable recognition rate.

Level	Mean	35*40	70*80	140*160
35*40	0.591	1.000	0.007	<0.001
70*80	0.692	0.007	1.000	0.252
140*160	0.732	<0.001	0.252	1.000

Table 6. Level comparison for image resolution based on RIU-LBP

In Table 7, we observe that both level 2 and level 3 are more important than level 1 for the sampling radius. However, there is no significant difference between level 2 and level 3.

Level	Mean	1	2	3
1	0.580	1.000	0.001	<0.001
2	0.705	0.001	1.000	0.452
3	0.730	<0.001	0.452	1.000

Table 7. Level comparison for sampling radius based on RIU-LBP

3.4 Analysis of factors of ULBP feature

In comparison with RIU-LBP feature, ULBP feature keeps the order of neighborhood coding and thus bears more direction information. Whether the analysis result with the RIU-LBP feature could be applicable to the ULBP features? Similar to the RIU-LBP, we perform MANOVA analysis to ULBP.

3.4.1 The significance and Interaction

Table 8 is the results for independent factors and their interactions. It shows that the four factors have significant effects independently for recognition rate. From large to small, the order of impact is respectively the image resolution, the blocking number, the sampling radius and the sampling density. The influence of sampling density is still the minimal. Compared with the RIULBP features, the image resolution is of the most notable effect for

	Df	Sum Sq	Mean Sq	*F* value	*P* value
I	2	0.0788	0.0394	**329.104**	<0.001
B	2	0.0672	0.0336	**280.965**	<0.001
R	2	0.0217	0.0108	90.517	<0.001
P	2	0.0107	0.0053	44.824	<0.001
B * *R*	4	0.0061	0.0015	12.680	<0.001
B * *I*	4	0.0043	0.00108	9.030	<0.001
R * *I*	4	0.0042	0.00106	8.856	<0.001
B * *P*	4	0.0018	<0.001	3.862	0.008
P * *I*	4	0.0011	<0.001	2.321	0.070
R * *P*	4	0.0007	<0.001	1.575	0.196
Residuals	48	0.0057			
Total	80	0.2026			

Table 8. MANOVA results based on Uniform LBP

ULBP feature instead of the blocking number. But for ULBP feature, the F values of the image resolution and the blocking number are very close and both are over 3 times larger than the other two factors. In the interaction part, we could see that the interaction have no obvious effect between the sampling density and the sampling radius. We could approximately believe that these two factors of ULBP operator are independent to each other. And similar to the RIU-LBP, the interactions between pairs of factors are much smaller than the independent factors of the ULBP feature.

3.4.2 Analysis of levels of each single factor

Similarly, we also analyze the difference among various levels of factors for ULBP feature, and the results are summarized in Table 9-12.

Based on Table 9, there are significant differences between the three levels of blocking numbers, and the 14×16 blocks is corresponding to the highest mean of recognition rate.

Level	Mean	14×16	7×8	14×16
7×8	0.735	1.000	<0.001	<0.001
7×8	0.781	<0.001	1.000	0.004
14×16	0.805	<0.001	0.004	1.000

Table 9. Level comparisons for blocking number based on Uniform LBP

As in Table 10, the three levels of sampling density are not significantly different. The average recognition rate is the highest when the sampling density is 8, which indicates that high sampling density is not necessary.

Level	Mean	4	8	16
4	0.759	1.000	0.120	0.420
8	0.788	0.120	1.000	0.420
16	0.776	0.420	0.420	1.000

Table 10. Level comparisons for sampling density based on Uniform LBP

For sampling radius, there is no apparent difference between level 2 and 3. When the sampling radius is 2, the mean of recognition rate is the highest.

Level	Mean	1	2	3
1	0.751	1.000	0.025	0.025
2	0.786	0.025	1.000	0.899
3	0.785	0.025	0.899	1.000

Table 11. Level comparisons for sampling radius based on Uniform LBP

Lastly, for the most prominent factor, i.e. the image resolution, there is no prominent difference between 70 * 80 and 140 * 160.

Level	Mean	35*40	70*80	140*160
35*40	0.730	1.000	<0.001	<0.001
70*80	0.791	<0.001	1.000	0.410
140*160	0.800	<0.001	0.410	1.000

Table 12. Level comparisons for image resolution based on Uniform LBP

3.5 More analysis on the blocking number

Based on the MANOVA analysis for both RIULBP and ULBP feature, we can conclude that the more the blocking number is, the higher the recognition rate will be. But should its value goes to the up-limit of 1 pixel per block? And what is the suitable blocking number? We do more experiments to analyze these problems for RIU-LBP features.

We pick two more levels of the blocking number, i.e. 35×40 and 70×80, and fixed sampling density 8, the sampling radius 2, and the image resolution 70*80 based on previous conclusions of MANOVA. The RIU-LBP features of these two groups of parameter setting are of dimension 14,000 and 56,000 respectively. And there is only one pixel in each block when the blocking number is 70×80. Figure 6 summarizes the variation of the face recognition rate with respect to the blocking number. It shows that the blocking number is not the higher the better.

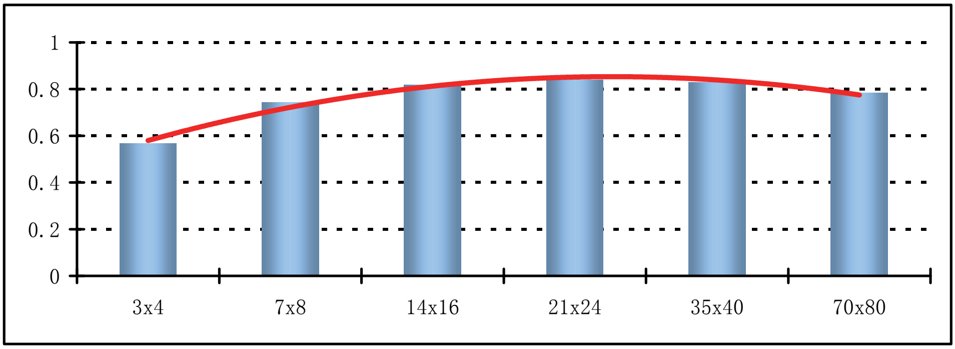


Fig. 6. Comparison recognition rate based on different the blocking number and RIU-LBP

(With sampling density 8, sampling radius 2 and image resolution 70*80)

We extend the experiments to 35*40 and 140*160 image resolution cases. The recognition rate of blocking number 14×16 is the highest for image resolution 35*40; the recognition rate of blocking number 35×40 is the highest for image resolution 140*160. Hence, the blocking number is not the more the better.

Based on the analysis result of ULBP and RIU-LBP feature, we could take the following steps in setting the parameters of the four factors. Although different setting might be necessary for the specific applications, the basic rule is that more bits should be assigned to the blocking numbers and less for the sampling density. Moreover, the appropriate blocking number should be selected in consideration of the image resolution together, and then to choose the proper value for sampling density and sampling radius.

I / B	3×4	7×8	14×16	21×24	35×40	70×80
35*40	0.460	0.689	**0.751**	0.750	0.751	N/A
70*80	0.567	0.743	0.817	**0.839**	0.830	0.784
140*160	0.581	0.770	0.831	0.847	**0.862**	0.849

Table 13. Comparison recognition rate based on different the blocking number and three image resolution (With RIU-LBP, sampling density 8, sampling radius 2)

3.6 Summary

We comprehensively analyze the four factors of the ULBP and RIU-LBP features and many useful conclusions are drawn. Firstly, the blocking number is the main factor that influences the recognition rate, which indicates that the contribution of local features in face recognition is more important than the global features. Secondly, the effect of sampling density for the recognition rate is small, but it severely affects the feature dimension. At the same time, such results mean the feasibility of using low sampling density to acquire features of high recognition rate. In addition, the sampling density and the sampling radius decide the setting of the LBP operator, but they have much less obvious affect to the recognition rate compared with the blocking numbers and the image resolution. These conclusions demonstrate that the complex encoding of the LBP operator is not important in face recognition. Finally, the interactions between the factors are of less effect to recognition rate compared with the independent factors.

4. Fusion of multi-directional RIU-LBP

The difference between the ULBP feature and the RIU-LBP feature lies in the way of LBP coding. The latter totally abandons the directional information and hence of much lower dimension but of less precision in face recognition. However, if introducing the direction information into the RIU-LBP feature in a linear complexity, a new feature of much lower dimension could result in similar precision as the ULBP feature.

4.1 Multi-directional RIU-LBP

The dimension explosion of the LBP features is mainly aroused by the sampling density. The basic LBP operator adopts an ordered way to code the variation in each direction around one pixel, thus it has to cost 2^P bits in the calculation of one LBP histogram and even the ULBP feature can only lower the cost to $P^2 - P + 3$. We propose a new LBP feature that fuses multi-directional low density RIU-LBP feature.

First, we split P neighbors of a pixel into several non-overlapped subsets with $P_1, P_2, ..., P_K (\sum_{i=1}^{k} P_i = P)$ uniformly distributed pixels respectively. In accordance with the mathematical coordinate system, set the angle of positive x axis as 0° and counterclockwise as positive direction, then each subset can be discriminate by its size $P_i (i \in \{1,2,...,k\})$ and the pixel with minimum positive angle $\theta_i (i \in \{1,2,...,k\})$, denoted as $S(\theta_i, P_i)$. An example is shown for a $P = 16$ case in Fig.7, in which the 16-point neighborhood is split into both two 8-point sets and four 4-point sets. Secondly, the regular

RIU-LBP feature is calculated for each neighbor $S(\theta_i, P_i)(i \in \{1,2,...,k\})$, denoted as $LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$, the so-called directional low density RIU-LBP feature. Lastly, a combination of all $LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)(i=1,...k)$ at feature level is taken as the final LBP features denoted as $\bigcup_{i=1}^{k} LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$, defined as multi-directional RIU-LBP feature. The dimension of $\bigcup_{i=1}^{k} LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$ is

$$D = (M \times N) \times \sum_{i=1}^{k} (P_i + 2) \qquad (4)$$

For the two division settings shown in Fig.7, the dimension of $\bigcup_{i=1}^{2} LBP^{Riu2}_{(8,\theta_i,1)}(8,7)(\theta_i \in \{0, \frac{\pi}{8}\})$ is 1120, and 1344 for $\bigcup_{i=1}^{4} LBP^{Riu2}_{(4,\theta_i,1)}(8,7)(\theta_i \in \{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}\})$. Both types of features are of the same computation complexity as $LBP^{Riu2}_{(16,1)}(8,7)$. In comparison with $LBP^{u2}_{(16,1)}(8,7)$, the dimension decreases to nearly 1/10.

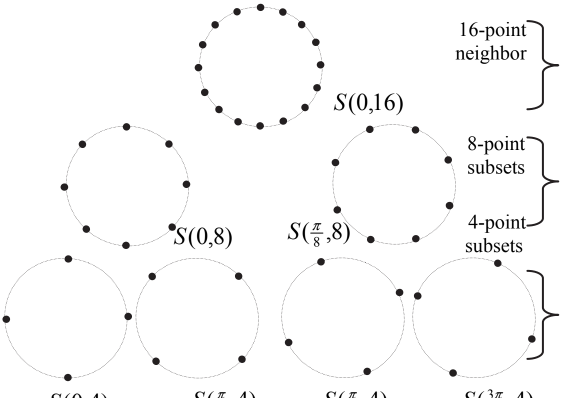


Fig. 7. Example of neighborhood split

With such simple way of feature fusion, the $\bigcup_{i=1}^{k} LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$ feature reserves the variation of intensity in a certain direction represented by each component $LBP^{Riu2}_{(P_i,\theta_i,R)}(M,N)$. Instead of the exponential or power 2 way of dimension growth with P

aroused by the basic LBP operator, the dimension of multi-directional RIU-LBP feature increases linearly with P growth.

4.2 Performance analysis

We perform the experimental analysis to the multi-directional RIU-LBP feature with the same face database described in Section 2.3. We also compare the proposed multi-directional RIU-LBP feature $\bigcup_{i=1}^{k} LBP_{(P_i,\theta_i,R)}^{Riu2}(M,N)$ with the uniform LBP feature and the RIU-LBP feature. Two examples are illustrated in Fig.8 and Fig.9. Fig.8 shows the comparison results of $LBP_{(8,1)}^{u2}(8,7)$, $LBP_{(8,1)}^{Riu2}(8,7)$ and $\bigcup_{i=1}^{2} LBP_{(4,\theta_i,1)}^{Riu2}(8,7)(\theta_i \in \{0,\frac{\pi}{8}\})$. All these three methods have adopted exactly the same 8 neighborhood in LBP coding. The four curves in Fig.9 respectively are results of $LBP_{(16,2)}^{u2}(8,7)$, $\bigcup_{i=1}^{2} LBP_{(8,\theta_i,2)}^{Riu2}(8,7)(\theta_i \in \{0,\frac{\pi}{8}\})$, $LBP_{(16,2)}^{Riu2}(8,7)$ and $\bigcup_{i=1}^{4} LBP_{(4,\theta_i,2)}^{Riu2}(8,7)(\theta_i \in \{0,\frac{\pi}{8},\frac{\pi}{4},\frac{3\pi}{8}\})$ with the same 16 neighborhood in LBP coding. Dimensions of all features are also put in the parenthesis. In Fig.8, where $P = 8, R = 1, M \times N = 8 \times 7$, the proposed multi-directional RIU-LBP features perform much better than the RIU-LBP features and the dimensions of both feature are very close. In Fig.9, where $P = 16, R = 2, M \times N = 8 \times 7$, two types of multi-directional RIU-LBP features all outperform the RIU-LBP features with very close length of feature. Moreover, both features have very close or even better CMS in comparison with that of the uniform LBP features though the dimension of both features are much lower.

Though all three types of LBP features have adopted exactly the same neighborhood in LBP coding, the curves in Fig.8 and Fig.9 prove the promising application of the proposed multi-directional RIU-LBP features in comparison with the uniform LBP feature and the RIU-LBP

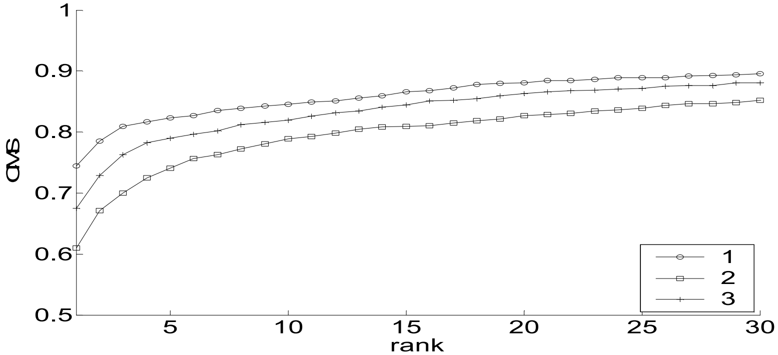


Fig. 8. Comparison results: 1- $LBP_{(8,1)}^{u2}(8,7)$ ($D = 3304$), 2- $LBP_{(8,1)}^{Riu2}(8,7)$ ($D = 560$) and 3- $\bigcup_{i=1}^{2} LBP_{(4,\theta_i,1)}^{Riu2}(8,7)(\theta_i \in \{0,\frac{\pi}{8}\})$ ($D = 672$)

feature. The RIU-LBP feature abandons the direction information of intensity variation around pixel. The uniform LBP feature uses a complex ordered coding way which bears of course more direction information of intensity variation. However, both work no better than the proposed algorithm.

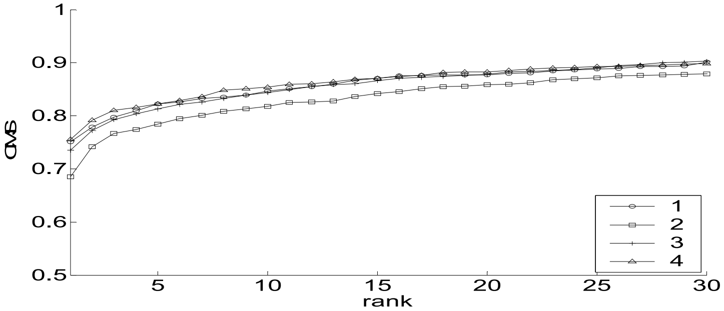


Fig. 9. Comparison results: 1- $LBP^{u2}_{(16,2)}(8,7)$ ($D=13608$), 2- $LBP^{Riu2}_{(16,2)}(8,7)$ ($D=1008$), 3- $\bigcup_{i=1}^{2} LBP^{Riu2}_{(8,\theta_i,2)}(8,7)(\theta_i \in \{0,\frac{\pi}{8}\})$ ($D=1120$) and 4- $\bigcup_{i=1}^{4} LBP^{Riu2}_{(4,\theta_i,2)}(8,7)(\theta_i \in \{0,\frac{\pi}{8},\frac{\pi}{4},\frac{3\pi}{8}\})$ ($D=1344$)

In one hand, the experiments reveal that the direction information of intensity variation is very important and useful in face recognition. In another hand, the proposed algorithm reserves such direction information through feature fusion. With much lower dimensional in comparison with the uniform LBP feature, the proposed algorithm gains very close and even better precision.

5. Conclusion

In this chapter, we first perform a thorough analysis of the four factors of the ULBP features and the RIU-LBP features. From a statistical point of view, we use MANOVA to study four factors the blocking numbers, the sampling density, the sampling radius and the image resolution that affect the recognition rate. Based on the analysis results, a multi-directional RIU-LBP, a modified LBP feature, is proposed through fusing the RIU-LBP features of various initial angels. Several conclusions are drawn as follows: 1) For the RIU-LBP features and the ULBP features, the impact of the blocking number is more important than the other factors. For example, for the RIU-LBP feature with constant values of other factors, when the blocking number changes from 7 *8 to 14 * 16, the recognition rate increased 9 percents. This result indicates that the accuracy of face recognition strongly relies on the localized LBP histograms. 2) The sampling density has little contribution to the recognition rate and high sampling density could not help to achieve high recognition rate. So complex LBP coding is not necessary. (3) The correlation between pairs of the factors is not important. (4) The multi-directional RIU-LBP feature preserve intensity variation around pixels in different directions at the cost of linearly increasing of feature dimension instead of the power 2 way of ULBP feature. Experiments not only show that the proposed scheme could result in better recognition rate than RIU-LBP feature after reserving the direction information in feature-level fusion, but also prove that with much lower dimension of features, the proposed multi-directional RIU-LBP feature could result in very close performance compared with the ULBP feature. In all, through the statistical analysis of the importance of four factors, an effective feature is proposed and the future directions to improve the LBP feature are presented.

6. Acknowledgment

This work is supported by the National Natural Science Foundation of China under Grant No.60605012, the Natural Science Foundation of Shanghai under Grant No.08ZR1408200, the Open Project Program of the National Laboratory of Pattern Recognition of China under Grant No.08-2-16 and the Shanghai Leading Academic Discipline Project under Grant No.J50103.

7. References

Ahonen, T; Hadid, A. & Pietik¨ainen, M. (2004). Face Recognition with Local Binary Patterns. *Proc. 8th European Conference on computer Vision,* pp. 469-481, ISBN 978-3-540-21984-2, Prague, Czech Republic, May 11-14, 2004

Ojala, T; Pietikäinen, M. & Mäenpää, T. (2002). Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. *Transactions on Pattern Analysis and Machine Intelligence,* Vol.24, No.7, (August 2002), pp.971-987, ISSN 0162-8828

Zhang, WC; Shan, SG; Zhang, HM; Chen, J; Chen, XL. & Gao, W. (2006). Histogram Sequence of Local Gabor Binary Pattern for Face Description and Identification. *Journal of Software*, Vol.17, No.12, (December 2006), pp.2508-2517, ISSN 1000-9825

Jin, HL; Liu, QS; Lu, HQ. & Tong, XF. (2004). Face Detection Using Improved LBP under Bayesian Framework. *Proc. 3th International Conference on Image and Graphics,* pp.306-309, ISBN 0-7695-2244-0, Hong Kong, China, December 18-20, 2004

Zhang, L; Chu, RF; Xiang, SM. & Liao, SC. (2007). Face Detection Based on Multi-Block LBP Representation. *Proc. International Conference on Advances in Biometrics,* pp.11-18, ISBN 978-3-540-74548-8, Seoul, Korea, August 27-29, 2007

Liao, SC ; Zhu, XX ; Lei, Z ; Zhang, L. & Li, SZ.(2007). Learning Multi-scale Block Local Binary Patterns for Face Recognition. *Proc. International Conference on Advances in Biometrics,* pp.828-837, ISBN 978-3-540-74548-8, Seoul, Korea, August 27-29, 2007

Zhao, S. & Gao, Y. (2008). Establishing Point Correspondence Using Multidirectional Binary Pattern for Face Recognition. *Proc. 19th International Conference on Pattern Recognition,* pp.1-4, ISBN 978-1-4244-2174-9, Tampa, Florida, USA, December 8-11, 2008

Yan, SC; Wang, H. & Tang, XO. (2007). Exploring Feature Descriptors for Face Recognition. *Proc. 30th International Conference on Acoustics, Speech and Signal Processing*, pp. I629-I632, ISBN 1-4244-0727-3, Honolulu, Hawaii, USA, April 15-20, 2007

He, LH; Zou, CR; Zhao, L. & Hu, D. (2005). An enhanced LBP feature based on facial expression recognition. *Proc. 27th Annual International Conference of the IEEE Engineering in Medicine and Biology,* pp. 3300-3303, ISBN 0-7803-8741-4, Shanghai, China, September 1-4, 2005

Wang, W; Huang, FF; Li, JW. & Feng HL. (2009). Face Description and Recognition by LBP Pyramid. *Journal of Computer-aided Design & Computer Graphics,* Vol. 21, No.1, (January 2009), pp.94-100, ISSN 1003-9775

Ojala, T; Pietikinen, M. & Harwood, D. (1996). A Comparative study of texture measures with classification based on featured distributions. *Journal of Pattern Recognition,* Vol.29, No.1, (January 1996), pp. 51-59 ISSN 0031-3203

Ahonen, T; Hadid, A. & Pietik¨ainen, M. (2006). Face Description with local binary patterns: Application to Face Recognition. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol.28, No.12, (December 2006), pp.2037-2041, ISSN 0162-8828

Chen, CJ. (2008). Decision Level Fusion of Hybrid Local Features for Face Recognition. *Proc. International Conference Neural Networks and Signal Processing,* pp.199-204, ISBN 978-1-4244-2310-1, Najing, China, June 7-11, 2008

Xie, SF; Shan, SG; Chen, SL; Meng, X. & Gao, W. (2009). Learned local Gabor patterns for face representation and recognition. *Journal of Signal Processing,* Vol.89, No.12, (December 2009), pp.2333-2344, ISSN 0165-1684

Wang, W; Huang, FF; Li JW. & Feng, HL. (2008). Face description and recognition using multi-scale LBP feature. *Journal of Optics and Precision Engineering,* Vol.16, No.4, (April 2008), pp.696-705, ISSN 1004-924X

Moitre, D. & Magnago, F. Using MANOVA Methodology in a Competitive Electric Market under Uncertainties. *Proc. Transmission & Distribution Conference and Exposition,* pp.1-6, ISBN 1-4244-0287-5, Caracas, Venezuela, August 15-18, 2006

Part 2

Face Recognition with Infrared Imaging

5

Recent Advances on Face Recognition Using Thermal Infrared Images

César San Martin[1,2], Roberto Carrillo[1], Pablo Meza[2],
Heydi Mendez-Vazquez[3], Yenisel Plasencia[3], Edel García-Reyes[3]
and Gabriel Hermosilla[4]
[1]*Information Processing Laboratory, DIE, University of La Frontera*
[2]*Center for Optics and Photonics, University of Concepción*
[3]*Advanced Technologies Application Center, CENATAV*
[4]*University of Chile*
[1,2,4]*Chile*
[3]*Cuba*

1. Introduction

Nowadays it is possible to find several works on face recognition, where principally the visible range spectra is used. In addition, many techniques are tested on different databases and it is possible to identify the best method to use in applications such as: surveillance, access control, human robots interaction, people searching, identification, among others. In practice, the performance of any face recognition system depends on conditions such as viewing angle of the face, lighting, sunglasses, occlusion of the face by other objects, low resolution of the images, different distance of the subject to the camera, focus of the scene, facial expressions, changes of the subject along the time, etc. This type of variations have been simulated in several works in order to estimate the limitations and advantages of the classification methods developed at the moment. One of the most important effects to consider is the kind of illumination that exists in the scene, difference in lighting condition produces variations in the recognition rate, decreasing the effectiveness of the methods. In this way, infrared technology is introduced in face recognition in order to eliminate the dependence of lighting conditions, achieving satisfactory results even when exist the complete absence of illumination. Bebis et al. (2006); Chen et al. (2003); Chen (2005); Dowdall (2005); Friedrich (2002); Heo et al. (2005); Li et al. (2007); Singh et al. (2008); Socolinsky and Selinger (2002); Socolinsky et al. (2001); Zou et al. (2005; 2007) Conventional visible cameras are composed of a lens and a focal plane array, a detectors matrix that collect the input information in the range 0.4-0.75 μm. These sensors can be composed by CCD or CMOS technology, and their function is to translate the incident light flux to an electrical signal and then it is digitalized using analog to digital conversion. The infrared camera is similar, but the lens and the sensor technology are changed, i.e., the focal plane array is composed of a matrix of infrared sensor or detector with response ranging from the near wave infrared (0.75-1.4 μm) to the long wave infrared (8-15 μm).

The development of infrared detectors has been evoked principally in the use of semiconductors or photodetectors Piotrowski and Rogalski (2004). In this type of detectors the radiation is absorbed by the material by means of the photon interaction with the electrons, presenting at the same time a good signal to noise ratio and a swift answer. In order to achieve it, the detectors require a cooling system, this is the main obstacle for a massive use of this type of systems, since it represents an increase of the weight, volume and system cost. In the last years another class of commercial arrays have appeared and are compound with thermic detectors, in which the incident radiation is absorbed, provoking changes on the physical properties of the material. In contrast to the photodetectors, the thermic detectors can operate at room temperature, they are cheap and easy-to-use, but present modest sensibility and lower answering velocity Liu et al. (2007).

The imaging system presents different undesired kinds of noise, being the principals temporal or electronic noise and the Spatial or Fixed-Pattern Noise (FPN) Mooney et al. (1989); Pron et al. (2000). Temporal noise is typically modeled by additive gaussian white noise, and by definition it varies with time. This type of noise is due to the photon noise and reset noise, and can be reduced using frame averaging. On the other hand, the FPN is due to the nonuniform response presented by the individual detectors and dark current non-uniformity, the principal characteristic is that not change in time and this kinds of noise can affect the final result of the classification. The FPN effect is stronger at longer wavelengths, such as in Infrared Focal-Plane Arrays (IRFPA), producing a severe mitigation on the quality and the effective resolution of the imaging device. Therefore, a nonuniformity correction (NUC) Perry and Dereniak (1993) is a mandatory task for properly using several imaging systems. The study associated of the NUC algorithms represents a broad area of work, in which we are supported by many developed articles.

This chapter presents the behavior of several state-of-art facial classification methods using the infrared spectra, considering temporal and fixed-pattern noise. The principal motivation is to introduce the recent advances in infrared face recognition field, and the use of different data sets built with at least two different infrared cameras. The performance of the recognition system is evaluated considering temporal and spatial noise, in real and simulated scenario, keeping in mind that a computer simulation allow the possibility of controlling the signal to noise ratio and then, obtain more representative results. An infrared scene, acquired from a IR camera,present only one level of noise, but it is possible to consider 20 degrees of maximum rotation of the face and different facial expression.

Two mechanism are used: full image and segmented image. The first one consist on using all pixels of the image, and then, build the classification. In a segmented image, the region corresponding to the face image is separated from the background and the process is introduced to design the classification rule, and then, the recognition process is performed. In some cases, from segmented images is possible to build a characteristic vector that represent the principal feature in order to perform the classification. The principal advantage of segmentation and feature representation corresponds to the reduction of the dimensionality and then, the computing time. In this chapter, we use full image in order to obtain more representative evaluations of the noise effect on recognition task.

The algorithms will be tested on a database with different frames for each subject considering vocalization and facial expressions, allowing to construct a symmetric database with and without noise. The results shown different behavior of different recognition techniques considering temporal or fixed noise pattern.

In this work, two IR face databases are used. The first one has been collected by Equinox Equinox (2009) and consists of images captured using multiple camera sensors: visible, long-wave infrared (LWIR), mid-wave infrared and short-wave infrared spectra. Moreover, they use a special visible-IR sensors capable of taking images with both visible CCD and LWIR microbolometer in $8 - 14\mu m$ spectral range. The database consists of $3,244$ face images from 90 individuals captured with left, right and frontal lighting, and neutral and varying facial expression. Also include an eyeglasses condition since eyeglasses block the thermal emission. Only LWIR imagery are used, with a resolution of $320x240$ pixels and 12 bits per pixel.The images contain a fixed-pattern noise from captured image and is removed using NUC scene-based-method.

The second database consists of images captured using the CEDIP JADE UC camera with an operating range in $8 - 14\mu m$ with microbolometer detector-based. The database consists of 612 images corresponding to 6 images per 102 persons. The images containing expression and vocalization faces, with $320x240$ pixels and 14 bits of resolutions. The images are captured and corrected using two-point radiometrically calibration by means black bodie radiator, located in the Center for Optics and Photonic CEFOP laboratory. This second database is noise free and this condition permit to simulate fixed and temporal noise allowing to gain more representative results.

We are interesting in to study both cases, the noise-behavior of correlation filters and LBP face recognition system, considering two scenarios: identification and verification. In the first case, the performance is evaluated using the correct classification (CC) percent, defined as the ability of the system to identify new pattern or signature. In a verification scenario typically it confirms if a subject is valid or not for a given data set, measured by the false acceptance rate (FAR) and a false rejection rate (FRR). CC, FAR, and FRR are obtained for the cases with/without fixed noise and temporal noise. In this work, the performance results are obtained with the top match approach.

2. Non-uniformity and FPN noise in IR-FPA sensors

In an IR-FPA system, the main noise source is the FPA temporal noise: the FPA NU noise and the readout noise due to the associated electronic Milton et al. (1985). The FPN corresponds to any spatial pattern that does not change in time. Then, the NU in IR-FPA is added to the readout signal forming a FPN that degrades the quality of acquired data. Visually, the FPN is a pixel-to-pixel variation when a uniform infrared radiator is captured by the IR sensors. Typically, each pixel on the IR-FPA can be modeled in the instant n using a first-order equation given by Perry and Dereniak (1993):

$$Y_{ij}(n) = A_{ij}(n)X_{ij}(n) + B_{ij}(n) + V_{ij}(n), \tag{1}$$

where $Y_{ij}(n)$ is the readout signal, $X_{ij}(n)$ is the photon flux collected by the ij sensor, $A_{ij}(n)$ and $B_{ij}(n)$ are the gain and offset of the detector respectively, and $V_{ij}(n)$ is the additive temporal noise, usually assumed to be white gaussian random process.

In order to solve this problem, several NU compensation techniques have been developed Amstrong et al. (1999); Averbuch et al. (2007); Hardie et al. (2000); Harris and Chiang (1999); Narendra (1981); Pezoa et al. (2006); Ratliff et al. (2002; 2005); Scribner et al. (1991; 1993); Torres and Hayat (2002); Torres et al. (2003); Zhou et al. (2005). They can be divided into calibration techniques and scene-based correction methods. The first group requires two uniform references from blackbody radiator at different temperatures, and by solving the system of equations, the gain and offset are obtained. The NU compensation is performed using the following equation:

$$\hat{X}_{ij}(n) = \frac{Y_{ij}(n) - \hat{B}_{ij}}{\hat{A}_{ij}}, \tag{2}$$

where $\hat{A}_{ij}$ and $\hat{B}_{ij}$ are the estimates of the gain and offset, and $\hat{X}_{ij}(n)$ is the estimated IR input irradiance.

The scene-based methods estimate gain and offset but the performance is limited by the amount of spatio-temporal information and the diversity of temperature in the image sequence. The condition of constant movement of the camera, to wich these methods are subject, is the main constraint for the scene-based methods. In addition, the requirement of a large number of frame also plays an important roll in a correct estimation.

A well-known scene-based NUC method is the constant statistics method proposed in Harris and Chiang (1999). The principal assumption of this method is that the the first and second moment of the input irradiance are equal to all sensors of FPA. Applying the mean and variance to equation (1), and assuming that the mean and standard deviation of X_n are 0 and 1, respectively, we can obtain the gain and offset from:

$$\hat{A}_{ij} = \sigma_{Y_{ij}} - \sigma_{v_{ij}}, \tag{3}$$

$$\hat{B}_{ij} = \mu_{Y_{ij}}. \tag{4}$$

In order to obtain a solution with low error estimation, it is required a good estimation of the mean and variance (large number of frames). To avoid this condition we assume that the readout data have a uniform distribution in a known range $[Y^{\min}, Y^{\max}]$, in such case the mean and variance can be obtained respectively from:

$$\mu_{Y_{ij}} = \frac{Y_{ij}^{\max} + Y_{ij}^{\min}}{2} \quad \wedge \quad \sigma_{Y_{ij}} = \frac{Y_{ij}^{\max} - Y_{ij}^{\min}}{\sqrt{12}}, \tag{5}$$

and then, the correction is performed using equation (2) with the values obtained from (3) and (4).

Another NUC method can by applied when the NU in the IR-FPA is mainly produced by the spatial variation in the offset, i.e., the gain $A_{ij}(n)$ is assumed a know constant given by the camera manufacturer. In this case, a one-point calibration is required in order to perform the NUC. If the offset is assumed constant for a particular time period t_c, it can be estimated recursively as:

$$\hat{B}_{ij}(n) = n^{-1} Y_{ij}(n) + n^{-1}(n-1)\hat{B}_{ij}(n-1), \tag{6}$$

and the FPN can be reduced by $Y_{ij}(n) - \hat{B}_{ij}(n)$. Also, the drift in time of the parameters t_d is such that satisfies $t_d \gg t_c$. This means that for a time greater than t_d a new estimation of the offset is necessary in order to perform an adaptive and continuous NUC.

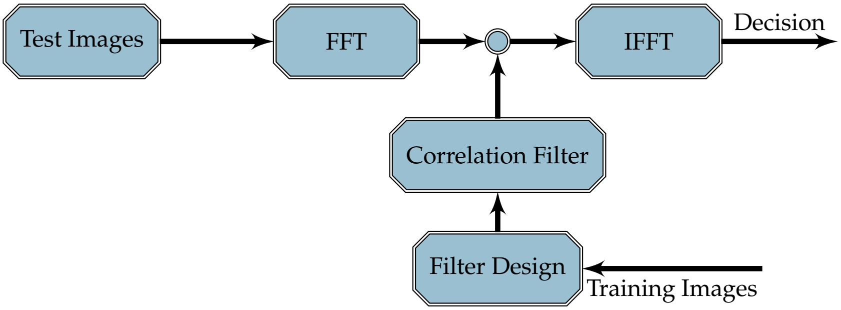


Fig. 1. The correlation process for face identification problems. Template correspond to the IR image set to design the correlation filters.

3. Face recognition methods

In this section, two traditional recognition methods are presented: correlation filters and local-binary pattern-based method. For both techniques a full image is used in order to build the classification rule. The aim of this is to evaluate the noise-tolerance of both methods when fixed and temporal noise are presented, degrading the quality of the acquired image in the input data.

3.1 Correlation filters

The correlation is a metric normally used to characterize the similarities between a reference pattern and a test pattern; this concept is widely use on recognition application, presenting a major degree of importance the use of the cross correlation obtaining the relative location of the object. Due to this analysis, the peak side lobe is consider as a classification metric. The cross correlation can by expressed as follow:

$$c(\tau_x, \tau_y) = \int\int T(f_x, f_y) R^*(f_x, f_y) e^{j2\pi(f_x \tau_x) + f_y \tau_y)} df_x df_y \tag{7}$$

$$= IFT\{T(f_x, f_y) R^*(f_x, f_y)\}, \tag{8}$$

where $R(f_x, f_y)$ and $T(f_x, f_y)$ are the 2D fourier transform of the reference and test pattern respectively. The equation 7 can be interpreted as the test pattern being filtered through a filter with a frequency response $H(f_x, f_y) = R(f_x, f_y)$, generating the $c(\tau_x, \tau_y)$ (Fig. 1)

One of the many application of this concept correspond to the face recognition Kumar et al. (2004), allowing to perform the classification at the speed of light in an optical laser-based system, generating a peak of information for a successful recognition, fixing a threshold of approval for a negative result Kumar (1992).

The fourier transform in the correlation allows the use of displaced or clipped images because it has not have influence on the discernment of the classification system, which favors the image preprocessing. The peak response just indicates the level of displacement without directly affecting the magnitude.

3.2 Filter types

The correlation filter generation procedure depend on the images training set included in the database, which must be selected maintaining a format that is as representative as possible, depending on the image size, certain variations in facial expressions, the face position, etc. There are several types of filters, including: phase-only filter, minimum average correlation energy, noise-tolerant, and optimal tradeoff filter.

3.2.1 POF (Phase-Only Filter)

One advantage of this filter is because it only uses one image per subject for the filter generation, represented by the following expression Horner & Gianino (1984):

$$POF \; : \; h = \left(\frac{FT\{r(x,y)\}}{|FT\{r(x,y)|\}} \right)^{*}, \tag{9}$$

but is necessary to consider that h has a low sensitivity to potential changes on the test images.

3.2.2 MACE (Minimum Average Correlation Energy)

The MACE filter Casasent & Ravichandran (1992) is designed with the objective of minimizing the average energy in the correlation plane resulting from the training images, restricting the value at the origin with respect to a preset value. The solution is represented by:

$$MACE \; : \; h = D^{-1}X(X^{*}D^{-1}X)^{-1}u, \tag{10}$$

where u is a row vector containing the desired values for the correlation peak, X is the complex matrix where each column is a vectorized training image and matrix D is the diagonal matrix containing the spectral power of training images.

3.2.3 NTC (Noise-Tolerant Correlation)

Because the MACE filter generates a correlation peak more notorious, tends to amplify the high spatial frequencies and any input noise. It is for this reason that the NTC filter aims to minimize the sensitivity of the filter represented by:

$$NTC \; : \; h = C^{-1}X(X^{*}C^{-1}X)^{-1}u, \tag{11}$$

diagonal matrix C is defined as the spectral power density of noise.

3.2.4 TOF (Optimal Tradeoff Filter)

The TOF algorithm Réfrégier (1993) provides a compromise between the features of the MACE filter, which accentuates the peak for positive results and the NTC filter, which aims to reduce the variance of the output noise. In this case, it is necessary to assume the presence of white noise in order to approximate the matrix C as the identity matrix, with a factor α equal to 0.99. Each filter is defined as:

$$NTC \; : \; h = T^{-1}X(X^{*}T^{-1}X)^{-1}u \tag{12}$$

$$T = \alpha D + (1-\alpha)C. \tag{13}$$

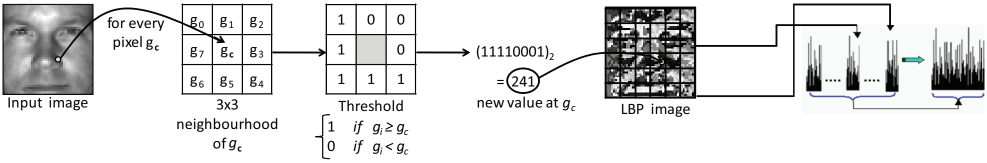


Fig. 2. Local Binary Patterns.

For analysis purposes, it will be use the TOF because of its robustness to evaluate the performance.

3.3 Local binary patterns

The use of the LBP operator in face recognition was introduced in Ahonen et al. (2004) and different extensions of the original operator have appeared afterwards Marcel et al. (2007). As it can be appreciated in Figure 2, the original LBP operator represents each pixel of an image by thresholding its 3x3- neighborhood with reference to the center pixel value, g_c, and considering the result as a binary number, called the LBP code. The image is then divided into rectangular regions and histograms of the LBP codes are calculated over each of them. Finally, the histograms of each region are concatenated into a single one that represents the face image. The Chi-square dissimilarity measure is used to compare the histograms of two different images.

3.4 Performance evaluation

In the diverse studies the analysis of the classification results depends on the characteristics of the system according to the structure that the database possesses; one of the most common is the peak-to-sidelobe ratio.

3.4.1 PSR (Peak-to-Sidelobe Ratio)

The correlation results can be compared by calculating the following Kumar & Hassebrook (1990):

$$PSR : p = \frac{peak - \mu}{\sigma}, \tag{14}$$

where the peak corresponds to the higher correlated output amplitude, the mean and standard deviation correspond to an outer area of fixed size around the peak resulting in the output image.

3.4.2 PCE (Peak-to-Correlation Energy)

A more accurate way of characterize the correlated output is through the calculation of Kumar & Hassebrook (1990):

$$PCE : p = \frac{|c(0,0)|^2}{\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |c(x,y)|^2 dxdy}, \tag{15}$$

which is defined as the ratio of the correlation peak $c(0,0)$ and the energy present in the correlation plane. The PCE value approaches infinity as c approaches a delta function. Thus, as desired, larger PCE values imply a sharper correlation peaks. However, the major reason for using the PCE over other available peak sharpness measures is its analytical convenience.

Since an identification measure only provides information if a classification is correct or incorrect, it is necessary to include some metric that allow to verify the performance of the classification itself. Typically face verification measures whether a subject is valid or not for a given set, that is why we define two sets of database, subjects valid and imposters, measured by the FAR and FRR.

3.4.3 FAR

The false acceptance rate points to the statistical calculation of those subjects classified correctly without belonging to the data base, represented by:

$$FAR = \left(\frac{\text{impostors accepted as valid subjects}}{\text{total number of invalid subjects}} \right). \tag{16}$$

3.4.4 FRR

The false rejection rate is a statistical complement of the previous measure, indicating the number of subjects rejected during the identification and belong to the database:

$$FRR = \left(\frac{\text{valid subjects rejected}}{\text{total number of valid subjects}} \right). \tag{17}$$

The FRR and FAR are directly affected by the threshold value defined in the identification. The variation of this value causes these rates vary inversely, being able to appreciate this effect in the distribution features seen in Fig. 3.

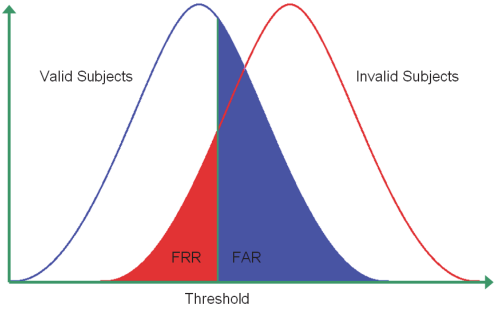


Fig. 3. Distribution of FAR and FRR for valid and invalid subjects.

For each threshold value, there is FAR and FRR defined as the identification, in an ideal case, the two curves should overlap at least as possible. Dependence between these two values can best be seen in Fig. 4, dependent on the statistical values calculated for value defined threshold, where the intersection on the curve indicates the critical threshold value to obtain an equal error rate (EER) for a $FRR = FAR$.

4. Face identification using Equinox infrared imagery

The Equinox Corporation database is composed of three 40 frame sequences from 90 persons, acquired in two days with three different light sources: frontal, left lateral and right lateral. The frame sequences were recorded while people were uttering vowels standing in a frontal pose, and three more images from each person were taken to capture the expression of,

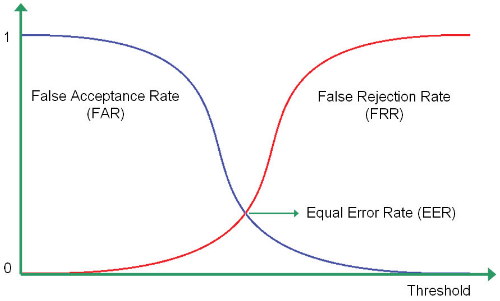


Fig. 4. EER as an intersection between FAR an FRR curves.

respectively to smile, frown and surprise. In addition, the complete process was repeated for those persons who wore glasses.

The LWIR images of Equinox are of 320x240 pixels size, they were represented as gray-scale images with 12 bits per pixels. The blackbody images are not available from Equinox (2009). Since much of the data is highly correlated, usually only a subset of the images is used for experimentation Heo et al. (2005); Socolinsky and Selinger (2002). In LWIR is not easy to precisely detect the facial features, so the images were not geometrically normalized and the different face images of one person are not aligned as can be appreciated in Figure 5. The size of the windows to divide the images for the LBP method was selected as 18x21 pixels taking into account this fact without decreasing the recognition performance.

4.1 FPN removal procedure

Considering the database Equinox (2009) it is possible to assume that the NU does not change in time. For instance, analyzing some images of two individuals (Fig. 5a and 5d) is possible to note that contain two typical distortions presents in IR imagery: dead-pixel and FPN. The first source of noise means that the detector always gives the same readout value independent of the input irradiance. The second is a FPN present in several sequences of individuals, but it is not possible to explain the nature of this FPN.

In order to remove the dead-pixel, it is possible to assume that the IR irradiance collected by the sensor ij is to be close to the neighbors around the sensor ij, and this value can be assumed the readout data. The FPN can be estimated using the equation (6) and then, reduced by perform the NUC process. Previously, the spatial offset of each image was removed resulting as shows Fig. 5b and 5e. The NUC images are shown in Fig. 5c and 5f, respectively. Note that the final images maintain the 320 × 240 resolution of the IR images of the database.

4.2 Data set for gallery and test

Following the procedure presented by Bebis et al. (2006); Socolinsky and Selinger (2002) we define multiple subsets using only three images of the vocal pronunciation frame sequence (vowel frames) and the three expression frames of each subject in each illumination condition:

VA: Vowel frames, all illuminations.
EA: Expression frames, all illuminations.
VF: Vowel frames, frontal illumination.
EF: Expression frames, frontal illumination.

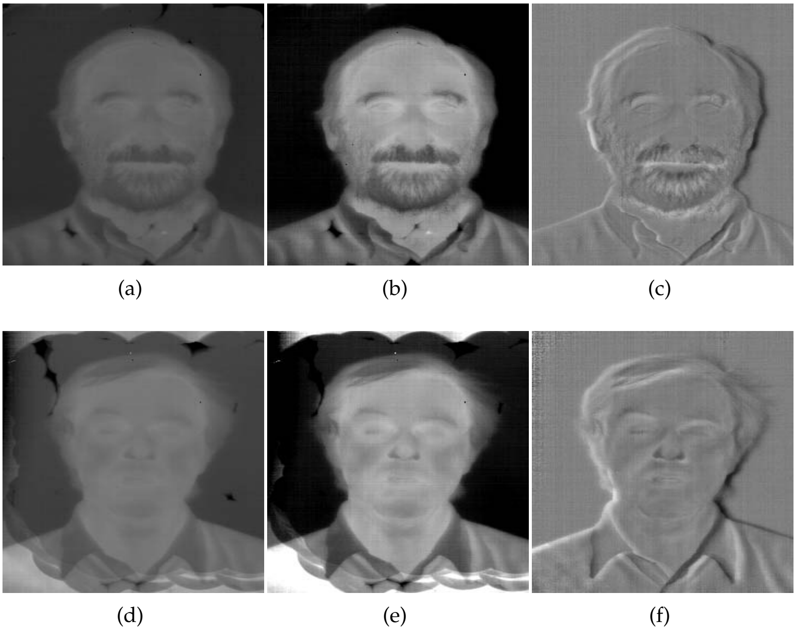


Fig. 5. Full size images of subject 2417 (a) and 2434 (d) from Equinox (2009) database. (b)(e) correspond to the spatial offset adjustment and (c) and (f) is with FPN removed. Its clear that the quality of the images are improved maintaining the 320x240 original spatial resolution.

VL: Vowel frames, lateral illuminations.
EL: Expression frames, lateral illuminations.
The performance of the correlation filter and LBP based matching are evaluated by using, each time, one set as gallery and another as a test set. Some of the subset combinations are not considered in the experiments since one subset is included in the other. Table 1 shows the performance of the evaluated methods.

4.3 Results using fixed-pattern-noise removal

From the first two rows of 1, corresponding to the use POF and TOF filter, is possible to appreciate the CC of 85.74% and 59.30% respectively, over the the original IR images. This implies that the performance of the POF filter in the original images is better than the TOF filter.

The LBP method on the other hand, presents an average of 97.3% of correct classification. Note that images are not geometrically normalized or cropped, and even a small localization error usually affects the appearance based methods. This performance is comparable with the best performance of an appearance based method obtained earlier with the same LWIR images by means of the Linear Discriminant Analysis (LDA) method Belhumeur et al. (1997), reported in Socolinsky and Selinger (2002) and summarized in Table 1. In this case, LDA achieves an average of 97.5% of correct classification. However, LBP has the advantage over the LDA method because it only needs one image per person in the gallery set, neither it requires a training set, which are very important properties of a face recognition system.

In order to improve the results and following the idea in San Martin et al. (2008), we applied the NU correction method discussed in section 2.2 prior to applying POF filter, TOF filter and the LBP representation, aiming to suppress the FPN present in the IR images. As can be appreciated in the second part of Table 1, the performance of POF and TOF filters with NUC in terms of CC are 97.93% and 99.60% respectively. Note that the performance of both filters is improved by using NUC method and the TOF filter CC is better than the one of POF filter. Surprisingly, with an average value of 93.3 percent, the performance with the NU correction using LBP was lower than without it. Inspecting the original and NU corrected images in Figure 5, it is apparent that although the NUC method suppresses fixed-pattern noise in the IR images, the random noise is magnified and the image texture is affected. Since LBP is a texture descriptor, it is sensitive to this kind of noise.

In order to support the hypothesis that the LBP method is sensitive to temporal random noise in LWIR images, we conducted the same experiments adding some random noise artificially to the original images. Table 5 displays the results of the experiments adding the random noise to the original LWIR images, with an average of 86.3 percent they confirm that LBP method decreases its performance in the presence of this kind of noise.

		without NUC						with NUC					
		VA	EA	VF	EF	VL	EL	VA	EA	VF	EF	VL	EL
POF filter	VA		84.64	83.77	84.39	82.40	84.14		98.13	99.13	98.50	98.50	98.88
	EA	84.08		85.58	89.33	87.08	89.33	97.94		97.94	98.13	97.94	97.75
	VF	89.76	81.90		84.27	80.77	84.27	100.00	98.25		98.75	97.50	99.25
	EF	97.19	80.52	91.39		85.58	91.76	99.44	96.44	97.75		96.07	97.94
	VL	83.52	85.27	82.77	84.52		83.77	99.75	98.50	98.88	98.75		99.00
	EL	81.27	90.07	84.46	86.70	87.64		95.13	96.44	95.51	95.69	95.88	
TOF filter	VA		54.56	55.31	54.81	53.43	55.68		100.00	99.75	100.00	99.88	99.88
	EA	52.25		56.74	67.79	64.42	60.30	99.63		99.63	100.00	99.81	99.81
	VF	60.05	48.69		52.18	51.31	55.93	100.00	100.00		100.00	100.00	100.00
	EF	88.20	53.18	73.22		60.49	78.65	99.81	99.81	99.44		99.25	100.00
	VL	52.93	55.81	54.68	54.81		53.93	99.88	100.00	99.75	100.00		99.88
	EL	51.12	60.30	67.60	54.68	76.04	98.13	98.13	98.31	98.88	97.75	98.88	
LBP	VA		98.13	97.75	98.50	99.25	95.13		88.89	95.63	87.89	95.38	90.14
	EA	97.67		93.77	98.44	97.67	100.00	97.67		94.94	93.77	98.05	97.67
	VF		98.13		97.00	97.38	95.51		85.77		92.51	90.26	86.02
	EF	99.24		99.62		99.24	95.08	98.48		98.86		98.48	94.70
	VL		98.13	96.63	96.63		96.63		91.26	93.63	83.65		90.76
	EL	97.74		92.83	95.09	98.49	96.60		93.96	93.21	98.11		
LDA	VA		99.60	98.30	96.20	99.60	99.30						
	EA	97.40		94.00	98.10	96.80	99.20						
	VF		100.00		97.00	98.80	98.60						
	EF	97.10		94.60		95.60	97.90						
	VL		99.50	97.40	95.80		99.60						
	EL	97.40		93.70	97.10	97.40							

Table 1. Correct classification percent with POF filter, TOF filter, LBP and LDA.

	VA	EA	VF	EF	VL	EL
VA		90.26	86.64	86.77	92.26	90.01
EA	87.03		81.06	80.16	87.55	91.05
VF		82.27		89.89	86.14	85.27
EF	89.14		88.76		85.61	83.33
VL		89.01	83.02	82.77		88.64
EL	86.79		80.50	82.64	87.92	

Table 2. Classification results with the LBP method adding random noise.

5. Face recognition performance using CEDIP JADE UC infrared imagery

The second database are collected using the CEDIP JADE infrared camera, with a focal plane array composed on type of detector capable of working without a cooling system known as microbolometer aSi and tuned at $8 - 14\mu m$. This data set consists of 6 images for 102 subjects considering vowel and expression variations, without illumination controls, generating the following codification:
V: Vowel frames.
E: Expression frames.
The infrared image are corrected using two-point calibration method, using the Mikron M345 black bodie radiatior. In this case, the goals is considering the following simulated scenarios:
- IR image without FPN and temporal noise.
- IR image with FPN with variance equal to 10, 20 and 30 percent.
- IR image with temporal noise with variance equal to 1, 5 and 10 percent.

In all cases, identification and verification problems are considered. In the first case, V (E) images are used as gallery and E (V) as test. In second experiment, the data set is divided in order to calculate the FAR, FRR and EER, and then, obtain the behavior of the classifications.

5.1 Identification results

In this experiment, a white noise is added in order to simulate the FPN and temporal noise. For all cases, five realization for each kind of noise and variance are considered, and the mean value is assumed as the mean correct classification using POF filter, TOF filter and LBP methods, respectively. In Table 3 the results considering FPN with 10, 20 and 30 percent are presented. In this case, the correlations filters strongly decrease their performance when the FPN noise variance increases. For the LBP method, the performance remains with low variability from the images without FPN noise. This mean that the LBP based method is robust to the FPN, not requiring a corrected infrared image.

5.2 Verification results

For this experiment, the database is divided into a training set composed of images of 34 subjects as clients, an evaluation set with images of the same subjects and a test set with 68 subjects. Three images per person are used for training, and the number of clients accesses is equal to $34x3 = 102$, i.e., the other three images of clients. The imposters accesses is given by the other 68 subjects of the database, and the number of accesses or comparisons can be summarized as:

		original image		FPN $\sigma = 10\%$		FPN $\sigma = 20\%$		FPN $\sigma = 30\%$	
		V	E	V	E	V	E	V	E
POF filter	V		91.50		91.50		91.44		91.57
	E	88.24		88.10		88.17		87.97	
TOF filter	V		98.69		84.12		81.18		76.47
	E	94.77		61.70		52.75		51.05	
LBP	V		99.35		97.32		97.19		97.52
	E	97.71		96.93		96.41		96.01	

Table 3. Correct classification percent with POF filter, TOF filter and LBP considering FPN.

		original image		noise $\sigma = 1\%$		noise $\sigma = 5\%$		noise $\sigma = 10\%$	
		V	E	V	E	V	E	V	E
POF filter	V		91.50		91.44		91.50		91.18
	E	88.24		88.24		88.24		88.17	
TOF filter	V		98.69		97.26		95.29		89.61
	E	94.77		94.44		92.81		88.17	
LBP	V		99.35		98.37		97.06		96.93
	E	97.71		96.86		96.80		95.23	

Table 4. Correct classification percent with POF filter, TOF filter and LBP considering temporal noise.

	Evaluation	Test
Clients accesses	102(34x3)	204(68x3)
Imposters accesses	3366(102x33)	13668(204x67)

Table 5. Distribution of the database to perform the verifications experiment.

		original image		FPN $\sigma = 10\%$		FPN $\sigma = 20\%$		FPN $\sigma = 30\%$	
		Evaluation	Test	Evaluation	Test	Evaluation	Test	Evaluation	Test
TOF	V	7.84	14.48	20.64	25.51	23.67	27.16	25.74	28.16
	E	11.76	14.72	34.63	43.00	34.62	43.01	38.29	44.45
LBP	V	3.92	5.42	7.45	7.15	7.50	7.74	8.48	8.41
	E	3.92	5.42	8.56	8.62	9.45	9.56	9.34	10.25

Table 6. Average Total Error Rate percent for the verifications experiment with TOF filter and LBP method considering FPN.

		original image		noise $\sigma = 1\%$		noise $\sigma = 5\%$		noise $\sigma = 10\%$	
		Evaluation	Test	Evaluation	Test	Evaluation	Test	Evaluation	Test
TOF	V	7.84	14.48	8.04	18.01	9.23	18.58	9.61	19.02
	E	11.76	14.72	12.54	15.41	12.44	15.46	16.09	20.60
LBP	V	3.92	5.42	3.92	5.10	6.00	6.77	9.34	10.51
	E	3.92	5.42	8.11	9.07	5.68	6.40	11.04	13.17

Table 7. Average Total Error Rate percent for the verifications experiment with TOF filter and LBP method considering temporal noise.

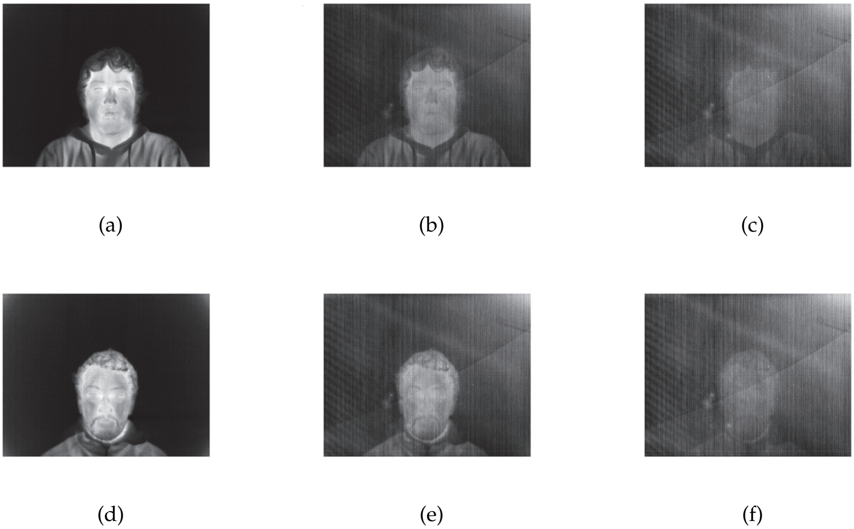


Fig. 6. Images of two different subjects from CEDIP JADE database: image without NUC (a)(d), with FPN 10% (b)(e) and FPN 30% (c)(f). The LBP method can be recognize the subject (c) and (f) with the same performance that the case (b) and (e).

The EER is the point at which the FRR is equal to the FAR in the evaluation set. The value obtained by the classification method at this point in the evaluation set is used as a threshold for the decision of acceptance or rejection in the test set. On the other hand, the Total Error Rate (TER) is the sum of FRR and FAR. The TER is used to evaluate the performance of the verification systems on the database. The lower this value, the better the recognition performance.

6. Remarks

From the results obtained using two infrared databases is possible to confirm that the LBP method is more robust than correlation filter when FPN is present, keeping in mind that this type of noise is natural in IR images, but from the results exposed is possible to consider that LBP does not require to apply NUC methods. An example is presented in Fig.6 (a) and (b), where two subjects images, initially noise-free, are afflicted due to the intrinsic properties of the FPN, drifting its parameters over time, decreasing the image quality, showing its effect on (b) and (e), and then (c) and (f). In this case, the results presented in this chapter allow to recognize the subject in image 6 (c) and (f) with the same error that (b) and (e), respectively. These results empirically demonstrate the ability of LBP method in order to perform good recognition task in infrared image with aggressive FPN.

7. Conclusion

In this chapter, is presented a review of various face recognition algorithms applied to an infrared database in order to evaluate the performance of each one, related with the problems associated by working on this range of the spectrum. In particular, the use of infrared technology allows to formulate a face recognition system invariant to the illumination. For this, the long-wave spectra located at $8 - 14\mu m$ correspond to the emitted energy contraries with the visible spectra, in which the sensor collected the reflected energy. But the long-wave infrared spectra generate an intrinsic and special kind of noise (called nonuniformity), the fixed-pattern-noise (FPN) that correspond to a fixed pattern superimposed to the infrared image. In order to solve this problem, several nonuniformity correction techniques has been proposed, been an active area in the last years. The FPN is a pattern that slowly change in time. In several study, this kind of noise can be considered fixed or constant for almost two hours, but is depending of the technology of the infrared sensors.

In pattern recognition fields, is not habitual to study the noise robustness of classification techniques. A little kind of study consider only temporal noise, i.e., a pattern noise that change frame to frame, typically modeled as white noise. Our group include optoelectronic group and pattern recognition group, and the principal contribution is to present an infrared sensor description and the behavior of face recognition techniques considering this kind of sensors. In this chapter, firstly is introduced the concepts of FPN noise, and the objective is to study the behavior of two classical face recognition techniques when the nonuniformity of infrared cameras is not reduced or compensated. Then, the variation on the performance of each technique with and without nonuniformity correction method is obtained using two infrared databases. For both face recognition techniques identification and verification are considered.

Two face recognition techniques are used: correlation filters and local-binary pattern (LBP) based method. The results show that the LBP algorithm is one of the most robust to fixed pattern noise due to the inherent features of the correlation filters, the presence of any persistent pattern noise in time diminish the performance of the classification. In other hand, the temporal noise does not affect on a comparable level with the FPN, but the LBP still shows outstanding results.

As a counterpart to the analysis obtained by the different realizations of noise, it is clear that nonuniformity corrections generate an improvement over the correlation filters discernment, achieving competitive results with the other algorithms presented. Considering the case when TOF filter is used in Table 6, the TER for IR face recognition is 25.74%, and when the FPN is removed from images, the TER is reduced to 7.84%. Under these conditions, the techniques of non-uniformity correction behave like a promising feature extraction technique, extracting features to effectively and efficiently differentiate a subject from another.

Future work considering to build another database with more images per subject and repeat the process considering two weeks between acquisitions. The objective is to evaluate the robustness of LBP when real FPN is presented in the camera considering the past of time. Also, more face recognition should be included in order to find more FPN robustness techniques to perform face identification.

8. Acknowledgments

This work was partially supported by Center for Optics and Photonics CEFOP, FB0824/2008, Fondecyt 11090034, Proyecto de Iniciación en Investigación 2009, CONICYT, Chile, and by Universidad de La Frontera DIUFRO N^o DI08-15.

9. References

Ahonen, T.; Hadid, A., & Pietikäinen, M. (2004) Face recognition with local binary patterns. *Lecture Notes on Computer Science* LNCS 3021, 469-481.

Armstrong, E.; Hayat, M.; Hardie, R.; Torres, S.; & Yasuda, B. (1999), Nonuniformity correction for improved registration and high-resolution image reconstruction in IR imagery. *Proceeding of SPIE*, 3808, 150-161.

Averbuch, A.; Liron, G. & Bobrovsky, B. (2008), Scene based non-uniformity correction in thermal images using Kalman filter. *Image and Vision Computing*, 25, 833-851.

Bebis, G.; Gyaourova, A.; Singh, A.; & Pavlidis, I. (2006), Face recognition by fusing thermal infrared and visible imagery. *Image and Vision Computing*24, 7, 727-724.

Belhumeur, P.; Hespanha, J. & Kriegman, D. (1997) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.*,19,7, 711Ű720.

Casasent, D.; & Ravichandran, G. (1992), Advanced distortion-invariant minimum average correlation energy (MACE) filters, *App. Opt.* 31, (8), 1109-1116.

Chen, X.; Flynn, P.; Bowyer, & K. (2003), PCA-based face recognition in infrared imagery: baseline and comparative studies. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG.* 127-134.

Chen X.; Flynn P. & Bowyer K. (2005). IR and visible light face recognition, *Computer Vision and Image Understanding*, Vol. 99, No. 3, 332-358, ISSN 1077-3142.

Dowdall J.; Pavlidis I. & Bebis G. (2003). Face Detection in the Near-IR Spectrum. *Image and Vision Computing*. Vol. 21, No. 7, 565-578, ISSN 0262-8856.

Equinox IR Database, (2009) http://www.equinoxsensors.com/products/HID.html.

Friedrich G. & Yeshurun Y. (2002).Seeing People in the Dark: Face Recognition in Infrared Images. *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, pp. 348-359, Vol. 252, 2002.ISBN 3-540-00174-3.

Hardie, R.; Hayat, M.; Armstrong, E.; & Yasuda, B. (2000), Scene-based nonuniformity correction using video sequences and registration. *Applied Optic*, 39, 1241-1250.

Harris, J.; & Chiang, Y. (1999), Nonuniformity correction of infrared image sequences using constant statistics constraint, *IEEE Trans. on Image Processing*, 8, 1148-1151.

Heo, J.; Savvides, M.; & Kumar, V. (2005), Performance Evaluation of Face Recognition using visual and thermal imagery with advanced correlation filters, *Conference on Computer Vision and Pattern Recognition, IEEE Computer Society*, pp.9-14.

Horner, J. & Gianino, P. (1984), Phase-only matched filter, *Applied Optics*, 23, 812-816 .

Kong, S.; Heo, J.; Abidi, B.; Paik, J.; & and Abidi, M. (2005), Recente advances in visual and infrared face recognition - a review. *Computer Vision and Image Undertanding*, 97, 1, 103-135.

Kumar, B. (1992), Tutorial survey of composite filter designs for optical correlators. *Appl. Opt.*, 31, 4774-4801.

Kumar, B.; Savvides, M.; Xie, C.; Venkataramani, K.; Thornton, J.; & Abhijit Mahalanobis (2004), Biometric Verification with Correlation Filters, *Appl. Opt.*, 43, 2, 391-402.

Kumar, B.; & Hassebrook, L. (1990), Performance measures for correlation filters. *App. Opt.* 29, 20, 2997-3006.

Li, S.; Chu, R.; Liao, S. & Zhang, L. (2007), Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29, 4, 627-639.

Liu, X.; Fang, H.; & Liu, L. (2007), Study in new structure uncooled a-Si microbolometer for infrared detection. *Microelectronics Journal*, 38, 735-739.

Marcel, S.; Rodriguez, Y., & Heusch, G. (2007), On the Recent Use of Local Binary Patterns for Face Authentication. *International Journal on Image and Video Processing Special Issue on Facial Image Processing*.

Milton, A.; Barone, F.; & Kruer, M. (1985), Influence of nonuniformity on infrared focal plane array performance. *Opctical Engineering*, 24, 855-862.

Mooney, J.; Shepherd, F.; Ewing, W.; Murguia, J.; & Silverman, J. (1989), Responsivity nonuniformity limited performance of infrared staring cameras. *Optical Engineering*, 28, 1151-1161.

Narendra P. & Foss N. (1981). Shutterless fixed pattern noise correction for infrared imaging arrays. *Proceeding of SPIE*, pp. 44-51, Vol. 282, Washington, DC, USA, April 1981.

Perry, D.; & Dereniak, E. (1993), Linear theory of nonuniformity correction in infrared staring sensors. *Optical Engineering*, 32, 1854-1859.

Pezoa, J.; Hayat M.; Torres, S. & Rahman, M. (2006), Multimodel kalman filtering for adaptive nonuniformity correction in infrared sensors, *The JOSA-A Opt. Soc. of America*, 23, 1282-1291.

Piotrowski, J.; & Rogalski, A. (2004), Uncooled long wavelength infrared photon detectors. *Infrared Physics and Technology*, 46, 151-131.

Pron, H.; Menanteau, W.; Bissieux, C.; Beaudoin, J. (2000), Characterization of a focal plane array (FPA) infrared camera. *Quantitativa Infrared Thermography QIRT Open Archives, QIRT 2000*, http://qirt.gel.ulaval.ca.

Ratliff, B.; Hayat, M.; & Hardie, R. (2002), An algebraic algorithm for nonuniformity correction in focal-plane arrays. *The JOSA-A Opt. Soc. of America*, 19, 1737-1747.

Ratliff, B.; Hayat, M.; & Tyo, J. (2005), Generalized algebraic scene-based nonuniformity correction algorihtm. *The JOSA-A Opt. Soc. of America*, 22, 239-249.

Réfrégier, R. (1993). Optimal trade-off filter for noise robustness, sharpness of the correlation peaks, and Horner efficiency. *Opt. Lett.*, 32, 1933-1935.

San Martin, C.; Meza, P.; Torres, S. & Carrillo, R. (2008), Improved Infrared Face Identification Performance Using Nonuniformity Correction Techniques. *Lecture Notes on Computer Science*, LNCS 5259, 1115-1123.

Scribner, D.; Sarkady, K.; & Kruer, M. (1991), Adaptive nonuniformity correction for infrared focal plane arrays using neural networks.*Proceeding of SPIE*, 1541, 100-109.

Scribner, D.; Sarkady, K.; & Kruer, M.; (1993), Adaptive retina-like preprocessing for imaging detector arrays. *Proceeding of the IEEE International Conference on Neural Networks*, 3, 1955-1960.

Singh, R.; Vatsa, M. & Noore, A. (2008), Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recogn.* 41, 3, 880-893.

Socolinsky, D. & Selinger, A. (2002), A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery. In Icpr'02: *Proceedings of the 16Th International Conference on Pattern Recognition*, 4, 4.

Socolinsky, D.; Wolff, L.; Neuheisel, J.; & Eveland, C. (2001) Illimination invariant face recognition using thermal infrared imagery. *Proc. IEEE CS Conf. Comp. Vision and Pattern recognition*, 1, 527-534.

Torres, S. & Hayat, M. (2002), Kalman filtering for adaptive nonuniformity correction in infrared focal plane arrays. *The JOSA-A Opt. Soc. of America*, 20, 470-480.

Torres, S.; Pezoa, J. & Hayat, M. (2003), Scene-based nonuniformity correction for focal plane arrays using the method of the inverse covariance form, *Applied Optics*, 42, 5872-5881.

Zhou, H.; Lai, R.; Qian, S.; & Jiag, G. (2005), New improved non uniformity correction for infrared focal plane arrays. *Opctis Communications*, 245, 49-53.

Zou, X.; Kittler J. & Messer K. (2005), Face recognition using active near-ir illumination. *British Machine Vision Conference Proceedings*.

Zou, X.; Kittler J. & Messer K. (2007), Illumination Invariant Face Recognition: A Survey. In *First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 1-8.

6

Thermal Infrared Face Recognition – a Biometric Identification Technique for Robust Security System

Mrinal Kanti Bhowmik[1], Kankan Saha[1], Sharmistha Majumder[1],
Goutam Majumder[1], Ashim Saha[1], Aniruddha Nath Sarma[1],
Debotosh Bhattacharjee[2], Dipak Kumar Basu[2] and Mita Nasipuri[2]
[1]*Tripura University (A Central University)*
[2]*Jadavpur University*
[1,2]*India*

1. Introduction

Face of an individual is a biometric trait that can be used in computer-based automatic security system for identification or authentication of that individual. While recognizing a face through a machine, the main challenge is to accurately match the input human face with the face image of the same person already stored in the face-database of the system. Not only the computer scientists, but the neuroscientists and psychologists are also taking their interests in the field of development and improvement of face recognition. Numerous applications of it relate mainly to the field of security. Having so many applications of this interesting area, there are challenges as well as pros and cons of the systems. Face image of a subject is the basic input of any face recognition system. Face images may be of different types like visual, thermal, sketch and fused images. A face recognition system suffers from some typical problems. Say for example, visual images result in poor performance with illumination variations, such as indoor and outdoor lighting conditions, low lighting, poses, aging, disguise etc. So, the main aim is to tackle all these problems to give an accurate automatic face recognition. These problems can be solved using thermal images and also using fused images of visual and thermal images. The image produced by employing fusion method provides the combined information of both the visual and thermal images and thus provides more detailed and reliable information which helps in constructing more efficient face recognition system. Objective of this chapter is to introduce the role of different IR spectrums, their applications, some interesting critical observations, available thermal databases, review works, some experimental results on thermal faces as well as on fused faces of visual and thermal face images in face recognition field; and finally sorting their limitations out.

2. Thermal face recognition

Any typical face image is a complex pattern consisting of hair, forehead, eyebrow, eyes, nose, ears, cheeks, mouth, lips, philtrum, teeth, skin, and chin. Human face has other additional features like expression, appearance, adornments, beard, moustache etc. The face is the feature which best distinguishes a person, and there are "special" regions of the human brain, such as the fusiform face area (FFA), which when get damaged prevent the recognition of the faces of even intimate family members. The patterns of specific organs such as the eyes or parts thereof are used in biometric identification to uniquely identify individuals.

Thermal face recognition deals with the face recognition system that takes thermal face as an input. In preceding description, the concept of thermal images will be made clearer. Thermal human face images are generated due to the body heat pattern of the human being. Thermal Infra-Red (IR) imagery is independent of ambient lighting conditions as the thermal IR sensors only capture the heat pattern emitted by the object. Different objects emit different range of Infra-red energy according to their temperature and characteristics. The range of human face and body temperature nearly same and quite uniform, varying from 35.5°C to 37.5°C providing a consistent thermal signature. The thermal patterns of faces are derived primarily from the pattern of superficial blood vessels under the skin. The vein and tissue structure of the face is unique for each person and, therefore, the IR images are also unique. Fig. 1 shows a thermal image corresponding to its visual one.

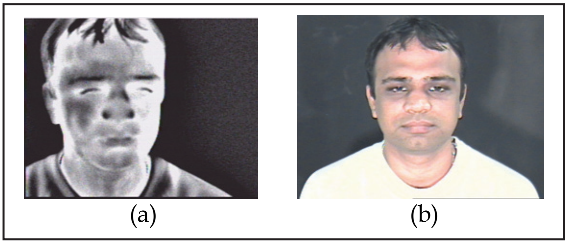


Fig. 1. (a) Thermal Image of Corresponding Visual One of (b).

2.1 Infrared

In Latin 'infra' means "below" and hence the name 'Infrared' means below red. 'Red' is the color of the longest wavelengths of visible light. Infrared light has a longer wavelength (and so a lower frequency) than that of red light visible to humans, hence the literal meaning of below red.

'Infrared' (IR) light is electromagnetic radiation with a wavelength between 0.7 and 300 micrometers, which equates to a frequency range between approximately 1 and 430 THz. IR wavelengths are longer than that of visible light, but shorter than that of terahertz radiation microwaves.

2.1.1 Infrared bands and thermal spectrum

Objects generally emit infrared radiation across a spectrum of wavelengths, but only a specific region of the spectrum is of interest because sensors are usually designed only to collect radiation within a specific bandwidth. As a result, the infrared band is often subdivided into smaller sections.

The *International Commission on Illumination* (CIE) recommended the division of infrared radiation into three bands namely, IR-A that ranges from 700 nm–1400 nm (0.7 µm – 1.4

µm), IR-B that ranges from 1400 nm–3000 nm (1.4 µm – 3 µm) and IR-C that ranges from 3000 nm–1 mm (3 µm – 1000 µm).

A commonly used sub-division scheme can be given as follows:

Near-infrared (NIR, IR-A DIN): This is of 0.7-1.0 µm in wavelength, defined by the water absorption, and commonly used in fiber optic telecommunication because of low attenuation losses in the SiO_2 glass (silica) medium. Image intensifiers are sensitive to this area of the spectrum. Examples include night vision devices such as night vision camera.

Short-wavelength infrared (SWIR, IR-B DIN): This is of 1-3 µm. Water absorption increases significantly at 1,450 nm. The 1,530 to 1,560 nm range is the dominant spectral region for long-distance telecommunications.

Mid-wavelength infrared (MWIR, IR-C DIN) or Intermediate Infrared (IIR): It is of 3-5 µm. In guided missile technology the 3-5 µm portion of this band is the atmospheric window in which the homing heads of passive IR 'heat seeking' missiles are designed to work, homing on to the IR signature of the target aircraft, typically the jet engine exhaust plume.

Long-wavelength infrared (LWIR, IR-C DIN): This infrared radiation band is of 8–14 µm. This is the "thermal imaging" region in which sensors can obtain a completely passive picture of the outside world based on thermal emissions only and require no external light or thermal source such as the sun, moon or infrared illuminator. Forward-looking infrared (FLIR) systems use this area of the spectrum. Sometimes it is also called "far infrared".

Very Long-wave infrared (VLWIR): This is of 14 - 1,000 µm.

NIR and SWIR is sometimes called "reflected infrared" while MWIR and LWIR is sometimes referred to as "thermal infrared". Due to the nature of the blackbody radiation curves, typical 'hot' objects, such as exhaust pipes, often appear brighter in the MW compared to the same object viewed in the LW.

Now, we can summarize the wavelength ranges of different infrared spectrums as in Table 1.

Spectrum	Wavelength range
Visible Spectrum	0.4-0.7 µm (micro meter / micron)
Near Infrared (NIR)	0.7-1.0 µm (micro meter / micron)
Short-wave Infrared (SWIR)	1-3 µm (micro meter / micron)
Mid-wave Infrared (MWIR)	3-5 µm (micro meter / micron)
Long-wave Infrared (LWIR)	8-14 µm (micro meter / micron)
Very Long-wave Infrared (VLWIR)	> 14 µm (micro meter / micron)

Table 1. Wavelength Range for Different Spectrums (Miller, 1994).

Developments in infrared technology (camera) over the last decade have given computer vision researchers a whole new diversity of imaging options, particularly in the infrared spectrum. Conventional video cameras use photosensitive silicon that is typically able to measure energy at electromagnetic wavelengths from 0.4µm to just over 1.0µm. Multiple technologies are currently available, with lessening cost and increasing performance, which are capable of image measurement in different regions of the infrared spectrum, as shown in Fig. 2.

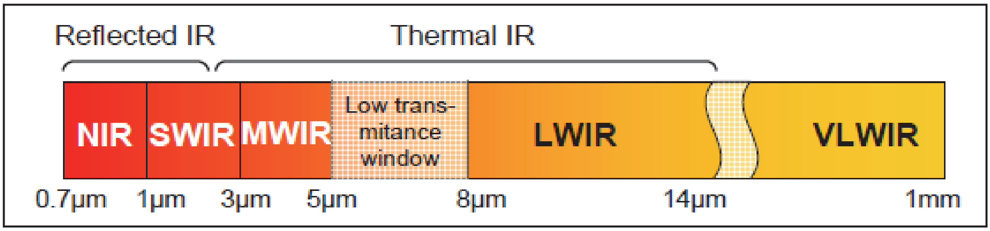


Fig. 2. Infrared band of the electromagnetic spectrum showing the different IR sub-bands.

At wavelengths of 3 µm and longer imaged radiation from objects become significantly emissive due to temperature, and is hence generally termed as thermal infrared. The thermal infrared spectrum is divided into two primary spectra (Wolff et al., 2006) the MWIR and LWIR. Between these spectra lies a strong atmospheric absorption band between approximately 5 and 8 µm wavelength, where imaging becomes extremely difficult due to nearly complete opaqueness of air. The range beyond 14µm is termed the very long-wave infrared (VLWIR) and although in recent years it has received increased attention. The amount of emitted radiation depends on both the temperature and the emissivity of the material. Emissivity in the thermal infrared is conversely analogous to the notion of reflective albedo used in the computer vision literature (Horn, 1977, Horn et al., 1979). For instance, a Lambertian reflector can appear white or grey depending on its efficiency for reflecting light energy. The more efficient it is in reflecting energy (more reflectance albedo) the less efficient it is in thermally emitting energy respective to its temperature (less emissivity). Objects with perfect emissivity of 1.0 are completely black. Many materials that are poor absorbers transmit most light energy while reflecting only a small portion. This applies to a variety of different types of glass and plastics in the visible spectrum.
Different types of thermal imaging are shown in Fig. 3.

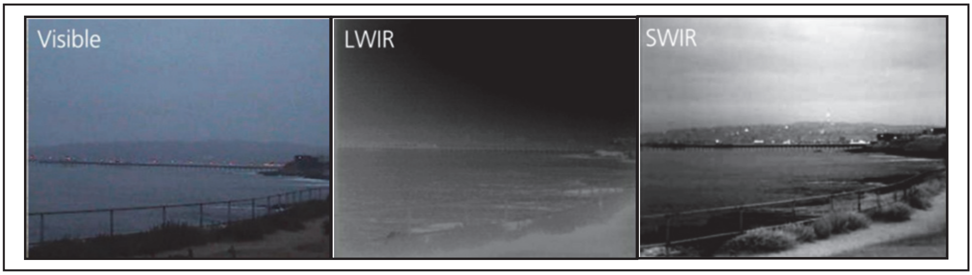


Fig. 3. At sunrise SWIR cameras provide effective imaging day and night thus eliminating thermal crossover.

2.1.2 Different types of thermal spectrums

Thermal Face Spectrums are captured mainly under three classes and those are

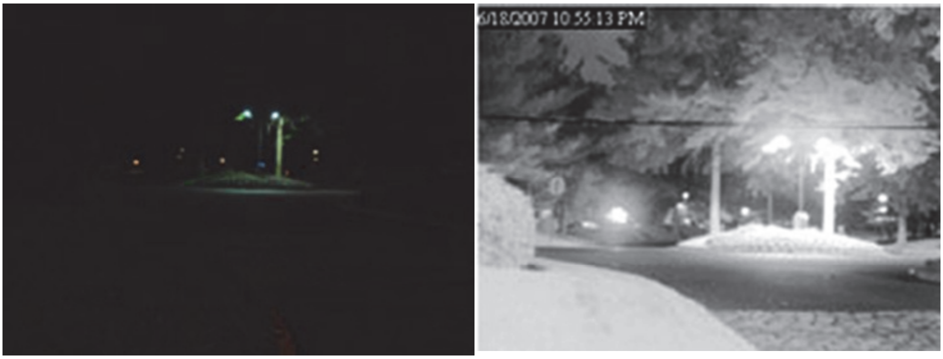
- SWIR
- MWIR
- LWIR

SWIR (Short-wave Infrared): Short-wave Infrared ranges from 1-3 μm (micro-meter/micron). SWIR has its own characteristics which are suitable for human face images. These characteristics are:

Light in the SWIR band is not visible to the human eye: The visible spectrum ranges from the wavelengths of 0.4 microns (blue, nearly ultraviolet to the eye) to 0.7 microns (deep red). Wavelengths longer than visible wavelengths can only be seen by dedicated sensors, such as InGaAs. Though shortwave infrared lights interact with the objects similar way the light of visual wavelength does, human eyes cannot track the object in short wavelength. SWIR light is reflective light; it bounces off objects much like visible light. As a result of its reflective nature, SWIR light has shadows and contrast in its imagery. Images from an InGaAs camera are comparable to visible images in resolution and detail; however, SWIR images are not in color. This makes objects easily recognizable and yields one of the tactical advantages of SWIR, namely, object or individual identification (http://www.sensorsinc/whyswir).

InGaAs sensors can be made extremely sensitive: literally counting individual photons. Thus, when built as focal plane arrays with thousands or millions of tiny point sensors, or sensor pixels, SWIR cameras will work in very dark conditions (http://www.sensorsinc/whyswir).

SWIR Works at Night: An atmospheric phenomenon called night sky radiance emits five to seven times more illumination than starlight, nearly all of it in the SWIR wavelengths. So, with a SWIR camera and this night radiance - often called nightglow - we can "see" objects with great clarity on moonless nights and share these images across networks as no other imaging device can do. Fig. 4 describes the clarity even at night can be obtained using SWIR cameras.



(a) (b)

Fig. 4. (a) The visible imagery of a parking lot at night. (b) The short wave infrared (SWIR) imagery (http://sensorsinc.com/nightvision) of the same scene under the same conditions using a Goodrich camera. of the electromagnetic spectrum showing the different IR sub-bands.

Low Power: InGaAs cameras can be small and use very little power, but give good results. InGaAs found early applications in the telecom industry as it is sensitive to the light used in long distance fiber optics communications, usually around 1550 nm.

Ability to Image through Glass: One major benefit of SWIR imaging that is unmatched by other technologies is the ability to image through glass. So, in short, SWIR images are basically used because of high sensitivity, high resolution, seeing in the light of night glow or night sky radiance, day-to-night imaging, covert illumination, able to see covert lasers and beacons, no cryogenic cooling required, conventional and low-cost visible spectrum lenses, small size and low power.

These foresaid characteristics have made SWIR useful even in the area of face recognition. Detecting Disguises at Border and Immigration Security Checkpoints Using Short Wave Infrared (SWIR) Cameras is one of such applications in Homeland-Defense (http://sensorsinc/border_security1). SWIR has a unique aspect of identifying artificial and natural materials. This helps to detect disguise, artificial hairs appear much darker than the original one, and sometimes it comes almost black. Fig. 5 shows how does an actor seen in disguise when he is captured under SWIR.

Translating this to a border crossing or checkpoint situation, one can assume that anyone wearing a disguise as they approach the border has some form of criminal or even hostile intent. Employed clandestinely, SWIR imaging can add a very valuable layer of protection in the defense of the homeland.

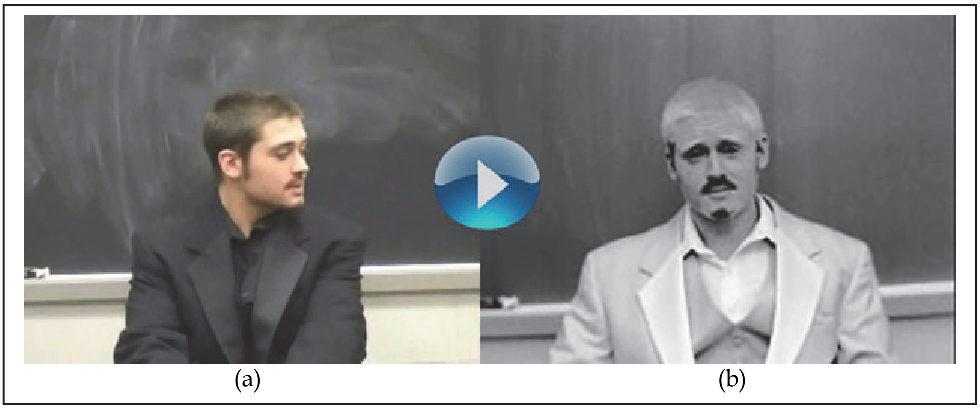


Fig. 5. (a) an actor wearing typical stage make up in color image, (b) same actor in SWIR band (http://sensorsinc.com/border_security1).

MWIR (Middle-Wave Infrared). MWIR ranges from 3-5 µm (micro meter / micron). For night - vision applications, the SWIR imaging with InGaAs technology is enhanced with thermal imaging cameras for MWIR and LWIR in the form of uncooled microbolometers or cooled infrared cameras. Their thermal detectors only show the presence of warm objects against a cooler background. In combination with thermal images, SWIR cameras thus simplify the identification of objects which in the thermal image alone are more difficult to recognize. LWIR is being discussed in the following point.

LWIR (Long-Wave Infrared). LWIR ranges from 8-14 µm (micro meter / micron). The sensor elements of microbolometer cameras for LWIR are made up of IR – absorbing conductors or semiconductors, whose radiation - dependent resistance is measured. Because polysilicon is also suitable as an absorber, they can also be made from polysilicon as MEMS and combined with evaluation circuits in CMOS technology. Fig. 6 shows an LWIR image.

2.1.3 Best thermal spectrum for face recognition purpose

The spectral distribution of energy emitted by an object is simply the product of the Planck distribution for a given temperature, with the emissivity of the object as function of wavelength (Siegal & Howell, 1981). In the vicinity of human body temperature (37° C), the Planck distribution has a maximum in the LWIR around 9μm, and is approximately one-sixth of this maximum in the MWIR. The emissivity of human skin in the MWIR is at least 0.91, and at least 0.97 in the LWIR. Therefore, face recognition in the thermal infrared favors the LWIR, since LWIR emission is much higher than that in the MWIR.

Fig. 6. LWIR Image.

Thermal IR and particularly Long Wave Infra-Red (LWIR) imagery is independent of illumination since thermal IR sensors operating at particular wavelength bands measure heat energy emitted and not the light reflected from the objects. More importantly IR energy can be viewed in any light conditions and is less subject to scattering and absorption by smoke or dust than visible light. Hence thermal imaging has great advantages in face recognition under low illumination conditions and even in total darkness (without the need for IR illumination), where visual face recognition techniques fail.

3D position variations naturally give rise to 2D image variations (Friedrich & Yeshurun, 2002). IR based face recognition is more invariant than CCD based one under various conditions, specifically varying head 3D orientation and facial expressions. Modification of facial expressions and head orientation cause direct 3D structural changes, as well as changes of shadow contours in CCD images that deteriorate the accuracy of any classification method. In an IR image this effect is greatly reduced.

Anatomical features of faces, useful for identification, can be measured at a distance using passive IR sensor technology with or without the cooperation of the subject (Gupta & Majumdar). The thermal infrared (IR) spectrum comprises of mid-wave infrared (MWIR), and long-wave infrared (LWIR), all longer than the visible spectrum.

Accelerated developments in camera technology over the last decade have given computer vision researchers a whole new diversity of imaging options, particularly in the infrared spectrum. Conventional video cameras use photosensitive silicon that is typically able to measure energy at electromagnetic wavelengths from 0.4μm to just over 1.0μm. Multiple technologies are currently available, with dwindling cost and increasing performance which are capable of image measurement in different regions of the infrared spectrum, as shown in Fig. 7 and Fig. 8, which shows the different appearances of a human face in the visible, shortwave infrared (SWIR), midwave infrared (MWIR) and long wave infrared (LWIR) spectra. Although in the infrared, the near-infrared (NIR) and SWIR spectra are still reflective and differences in appearance between the visible, NIR and SWIR are due to reflective material properties. Both NIR and SWIR have been found to have advantages over imaging in the visible for face detection (Dowdall et al., 2002) and detecting disguise (Pavlidis & Symosek, 2000).

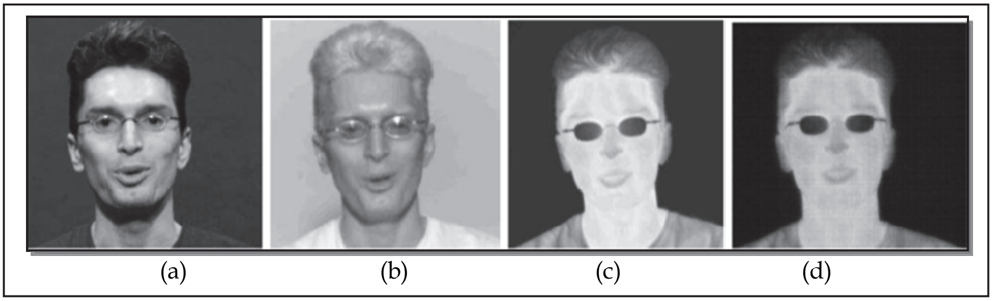


Fig. 7. A face simultaneously imaged in the (a) visible spectrum, 0.4–0.7 µm, (b) short-wave infrared, 1.0–3.0µm, (c) mid-wave infrared, 3.0–5.0µm, and (d) long-wave infrared, 8.0-14.0µm.

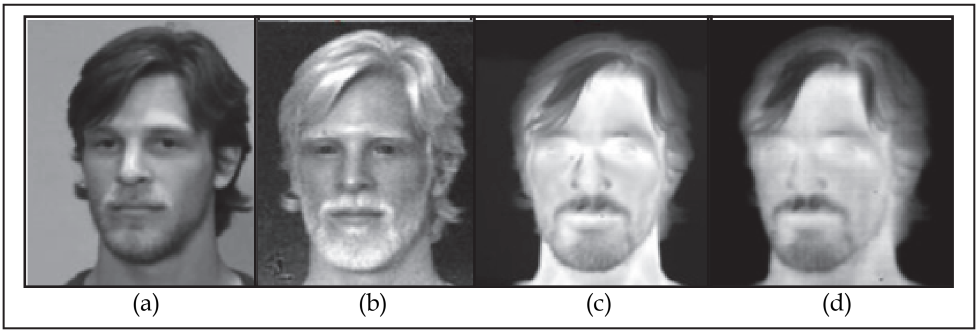


Fig. 8. A face simultaneously imaged in the (a) visible spectrum, 0.4–0.7 µm, (b) short-wave infrared, 1.0–3.0µm, (c) mid-wave infrared, 3.0–5.0µm, and (d) long-wave infrared, 8.0-14.0µm.

Thermal infrared imaging for face recognition first used MWIR platinum silicide detectors in the early 1990s (Prokoski, 1992). At that time, cooled LWIR technology was very expensive. By the late 1990s, uncooled micro bolometer imaging technology in the LWIR became more accessible and affordable, enabling wider experimental applications in this regime. At that time, cooled MWIR technology was about ten times more sensitive than uncooled micro bolometer LWIR technology, and even though faces are more emissive in the LWIR, in the late 1990s MWIR could still discern more image detail of the human face. At present, uncooled micro bolometer LWIR technology coming off the assembly lines is rapidly approaching one-half of the sensitivity of cooled MWIR. For face recognition in the thermal infrared, this is a turning point as for the first time the most appropriate thermal infrared imaging technology (i.e. LWIR) for studying human faces is also the most affordable.

2.2 Advantages of thermal face over visual face

Visual images are considered to be the best in some cases like extracting and locating facial features easily. Another advantage of visual image is that the visual cameras are less expensive. But Visual images have some problems (Kong et al., 2005) with themselves:

- Visual images results in poor performance with illumination variations, such as in indoor and outdoor lighting conditions.
- Again it is not efficient enough to distinguish different facial expressions.
- It is difficult to segment out faces from cluttered scene.
- Visual images are useless in very low lighting.
- Visual images are unable to detect disguise.

The challenges are even more profound when one considers the large variations in the visual stimulus due to

- illumination conditions,
- viewing directions or poses,
- facial expressions,
- aging, and
- disguises such as facial hair, glasses, or cosmetics.

Unlike using the visible spectrum, recognition of faces using different multi-spectral imaging modalities, particularly infrared (IR) imaging sensors (Yoshitomi et al., 1997; Prokoski, 2000; Selinger & Socolinsky, 2001; Wolff et al., 2006; Wolff et al., 2001; Heo et al., 2004) has become an area of growing interest. So, to solve different challenges faced while using visual images in face recognition systems, thermal images are used because of (Kong et al., 2005):

- Face (and skin) detection, location, and segmentation are easier when using thermal images.
- Within-class variance smaller.
- Nearly invariant to illumination changes and facial expressions.
- Works even in total darkness.
- Useful for detecting disguises.

Fig. 9 shows that thermal face images of a human being have no effect of illumination.

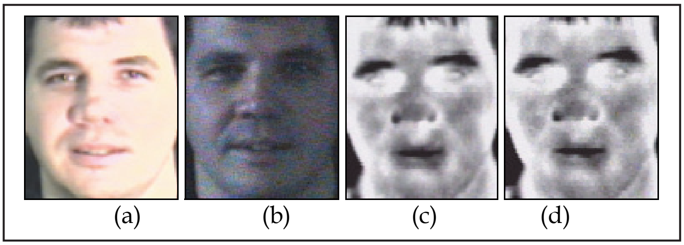


Fig. 9. Thermal images have no effect of illumination: (c) and (d) are the corresponding thermal images of the visual images shown in (a) and (b) respectively.

2.3 Some critical observations on thermal face

However, thermal imaging needs to solve some challenging problems, which are discussed next.

2.3.1 Identical twins

Though identical twins have some different thermal patterns, there are some exceptions too. The twins' images in Fig. 10 are not necessarily substantially different from each other.

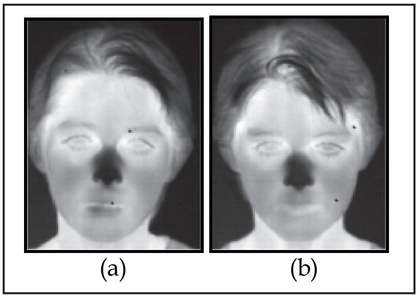


Fig. 10. A pair of twins' IR images.

2.3.2 Exhale-inhale effect

High temporal frequency thermal variation is associated with breathing. The nose or mouth appears cooler as the subject is inhaling and warmer as he or she exhales, since exhaled air is at core body temperature, which is several degrees warmer than skin temperature. Fig. 11 shows the different thermal images of same person while exhaling and inhaling.

2.3.3 Metabolism effect

Symptoms such as alertness and anxiety can be used as a biometric, which is difficult to conceal as redistribution of blood flow in blood vessels causes abrupt changes in the local skin temperature. Thermal signatures can be changed significantly according to different body temperatures caused by physical exercise or ambient temperatures (Heo et al., 2005).

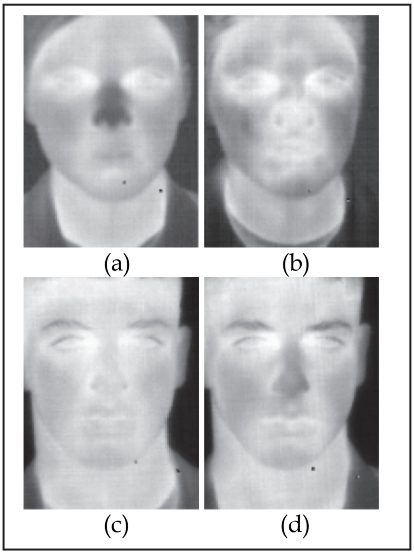


Fig. 11. Variation in facial thermal emission from two subjects in different sessions. For first subject (a) and (b) are different and for second subject (c) and (d) are different.

Fig. 11 shows comparable variability within data collected with LWIR sensor. Each column shows images acquired in different sessions. It is clear that thermal emission patterns around the eyes, nose and mouth are rather different in different sessions. Such variations can be induced by changing environmental conditions. For example, exposed to cold or wind, capillary vessels at the surface of the skin contract, reducing the effective blood flow and thereby the surface temperature of the face. Also, when a subject transitions from a cold outdoor environment to a warm indoor one, a reverse process occurs, whereby capillaries dilate suddenly flushing the skin with warm blood in the body's effort to regain normal temperature.

Additional fluctuations in thermal appearance are unrelated to ambient conditions, but are rather related to the subject's metabolism. Vigorous physical activity, consumption of food, alcohol or caffeine may all affect the thermal appearance of a subject's face.

2.3.4 Effects of using glasses

Thermal images of a subject wearing eyeglasses may lose information around the eyes since glass blocks a large portion of thermal energy; in fact most of the thermal energy is blocked. Thermal imaging has difficulty in recognizing people inside a moving vehicle (because of speed, glass).

2.3.5 Liveness solution

Spoofing attack (or copy attack) is a fatal threat for biometric authentication systems. Liveness detection, which aims at recognition of human physiological activities as the liveness indicator to prevent spoofing attack, is becoming a very active topic in field of fingerprint recognition and iris recognition. In face recognition community, although numerous recognition approaches have been presented, the effort on anti-spoofing is still very limited. Liveness detection methods allow differentiating live human characteristics from characteristics coming from other sources. Spoofing, now-a-days has become a big threat for biometrics, especially in the field of face recognition. Therefore anti-spoof problem should be well solved before face recognition could widely be applied in our life. If we are thinking of keeping image as a part of biometric identity then it is very easy for someone to spoof an image and retrieve sensitive information from somebody's account and retrieve their valuable information by hacking their account. There are various methods to spoof an image. Photo attack is the cheapest and easiest spoofing approach, since one's facial image is usually very easily available in the public. Video spoofing is another big threat to face recognition systems, because it is very similar to live face and can be shot in front of legal user's face by a needle camera. It has many physiological clues that photo does not have, such as head movement, facial expression, blinking etc.

Thermal images can be a solution to the spoofing problem and detecting live faces as it captures only the heat emitted and so the thermal images generated from the emitted heat by a photograph or a video will be totally different from the thermal image of an original human face.

2.4 Drawbacks of thermal images

Though thermal images have found to be useful in accurate recognition of human face; there are some limitations of them.

- Redistribution of blood flow due to alertness and anxiety causes abrupt changes in the local skin temperature.
- Thermal calibration is mandatory in ambient temperature or activity level may change thermal characteristics.
- Energetic physical activity, consumption of food, alcohol, caffeine etc. may also affect the thermal characteristics.
- While breathing exhaling results big change in skin temperature.
- Glasses block most of thermal energy.
- Not appropriate for recognition of vehicle occupants (because of speed, glass).
- Thermal images have low resolution.
- Thermal cameras are expensive.

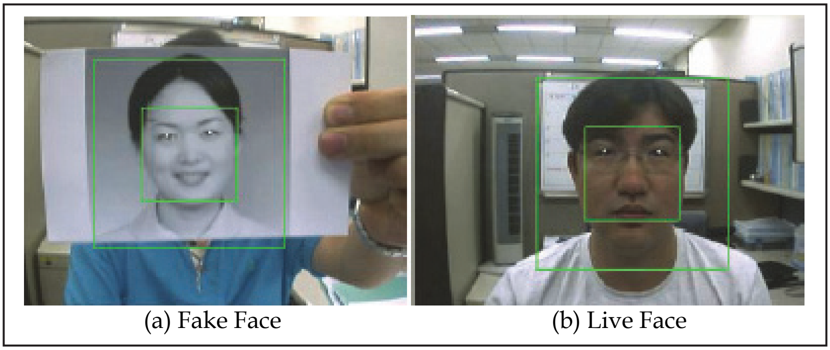

(a) Fake Face (b) Live Face

Fig. 12. Visual images may fail to detect fake faces.

2.5 Problems of face recognition solvable by thermal images

In contrast to visual face images thermal face images are having certain characteristics by which they are able to handle some difficulties mentioned earlier. But they do have some limitations like pose variation; aging etc can't be solved by thermal images. Table 2 describes the problems in face recognition that can or cannot be solved using thermal images.

Different Problems of Face Recognition	Problems that can be Solved using Thermal Face Images
Illumination Variation	√
Pose Variation	×
Variation in Expression	√
Different Scaling	×
Disguises	√ (except use of glasses)
Aging Problem	×
Variation in Temperature	×

Table 2. Different face recognition problems and use of thermal face images to solve them.

2.6 Thermal face image databases

There are only two publicly available thermal face image databases. These are IRIS (Imaging, Robotics and Intelligent System) Thermal/Visible Face Database and Terravic Facial IR Database. Among them IRIS database comprises of almost all the features that should be present in a standard face database. A short overview of these two databases is given in Table 3.

Name of the Database	No. of Subjects	Pose	Illumination	Facial Expression	Time
IRIS Thermal/Visible Database	30	11	6	3	1
Terravic Facial Infrared Database	20	3	---	1	1

Table 3. Overview of the recording conditions for all databases discussed in this section.

These two databases are publicly available benchmark dataset for testing and evaluating novel and state-of-the-art thermal face recognition algorithms. The benchmark contains videos and images recorded in and beyond the visible spectrum and are available for free to all researchers in the international computer vision communities. It also allows a large spectrum of IEEE and SPIE vision conference and workshop participants to explore the benefits of the non-visible spectrum in real-world applications, contribute to the OTCBVS workshop series, and boost this research field significantly.

2.6.1 IRIS (Imaging, Robotics and Intelligent System) thermal/visual database

In the IRIS database unregistered thermal and visible face images are acquired simultaneously under variable illuminations, expressions, and poses. Total no. of 30 individuals of RGB color image type with Exp1 (surprised), Exp2 (laughing) and Exp3 (Anger) are available. Resolution of each image is 320 x 240. Illumination types available in this database are left light on, right light on, both lights on, dark room, left and right lights off with varying poses like left, right, mid, mid-left, mid-right. Two different sensors are used to capture this database. One is Thermal - Raytheon Palm-IR-Pro and another is Visible - Panasonic WV-CP234. Table 4 is furnished with the IRIS database (http://cse.ohio-state.edu/otcbvs-bench/Data/02/download) overview.

No. of Subjects	Conditions		Image Resolution	Total Number of Images	
30	Facial Expression	3	320 x 240	Thermal	1529
	Illumination	5		Visual	1529

Table 4. IRIS (Imaging, Robotics and Intelligent System) Database Information.

Sensors used

- Thermal - Raytheon Palm-IR-Pro.
- Visible - Panasonic WV-CP234.

Setup of the Camera. Fig. 13 shows the camera set up used for preparing the IRIS Thermal/Visible database.

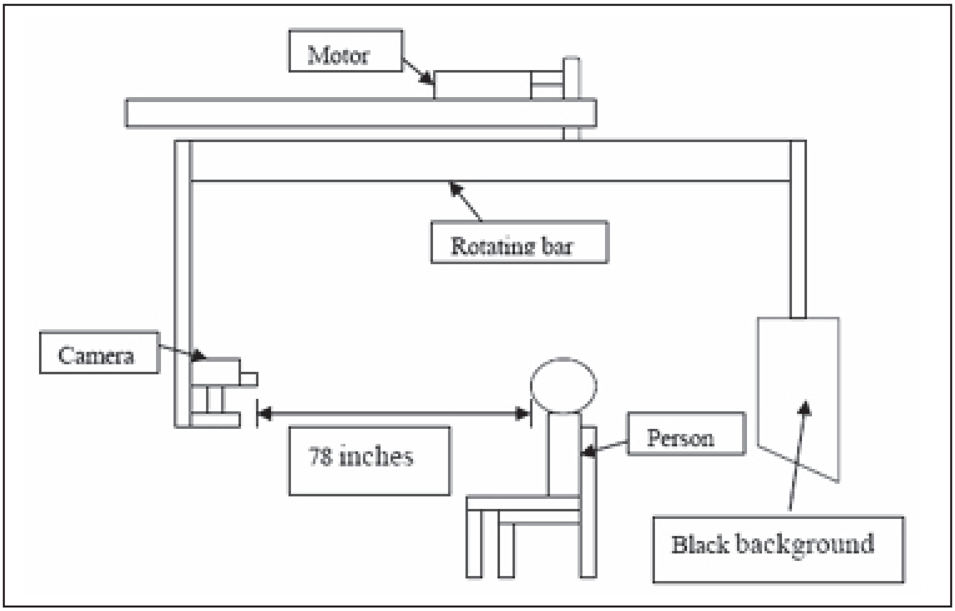


Fig. 13. Camera Setup for IRIS (Imaging, Robotics and Intelligent System) Database.

Database description

- It contains images of 30 individuals (28 men and 2 women).
- The imaging and recorded condition (camera parameters, illumination setting, camera distance).
- Total 176-250 images per person and 11 images per rotation (poses for each expression and each illumination) are captured.
- Database images having the format of bmp color images are 320 x 240 pixels in size.
- The subjects were recorded in 3 different expressions Exp-1 (Surprised), Exp-2 (laughing), Exp-3 (Anger) and 5 different illumination Lon (left light on), Ron (right light on), 2on (both lights on), dark (dark room), off (left and right lights off) with varying poses.
- Size of this database is 1.83 GB.
- Variable numbers of images are available per class.
- Total 3058 images are available; 1529 images are thermal and other images are visual.
- All the classes don’t contain each type of illumination.

This database has disguise faces too. Samples of images with different facial expression and different illumination conditions, and different disguise faces are given in Fig. 14, Fig. 15, and Fig. 16, respectively.

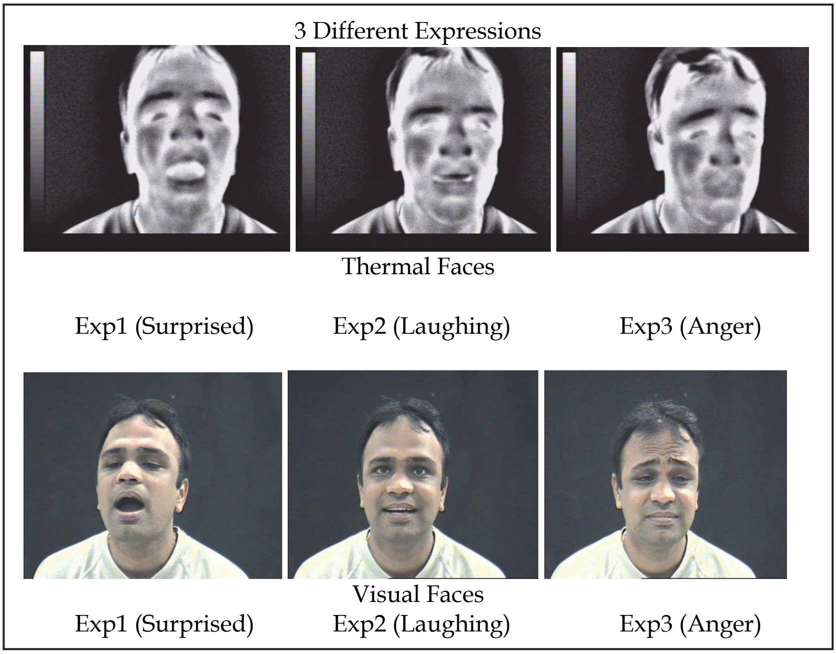


Fig. 14. Faces of Expression in IRIS Thermal/Visible database.

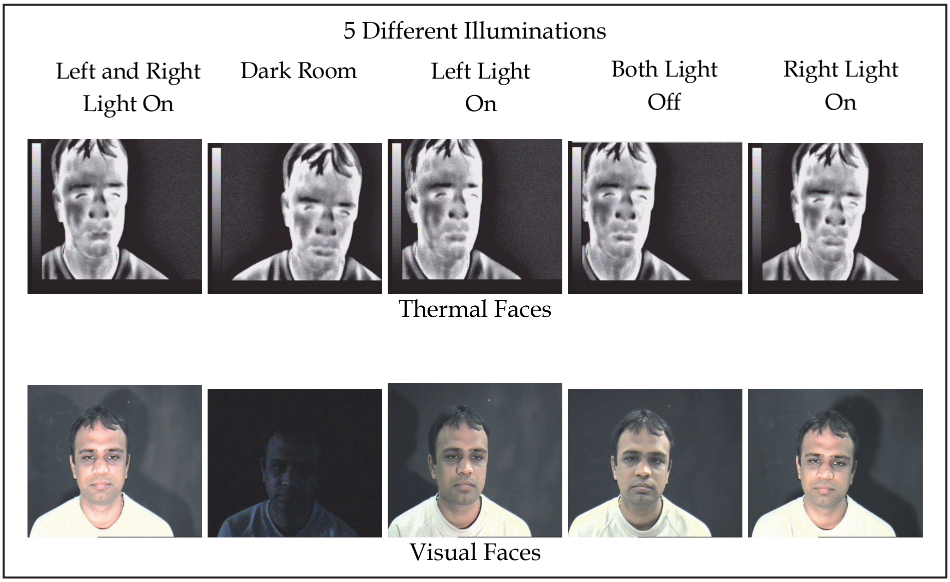


Fig. 15. Faces of Illumination in IRIS Thermal/Visible database.

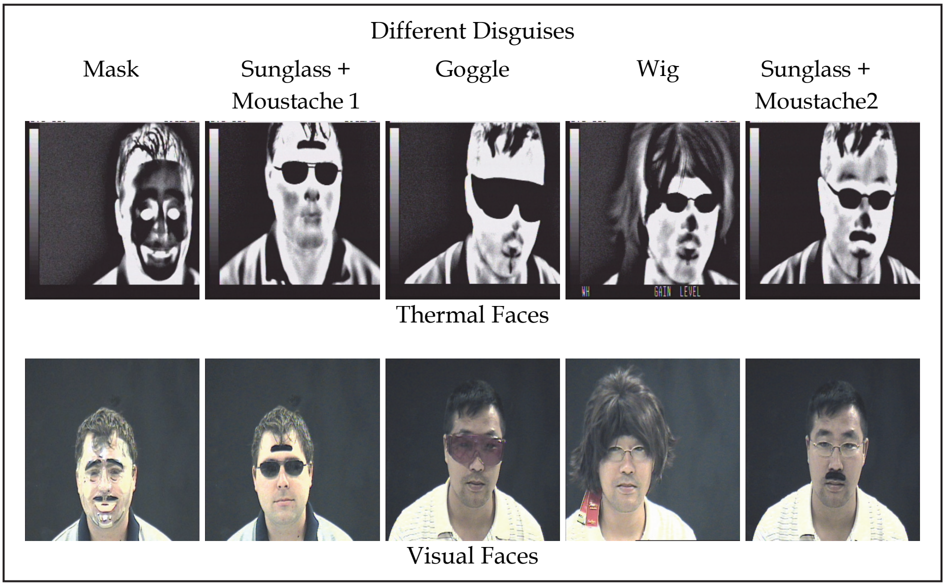


Fig. 16. Sample of Disguised Faces in IRIS Thermal/Visible database.

2.6.2 Terravic facial infrared database

The Terravic Facial Infrared database contains total no. of 20 classes (19 men and 1 woman) of 8-bit gray scale JPEG thermal faces. Size of the database is 298MB and images with different rotations are left, right and frontal face images also available with different items like glass and hat. Table 5 is furnished with the Terravic Facial Infrared database (http://cse.ohio-state.edu/otcbvs-bench/Data/02/download) overview.

No. of Subjects	Conditions	Image Resolution
20	Front	320 x 240
	Left	
	Right	
	Indoor/Outdoor	
	Hat	
	Glasses	

Table 5. IRIS (Imaging, Robotics and Intelligent System) Database Information.

Sensors Used. Raytheon L-3 Thermal-Eye 2000AS.

Database description

- In this database, they provide total 20 classes.
- They use different poses for this database like front, left, right, indoor, outdoor; glasses, hat, both.

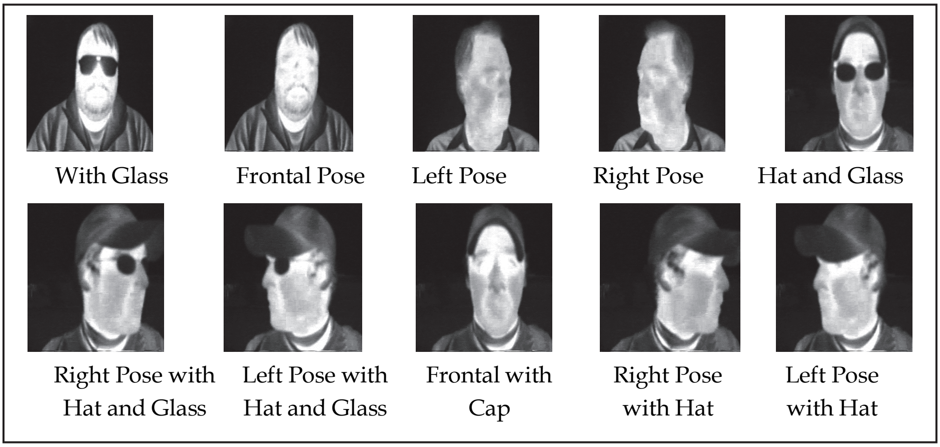
- Type of that database is thermal in JPEG format.
- Size of that database is 298 MB.
- Total no. of images is 21,308.

In Fig. 17 some image samples of Terravic Facial Infrared database are shown.

2.7 Some review work on thermal face recognition

Over the last few years, many researchers have investigated the use of thermal infrared face images for person identification to tackle illumination variation, facial hair, hairstyle etc. (Chen et al. 2003; Buddharaju et al., 2004; Socolinsky & Selinger, 2004, Singh et al., 2004, Buddharaju et al., 2007).

It is found that while face recognition using different expressions with visible-light imagery outperforms that with thermal imagery when both gallery and probe images are acquired indoors, if the probe image or the gallery and probe images are acquired outdoors, then it appears that the performance possible with IR can exceed that with visible light. IR imagery represents a feasible substitute to visible imaging in the search for a robust and practical identification system. Leonardo Trujillo et.al. proposed an unsupervised local and global feature extraction paradigm to the problem of facial expression using thermal images (Trujillo et al., 2005).



With Glass
Frontal Pose
Left Pose
Right Pose
Hat and Glass
Right Pose with Hat and Glass
Left Pose with Hat and Glass
Frontal with Cap
Right Pose with Hat
Left Pose with Hat

Fig. 17. Sample of Terravic Facial Infrared Database.

First they have localized facial features by novel interest point detection and clustering approach and after that they apply PCA for feature extraction and at last they use SVD for facial expression classification. For facial expression recognition they use the IRIS dataset. Their experimental results show that their FER system clearly degrades when classifying the "happy" expression of the dataset. Xin Chen et al. developed a face recognition technique with PCA. They used PCA to study the comparisons and combination of infrared and visible images to the effects of lighting, facial expression change and the time difference between gallery and probe images (Chen et al., 2003).

The techniques developed by Socolinsky & Selinger (Socolinsky & Selinger, 2004) show performance statistics for outdoor face recognition and recognition across multiple sessions. A few experimental results with thermal images in face recognition are being recorded in Table 6. All the result support the conclusion that face recognition performance with thermal infrared imagery is stable over multiple sessions.

Method	Recognition rate
Fusion of Thermal and Visual (Singh et al., 2004)	90%
Segmented Infrared Images via Bessel forms (Buddharaju et al., 2004)	90%
PCA for Visual indoor Probes (Socolinsky & Selinger, 2004)	81.54%
PCA+LWIR(Indoor probs) (Socolinsky & Selinger, 2004)	58.89% (Maximum)
LDA+LWIR (Indoor probs) (Socolinsky & Selinger, 2004)	73.92% (Maximum)
Equinox +LWIR (Indoor probs) (Socolinsky & Selinger, 2004)	93.93% (Maximum)
PCA+LWIR(Outdoor probs) (Socolinsky & Selinger, 2004)	44.29% (Maximum)
LDA+LWIR (Outdoor probs) (Socolinsky & Selinger, 2004)	65.30% (Maximum)
Equinox+LWIR (Outdoor probs) (Socolinsky & Selinger, 2004)	83.02% (Maximum)
ARENA+LWIR (Different illumination but same expression) (Socolinsky et al., 2001)	99.3% (Average), 99% (Minimum)
Eigenfaces +LWIR (Different illumination but same expression) (Socolinsky et al., 2001)	95.0% (Average), 89.4% (Minimum)
ARENA+LWIR (Different illumination and expression) (Socolinsky et al., 2001)	99% (Average), 98% (Minimum)
Eigenfaces +LWIR(Different illumination and expression) (Socolinsky et al., 2001)	93.3% (Average), 86.8% (Minimum)

Table 6. Comparison of recognition rate of different methods.

3. Fused images

There is another type of face image used in face recognition field which is known as fused image. Fused image comprises of more than one image. So, the resulted fused image is more informative than any of the individual image. In face recognition, sometimes visual images are found to be more helpful than thermal images. For example, problem of variation in temperature can be solved by visual images which cannot be in case of thermals. Again, thermal images are best to detect liveliness. So, we can apply the concept of fusion in face recognition to get better result of person recognition. So, fusion of both thermal and visual images will generate a new fused image that will store the information of both thermal and visual faces. Fig. 18 shows a pictorial example of a fused image.

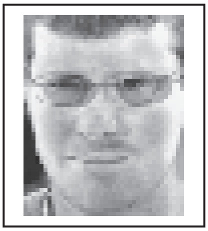


Fig. 18. Fused image.

Image fusion methods mainly fall under two categories viz. spatial domain fusion and transform domain fusion. Many scientists have been involved in research and development in the area of data fusion for over a decade.

3.1 Different types of fusion techniques

Over a decade, researchers are working on image fusion. Generally, in face recognition three types of fusion techniques can be used. Those are:

- Feature Level Fusion
- Decision Level Fusion
- Pixel / Data Level Fusion

3.1.1 Feature level fusion

Before merging the features of the source data, all the features are required to be merged together. Fusion at feature level involves the integration of feature sets corresponding to different sensors. These feature vectors are often fused to form joint feature vectors from which the classification is made. The first hurdle towards feature level fusion is effective feature detection. Once features are selected, the role of feature-level fusion is to establish boundaries in feature space and separate patterns belonging to different classes. Thus two main issues are involved: feature detection and the use of distance metrics for clustering. In general, signal features can be classified into the following 3 categories:

- Time Domain Features that describe waveform characteristics (slopes, amplitude values, maxima/minima and zero crossing rates) and statistics (mean, standard deviation, energy, kurtosis, etc).
- Frequency Domain Features (periodic structures, Fourier coefficients, spectral density)
- Hybrid Features that cover both time and frequency domains (Wavelet representations, Wigner-Ville distributions, etc.)

Since the feature set contains richer information about the raw biometric data than the match score or the final decision, integration at this level is expected to provide better recognition results. However, fusion at this level is difficult to achieve in practice because of the following reasons:

- The feature sets of multiple modalities may be incompatible. For example, minutiae set of finger prints and eigen-coefficients of faces.
- The relationship between the feature spaces of different biometric systems may not be known.
- Concatenating two feature vectors may result in a feature vector with very large dimensionality leading to the 'curse of dimensionality' problem.

3.1.2 Decision level fusion

Decision level fusion combines the results from multiple algorithms to yield a final fused decision. Decision level fusion is generally based on a joint declaration of multiple single source results (or decisions) to achieve an improved classification or event detection. At the decision level, prior knowledge and domain specific information can also be incorporated. Widely used methods for decision level fusion include the following:

- Bayesian inference (Information theory, inference and learning algorithms, book by David Mackay).
- Classical inference; computing a joint probability given an assumed hypothesis usually using Maximum A Posteriori (MAP) or maximum likelihood decision rules.
- Dempster-Shafer's method.

Decision-level fusion schemes can be broadly categorized according to the type of information the decision makers output. Abstract-level fusion algorithms are used to fuse the individual experts that produce only class labels. In this category, plurality voting is the most commonly used one that just outputs the class label having the highest vote.

3.1.3 Pixel/data level fusion

Pixel level fusion is the combination of the raw data from multiple source images into a single image. It combines information into a single image from a set of image sources using pixel, feature or decision level techniques.

The task of interpreting images, either visual images alone or thermal images alone, is an unconstraint problem. Thermal image can at best yield estimates of surface temperature that, in general, is not specific in distinguishing between object classes. The features extracted from visual intensity images also lack the specificity required for uniquely determining the identity of the imaged object.

The interpretation of each type of image thus leads to ambiguous inferences about the nature of the objects in the scene. The use of thermal data gathered by an infrared camera, along with the visual image, is seen as a way of resolving some of these ambiguities. On the other hand, thermal images are obtained by sensing radiation in the infrared spectrum. The radiation sensed is either emitted by an object at a non-zero absolute temperature, or reflected by it. The mechanisms that produce thermal and visual images are different from each other. Thermal image produced by an object's surface can be interpreted to identify these mechanisms. Thus, thermal images can provide information about the object being imaged which is not available from a visual image (Yin & Malcolm, 2000).

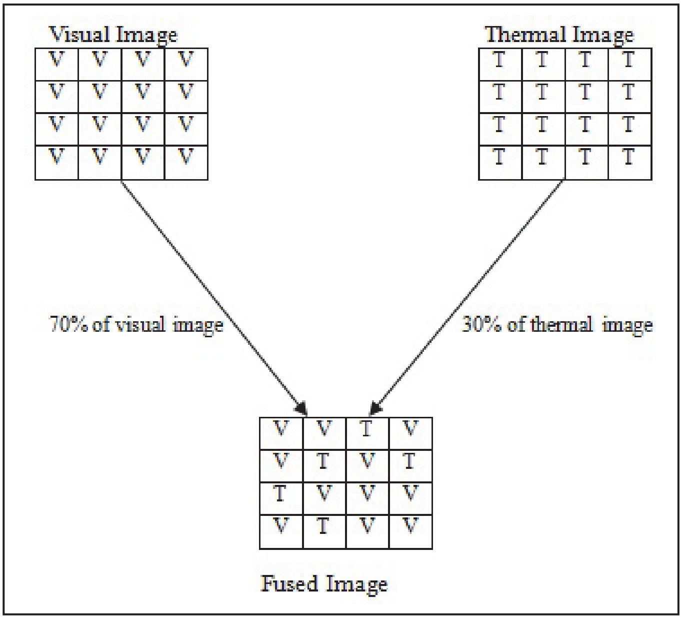
A great deal of effort has been expended on automated scene analysis using visual images, and some work has been done in recognizing objects in a scene using infrared images. However, there has been little effort on interpreting thermal images of outdoor scenes based on a study of the mechanism that gives rise to the differences in the thermal behavior of object surfaces in the scene. Also, nor has been any effort been made to integrate information extracted from the two modalities of imaging.

The process of image fusion may be where pixel data of 70% of visual image and 30% of thermal image of same class or same image is brought together into a common operating image or now commonly referred to as a Common Relevant Operating Picture (CROP) (Hughes, 2006). This implies that an additional degree of filtering and intelligence is to be applied to the pixel streams to present pertinent information to the user. So image pixel fusion has the capacity to enable seamless working in a heterogeneous work environment with more complex data. For accurate and effective face recognition we require more informative images. Image by one source may lack some information which might be available in images by other source (i.e. visual). So if it becomes possible to combine the features of both the face images then efficient, robust, and accurate face recognition can be developed.

Ideally, the fusion of common pixels can be done by pixel-wise weighted summation of visual and thermal images (Horn & Sjoberg, 1979), as below:

$$F(x, y) = a(x, y)V(x, y) + b(x, y)T(x, y) \tag{1}$$

where, *F(x, y)* is a fused output of a visual image, *V(x, y),* and a thermal image, *T(x, y)*, while *a(x, y)* and *b(x, y) represent* the weighting factors for visual and thermal images respectively.



Visual Image

V	V	V	V
V	V	V	V
V	V	V	V
V	V	V	V

Thermal Image

T	T	T	T
T	T	T	T
T	T	T	T
T	T	T	T

Fused Image

V	V	T	V
V	T	V	T
T	V	V	V
V	T	V	V

Fig. 19. Fusion Technique.

4. Discussions

Being inspired by the characteristics of thermal and fused images that are applicable for solving different problems of face recognition, we have been using thermal and fused images in our so far research-works. In this section some important observations and outcome of these research works are briefly discussed.

To handle the challenges of face recognition that include pose variations, changes in facial expression, partial occlusions, variations in illumination, rotation through different angles, change in scale etc., two techniques have been applied. In the first method log-polar transformation is applied to the fused images, which are obtained after fusion of visual and thermal images, whereas in second method fusion is applied on log-polar transformed individual visual and thermal images. Log-polar transformed images are capable of handling complicacies introduced by scaling and rotation. The second method has shown better performance, which is 95.71% (maximum) and on an average 93.81% as correct recognition rate on Object Tracking and Classification Beyond Visible Spectrum (OTCBVS) database (Bhowmik et al, April 11 – 15, 2011).

In another experiment, the aim was to recognize thermal face images for face recognition using line features and Radial Basis Function (RBF) neural network as classifier for them. The proposed method works in three different steps. In the first step, line features are extracted from thermal polar images and feature vectors are constructed using these lines. In the second step, feature vectors thus obtained are passed through eigenspace projection for the dimensionality reduction of feature vectors. Finally, the images projected into eigenspace are classified using a Radial Basis Function (RBF) neural network.

Experimental results of verification and identification is performed in the OTCBVS database and the maximum success rate is 100% whereas on an average it is 94.44% (Bhowmik et al, April 25 – 29, 2011a).

For achieving better recognition rate, an image fusion technique based on weighted average of Daubechies wavelet transform (db2) coefficients from visual face image and their corresponding thermal images have been conducted. Both PCA and ICA have separately been applied for dimension reduction (Bhowmik et al, April 25 – 29, 2011b). The resulted fused images have then been classified using multi-layer perceptron (MLP). Experimental results show that the performance of ICA architecture-I is better than the other two approaches i.e. PCA and ICA-II. The average success rate for PCA, ICA-I and ICA-II are 91.13%, 94.44% and 89.72% respectively. However, approaches presented here achieves maximum success rate of 100% in some cases, especially in case of varying illumination over the IRIS Thermal/Visual Face Database.

Thermal images minimize the affect of illumination changes and occlusion due to moustache, beards, adornments etc. The training and testing sets of thermal images are registered in a polar coordinate that is capable to handle complicacies introduced by scaling and rotation and then polar images are projected into eigenspace and finally classified using a multi-layer perceptron. The results improve significantly in the verification and identification performance and the success rate is 97.05% over the OTCBVS database (Bhowmik et al., 2008).

Fused images generated from visual and thermal ones are projected into eigenspace and finally classified using a radial basis function neural network that is useful to recognize unknown individuals with a maximum success rate of 96% of the Object Tracking and Classification Beyond Visible Spectrum (OTCBVS) database (Bhowmik et al., 2009).

Biometric is a unique identity for each person where security as well as authentication is concerned. Based on machine learning algorithms, a general framework is designed to show the effectiveness of a biometric system using different levels of pixel fusion. One of the most popular feature extraction algorithms along with multilayer perceptron (MLP) has been used for classification purpose over the OTCBVS database that leads to an optimal recognition result (Bhowmik et al., 2011).

When only one-sided semi profile thermal images of an individual are available, a mosaicing technique can be applied to build an apparent 2-D front profile view of that person (Majumder et al., 2011). The available semi-profile image is converted into a mirror image which is similar to the opposite half semi-profile thermal image of that person. Human face is symmetric. So, simple concatenation of these two images (one is the original side view image and another is the mirror or opposite side view image) produces an apparent 2-D front face thermal mosaiced image of that person. This mosaiced image can further be imputed in any simple thermal face recognition system.

As a part of future work of the foresaid thermal face mosaicing, the mosaiced faces are experimented for face recognition purpose where we are checking the similarity between mirror face images and normal face images. Here thermal and visual faces are fused together and then mosaicing is applied to construct apparent or complete profile view of an individual and then these mosaiced faces are classified using support vector machine classifier.

5. Acknowledgment

Authors are thankful to the projects entitled "Development of Techniques for Human Face Based Online Authentication System Phase-I", Vide No. 12(14)/08-ESD, dated: 27.01.2009 and "Creation of a Visual Face Database of North-Eastern People and Implementation of Techniques for Face Identification", Vide No. 12(2)/2011-ESD,dated 29/3/11, both sponsored by Department of Information Technology under the Ministry of Communications and Information Technology, Govt. of India and "Design and Development of Facial Thermogram Technology for Biometric Security System," funded by University Grants Commission(UGC), India for providing necessary infrastructure facilities to conduct experiments relating to different works mentioned in this chapter.

6. References

Bhowmik M.K., Bhattacharjee D., Basu D. K. & Nasipuri M. (2008). Classification of Polar-Thermal Eigenfaces Using Multilayer Perceptron for Human Face Recognition, *Proceedings of 2008 IEEE Region 10 Colloquium and the Third ICIIS*, PI-382, Kharagpur, India, December 8-10, 2008

Bhowmik M.K., Bhattacharjee D., Nasipuri M., Basu D. K. & Kundu M. (2009). Classification of Fused Images using Radial Basis Function Neural Network for Human Face Recognition, *Proceedings of The World congress on Nature and Biologically Inspired Computing (NaBIC-09) published by IEEE Explore,* ISBN 978-1-4244-5053-4, PSG college, Peelamedu, Coimbatore, India, Dec. 9-11, 2009

Bhowmik M.K., Bhattacharjee D., Basu D. K. & Nasipuri M. (2011). A Comparative Study on Fusion of Visual and Thermal Face Images at Different Pixel Level. *International Journal of Information Assurance and Security (JIAS)*, Vol. 6, No. 1, (2011), pp. 80-86, ISSN 1554-1010

Bhowmik M.K., Bhattacharjee D., Basu D. K. & Nasipuri M. (April 11 – 15, 2011). Polar Fusion Technique Analysis for Evaluating the Performances of Image Fusion of Thermal and Visual Images for Human Face Recognition. *Proceedings of IEEE Symposium Series on Computational Intelligence,* Paris, France, pp. 62-69, April 11 – 15, 2011

Bhowmik M.K., Bhattacharjee D., Basu D. K. & Nasipuri M. (April 25 – 29, 2011a). Classification of Thermal Face Images using Radial Basis Function Neural Network. *Proceedings of SPIE Defense, Security, and Sensing 2011, IR Sensors and Systems, Conference DS100,* SPIE and SPIE Digital Library, Orlando World Center Marriott Resort & Convention Center, Orlando, Florida, USA, April 25 – 29, 2011

Bhowmik M.K., Bhattacharjee D., Basu D. K. & Nasipuri M. (April 25 – 29, 2011b). Independent Component Analysis (ICA) of fused Wavelet Coefficients of thermal and visual images for human face recognition. *Proceedings of SPIE Defense, Security, and Sensing 2011, IR Sensors and Systems, Conference DS100, SPIE and SPIE Digital Library,* Orlando World Center Marriott Resort & Convention Center, Orlando, Florida, USA, April 25 – 29, 2011

Buddharaju P., Pavlidis I. & Kakadiaris I. (2004). Face Recognition in the Thermal Infrared spectrum. *Proceeding of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshops (CVPRW'04)*, 2004

Buddharaju P., Pavlidis I. T., Tsiamyrtzis P. & Bazakos M. (April 2007). Physiology-Based Face Recognition in the Thermal Infrared Spectrum. *IEEE transactions on pattern analysis and machine intelligence*, Vol.29, No.4., April 2007

Chen X., Flynn P.J. & Bowyer K.W. (2003). PCA-Based Face Recognition in Infrared Imagery: Baseline and Comparative Studies. *Proceedings of the IEEE International Workshop on Analysis and Modelingof Faces and Gestures (AMFG'03)*, 2003

Chen X., Flynn P.J. & Bowyer K.W. (2005) *IR and Visible light face Recognition*, University of NotreDame, USA <http://identix.com/products>

Dowdall J., Pavlidis I. & Bebis G. (2002). A face detection method based on multib and feature extraction in the near-IR spectrum, *Proceedings of IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Kauai, Hawaii, 2002

Friedrich G. & Yeshurun Y. (2002). Seeing People in the Dark: Face Recognition in Infrared Images, *Biologically Motivated Computer Vision*, Lecture Notes in Computer Science, Springer, Berlin / Heidelberg, 2002

Gupta A. and Majumdar S.K., Machine Recognition of Human Face, http://anshulg.com/index_files/5.pdf

Heo J., Kong S., Abidi B. & Abidi M. (2004). Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition, *IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR 2004*, pp. 94-99, Washington, D.C., 2004

Heo J., Savvides M. & Vijayakumar B.V.K. (2005). Performance Evaluation of Face Recognition using Visual and Thermal Imagery with Advanced Correlation Filters, *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005

Horn B. (1977). Understanding image intensities, *Artificial Intelligence*, pp. 1–31, 1977

Horn B. & Sjoberg R. (1979). Calculating the reflectance map, *Applied Optics 18*, pp. 1770–1779, 1979

Hughes D. (2006). Sinking in a Sea of Pixels- The Case for Pixel Fusion, Silicon Graphics, Inc., http://sgi.com/pdfs

Kong S.G., Heo J., Abidi B.R., Paik J. & Abidi M.A. (2005). Recent advances in visual and infrared face recognition - a review. *Published by Computer Vision and Image Understanding,* Vol. 97, Issue 1, pp. 103 – 135

Ling B., Trang A.H. & Phan C. (2007). Multi-classifier buried mine detection using MWIR images, *Proceedings of SPIE, the International Society for Optical Engineering*, Vol. 6553, pp. 655310.1-655310.12, ISSN 0277-786X

Majumder S., Majumder G. & Bhowmik M.K. (2011). Human Face Mosaicing using Thermal Image, *Proceedings of National Conference on Mathematical Analysis and its Applications (NCMAA' 11)*, Tripura University (A Central University), India, January 5-6, 2011

Miller J.L. (1994). *Principles of Infrared Technology: A practical guide to the state of the art*, Van Nostrand Reinhold, ISBN 10: 0442012101 / 0-442-01210-1, ISBN 13: 9780442012106, New York, 1994

Pavlidis I. & Symosek P. (2000). The imaging issue in an automatic face/disguise of detection system, *Proceedings of IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, Hilton Head, 2000

Prokoski F. (1992). Method for identifying individuals from analysis of elemental shapes derived from biosensor data. *U.S. Patent 5,163,094*

Prokoski F. (2000). History, Current Status, and Future of Infrared Identification, *Proceedings of . IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp.514

Selinger A. & Socolinsky D.A. (2001). Appearance-Based Facial Recognition Using Visible and Thermal Imagery: A Comparative Study, Technical Report, Equinox Corporation.

Siegal R. & Howell J. (1981). Thermal Radiation Heat Transfer. *McGraw-Hill,* New York

Singh S., Gyaourva A., Bebis G. & Pavlidis I. (2004). Infrared and Visible Image Fusion for Face Recognition. *Proceedings of. SPIE*, Vol.5404, pp.585-596, Aug.2004

Socolinsky D.A. & Selinger A. (2004). Thermal Face Recognition in an Operational Scenario. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 2004

Socolinsky D.A., Wolff L.B., Neuheisel J.D. & Eveland C.K. (2001). Illumination Invariant Face Recognition Using Thermal Imagery. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01)*, Hawaii, 2001

Trujillo L., Olague G., Hammoud R. & Hernandez B. (2005). Automatic Feature Localization in Thermal Images for Facial Expression Recognition. *Proceeding of CVPR '05 Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, ISBN:0-7695-2372-2-3, Washington, DC, USA, 2005

Vizgaitis J., (2006). Selecting Infrared Optical Materials, Retrieved from < http://optics.arizona.edu/optomech/student%20reports/tutorials/VizgaitisTutorial1.doc>

Wilder J., Phillips P.J., Jiang C. & Wiener S. (1996). Comparison of visible and infrared imagery for face recognition. *Proceedings of 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 182-187, Killington, VT

Wolff L.B., Socolinsky D.A. & Eveland C.K. (2001). Quantitative measurement of illumination invariance for face recognition using thermal infrared imagery, *Proceedings of CVPR Workshop on Computer Vision Beyond the Visible Spectrum.*

Wolff L.B., Socolinsky D.A. & Eveland C.K. (2006). Face Recognition in the Thermal Infrared, *Computer Vision Beyond the Visible Spectrum Book*, pp. 167-191, Springer London

Yin Z. & Malcolm A.A. (2000). Thermal and Visual Image Processing and Fusion, SIMTech Technical Report

Yoshitomi Y., Miyaura T., Tomita S, & Kimura S. (1997). Face identification using thermal image processing, *Proceedings of IEEE Int. Workshop on Robot and Human Communication*, pp.374-379

http://cse.ohio-state.edu/otcbvs-bench/Data/02/download

http://sensorsinc.com/border_security1

http://sensorsinc.com/nightvision

http://sensorsinc.com/whyswir

Part 3

Refinements of Classical Methods

7

Dimensionality Reduction Techniques for Face Recognition

Shylaja S S, K N Balasubramanya Murthy and S Natarajan
Department of Information Science and Engineering, P E S Institute of Technology
India

1. Introduction

High level of image content analysis is required for several applications. This is taking more significance as the number of digital images stored is growing exponentially. On the one hand the technology should help store these images, on the other, enable us to develop newer algorithmic models aimed at efficient and quick retrieval of images. The entire captured data may not be applicable for an application and hence deriving a subset of data to achieve objective function is desirable.

Face detection and recognition are preliminary steps to a wide range of applications such as personal identity verification, video-surveillance, facial expression extraction, gender classification, advanced human and computer interaction. A face recognition system would allow user to be identified by simply walking past a surveillance camera. Research has been devoted to facial recognition for years and has brought forward algorithms in an attempt to be as accurate as humans are.

A face recognition system is expected to identify faces present in images and videos automatically. It can operate in either or both of two modes:

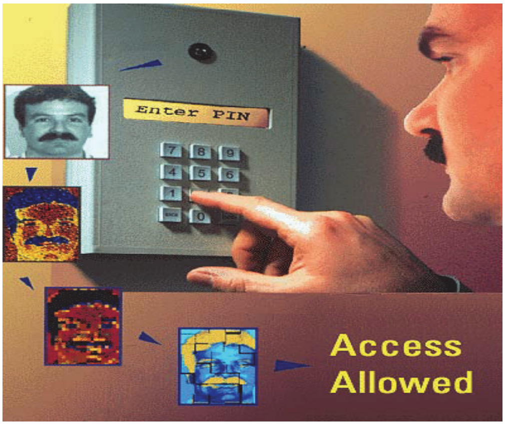


Fig. 1. Face Verification System

- Face verification or authentication,(fig above)

- Face identification or recognition.

Face verification involves a one-to-one match that compares a query face image against a template face image whose identity is being claimed. Face identification involves one-to-many matches that compare a query face image against all the template images in the database to determine the identity of the query face. Another face recognition scenario involves a watch-list check, where a query face is matched to a list of suspects (one-to-few matches). As per Hietmeyer, face recognition is one of the most effective biometric techniques for travel documents and scored higher on several evaluation parameters.

Computational models of face recognition must address several difficult problems. This difficulty arises from the fact that faces must be represented in a way that best utilizes the available face information to distinguish a particular face from all other faces. The problem of dimensionality reduction arises in face recognition because an m X n face image is reconstructed to form a column vector of mn components, for computational purposes. As the number of images in the data set increases, the complexity of representing data sets increases. Analysis with a large number of variables generally consumes a large amount of memory and computation power.

2. Dimensionality reduction

Efforts are on for efficient storage and retrieval of images. Considerable progress has happened in face recognition with newer models especially with the development of powerful models of face appearance. These models represent faces as points in high-dimensional image spaces and employ dimensionality reduction to find a more meaningful representation, therefore, addressing the issue of the ”curse of dimensionality”. Dimension reduction is a process of reducing the number of variables under observation. The need for dimension reduction arises when there is a large number of univariate data points or when the data points themselves are observations of a high dimensional variable. The key observation is that although face images can be regarded as points in a high-dimensional space, they often lie on a manifold (i.e., subspace) of much lower dimensionality, embedded in the high-dimensional image space. The main issue is how to properly define and determine a low-dimensional subspace of face appearance in a high-dimensional image space.

Dimensionality reduction techniques using linear transformations have been very popular in determining the intrinsic dimensionality of the manifold as well as extracting its principal directions. Dimensionality reduction is an effective approach to downsizing data. In statistics, dimension reduction is the process of reducing the number of random variables under consideration, $R^N \rightarrow R^M$ $(M<N)$ and can be divided into feature selection and feature extraction.

Feature selection is choosing a subset of all the features

$$[x_1\ x_2 \dots\ x_n] \xrightarrow{\text{Feature selection}} [\, x_{i1}\, x_{i2} \dots x_{im}\,]$$

Feature extraction is creating new features from existing ones

$$[x_1\ x_2 \dots\ x_n] \xrightarrow{\text{Feature extraction}} [\, y_1\ y_2 \dots\ y_m\,]$$

In either case, the goal is to find a low-dimensional representation of the data while still describing the data with sufficient accuracy.

For reasons of computational and conceptual simplicity, the representation is often sought as a linear transformation of the original data. In other words, each component of the representation is a linear combination of the original variables. Well-known linear transformation methods include principal component analysis, factor analysis, and projection pursuit. Independent component analysis (ICA) is a recently developed method in which the goal is to find a linear representation of nongaussian data so that the components are statistically independent, or as independent as possible. Such a representation seems to capture the essential structure of the data in many applications, including feature extraction and signal separation.

Several techniques exist to tackle the curse of dimensionality out of which some are linear methods and others are nonlinear. PCA, LDA, LPP are some popular linear methods and nonlinear methods include ISOMAP & Eigenmaps. PCA and LDA are the two most widely used subspace learning techniques for face recognition. These methods project the training sample faces to a low dimensional representation space where the recognition is carried out. The main supposition behind this procedure is that the face space (given by the feature vectors) has a lower dimension than the image space (given by the number of pixels in the image), and that the recognition of the faces can be performed in this reduced space**.** PCA has the advantage of capturing holistic features but ignore the localized features. Fisher faces from LDA technique extracts discriminating features between classes and is found to perform better for large data sets. Its shortcoming is that of Small Sample Space (SSS) problem. LPPs are linear projective maps that arise by solving variational problem that optimally preserves the neighborhood structure of the data set.

In many cases, face images may be visualized as points drawn on a low-dimensional manifold hidden in a high-dimensional ambient space. Specially, we can consider that a sheet of rubber is crumpled into a (high-dimensional) ball. The objective of a dimensionality-reducing mapping is to unfold the sheet and to make its low-dimensional structure explicit. If the sheet is not torn in the process, the mapping is topology-preserving. Moreover, if the rubber is not stretched or compressed, the mapping preserves the metric structure of the original space.

PCA is guaranteed to discover the dimensionality of the manifold and produces a compact representation. Turk and Pentland use Principal Component Analysis to describe face images in terms of a set of basis functions, or "eigenfaces". LDA is a supervised learning algorithm. LDA searches for the project axes on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Unlike PCA which encodes information in an orthogonal linear space, LDA encodes discriminating information in a linear separable space using bases are not necessarily orthogonal. It is generally believed that algorithms based on LDA are superior to those based on PCA. However, some recent work shows that, when the training dataset is small, PCA can outperform LDA, and also that PCA is less sensitive to different training datasets.

Recently, a number of research efforts have shown that the face images possibly reside on a nonlinear submanifold. However, both PCA and LDA effectively see only the Euclidean structure. They fail to discover the underlying structure, if the face images lie on a nonlinear submanifold hidden in the image space. Some nonlinear techniques have been proposed to discover the nonlinear structure of the manifold, *e.g.* Isomap, LLE and Laplacian Eigenmap. These nonlinear methods do yield impressive results on some benchmark artificial data sets. However, they yield maps that are defined *only* on the training data points and how to evaluate the maps on novel test data points remains unclear.

3. Singular Value Decomposition (SVD)

Singular value decomposition (SVD) is an important factorization of a rectangular real or complex matrix, with many applications in signal processing and statistics. As applied to face recognition this technique is used to extract the holistic global features of the training set SVD is the best, in the mean-square error sense, linear dimension reduction technique. Being based on the covariance matrix of the variables, it is a second-order method. SVD seeks to reduce the dimension of the data by finding a few orthogonal linear combinations of the original variables with the largest variance.

The basic idea behind SVD is taking a high dimensional, highly variable set of data points and reducing it to a lower dimensional space that exposes the substructure of the original data more clearly and orders it from most variation to the least. What makes SVD practical for pattern recognition applications is that one can simply ignore variation below a particular threshold to massively reduce the data but be assured that the main relationships of interest have been preserved.

Singular value decomposition (SVD) can be looked at from three mutually compatible points of view. On the one hand, we can see it as a method for transforming correlated variables into a set of uncorrelated ones that better expose the various relationships among the original data items. At the same time, SVD is a method for identifying and ordering the dimensions along which data points exhibit the most variation. This ties into the third way of viewing SVD, which is that once we have identified where the most variation is, it's possible to find the best approximation of the original data points using fewer dimensions. Hence, SVD can be seen as a method for data reduction.

As said earlier Singular Value Decomposition is a way of factoring matrices into a series of linear approximations that expose the underlying structure of the matrix. If A is the input matrix, calculating the SVD consists of finding the eigenvalues and eigenvectors of AA^T and A^TA. This yields three matrices U,V & S where the eigenvectors of A^TA make up the columns of V , the eigenvectors of AA^T make up the columns of U. and the singular values in **S** are square roots of eigenvalues from AA^T or A^TA. The singular values are the diagonal entries of the S matrix and are arranged in descending order. The singular values are always real numbers. If the matrix A is a real matrix, then U and V are also real.

In the factorization, the first principal component is s1, with the largest variance is the linear combination with T T . We have $S_1 = XW_1$, where the p-dimensional coefficient vector solves $W_1 = (W_{11},....,W_{1P})$ where

$$W_1 = \underset{\|w=1\|}{\arg\max} Var\{x,w\} \tag{1}$$

The second PC is the linear combination with the second largest variance and orthogonal to the first PC, and so on. There are as many PCs as the number of the original variables. PCs explain most of the variance, so that the rest can be disregarded with minimal loss of information. Since the variance depends on the scale of the variables, it is customary to first standardize each variable to have mean zero and standard deviation one. After the standardization, the original variables with possibly different units of measurement are all in comparable units.

The mathematical model formulated is given below:

Let A is m′ X n′ real matrix and N=A^{T}A

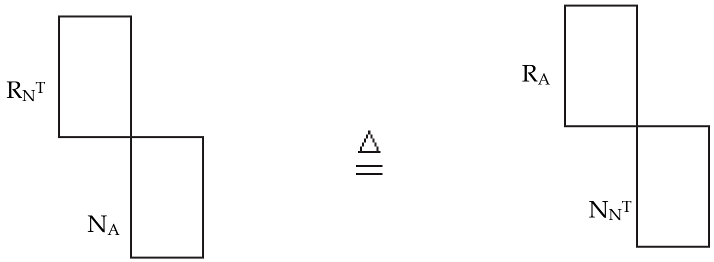


Fig. 2. Range and Null space of matrix

R denotes the range space and N denotes the null space of a matrix. Rank of A, A^T, A^TA, AA^T is equal and is denoted by ρ orthonormal basis v_i $1 \le i \le \rho$ are sought for R_A^T where ρ is the rank of R_A^T& u_i $1 \le i \le \rho$ for R_A such that,

$$AV_j = S_jU \tag{2}$$

$$A^TU_j = S_jV_j \,, \qquad 1 \le j \le \rho \tag{3}$$

Advantages of having such a basis are that geometry becomes easy and gives a decomposition of A into ρ one-ranked matrices. Combining the equations (2) & (3)

$$A = \sum S_jV_{\mathrm{j}}\,\mathrm{U_j}^{\mathrm{T}} \,, \qquad 1 \le j \le \rho \tag{4}$$

If V_j is known then , $U_j = (1/S_j)AV_j$ $|S_j| = \| AV_j \|$ therefore ,$s_j \neq 0$, choosing $s_j > 0$,

$$AV_j = S_jU_j \tag{5}$$

$$A^TAV_j = S_jA^TU_j \tag{6}$$

$$A^TAV_j = S_j^2V_j \tag{7}$$

Let $S_j^2 = \mu_i$, $NV_j = \mu_iV_i$ is required U_i's as orthonormal eigenvectors of $N=A^TA$ are found and $S_j = \sqrt{\mu_i}$ Where μ_i >0 are eigen values corresponding to V_j. The resulting Ui span the Eigen subspace. When SVD is applied to the sample set below in figure 3, the corresponding eigen faces obtained are shown in figure 4. The figure is highlighting the holistic features from the given sample set.

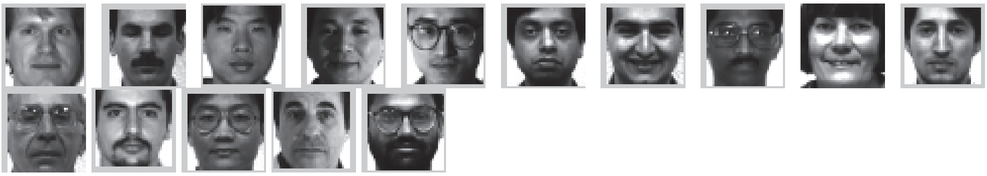


Fig. 3. Training set example faces

Fig. 4. Eigen faces from SVD

Basis selection from SVD

If A is the face Space, then x vectors are drawn from $[X_1\ldots..X_x] = \Pi_{<1..x>}(A^{-1}UD)$ Where U & D are the unitary and diagonal matrices of SVD of A.

5. Linear Discriminant Analysis

Fisher Linear Discriminant also referred as Linear Discriminant Analysis is a classical pattern recognition method, which was introduced by Fisher (1934). It is a very effective feature extraction method but facing issues for Small Sample Space problem.

The Dimensionality Reduction technique SVD searches for directions in the data that have largest variance and subsequently project the data onto it. In this way, one can obtain a lower dimensional representation of the data, that removes some of the ″noisy″ directions. There are many difficult issues with how many directions one needs to choose. It is an unsupervised technique and as such does not include label information of the data. For instance, if we imagine 2 clusters in 2 dimensions, one clusters has y = 1 and the other y = ¡1. The clusters are positioned in parallel and very closely together, such that the variance in the total data-set, ignoring the labels, is in the direction of the clusters. For classification, this would be a terrible projection, because all labels get evenly mixed and will destroy the useful information.

A much more useful projection is orthogonal to the clusters, i.e. in the direction of least overall variance, which would perfectly separate the data-cases (obviously, we would still need to perform classification in this 1-D space).

The conventional solution to misclassification for small sample size problem and large data set with similar faces is the use of PCA into LDA i.e. fisher faces. PCA is used for dimensionality reduction and then LDA is performed on the lower dimensional space. Discriminant analysis often produces models whose accuracy approaches complex modern methods. The target variable may have two or more categories. The following figure 5 shows a plot of the two categories with the two predictors on orthogonal axes:

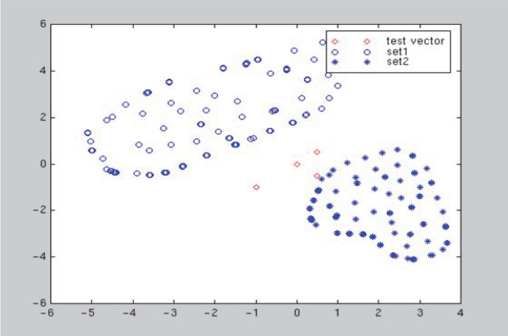


Fig. 5. A plot of the two categories with the two predictors on orthogonal axes

A transformation function is found that maximizes the ratio of between-class variance to within-class variance as illustrated by this figure 6 produced by Ludwig Schwardt and Johan du Preez:

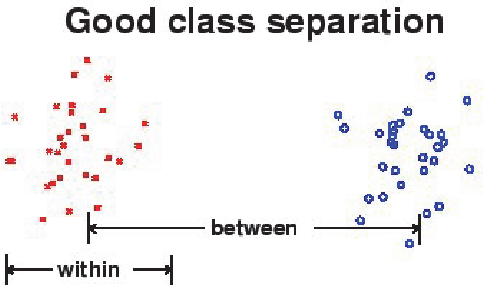


Fig. 6. Output of applying transformation function

The transformation seeks to rotate the axes so that when the categories are projected on the new axes, the differences between the groups are maximized. So the question is, how do we utilize the label information in finding informative projections?
To that purpose Fisher-LDA considers maximizing the following objective:

$$J(w)=\frac{w^T S_B w}{w^T S_w w} \tag{8}$$

The second use of the term LDA refers to a discriminative feature transform that is optimal for certain cases [10]. This is what we denote by LDA throughout this paper. In the basic formulation, LDA finds eigenvectors of matrix

$$T=S_w^{-1}\ S_b \tag{9}$$

Here S_b is the between-class covariance matrix, that is, the covariance matrix of class means. Sw denotes the within-class covariance matrix, that is equal to the weighted sum of covariance matrices computed for each class separately. S_w^{-1} captures the compactness of each class, and S_b represents the separation of the class means. Thus T captures both. The eigenvectors corresponding to largest k eigenvalues of T form the rows of the transform matrix w, and new discriminative features d_k are derived from the original ones d simply by

$$d_k=W_d \tag{10}$$

The straightforward algebraic way of deriving the LDA transform matrix is both a strength and a weakness of the method. Since LDA makes use of only second-order statistical information, covariances, it is optimal for data where each class has a unimodal Gaussian density with well separated means and similar covariances. Large deviations from these assumptions may result in sub-optimal features.

Also the maximum rank of S_b in this formulation is $N_c - 1$ where N_c the number of different classes is. Thus basic LDA cannot produce more than $N_c - 1$ features. This is, however, simple to remedy by projecting the data onto a subspace orthogonal to the computed eigenvectors, and repeating the LDA analysis in this space.

However, the classification performance of traditional LDA is often degraded by the fact that their separability criteria are not directly related to their classification accuracy in the output space. A solution to the problem is to introduce weighting functions into LDA. Object classes that are closer together in the output space, and thus can potentially result in misclassification, should be more heavily weighted in the input space. This idea has been further extended in with the introduction of the fractional-step linear discriminant analysis algorithm (F-LDA), where the dimensionality reduction is implemented in a few small fractional steps allowing for the relevant distances to be more accurately weighted. Although the method has can be applied on low dimensional patterns it cannot be directly applied to high-dimensional patterns, such as those face images due to two factors: (1) the computational difficult of the eigen-decomposition of matrices in the high-dimensional image space; (2) the degenerated scatter matrices caused by the small sample size, which widely exists in the FR tasks where the number of training samples is smaller than the dimensionality of the samples.

The traditional solution to the SSS problem requires the incorporation of a PCA step into the LDA framework. In this approach, PCA is used as a pre-processing step for dimensionality reduction so as to discard the null space of the within-class scatter matrix of the training data set. Then LDA is performed in the lower dimensional PCA subspace. However, it has been shown that the discarded null space may contain significant discriminatory information. To prevent this from happening, solutions without a separate PCA step, called direct LDA (D-LDA) methods have been presented recently. In the D-LDA framework, data are processed directly in the original high-dimensional input space avoiding the loss of significant discriminatory information due to the PCA pre-processing step.

Firstly dimensionality of the original input space is lowered by introducing a new variant of D-LDA that results in a low-dimensional SSS-free subspace where the most discriminatory features are preserved. The variant of D-LDA utilizes a modified Fisher's criterion to avoid a problem resulting from the wage of the zero eigenvalues of the within-class scatter matrix as possible divisors. Also, a weighting function is introduced into the variant of D-LDA, so that a subsequent F-LDA step can be applied to carefully re-orient the SSS-free subspace resulting in a set of optimal discriminant features for face representation.

The DF-LDA is a linear pattern recognition method. Compared with nonlinear models, a linear model is rather robust against noises and most likely will not over fit. Although it has been shown that distribution of face patterns is highly non convex and complex in most cases, linear methods are still able to provide cost effective solutions to the FR tasks through integration with other strategies, such as the principle of "divide and conquer," in which a large and nonlinear problem is divided into a few smaller and local linear sub problems.

Let S_{BTW} and S_{WTH} denote the between- and within-class scatter matrices of the training image set, respectively. LDA-like approaches such as the Fisherface method find a set of basis vectors, denoted by that maximizes the ratio between S_{BTW} and S_{WTH} is

$$\Psi = \arg\max_{\Psi} \frac{\left|\left(\Psi^T S_{BTW} \Psi\right)\right|}{\left|\left(\Psi^T S_{WTH} \Psi\right)\right|} \tag{10}$$

The maximization process in (3) is not directly linked to the classification error which is the criterion of performance used to measure the success of the Face Recognition procedure. Thus, the weighted between-class scatter matrix can be expressed as:

$$S_{BTW} = \sum^{C} \Phi_i \Phi_i^T \tag{11}$$

where

$$\Phi_i = \left(L_i/L\right)^{1/2} \sum_{i=1}^{C} \left(w\left(d_{ij}\right)\right)^{1/2} \left(\overline{Z}_i - \overline{Z}_j\right), \overline{Z}_i \tag{12}$$

is the mean of class Z_i, L_i is the number of elements in Z_i, and

$$d_{i,j} = \left\|\overline{Z_i} - \overline{Z_j}\right\| \tag{13}$$

is the Euclidean distance between the means of class i and j .

Basis selection from DF-LDA

The set Y vectors are chosen by the equation $[y_1\ldots..y_y] = \Pi_{<1..y>}(U^T S_{TOT} U^T)$ Where S_{TOT} is the sum of between and within class scatter matrices, U is a diagonal matrix from Eigen values and vectors. Fisher faces are shown in figure 7 below.

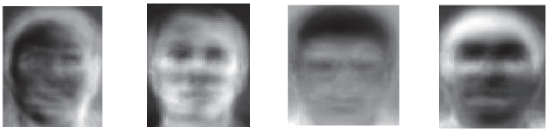


Fig. 7. Fisher Faces from DF-LDA

We can clearly see from fisher faces that more pronounced features are highlighted than the rest of the face point like hair, eyebrows etc.

6. Locality preserving projections

Different from Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) which effectively see only the Euclidean structure of face space, LPP finds an embedding that preserves local information, and obtains a face subspace that best detects the essential face manifold structure. The Laplacianfaces are the optimal linear approximations to the eigen functions of the Laplace Beltrami operator on the face manifold. In this way, the unwanted variations resulting from changes in lighting, facial expression, and pose may be eliminated or reduced. Theoretical analysis shows that PCA, LDA and LPP can be obtained from different graph models. By using Locality Preserving Projections (LPP), the face images are mapped into a face subspace for analysis.

LPP shares many of the data representation properties of nonlinear techniques such as Laplacian Eigen maps or Locally Linear Embedding. Yet LPP is linear and more crucially is defined everywhere in ambient space rather than just on the training data points. It builds a graph incorporating neighborhood information of the data set. Using the notion of the Laplacian of the graph, transformation matrix is computed which maps the data points to a subspace.

This linear transformation optimally preserves local neighborhood information in a certain sense. The representation map generated by the algorithm may be viewed as a linear discrete approximation to a continuous map that naturally arises from the geometry of the manifold. In the meantime, there has been some interest in the problem of developing low dimensional representations through kernel based techniques for face recognition. These methods can discover the nonlinear structure of the face images. However, they are computationally expensive. Moreover, none of them explicitly considers the structure of the manifold on which the face images possibly reside.

While the Eigen faces method aims to preserve the global structure of the image space, and the Fisher faces method aims to preserve the discriminating information; our Laplacianfaces method aims to preserve the local structure of the image space. In many real world classification problems, the local manifold structure is more important than the global Euclidean structure, especially when nearest neighbor like classifiers are used for classification.

LPP seems to have discriminating power although it is unsupervised. An efficient subspace learning algorithm for face recognition should be able to discover the nonlinear manifold structure of the face space. LPP shares some similar properties to LLE, such as a locality preserving character. However, their objective functions are totally different. LPP is obtained by finding the optimal linear approximations to the eigen functions of the Laplace Beltrami operator on the manifold. LPP is linear, while LLE is nonlinear. Moreover, LPP is defined everywhere, while LLE is defined only on the training data points and it is unclear how to evaluate the maps for new test points. In contrast, LPP may be simply applied to any new data point to locate it in the reduced representation space. LPP seeks to preserve the intrinsic geometry of the data and local structure.

The objective function of LPP is as follows:

$$\min \sum_{ij} \left(y_i - y_j \right)^2 S_{ij} \tag{14}$$

Where

$$S_{ij} = \begin{cases} \exp\left(\left\| x_i - x_j \right\|^2 / t \right), & \left\| x_i - x_j \right\|^2 < \varepsilon \\ 0 \end{cases} \tag{15}$$

7. Statistical view of LPP

LPP can also be obtained from statistical viewpoint. Suppose the data points follow some underlying distribution. Let d be the number of non-zero S_{ij}, and D be a diagonal matrix whose entries are column (or row, since S is symmetric) sums of S, $D_{ii}=\sum_j S_{ji}$. By the Strong Law of Large Numbers, $E(zz^T \mid ||z||< \varepsilon)$ can be estimated from the sample points as follows:

$$E\left(ZZ^T \middle| \ \|z\|<\varepsilon \ \right)$$

$$\approx \frac{1}{d}\sum_{\|Z\|<\varepsilon} ZZ^T$$

$$= \frac{1}{d}\sum_{\|x_i - x_j\|<\varepsilon} \left(x_i - x_j\right)\left(x_i - x_j\right)^T$$

$$= \frac{1}{d}\sum_{i,j}\left(x_i - x_j\right)\left(x_i - x_j\right)^T S_{ij}$$

$$= \frac{1}{d}\left(\sum_{i,j} x_i x_i^T S_{ij} + \sum_{i,j} x_j x_j^T S_{ij} - \sum_{i,j} x_i x_j^T S_{ij} - \sum_{i,j} x_j x_i^T S_{ij}\right)$$

$$= \frac{2}{d}\left(\sum_{i,j} x_i x_i^T D_{ii} - \sum_{i,j} x_i x_j^T S_{ij}\right)$$

$$= \frac{2}{d}\left(XDX^T - XSX^T\right)$$

$$= \frac{2}{d} XLX^T \tag{16}$$

where $L = D - S$ is the Laplacian matrix. The *ith* column of matrix X is x*i*.

8. Theoretical analysis of LPP, PCA AND LDA

In this section, we present a theoretical analysis of LPP and its connections to PCA and LDA.

8.1 Connections to PCA

It is worthwhile to point out that XLX^T is the data covariance matrix, if the Laplacian matrix *L is*

$$\frac{1}{n} I - \frac{1}{n^2} ee^T \tag{17}$$

where n is the number of data points, I is the identity matrix and e is a column vector taking 1 at each entry. In fact, the Laplacian matrix here has the effect of removing the sample mean from the sample vectors.

In this case, the weight matrix S takes $1/n^2$ at each entry, i.e

$$S_{ij} = 1/n^2, \forall i, j \tag{18}$$

$$D_{ii} = \sum_j S_{ji} = 1/n \tag{19}$$

Hence the Laplacian matrix is

$$L = D - S = \frac{1}{n} I - \frac{1}{n^2} ee^T \tag{20}$$

Let m denote the sample mean i.e.

$$m = 1/n \sum\nolimits_i x_i \tag{21}$$

we have

$$XLX^T = \frac{1}{n} X\left(I - \frac{1}{n^2} ee^T \right) X^T$$

$$= \frac{1}{n} XX^T - \frac{1}{n^2} (X_e)(X_e)^T$$

$$= \frac{1}{n} \sum_i x_i x_i^T - \frac{1}{n^2} (nm)(nm)^T$$

$$= \frac{1}{n} \sum_i (x_i - m)(x_i - m)^T + \frac{1}{n} \sum_i x_i m^T + \frac{1}{n} \sum_i m x_i^T - \frac{1}{n} \sum_i mm^T - mm^T$$

$$= E[(x - m)(x - m)]^T + 2mm^T - 2mm^T$$

$$= E[(x - m)(x - m)]^T \tag{22}$$

Where $E[(x - m)(x - m)]^T$, is just the covariance matrix of the data set

The above analysis shows that the weight matrix S plays a key role in the LPP algorithm. When we aim at preserving the global structure, we take ε (or k) to be infinity and choose the eigenvectors (of the matrix XLX^T) associated with the largest eigenvalues. Hence the data points are projected along the directions of maximal variance. ε should be sufficiently small to preserve the local structure and choose the Eigen vectors associates with smallest Eigen values.

Hence the data points are projected along the directions preserving locality. It is important to note that, when ε (or k) is sufficiently small, the Laplacian matrix is no longer the data covariance matrix, and hence the directions preserving locality are not the directions of minimal variance. In fact, the directions preserving locality are those minimizing *local* variance.

8.2 Connections to LDA

LDA seeks directions that are efficient for discrimination. The projection is found by solving the generalized Eigen value problem

$$S_B w = \lambda S_W w \tag{23}$$

where S_B and S_W are between and within class scatter matrices. Suppose there are *l* classes. The ith class contains *ni* sample points. Let m(*i*) denote the average vector of the *i*th class. Let x(*i*) denote the random vector associated tothe *i*th class and) (i *j* x denote the *j*th sample point in the *i*t*h* class. We can rewrite the matrix S_w as follows:

$$S_W = \sum_{i=1}^{l}\left(\sum_{j=1}^{ni}\left(x_j^{(i)} - m^{(i)}\right)\left(x_j^{(i)} - m^{(i)}\right)^T\right)$$

$$= \sum_{i=1}^{l}\left(\sum_{j=1}^{ni}\left(x_j^{(i)}\left(x_j^{(i)}\right)^T - m^{(i)}\left(m^{(i)}\right)^T - x_j^{(i)}\left(m^{(i)}\right)^T + m^{(i)}\left(m^{(i)}\right)^T\right)\right)$$

$$= \sum_{i=1}^{l}\left(\sum_{j=1}^{ni}\left(x_j^{(i)}\left(x_j^{(i)}\right)^T - n_i m^{(i)}\left(m^{(i)}\right)^T\right)\right)$$

$$= \sum_{i=1}^{l}\left(X_iX_i^T - \frac{1}{n_i}\left(x_1^{(i)} + \ldots + x_{ni}^{(i)}\right)\left(x_1^{(i)} + \ldots + x_{ni}^{(i)}\right)^T\right)$$

$$= \sum_{i=1}^{l}\left(X_iX_i^T - \frac{1}{n_i}X_i\left(e_ie_i^T\right)X_i^T\right)$$

$$= \sum_{i=1}^{l} X_iL_iX_i^T \tag{24}$$

Where,

$X_iL_iX_i^T$ is the data covariance matrix of the ith class and

X_i = [$X_1^{(i)}$, $X_2^{(i)}$, $X_3^{(i)}$,…. $X_{ni}^{(i)}$] is a d x n_i matrix.

$L_i = I - 1/n_ie_ie_i^T$ is a n_i x n_i matrix where *I* is the identity matrix and $e_i\left(1,1,1....1\right)^T$ is an *ni* dimensional vector.

To further simplify the above equation, we define

$$W_{ij} = \begin{cases} 1/n_k & \text{if } x_i \text{ and } x_j \text{ both belong to the kth class} \\ 0 & \text{otherwise,} \end{cases} \tag{25}$$

It is interesting to note that we could regard the matrix *W* as the weight matrix of a graph with data points as its nodes. Specifically, *Wij* is the weight of the edge (x*i*, x*j*). *W* reflects the class relationships of the data points. The matrix *L* is thus called *graph Laplacian*, which plays key role in LPP.

Similarly, we can compute the matrix *SB* as follows:

$$S_B = \sum_{i=1}^{l} n_i\left(m^{(i)} - m\right)\left(m^{(i)} - m\right)^T$$

$$=\left(\sum_{i=1}^{l} n_i m^{(i)}\left(m^{(i)}\right)^T\right)-m\left(\sum_{i=1}^{l} n_i\left(m^{(i)}\right)^T\right)-\left(\sum_{i=1}^{l} n_i m^{(i)}\right)m^T+\left(\sum_{i=1}^{l} n_i\right)mm^T$$

$$=\sum_{i=1}^{l}\left(\frac{1}{n_i}\left(x_1^{(i)}+\ldots+x_{ni}^{(i)}\right)\left(x_1^{(i)}+\ldots+x_{ni}^{(i)}\right)^T\right)-2nmm^T+nmm^T$$

$$=\left(\sum_{i=1}^{l}\sum_{j,k=1}^{n_i}\frac{1}{n_i}x_j^{(i)}\left(x_j^{(i)}\right)^T\right)-2nmm^T+nmm^T$$

$$=XWX^T-2nmm^T+nmm^T$$

$$=XWX^T-nmm^T$$

$$=XWX^T-X\left(\frac{1}{n}ee^T\right)X^T$$

$$=X\left(W-\frac{1}{n}ee^T\right)X^T$$

$$=X\left(W-I+I-\frac{1}{n}ee^T\right)X^T$$

$$=-XLX^T+X\left(I-\frac{1}{n}ee^T\right)X^T$$

$$=-XLX^T+C \tag{26}$$

where e = (1,1,…,1)^T is a n dimensional vector and $C=X\left(I-\frac{1}{n}ee^T\right)X^T$ is the data covariance matrix.

Thus, the generalized eigenvector problem of LDA can be written as follows:

$$S_B w=\lambda S_W w \tag{27}$$

$$\Rightarrow\left(C-XLX^T\right)w=\lambda XLX^T w$$

$$\Rightarrow Cw=\left(1+\lambda\right)XLX^T w$$

$$\Rightarrow XLX^T w=\frac{1}{1+\lambda}Cw$$

Thus, the projections of LDA can be obtained by solving the following generalized eigenvalue problem,

$$\Rightarrow XLX^T w=\lambda Cw \tag{28}$$

The optimal projections correspond to the eigenvectors associated with the smallest eigenvalues. If the sample mean of the data set is zero, the covariance matrix is simply XX^T which is close to the matrix XDX^T in the LPP algorithm. Our analysis shows that LDA actually aims to preserve discriminating information and global geometrical structure. Moreover, LDA has a similar form to LPP. However, LDA is supervised while LPP can be performed in either supervised or unsupervised manner.

8.3 Learning laplacian faces for representation

LPP is a general method for manifold learning. It is obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Betrami operator on the manifold. Therefore, though it is still a linear technique, it seems to recover important aspects of the intrinsic nonlinear manifold structure by preserving local structure. Based on LPP, Laplacianfaces method for face representation is a locality preserving subspace. In the face analysis and recognition problem one is confronted with the difficulty that the matrix XDX^T is sometimes singular. This stems from the fact that sometimes the number of images in the training set (n) is much smaller than the number of pixels in each image (m). In such a case, the rank of XDX^T is at most n, while XDX^T is an $m \times m$ matrix, which implies that XDX^T is singular. To overcome the complication of a singular XDX^T, we first project the image set to a PCA subspace so that the resulting matrix XDX^T is nonsingular. Another consideration of using PCA as preprocessing is for noise reduction. This method, we call *Laplacianfaces*, can learn an optimal subspace for face representation and recognition.

The algorithmic procedure of Laplacianfaces is formally stated below:

1. PCA projection: We project the image set {x*i*} into the PCA subspace by throwing away the smallest principal components.
2. Constructing the nearest-neighbor graph: Let G denote a graph with n nodes. The *ith* node corresponds to the face image x_i. We put an edge between nodes i and j if x_i and x_j are "close", i.e. x*i* is among k nearest neighbors of x*i* or x*i* is among k nearest neighbors of x_j. The constructed nearest neighbor graph is an approximation of the local manifold structure. Note that, here we do not use the ε - neighborhood to construct the graph. This is simply because it is often difficult to choose the optimal ε in the real world applications, while k nearest neighbor graph can be constructed more stably. The disadvantage is that the k nearest neighbor search will increase the computational complexity of our algorithm. When the computational complexity is a major concern, one can switch to the ε -neighborhood.
3. Choosing the weights: If node i and j are connected, put

$$S_{ij} = e\frac{\left\|x_i - x_j\right\|^2}{t} \tag{29}$$

where t is a suitable constant. Otherwise, put $S_{ij} = 0$. The weight matrix S of graph G models the face manifold structure by preserving local structure.

4. Eigenmap: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem:

$$XLX^T w = \lambda XDX^T w \tag{30}$$

where D is a *k*-dimensional vector. *W* is the transformation matrix. This linear mapping best preserves the manifold's estimated intrinsic geometry in a linear sense. The column vectors of *W* are the so called *Laplacianfaces*.

9. Face representation using laplacianfaces

As we described previously, a face image can be represented as a point in image space. A typical image of size *m*×*n* describes a point in *m*×*n*-dimensional image space. However, due to the unwanted variations resulting from changes in lighting, facial expression, and pose, the image space might not be an optimal space for visual representation, we have discussed how to learn a locality preserving face subspace which is insensitive to outlier and noise. The images of faces in the training set are used to learn such a locality preserving subspace. The subspace is spanned by a set of eigenvectors of equation (1), i.e. w0, w1, …, wk-1.

Eigenmaps are obtained from the generalized eigenvector problem as ALA^T a = λADA^T a where D is a diagonal matrix whose entries are column or row, since W is symmetric sums of W, D_{ii} = ΣjW_{ji}., L = D -W is the Laplacian matrix is equivalent nonlinear Laplace Beltrami opearator. The ith column of matrix A is xi. Let the column vectors a_0; _ _ _ ; a_{l-1} be the solutions of equation (), ordered according to their eigenvalues, in ascending order Thus, the embedding is as follows: yi = E^T x_i; E = (a_0; a_1; _ _ _ ; a_{l-1}) where y_i is a l-dimensional vector, and E is a n x l matrix.. The y_i represent the Laplacian faces.

9.1 Basis selection from LPP

Locality information can be preserved by the following transformation on A, the input face space [z_1 …. Z_z] = $\Pi_{<1..z}$(A^TL A) Where L =D-W gives the Laplacian matrix. D is the diagonal matrix and W is the weight matrix of the K nearest neighbors clustering.

Basis for the face space is obtained as, $B = [X_1\ X_2\ ....X_x\ Y_1\ Y_2......Y_y\ Z_1\ Z_2.....Z_z]$, such that

$$x + y + z = \frac{M}{3} \tag{31}$$

and

$$x + y + z = \frac{2M}{3} \tag{32}$$

where M is the dimension of the original face space

9.2 Projection onto reduced subspace

Each face in the training set Φ_i can be represented as a linear combination of these vectors, U_i ε B , 1 ≤ *i* ≤ K such that $\Phi_i = \sum_{j=1}^{k} w_j u_j$, where u_j's are Eigenfaces. These weights are calculated as: $w_j = u_j^T \Phi_i \Omega_i = [w_1\ w_2 ... w_k]$ i.e. the orthogonal projection of a face vector on each basis vector.

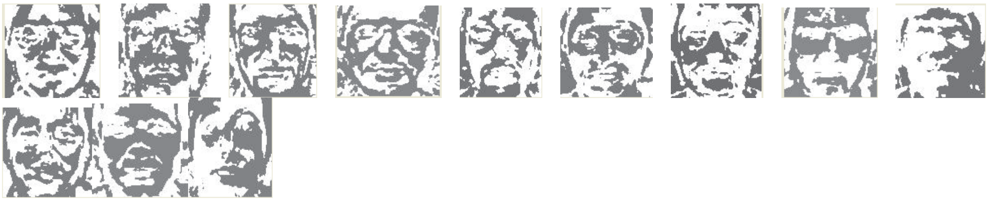


Fig. 8. Laplacian Faces from LPP

10. Independent component analysis

Independent component analysis (ICA) is a statistical method, the goal of which is to decompose multivariate data into a linear sum of non-orthogonal basis vectors with coefficients (encoding variables, latent variables, hidden variables) being statistically independent.

ICA generalizes a widely-used subspace analysis method such as principal component analysis (PCA) and factor analysis, allowing latent variables to be non-Gaussian and basis vectors to be non-orthogonal in general. ICA is a density estimation method where a linear model is learned such that the probability distribution of the observed data is best captured, while factor analysis aims at best modeling the covariance structure of the observed data.

The ICA model is a generative model, which means that it describes how the observed data are generated by a process of mixing the components *si*. The independent components are latent variables, meaning that they cannot be directly observed. Also the mixing matrix is assumed to be unknown. All we observe is the random vector X, and we must estimate both A and S using it. This must be done under as general assumptions as possible. The starting point for ICA is the very simple assumption that the components S_i are statistically *independent*. It will be seen below that we must also assume that the independent component must have *nongaussian* distributions. However, in the basic model we do *not* assume these distributions known (if they are known, the problem is considerably simplified.) For simplicity, we are also assuming that the unknown mixing matrix is square, but this assumption can be sometimes relaxed. Then, after estimating the matrix A, we can compute its inverse, say W, and obtain the independent component simply by: s=Wx

ICA is very closely related to the method called *blind source separation* (BSS) or blind signal separation. A "source" means here an original signal, i.e. independent component, like the speaker in a cocktail party problem. "Blind" means that we know very little, if anything, on the mixing matrix, and make little assumptions on the source signals. ICA is one method, perhaps the most widely used, for performing blind source separation.

The task of ICA is to estimate the mixing matrix A or its inverse $W = A^{-1}$ such that elements of the estimate $y = A^{-1}x = Wx$ are as independent as possible. For the sake of simplicity, we often leave out the index t if the time structure does not have to be considered.

PCA makes one important assumption: the probability distribution of input data must be Gaussian. When this assumption holds, covariance matrix contains all the information of (zero-mean) variables. Basically, PCA is only concerned with second-order (variance) statistics. The mentioned assumption need not be true. If we presume that face images have more general distribution of probability density functions along each dimension, the representation problem has more degrees of freedom. In that case PCA would fail because the largest variances would not correspond to meaningful axes of PCA.

$$x_i(t) = a_i1*s_1(t) + a_i2*s_2(t) + a_i3*s_3(t) + a_i4*s_4(t) \ldots \tag{33}$$

Here, i =1:4.
In vector-matrix notation, and dropping index t, this is

$$x = A\ s \tag{34}$$

$$s = A^{-1}\ x \tag{35}$$

$$s = W\ x \tag{36}$$

$$W = A^{-1} \tag{37}$$

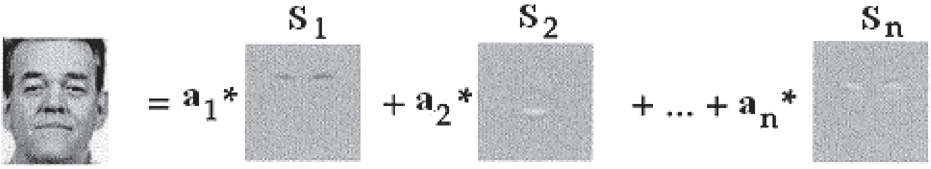


Fig. 9. Mixture Matrix forming face

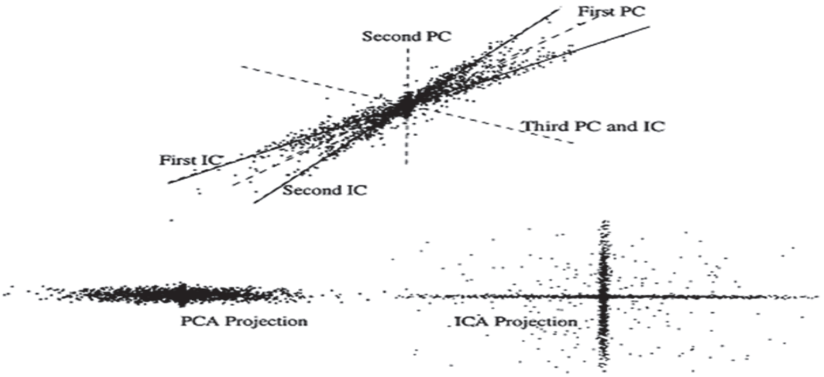


Fig. 10. Different Principal Component(PC) directions & PCA vs. ICA Projections

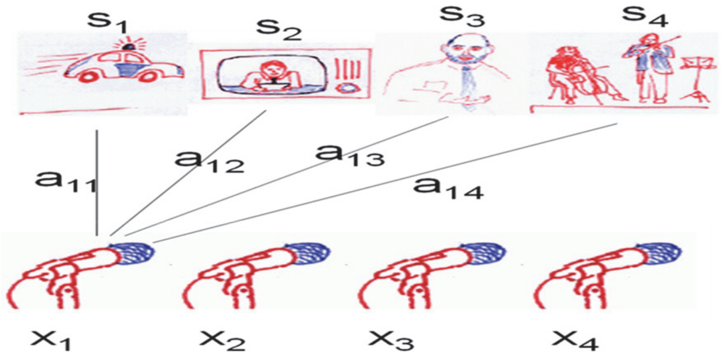


Fig. 11. x=As (Blind Source Separation)

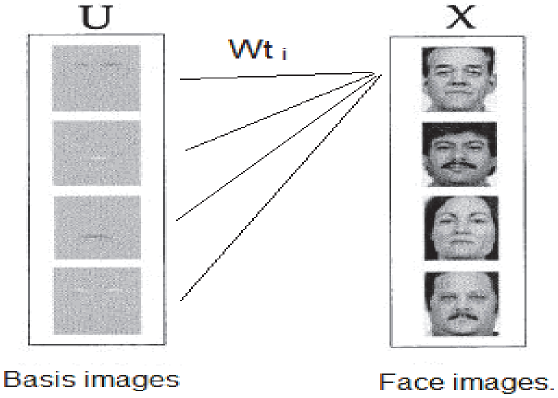


Fig. 12. Construction of face from Basis

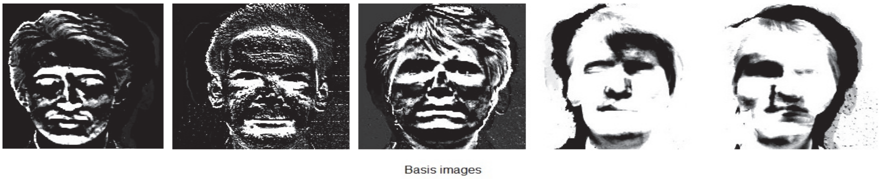


Fig. 13. Basis Images from ICA

11. Random projections

There has been a strong trend lately in face processing research away from geometric models towards appearance models. Appearance-based methods employ dimensionality reduction to represent faces more compactly in a low-dimensional subspace which is found by optimizing certain criteria. Recently, Random Projection (RP) has emerged as a powerful method for dimensionality reduction. It represents a computationally simple and efficient method that preserves the structure of the data without introducing significant distortion. D $O(\log n / \varepsilon^2)$ dimensional subspace such that the distances between the points are approximately preserved.

Transforms Γ_i to a lower dimension *d*, with d<<p via the following transformation: $\Gamma_i = R\Gamma_i$ where *R* is orthonormal and its columns are realizations of independent and identically distributed (i.i.d.) zero-mean normal variables, scaled to have unit length. RP is motivated by the *Johnson-Lindenstrauss* lemma that states that a set of *M* points in a high dimensional.

Euclidean space can be mapped down onto a $d \geq O(\log n / \varepsilon^2)$ dimensional subspace such that the distances between the points are approximately preserved.

The main reason for orthogonalizing the random vectors is to preserve the similarities between the original vectors in the low-dimensional space. In high enough dimensions, however, it is possible to save computation time by avoiding the orthogonalization step without affecting much the quality of the projection matrix. This is due to the fact that, in high-dimensional spaces, there exist a much larger number of almost orthogonal vectors than orthogonal vectors. Thus, high-dimensional vectors having random directions are very likely to be close to orthogonal.

12. Mixture of components

One can use different ratios of feature vectors drawn from SVD, DF-LDA & LPP Techniques. The first step can be normalizing the images in the training set to compensate for the illumination effects. These processed images should be subjected to dimensionality reduction using each of the methods mentioned in the chapter. Basis selection can be carried out using these independent sets of dimension reduced vectors in different proportions aimed at enhancing the efficiency and accuracy of recognition task. Below is a sample example mentioned with two trials, one with for 1/3rd dimensionality reduction and another with 2/3rd reduction. In each of the trials, several iterations are performed by taking different combinations of the feature vectors. The iterations will converge when the desired precision of recognition rate is obtained.

12.1 Example

12.1.1 Preprocessing

The Face Space: For the recognition task , each m X n I_i image in the training set is transformed into a column vector of mn components. A matrix S (mn X M) is constructed such that S =[$I_1 I_2 \ldots I_M$] , where M is number of face images in training set It is found that all N vectors are linearly independent, which implies that the range space of matrix S is the entire region spanned by the columns of S. i.e Range space of S R(S)=[S]

Normalization: Normalize the images ,to reduce illumination effects and lighting conditions as,

$$Ai = (\Phi - \mu_i)X\frac{\delta'}{\delta} + \mu' \tag{40}$$

For i=1,2,3….., M
Where,

$$\mu_i = \frac{1}{N^2} \sum_{j=1}^{N^2} x_j \tag{41}$$

$$\mu'_i = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{N^2} \sum_{j=1}^{N^2} A_{ij} \tag{42}$$

$$\delta_i = \sqrt[2]{\frac{1}{N^2 - 1} \sum_{j=1}^{N^2} \left(x_j - \mu_i\right)} \tag{43}$$

$$\delta'_i = \frac{1}{N^2} \sum_{i=1}^{N^2} \sqrt[2]{\frac{1}{N^2 - 1} \sum_{j=1}^{N^2} A_{ji}} \tag{44}$$

12.1.2 Basis selection

Recognition Task: Unknow probeface is normalized (ϕ) and projected on to the subspace to get weight for the probe image $w_i = u_j^T \Phi$ Euclidean Distance measure is used in classification given by $e_r = \min \|\Omega - \Omega_i\|$. And if $e_r < \Theta$ where Θ is a threshold chosen heuristically, then we the probe image is recognized as the image with which it gives the lowest score. If however $e_r > \Theta$ then the probe does not belong to the database.

Deciding on the Threshold: A set of 150 known images other than the ones in the data set is used in the computation of threshold θ given by $\theta = \mu + \eta\sigma$. Where,

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \tag{45}$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} \left(x_j - \mu\right)^2 \tag{46}$$

$\eta \in I$ is chosen according to level of precision required in the results. x_i Є γ

The method of choosing right combination of right proportion of feature vectors has been applied on a large database consisting of a variety of still images with illumination, expression variations as well as partially occluded images. The ratio 3:2:5:: SVD:DF-LDA:LPP has yielded highest accuracy in recognition. The example is tried on a total test set of 165 images drawn from YALE dataset and the training set consisting 15 classes having a class count of five images.

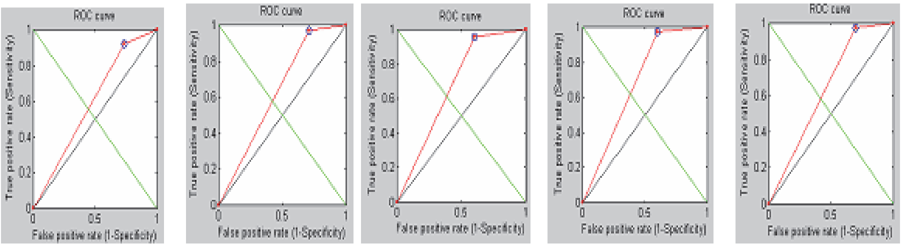
An ROC graph is plotted to visualize and analyze the working of face recognition efficiency. It is a two dimensional graph in which TP rate, true positive rate, is plotted on the Y axis and FP rate, false positive rate, is plotted on the X axis. Given a set of test images a two by two contingency table is constructed representing the dispositions of the set of images.

SVD (no. of vectors)	DFLDA (no. of vectors)	LPP (no. of vectors)	EFFICIENCY (in %)
15	5	5	80.00
5	5	15	81.21
8	9	8	81.81
15	5	15	87.27
5	15	5	81.21

Table 1. Iterations for subspace of dimension M/3

Graph No.	True Positive	False Negative	False Positive	True Negative
1	122	28	5	10
2	123	27	4	11
3	125	15	5	10
4	132	18	3	12
5	122	28	3	12

Table 2. Comparative results with Iteration Trial of M/3



3:1:1 1:1:3 8:9:8 3:1:3 1:3:1

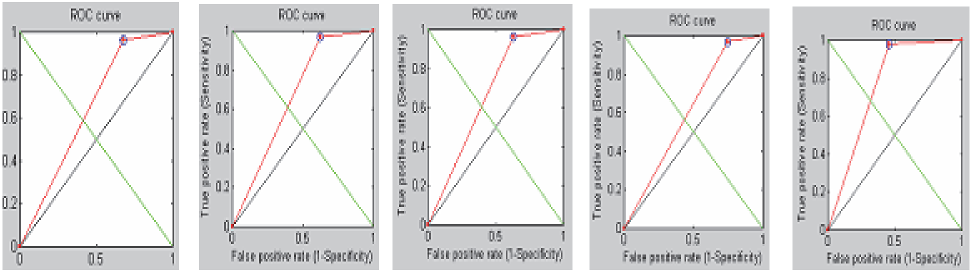
Fig. 14. ROC's Indicating the True Positive VS False Positive for M/3

SVD (no. of vectors)	DFLDA (no. of vectors)	LPP (no. of vectors)	EFFICIENCY (in %)
30	10	10	84.24
10	10	30	85.45
20	15	15	86.67
25	15	10	84.84
15	10	25	92.12

Table 3. Iterations for subspace of dimension 2M/3

Graph No.	True Positive	False Negative	False Positive	True Negative
1	129	21	5	10
2	132	18	4	11
3	133	17	5	10
4	128	32	4	11
5	140	10	3	12

Table 4. Comparative results with Iterartion Trial of M/3



3:1:1 1:1:3 4:3:3 5:3:2 3:2:5

Fig. 15. ROC's Indicating the True Positive VS False Positive for 2M/3

13. Conclusion

In this chapter several linear and non linear dimensionality reduction techniques were discussed from the perspective of face recognition. Since the face images contain several characteristic features both global and local, using any one method alone may not yield better recognition accuracy. It may be good to have combinations of the basis vectors from several approaches to achieve higher accuracy. Underlying manifold structure in image space will get face subspace and is possible with LPP, ICA methods. More pronounced features can be drawn from the space in case of LDA based algorithms. Random and PCA projections give appearance models which are holistic in nature.

Future of face recognition can also look at increase in dimension like depth information for recognition purposes. Algorithmic models should aim at addressing scale invariance feature vectors which can hopefully solve recognition task even under extreme variations in images.

The approach to face recognition was motivated by information theory, leading to the idea of basing face recognition on a small set of image features that best approximate the set of known face images, without requiring that they correspond to our intuitive notions of facial parts and features. The approach does provide a practical solution to the problem of face recognition and is relatively simple and has been shown that it can work well in a constrained environment. Anecdotal experimentation with acquired image sets indicates that profile size, complexion, ambient lighting and facial angle play significant parts in the recognition of a particular image.

14. References

R.Hietmeyer.Biometric identification promises fast and secure processing of airline passengers.The International Civil Aviation Organization Journal ,55(9):10 – 11,2000.

Michael E. Wall, Andreas Rechtsteiner, Luis M. Rocha, "Singular Value Decomposition And Principal Component Analysis" , A Practical Approach to Microarray Data Analysis , Chapter 5, pp. 91-109, 2003

M.A. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, pp.71-86, 1991.

Xiaofei He, Partha Niyogi, " Locality Preserving Projections", IN the proceedings of Advances in Neural Information Processing Systems, 2003.

Juwei Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face Recognition Using LDA Based Algorithms", IEEE Transactions on Neural Networks, VOL. 14, NO. 1,pp. 195-200, 2003.

Yu, H., Yang, J.,"A Direct LDA #algorithm for High-Dimensional Data with Application to Face Recognition. Pattern Recognition, Vol.34, pp.2067–2070., 2001

Sam Roweis, and Lawrence K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," In the proceedings of Science Journal, vol 290, pp. 2323—2326, 2000.

Joshua B. Tenenbaum,* Vin de Silva,John C. Langford ,"A Global Geometric Framework for Nonlinear Dimensionality Reduction" , In the proceedings of Science Journal, vol 290, pp. 2319-2323, 2000.

Xiaofei He, Shuicheng Yan, Yuxiao Hu, Partha Niyogi, and Hong-Jiang Zhang, "Face Recognition Using Laplacianfaces", In the proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 3, pp. 328- 340, 2005

Berry, M. W., Dumais, S. T., and O'Brian, G. W.(1995). Using Linear Algebra for Intelligent Information Retrieval., In the proceedings of SIAM Review Journal, 37(4).pp. 573—595, 1995.

J. Liu and S. Chen. Discriminant common vecotors versus neighbourhood components analysis and laplacianfaces: A comparative study in small sample size problem. Image and Vision Computing, 24(3):249–262, 2006.

Navin Goel*a*, George Bebis*a*, Ara Nefian*b* "Face Recognition Experiments with Random Projection" ,

Muhammad Imran Razzak, Muhammad Khurram Khan, Khaled Alghathbar, Rubiyah Yousaf, "Face Recognition using Layered Linear Discriminant Analysis and Small Subspace", In the 10th IEEE International Conference on Computer and Information Technology (CIT 2010), pp. 1407-1412, 2010.

MaxWelling, "Fisher Linear Discriminant Analysis", Classnotes in Machine Learning.

Kari Torkkola, "Discriminative Features for Text Document Classification", In the proceedings of International Conference on Pattern Recognition, pp. 472-475, 2002.

Shermina.J, “Application of Locality Preserving Projections in Face Recognition”, In the proceedings of International Journal of Advanced Computer Science and Applications, Vol. 1, No. 3, pp. 82 -85 September 2010

Deng Cai, Xiaofei He, Jiawei Han, Hong-Jiang Zhang,” Orthogonal Laplacianfaces for Face Recognition “, In the proceedings of IEEE Transactions on Image Processing,pp.1-7 2010

S.Sakthivel, R.Lakshmipathi, “Enhancing Face Recognition Using Improved Dimensionality Reduction And Feature Extraction Algorithms –An Evaluation With ORL Database”, International Journal of Engineering Science and Technology, Vol. 2(6), pp.2288-2295, 2010.

Ruba Soundar Kathavarayan,, Murugesan Karuppasamy, “Preserving Global and Local Features for Robust Face Recognition under Various Noisy Environments”, In the proceedings of International Journal of Image Processing (IJIP) Volume(3), Issue(6), pp. 328-340, 2010.

Neeta Nain, Prashant Gour, Nitish Agarwal, Rakesh P Talawar, Subhash Chandra, “Face Recognition using PCA and LDA with Singular Value Decomposition(SVD) using 2DLDA”, Proceedings of the World Congress on Engineering Vol I, ISBN:978-988-98671-9-5, pp. 1-4, 2008.

Tat-Jun Chin, Konrad Schindler, David Suter, “Incremental Kernel SVD for Face Recognition with Image Sets”, In the proceedings of Seventh IEEE International Conference on Automatic Face and Gesture Recognition pp.461-466, 2006.

Marian Stewart Bartlett, , Javier R. Movellan, Terrence J. Sejnowski, “Face Recognition by Independent Component Analysis”, IEEE Transactions on Neural Networks, Vol. 13, NO. 6, , pp. 1450-1464, 2002.

K.J. Karande, S.N. Talbar, ‘Independent Component Analysis of Edge Information for Face Recognition’, In the Proceedings of International Journal of Image Processing vol.3, issue 3, 120-130, 2009.

M. Belkin, P. Niyogi, “Using Manifold Structure for Partially Labeled Classification”, In the Proceedings of Conference on Advances in Neural Information Processing System, 2002.

W. Zhao, R. Chellappa, P.J. Phillips, ‘Subspace Linear Discriminant Analysis for Face Recognition’, Technical Report CAR-TR-914, Center for Automation Research, University of Maryland, 1999.

P. C. Yuen and J. H. Lai, “Independent Component Analysis of Face Images,”, In the proceedings of IEEE Workshop Biologically Motivated Computer Vision, Seoul, Korea, 2000.

A. M. Martinez and A. C. Kak, “PCA versus LDA,” In the proceedings of IEEE Transaction on Pattern Analysis and Machine Intelligence,, vol. 23, no. 2, pp. 228–233, 2001.

T. Shakunaga and K. Shigenari, “Decomposed Eigenface for Face Recognition under Various Lighting Conditions,” In the proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001.

Zhonglong Zheng, Fan Yang, Wenan Tan, Jiong Jia and Jie Yang, “Gabor feature-based face recognition using supervised locality preserving projection”, In the proceedings of *Signal Processing,* Vol. 87, Issue 10, pp. 2473-2483, 2007.

Books

Santosh S. Vempala ,"The Random Projection Method",DIMACS 5th Edition

Rafael C. Gonzalez, Richard E. Woods,"Digital Image Processing" 2nd Edition

8

Face and Automatic Target Recognition Based on Super-Resolved Discriminant Subspace

Widhyakorn Asdornwised
Department of Electrical Engineering, Chulalongkorn University, Bangkok
Thailand

1. Introduction

Recently, super-resolution reconstruction (SRR) method of low-dimensional face subspaces has been proposed for face recognition. This face subspace, also known as *eigenface,* is extracted using *principal component analysis* (PCA). One of the disadvantages of the reconstructed features obtained from the super-resolution face subspace is that no class information is included. To remedy the mentioned problem, at first, this chapter will be discussed about two novel methods for super-resolution reconstruction of discriminative features, i.e., *class-specific* and *discriminant analysis of principal components*; that aims on improving the discriminant power of the recognition systems. Next, we discuss about *two-dimensional principal component analysis* (2DPCA), also refered to as *image PCA*. We suggest new reconstruction algorithm based on the replacement of PCA with 2DPCA in extracting super-resolution subspace for face and automatic target recognition. Our experimental results on Yale and ORL face databases are very encouraging. Furthermore, the performance of our proposed approach on the MSTAR database is also tested.

In general, the fidelity of data, feature extraction, discriminant analysis, and classification rule are four basic elements in face and target recognition systems. One of the efficacies of recognition systems could be improved by enhancing the fidelity of the noisy, blurred, and undersampled images that are captured by the surveillance imagers. Regarding to the fidelity of data, when the resolution of the captured image is too small, the quality of the detail information becomes too limited, leading to severely poor decisions in most of the existing recognition systems. Having used super-resolution reconstruction algorithms (Park et al., 2003), it is fortunately to learn that a high-resolution (HR) image can be reconstructed from an undersampled image sequence obtained from the original scene with pixel displacements among images. This HR image is then used to input to the recognition system in order to improve the recognition performance. In fact, super-resolution can be considered as the numerical and regularization study of the ill-conditioned large scale problem given to describe the relationship between low-resolution (LR) and HR pixels (Nguyen et al., 2001).

On the one hand, feature extraction aims at reducing the dimensionality of face or target image so that the extracted feature is as representative as possible. On the other hand, super-resolution aims at visually increasing the dimensionality of face or target image. Having applied super-resolution methods at pixel domain (Lin et al., 2005; Wagner et al., 2004), the performance of face and target recognition applicably increases. However, with the emphases on improving computational complexity and robustness to registration error and noise, the continuing research direction of face recognition is now focusing on using eigenface super-resolution (Gunturk et al., 2003; Jia & Gong, 2005; Sezer et al., 2006).

The essential idea of eigen-domain based super-resolution using 2D eigenface instead of the conventional 1D eigenface is to overcome the three major problems in face recognition system, i.e., the curse of dimensionality, the prohibited computing processing of the singular value decomposition at visually improved high-quality image, and natural structure and correlation breaking in the original data.

In Section 2, the basic of super-resolution for low-dimensional framework is briefly explained. Then, discriminant approaches are detailed in Section 3 with the purpose of increasing the discrimination power of the eigen-domain based super-resolution. In Section 4, the implement of the two dimensional eigen-domain based super-resolution is addressed. We also discuss the possibility of the extension of two dimensional eigen-domain based super-resolution with discriminant information in Section 5. Finally, Section 6 provides the experimental results on the Yale and ORL face databases and MSTAR non-face database.

2. Eigenface-domain super-resolution

The fundamental of the super-resolution for in low-dimensional face subspace is formulated here. The important of the image super-resolution model and its eigenface-domain based reconstruction is that they can be used for practical extensions of one- and two-dimensional super-resolved discriminant face subspaces in the next sections, respectively.

2.1 Image super-resolution model

According to the numerically computational SRR framework (Nguyen et al., 2001), the relationship between an HR image and a set of LR images can be formulated in matrix form as follows:

$$\mathbf{f}_k = D_k B_k E_k \mathbf{x} + \mathbf{n}_k, \quad 1 \le k \le p \tag{1}$$

where p is the number of available frame, $\mathbf{f}_k$ and $\mathbf{x}$ are vectors extracted from the k^{th} LR image frame and HR image in lexicographical order, respectively, and D is the down-sampling operator, B is the blurring or averaging operator, and E_{k} is the affine transform, and $\mathbf{n}_{\text{k}}$ is noise of the frame k, respectively.

Thus, we can reformulate (1) as

$$\begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_p \end{bmatrix} = \begin{bmatrix} D_1 B_1 E_1 \\ \vdots \\ D_p B_p E_p \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{n}_1 \\ \vdots \\ \mathbf{n}_p \end{bmatrix} \tag{2}$$

or

$$\begin{gathered} \mathbf{f}_k = H_k \mathbf{x} + \mathbf{n}_k \\ \mathbf{f} = \mathbf{H}\mathbf{x} + \mathbf{n}. \end{gathered} \tag{3}$$

The above equation can be solved as an inverse problem with a regularization term, or

$$\mathbf{x} = \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{H}^{\mathrm{T}} + \lambda \mathbf{I})^{-1} \mathbf{f} \ . \tag{4}$$

It should be noted that the matrix **H** is a very large sparse matrix. As a result, the analytic solution of **x** is very hard to find. One of the popular methods used for finding the solution of this kind of the inverse problem is by using conjugate gradient method.

2.1 Reconstruction algorithm

Common preprocessing step used for pattern recognition and in compression schemes is dimensionality reduction of data. In image analysis, PCA is one of the popular methods used for dimensionality reduction. Let $\mathbf{\Phi}$ be an optimal eigenface that removes the redundancy by decorrelating the image data **x**. The optimal eigenfaces are coded in its columns. Face image **x** is assumed to be vectored. Thus, the optimal image representation of **x** can be written as

$$\mathbf{x} = \mathbf{\Phi a} + \mathbf{e}_x , \tag{5}$$

where **a** is the $L \times 1$ dimensional feature that represents **x**, and $\mathbf{e_x}$ is its representation error. Given that $\mathbf{\Psi}$ is the $\beta^2 N^2 \times L$ matrix that contains eigenfaces of the k^{th} LR image frame, where the scaling resolution factor β is within the range 0 to 1 and N is the total face image pixels. We can formulate the low-resolution image representation as

$$\mathbf{f}_k = \mathbf{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k} . \tag{6}$$

By substituting (5) and (6) in (3), we obtain

$$\mathbf{\Psi}\hat{\mathbf{a}}_k + \mathbf{e}_{f_k} = H_k \mathbf{\Phi a} + H_k \mathbf{e}_x + \mathbf{n}_k . \tag{7}$$

Since

$$\mathbf{\Psi}^T \mathbf{e}_{f_k} = 0 \tag{8}$$

and

$$\mathbf{\Psi}^T \mathbf{\Psi} = 1 . \tag{9}$$

It is easy to derive the following equation

$$\hat{\mathbf{a}}_k = \mathbf{\Psi}^T H_k \mathbf{\Phi a} + \mathbf{\Psi}^T H_k \mathbf{e}_x + \mathbf{\Psi}^T \mathbf{n}_k . \tag{10}$$

By considering the second and third terms as the observation noise with Gaussian distribution (Gunturk et al., 2003), we can obtain

$$\hat{\mathbf{a}}_k = \Lambda_k \mathbf{a} + \mathbf{\Psi}^T \zeta_k , \tag{11}$$

where

$$\zeta_k = H_k \mathbf{e}_x + \mathbf{n}_k , \tag{12}$$

and

$$\Lambda_k = \mathbf{\Psi}^T H_k \mathbf{\Phi} . \tag{13}$$

Without loss of generality, we can numerically solve for the true super-resolution feature vector at the eigen-domain level as in (5), or

$$\mathbf{a} = \mathbf{\Lambda}^{\mathrm{T}}(\mathbf{\Lambda}\mathbf{\Lambda}^{\mathrm{T}} + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}}\,, \tag{14}$$

where γ is the regularization term. In particular, we introduce the notation p in (14) in order to differentiate the PCA-domain based super-resolution approach (Gunturk et al., 2003) from our proposed approaches which will be presented in the upcoming sections,

$$\mathbf{a}_p = \mathbf{\Lambda}_p^T(\mathbf{\Lambda}_p\mathbf{\Lambda}_p^T + \gamma\mathbf{I})^{-1}\hat{\mathbf{a}}_p\,. \tag{15}$$

3. Discriminant face subspaces

PCA and its eigenface extension are constructed around the criteria of preserving the data distribution. Hence, it is well suited for face representation and reconstruction from the projected face feature. However, it is not an efficient classification method because the between classes relationship has been neglected. Here, we discuss on the possibilities that how we can embed discriminant information into eigenface-domain based super-resolution.

3.1 Face-specific subspace super-resolution

As widely known, the eigen-domain based face recognition methods use the subspace projections that do not consider class label information. The eigenface's criterion chooses the face subspace (coordinates) as the function of data distribution that yields the maximum covariance of all sample data. In fact, the coordinates that maximize the scatter of the data from all training samples might not be so adequate to discriminate classes. In recognition task, a projection is always preferred to include discrimination information between classes. One of the extensions of eigenface, called face-specific subspace (FSS) (Shan, 2003), is proposed as an alternative feature extraction method to include class information for face recognition application. According to FSS, each reduced dimensional basis of class-specific subspace (CSS) is learned from the training samples of the same class. Actually, each individual set of CSS optimally represents the data within its own class with negligible error. As a result, large representation error occurs, when the input data is projected and then reconstructed using a reduced set with less maximum covariance coordinates (or equivalently, using a set of principal components that does not belong to the input class). This way, by using reconstruction error obtained from projection-reconstruction process between classes, also called distance from CSS (DFCSS), a new metric can be suitably used as the distance for classifying the input data. In other words, the smaller the DFCSS is, the higher the probability that the input data belongs to the corresponding class will be. Similar work based on FSS (Belhumeur, 1997) attacking wide attentions in face recognition society is also published recently.

The original face-specific subspace (FSS) was proposed to manipulate the conventional eigenface in order to improve the recognition performance. According to FSS, the difference between FSS and the traditional method is that the covariance matrix of the p^{th} class is individually evaluated from training samples of the p^{th} class. Thus, the p^{th} FSS is represented as a 4-tuple, i.e., the projection matrix, the mean of the p^{th} class, the eigenvalues of covariance matrix, and the dimension of the p^{th} CSS. For identification, the input sample is projected using all CSSs and then reconstruct by those CSSs. If reconstruction error which obtained from the p^{th} CSS is minimum then the input sample is belong to the p^{th} class, also called distance from CSS (DFCSS).

There are many advantages of using CSS in face and target recognition. For example, the transformation matrices are trained from samples within their own classes, thus it is more optimum (using fewer components) to represent each sample in its own class than a transformation matrix trained by samples in all classes. Additionally, since DFCSS is the distance between the original image and its reconstruction image obtained from CSS, the memory space needed is only for storing the C transformation matrices, where C is the number of classes. This is far less than the conventional subspace methods, where we need to store both a single all-classes transformation matrix and also its prototypes (a large set of feature vectors calculated for all training samples). Moreover, the number of distance calculation in CSS is less than the number of distance calculation in conventional methods, since the number of classes is usually less than the number of training samples.

By combining super-resolution reconstruction approach with class-specific idea, a new method for face and automatic target recognition is proposed.

3.2 Discriminant analysis of principal components

The PCA's criterion chooses the subspace as the function of data probability distribution while *linear discriminant analysis* (LDA) chooses the subspace which yields maximal inter-class distance, and at the same time, keeping the intra-class distance small. In general, LDA extracts features which are better suitable for classification task. Both techniques intend to project the vector representing face image onto lower dimensional subspace, in which each 2D face image matrix must be first transformed into vector and then a collection of the transformed face vectors are concatenated into a matrix.

The PCA and LDA implementation causes three major problems in pattern recognition. First of all, the covariance matrix, which collects the feature vectors with high dimension, will lead to *curse of dimensionality*. It will further cause the very demanding computation both in terms of memory and time. Secondly, the spatial structure information could be lost when the column-stacking vectorization and image resize are applied. Finally, especially in face recognition task, the available number of training samples is relatively small compared to the feature dimension, so the covariance matrix which estimated by these features trends to be singular, which is addressed ased *singularity problem* or *small sample zize* (SSS) problem. Especially, as a supervised technique, LDA has a tendency to overfitting because of the SSS problems.

Various solutions have been proposed for solving the SSS problem. Among these LDA extensions, Fisherface and the discriminant analysis of principal components framework (Zhao, 1998) demonstrate a significant improvement when applying LDA over principal components subspace. Since both PCA and LDA can overcome the drawbacks of each other.

It has also been noted that LDA faces two certain drawbacks when directly applied to the original input space. First of all, some non-face information such as image background has been regarded by LDA as the discriminant information. This causes misclassification when the face of the same subject is presented on different background. Secondly, the within-class scatter matrix trends to be singular when SSS problem has occurred. Projecting the high dimensional input space into low dimensional subspace via PCA first can solve the shortcomings of the LDA problems. In other words, class information should be included to PCA by incorporating LDA.

3.2.1 Proposed reconstruction algorithm

Here, we can obtain a linear projection which maps the HR input image **x** first into the face subspace, and finally into the classification space **z**. Thus, we can modify the equation (5) to be

$$\mathbf{x} = \mathbf{W}_z \mathbf{\Phi} \mathbf{a}_{dp} + \mathbf{e}_x + \mathbf{e}_z \,, \tag{16}$$

where $\mathbf{W}_z$ is the optimal discrimination projection obtained from solving the generalized eigenvalue problem:

$$\mathbf{S_B W}_z = \boldsymbol{\lambda} \mathbf{S_w W}_z \,, \tag{17}$$

and $\mathbf{S_B}$, $\mathbf{S_W}$ are the *between-class* and *within-class scatter matrices*, respectively. Similarly, we can find the optimal discriminant project of the LR image frame $\mathbf{W}_z^{'}$, by little manipulating on (16)-(17) with corresponding LR images.

With little manipulations, we can reconstruct discriminant analysis of principal components based super-resolution as

$$\mathbf{a}_{dp} = \mathbf{\Lambda}_{dp}^{T} (\mathbf{\Lambda}_{dp} \mathbf{\Lambda}_{dp}^{T} + \gamma \mathbf{I})^{-1} \hat{\mathbf{a}}_{dp} \,, \tag{18}$$

where

$$\mathbf{\Lambda}_{lp} = \mathbf{W}_z^{'} \mathbf{\Psi}^{\mathbf{T}} \mathbf{H} \mathbf{W}_z \mathbf{\Phi} \,, \tag{19}$$

and

$$\boldsymbol{\zeta} = \mathbf{H e_x} + \mathbf{H e}_z + \mathbf{n}_k \,. \tag{20}$$

4. Two-dimensional eigen-domain based super-resolution

Recently, Yang (Yang et al., 2004) proposed an original technique called *two-dimensional principal component analysis* (2DPCA), in which the image covariance matrix is computed directly on image matrices so the spatial structure information can be preserved. One of the benefits of this method is that the dimension of the covariance matrix just equals to the width of the face image or the height in case of 2DPCA variant. This size is much smaller than the size of covariance matrix estimated in PCA. Therefore, the image covariance matrix can be better estimated with full rank in case of few training examples, like in face recognition.

We now consider linear projection of the form

$$\tilde{\mathbf{x}} = \mathbf{\Theta v} + \tilde{\mathbf{e}}_{\mathbf{x}} \,, \tag{21}$$

where $\tilde{\mathbf{x}}$ represents any face image in its original matrix form, $\{\theta_1, \cdots, \theta_d\}$, be the *d* largest eigenvectors that can be form to be $\mathbf{\Theta}$,and **v** is the projected HR feature of this image on $\mathbf{\Theta}$, called *principal component matrix*. The criterion used for obtaining the eigenvectors in (21) has been descriptively shown in Yang and Sanguangsat (Yang et al., 2004; Sanguangsat, 2006).

4.1 Alternative image super-resolution model

LR and HR images can be simply related as (Vijay, 2008)

$$\tilde{\mathbf{f}}_k = L_k \tilde{\mathbf{x}} R_k + \tilde{\mathbf{n}}_k, \quad 1 \le k \le p. \tag{22}$$

where p is the number of available frame; L_k , R_k are downsampling matrices, and $\tilde{\mathbf{f}}_k$, $\tilde{\mathbf{x}}$ are image matrices from the k^{th} LR image frame and HR image, respectively. It should be noted that two-dimensional Gaussian blur can be represented by using together the two separate L_k and R_k . An extension to downsampling and affine transform can also be easily conducted by placing the elements of the matrices properly (Gsmooth, n.d.). It should also be noted that both the input LR and HR image are represented in its original matrix form. We do not transform the LR and HR images to be vectors in lexicography order as in (1).

4.2 Proposed reconstruction algorithm

Thus,

$$\mathbf{\Gamma}\hat{\mathbf{v}}_k + \mathbf{e}_{\tilde{f}_k} = L_k \mathbf{\Theta} \mathbf{v} R_k + L_k \mathbf{e}_x R_k + \tilde{\mathbf{n}}_k, \tag{23}$$

where $\{\Gamma_1, \cdots, \Gamma_d\}$, be the d largest eigenvectors that can be form to be $\mathbf{\Gamma}$,and $\hat{\mathbf{v}}_k$ is the projected LR feature of the image on $\mathbf{\Gamma}$.
Without loss of generality,

$$\mathbf{\Gamma}^T \mathbf{e}_{f_k} = 0 \tag{24}$$

and

$$\mathbf{\Gamma}^T \mathbf{\Gamma} = 1. \tag{25}$$

It is easy to derive the following equation

$$\hat{\mathbf{v}}_k = \mathbf{\Gamma}^T L_k \mathbf{\Theta} \mathbf{v} R_k + \mathbf{\Gamma}^T L_k \mathbf{e}_x R_k + \mathbf{\Gamma}^T \tilde{\mathbf{n}}_k. \tag{26}$$

It should be noted that $\hat{\mathbf{v}}_k$ is a feature matrix, unlike $\hat{\mathbf{a}}_k$ which is a feature vector. Thus, it is a little more complicated to solve the inverse problem for super-resolution feature matrix $\mathbf{v}_k$. By applying vector operator as presented in Kumar and Schott (Kumar, 2008; Schott, 2005),
(26) can be rewritten as

$$\hat{\beta}_k = \Xi_k \beta + \eta_k, \tag{27}$$

where $\hat{\beta}_k = vec(\hat{\mathbf{v}}_k)$, $\beta_k = vec(\mathbf{v})$, $\Xi_k = R_k^T \otimes \mathbf{\Gamma}^T L_k \mathbf{\Theta}$ and $\eta_k = vec(\mathbf{\Gamma}^T L_k \mathbf{e}_x R_k + \mathbf{\Gamma}^T \tilde{\mathbf{n}}_k)$. Here $\otimes$ is Kronecker operator. This way, we can solve for the two-dimensional feature matrix at the eigen-domain level similarly to (15) and (18), or

$$\boldsymbol{\beta} = \boldsymbol{\Xi}^{\mathrm{T}} (\boldsymbol{\Xi}\boldsymbol{\Xi}^{\mathrm{T}} + \gamma \mathbf{I})^{-1} \hat{\boldsymbol{\beta}}, \tag{28}$$

where γ is the regularization term. Thus, after we convert $\boldsymbol{\beta}$ back to matrix, we will obtain the desired super-resolution feature matrix.

5. Extensions to two-dimensional linear discriminant analysis of principal component matrix

Similarly to PCA, 2DPCA is more suitable for face representation than face recognition. For better performance in recognition task, LDA is necessary. Unfortunately, the linear transformation of 2DPCA reduces only the size of rows. However, if we apply LDA directly to 2DPCA, the number of the rows still equals to the height of original image. As a result, we are still facing the singular problem in LDA. Thus, a modified LDA, called *two-dimensional linear discriminant analysis* (2DLDA), based on the 2DPCA concept is proposed to overcome the SSS problem. Applying 2DLDA to 2DPCA not only can solve the SSS problem and the curse of dimensionality dilemma but also allows us to work directly on the image matrix in all projections. This way, the spatial structure information is still maintained. Moreover, the SSS problem has been remedy since the size of all scatter matrices cannot be greater than the width of face image. Our research group (Sanguangsat, 2006) are the first group that focus on the extension of discriminant analysis of principal component of Section 3.1 by two-dimensional projection, called *two-dimensional linear discriminant of principal component matrix*

6. Experimental results

Having assumed that we can perfectly obtain the information regarding to frame to frame motion, hence we can use these information to form the proper super-resolution matrix equation in (5). In our experiment settings, evaluation images were shifted by a uniform random integer, blurred with 4×4 Gaussian point spreading function with standard deviation 1, and downsampled by a factor of four to produce 16 low-resolution images for each high-resolution image. Using 9 (preselected) out of 16 complete set of frames of each image, we can construct the super-resolution subspaces and also super-resolution images, respectively. Our super-resolution subspace approach is then compared with pixel-domain super-resolution approach using the class-specific subspace for face and automatic target recognition. Here, we conduct and show experiments according to the algorithm proposed in Subsection 3.1 only. Ongoing experiments on the other reconstruction algorithms, i.e., *discriminant analysis of principal components, two-dimensional eigenface-domain based super-resolution, and 2DLDA of 2DPCA*, are conducting. Essentially, we expect very encouraging the recognition results.

6.1 Evaluation databases

Eigenface-domain super-resolution method is used as the baseline for comparison based on the well-known Yale and AR face databases (Yale, 1997; Martinez, 1998) and MSTAR non-face database (Center, 1997), respectively.

6.1.1 Yale database

The Yale database contains 165 images of 15 subjects. There are 11images per subject, one for each of the following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink. All sample images of one person from the Yale database are shown in Fig. 1. Each image was manually cropped and resized to 100×80 pixels. In all experiments, the five image samples (centerlight, glasses, happy, leftlight, and noglasses) are used for training, and the six remaining images (normal, rightlight, sad, sleepy, surprise and wink) for test.

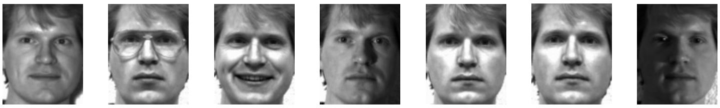


Fig. 1. The sample images of one subject in the Yale database

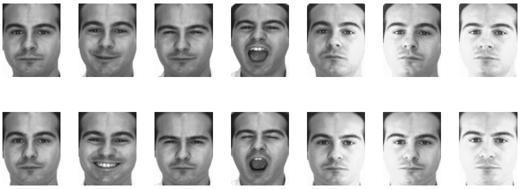


Fig. 2. The sample images of one subject in the AR database

6.1.2 AR database

The AR face database was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) at the U.A.B. It contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women). Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). The pictures were taken at the CVC under strictly controlled conditions. No restrictions on wear (clothes, glasses, etc.), make-up, hair style, etc. were imposed to participants. Each person participated in two sessions, separated by two weeks (14 days) time. The same pictures were taken in both sessions.

In our experiments, only 14 images without occlusions (sun glasses and scarf) are used for each subject, as shown in Fig. 2. All images were manually cropped and resized to 112×92 pixels, and then convert to 256 level gray scale images. The first five images per subject are used to train, and the remaining images to test.

6.1.3 MSTAR database

The MSTAR public release data set contains high resolution synthetic aperture radar data collected by the DARPA/Wright laboratory Moving and Stationary Target Acquisition and Recognition (MSTAR) program. The data set contains SAR images with size 128×128 of three difference types of military vehicles, i.e., BMP2 armored personal carriers (APCs), BTR70 APCs, and T72 tanks. The sample images from the MSTAR database are shown in Fig. 3. Because the MSTAR database is large, at this time, all images were centrally cropped to 32×32 pixels for evaluation purpose.

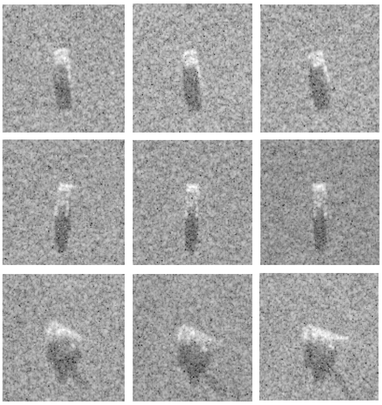


Fig. 3. Sample SAR images of MSTAR database: the upper row is BMP2 APCs, the middle row is BTR70 APCs, and the lower row is T72 tank.

Tables 1 and 2 detail the training and testing sets, where the depression angle means the look angle pointed at the target by the antenna beam at the side of the aircraft. Based on the different depression angles SAR images acquired at different times, the testing set can be used as a representative sample set of the SAR images of the targets for testing the recognition performance.

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	17°	233
	2	9566		231
	3	C21		233
BTR-70	1	C71	17°	233
T-72	1	132	17°	232
	2	812		231
	3	S7		228

Table 1. MSTAR images comprising training set

	Vehicle No.	Serial No.	Depression Angle	Images
BMP-2	1	9563	15°	195
	2	9566		196
	3	C21		196
BTR-70	1	C71	15°	196
T-72	1	132	15°	196
	2	812		195
	3	S7		191

Table 2. MSTAR images comprising testing set

6.2 Class-specific subspace results

The class-specific super-resolution images reconstructed for classification with pixel-domain and eigen-domain based approaches are shown in Fig. 4 and 5, respectively. The first images in the first column are the input testing images. The images from the second to the sixth columns are corresponding to the class-specific super-resolution reconstruction obtained from the corresponding five different set of class-specific eigenfaces. Here, we show five class-specific units. Thus, five reconstructed images are obtained from each input image. Image with least error at i^{th} class-specific unit will be identified to i^{th} class. It should be noted that the images reconstructed using pixel-domain based super-resolution approach give us good perceptual view. However, as shown for eigen-domain based approach, the fourth and fifth input images also give us good perceptual views, while others give comparable reconstruction results. Thus, the reconstruction images based on class-specific super-resolution subspace are more dependent to its corresponding eigen-vectors.

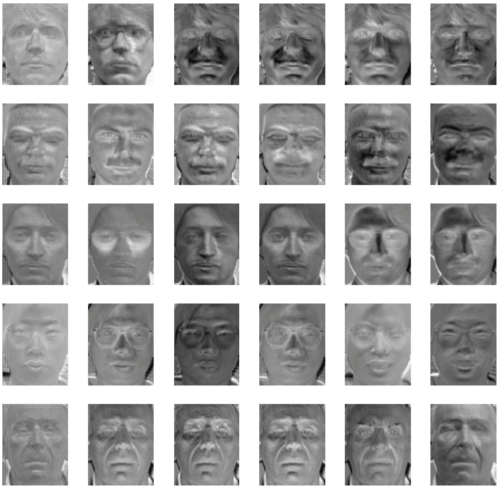


Fig. 4. Samples of class-specific pixel-domain based super-resolution reconstruction images

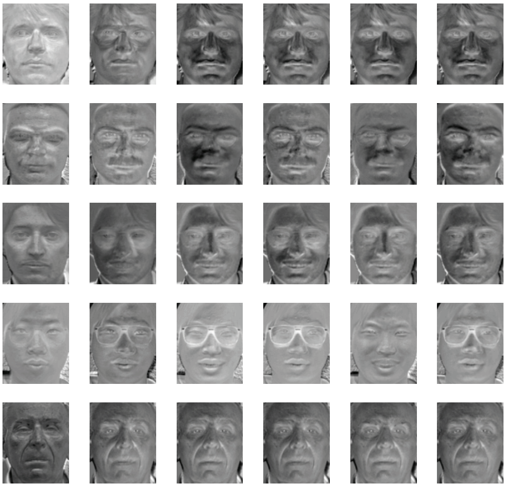


Fig. 5. Samples of class-specific eigen-domain based super-resolution reconstruction images

	BMP-2	BTR-70	T-72	Recognition Acc.
BMP-2	526	0	61	89.61
BTR-70	103	0	93	0
T-72	19	0	563	96.74
Average	-	-	-	79.78

Table 3. The pixel-domain super-resolution based class-specific subspace method: Recognition test of a three class problem for 32 x 32 images.

	BMP-2	BTR-70	T-72	Recognition Acc.
BMP-2	526	0	61	89.61
BTR-70	116	0	80	0
T-72	41	0	541	92.96
Average	-	-	-	78.17

Table 4. Our proposed method: Recognition test of a three class problem for 32 x 32 images.

Database	Pixel-Domain	Eigen-Domain
Yale	88.56	81.11
AR	88.00	87.50
MSTAR	79.78	78.17

Table 5. Comparison Results for 32 x 32 images.

Table 3 and 4 show the confusion matrices of the MSTAR target recognition. As shown in Table 5, the performance of the pixel-domain based super-resolution method is slightly better than our proposed method. However, our method is greatly benefits in term of computation. Additionally, we can derive principal component coefficients of the face databases using simple matrix inversion of very small size, which is 36×36 only. This is because of the reason we use inner product approach to calculate the PCA coefficients. Thus, our algorithm is far faster than implementing super-resolution at pixel-domain. In pixel-domain based super-resolution approach, they have to solve a very large and sparse matrix using conjugate gradient method. In the MSTAR database, we found that the class 2 target cannot be recognized at all. This may be because the size of the low-resolution test image is too small. If we increase the size of the test images to 48×48 or larger, we think that we can have better recognition accuracy.

7. Conclusion

In this chapter we have conducted experiments on face and automatic target recognition by focusing on the eigenface-domain based super-resolution implementations. We have also presented an extensive literature survey on the subject of more advanced and/or discriminant eigenface subspaces. From our discussion, several new super-resolution reconstruction algorithms have been proposed here.

In particular, several new eigenface-domain super-resolution algorithms are suggested as follows

1. Class-specific face subspace based super-resolution is proposed in Subsection 3.1
2. Equation (18) is used for including discriminant analysis of principal components for extracting face feature for eigenface-domain super-resolution
3. Equation (28) is used for two-dimensional eigenface-domain super-resolution
4. Two-dimensional eigenface in Equation (28) is proposed to be replaced by two-dimensional linear discriminant analysis of principal component matrix

Current research in face and automatic target recognition is yet to utilize the full potential of these techniques. During preparing this chapter, we have just realized that there many aspects of studies and comparisons that should be conducted to gain more understanding on the variants of the eigenface-domain based super-resolution. For example, recognition accuracy should be compared between majority-voting using multiple low-resolution eigenfaces VS one super-resolved eigenface. This way, we can relate a set of LR face recognition with multiple classifier system. Furthermore, all of the proposed algorithms use a two-stage approach, that is, dimensionality reduction is first implemented, after that the super-resolution enhancement is performed. It may be a little more encouraging if we can further conduct the study on *joint dimensionality reduction-resolution enhancement.* This idea is quite similar to *joint source-channel coding,* which is a very popular approach studied for transmitting data over network. Evidently, we are thinking about computing certain desired eigenfaces and then super-resolve the computed eigenfaces on the fry. This approach trends to be quite a more biological plausible.

8. Acknowledgments

The MSTAR data sets provided through the Center for Imaging Science, John Hopkin University, under the contact ARO DAAH049510494. This work was partially supported by the Thailand Research Fund (TRF) under grant number: MRG 5080427.

9. References

Park, S. C.; Park, M. K. & Kang, M. G. (2003). Super-Resolution Image Reconstruction: A Technical Overview. IEEE Signal Processing Magazine, Vol. 20, pp. 21-36

Nguyen, N.; Milanfar, P. & Golub, G. H (2001). A Computationally Efficient Image Super-Resolution Algorithm. IEEE Trans. Image Proc., Vol. 4, pp. 573-583

Lin, F.; Cook, J.; Chandran, V. & Sridharan, S. (2005). Face Recognition from Super-Resolved Images, ISSPA, Australia, August 2005

Wagner, R.; Waagen, D. & Cassabaum, M. (2004). Image Super-Resolution for Improved Automatic Target Recognition, SPIE Defense & Security Symposium, Orlando, USA, April 2004

Gunturk, B. K.; Batur, A. U.; Altunbasak, Y.; Hayes, M. H. & Mersereau, R. M. (2003). Eigenface-Domain Super-Resolution for Face Recognition. IEEE Trans. Pattern Anal. and Mach. Intell., Vol. 12, pp. 597-606

Jia, K. & Gong, S. (2005). Multi-Modal Tensor Face for Simultaneous Super-Resolution and Recognition, Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV), pp. 1683-1690, Washington, DC, USA, 2005

Sezer, O.G.; Altunbasak, Y. & Ercil, A. (2006). Face Recognition with Independent Component-based Super-Resolution, Proc. of SPIE: Visual Communications and Image Processing, Vol. 6077, 2006

Shan, S.; Gao, W. & Zhao, D. (2003). Face Recognition based on Face-Specific Subspace. International Journal of Imaging Systems and Technology, Vol.13, No.1, pp.23-32

Belhumeur, P. N.; Hespanha, J. P. & Kriegman, D. J. (1997). Eigenface vs. Fisherfaces: Recognition using Class Specific Linear Projection, IEEE Trans. Pattern Anal. Machine Intell, Vol.19 (May 1997), pp.711-720

Zhao, W.; Chellappa, R. & Krishnaswamy, A. (1998) Discriminant Analysis of Principal Components for Face Recognition, Proceedings of the 3rd IEEE International Conference on Face and Gesture Recognition (FG), pp. 336-341, Nara, Japan, 14-16 April 1998

Yang, J.; Zhang, D.; Frangi, A.F. & Yang, J. yu (2004) Two-dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Trans. Pattern Anal. and Mach. Intell., Vol.26, pp.131-137, Jan. 2004.

Sanguansat, P.; Asdornwised, W.; Jitapunkul, S. & Marukatat, S. (2006).Two-Dimensional Linear Discriminant Analysis of Principle Component Vectors for Face Recognition. IEICE Transaction on Information and System: Special Issue on Machine Vision Applications, Vol. E89- D, No. 7, pp. 2164-2170

Vijay Kumar, B. G. & Aravind, R. (2008). A 2D Model for Face Superresolution. Proceedings of the 17th International Conference on Pattern Recognition (ICPR), pp. 1–4, Tampa, Florida, USA, 2008

Gsmooth (n.d.). Gaussian Smoothing. Available from http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm

Schott, J. R. (2005), Matrix Analysis for Statistics, John Wiley & Sons, ISBN 0-471-66983-0, New Jersey, USA

Yale University (1997). The Yale Face Database. Available from http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

Martinez, A. & Benavente, R. (1998). The AR Face Database. Available from http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html.

The Center for Imaging Science (CIS) (1997). DARPA Moving and Stationary Target Recognition Program (MSTAR). Available from http://cis.jhu.edu/data.sets/MSTAR.

Efficiency of Recognition Methods for Single Sample per Person Based Face Recognition

Miloš Oravec, Jarmila Pavlovičová, Ján Mazanec,
Ľuboš Omelina, Matej Féder and Jozef Ban
Faculty of Electrical Engineering and Information Technology
Slovak University of Technology in Bratislava
Slovakia

1. Introduction

Even for the present-day computer technology, the biometric recognition of human face is a difficult task and continually evolving concept in the area of biometric recognition. The area of face recognition is well-described today in many papers and books, e.g. (Delac et al., 2008), (Li & Jain, 2005), (Oravec et al., 2010). The idea that two-dimensional still-image face recognition in controlled environment is already a solved task is generally accepted and several benchmarks evaluating recognition results were done in this area (e.g. Face Recognition Vendor Tests, FRVT 2000, 2002, 2006, http://www.frvt.org/). Nevertheless, many tasks have to be solved, such as recognition in unconstrained environment, recognition of non-frontal images, single sample per person problem, etc.

This chapter deals with single sample per person face recognition (also called one sample per person problem). This topic is related to small sample size problem in pattern recognition. Although there are also advantages of single sample – fast and easy creation of a face database and modest requirements for storage, face recognition methods usually fail to work if only one training sample per person is available.

In this chapter, we concentrate on the following items:

- Mapping the state-of-the-art of single sample face recognition approaches after year 2006 (the period till 2006 is covered by the detailed survey (Tan et al., 2006)).
- Generating new face patterns in order to enlarge the database containing single samples per subject only.

Such approaches can include modifications of original face samples using e.g. noise, mean filtering, suitable image transform (forward transform, then neglecting some coefficients and image reconstruction by inverse transform), or generating synthetic samples by ASM (active shape method) and AAM (active appearance method).

- Comparing recognition efficiency using single and multiple samples per subject.

We illustrate the influence of number of training samples per subject to recognition efficiency for selected methods. We use PCA (principal component analysis), MLP (multilayer perceptron), RBF (radial basis function) network, kernel methods and LBP (local binary patterns). We compare results using single and multiple training samples per person for images taken from FERET database. For our experiments, we selected large image set from FERET database.

- Highlighting other relevant important facts related to single sample recognition.

We analyze some relevant facts that can influence further development in this area. We also outline possible directions for further research.

2. Face recognition based on a single sample per person

2.1 General remarks

Generally, we can divide the face recognition methods into three groups (Tan et al., 2006): holistic methods, local methods and hybrid methods.

Holistic methods like PCA (eigenfaces), LDA (fisherfaces) or SVM need principally more image samples per person in the training phase. To solve the one sample problem there are basically two ways how to deal with it:

- To extend the classical methods to be trained from single sample more efficiently – e.g. 2D-PCA (Yang et al., 2004), (PC)2A (Wu & Zhou, 2002), E(PC)2A (Chen et al., 2004a), SPCA (Zhang, et al., 2005), APCA (Chen & Lovell, 2004), FLDA (Chen, et al., 2004b), Gabor+PCA+WSCM (Xie & Lam, 2006).
- To enlarge the training set by new representations or generating new views.

Local methods can be divided into 2 groups:

- Local feature based, which mostly work with some type of graph spread over the face regions with corners in important face features – face recognition is formulated as a problem of graph matching. These methods deal with the one sample problem better than the typical holistic methods (Tan et al., 2006). EBGM (Elastic Bunch Graph Matching) or DCP (directional corner points) are examples of this type of methods.
- Local appearance-based methods extract information from defined local regions. The features are extracted by known methods for texture classification (Gabor wavelets, LBP, etc.) and the feature space is reduced by known methods like PCA or LDA.

An excellent introduction to the single sample problem and survey of related methods mapping state-of-the-art till 2006 is described and discussed in (Tan et al., 2006).

2.2 State-of-the-art in single sample per person face recognition from 2006

After year 2006, new approaches were proposed. They are based mainly on enhancement of various conventional methods.

Principal Component Analysis (PCA) is still one of the most popular methods used to deal with one sample problem. Despite of its popularity, calculating of representative covariance matrix from one sample is very difficult task. In contrast to conventional application of PCA, 2DPCA (Yang et al. 2004) is based on two dimensional matrices, where the image does not need to be previously transformed into a 1D vector.

In (Que et al., 2008) a new face recognition algorithm MW(2D)2PCA was proposed. Modular Weighted (2D)2PCA (MW(2D)2PCA) is based on the study of (2D)2PCA. Weighting method (W) emphasizes the different influence of different eigenvectors and image blocking method (M) can extract detailed information of face image more effectively. Modularization of image into several blocks according to face elements provides more detailed information of face and assigns this approach rather to local appearance than holistic methods. The best recognition rate achieved by this method was 74.14%.

Similar approach, that deals with the single sample problem from human perception point of view, was proposed in (Zhan et al., 2009) where modularized image was processed by 2D

DCT to extract features, instead of (2D)2PCA. Gabor filters can be applied even to the image divided into several areas to reduce illumination impact as it is shown in (Nguyen & Bai, 2009).

Standard way to solve single sample problem is to use local facial representations. Conventional procedure in local methods is face image partitioning into several segments. In (Akbari et al., 2010), an algorithm based on single image per person, with input images segmented into 7 partitions was proposed. The moment feature vectors of a definite order for all images are extracted and distance measure is used to recognize the person.

Another way to get better results of recognition is a fusion of more biometrics. In (Ma et al., 2009) a new multi-modal biometrics fusion approach was presented. They used face and palmprint biometrics and combined the normalized Gaborface and Gaborpalm images at the pixel level. They presented a kernel PCA plus RBF classifier (KPRC) to classify the fused images. Using both face and palmprint samples, the average recognition results were improved from 42.60% and 52.36% (single-modal biometrics) to 87.01% (multi-modal biometrics).

In (Xie & Lam, 2006) novel Gabor-based kernel principal component analysis with doubly nonlinear mapping for human face recognition was proposed. The algorithm is evaluated using 4 databases: Yale, AR, ORL and YaleB database. The best of the proposed variations of the algorithm GW+DKPCA get very good results even under varying lighting, expression and perspective conditions.

(Kanan & Faez, 2010) presents a new approach for face representation and recognition based on Adaptively Weighted Sub-Gabor Array (AWSGA). The proposed algorithm utilizes a local Gabor array to represent faces partitioned into sub-patterns. It employs an adaptively weighting scheme to weight the Sub-Gabor features extracted from local areas based on the importance of the information they contain and their similarities to the corresponding local areas in the general face image. Experiments on AR and Yale databases show, that the proposed method significantly outperforms eigenfaces and modular eigenfaces in most of the benchmark scenarios under both ideal conditions and varying expressions and lighting conditions and this method achieves better results under partial occlusion conditions than the local probabilistic approach.

A novel feature extraction method named uniform pursuit (UP) was proposed in (Deng et al., 2010). A standardized procedure on the large-scale FERET and FRGC databases was applied to evaluate the one sample problem. Experimental results show that the robustness, accuracy and efficiency of the proposed UP method can compete successfully with the state-of-the-art one sample based methods.

In (Qiao et al., 2010), a new graph-based semi-supervised dimensionality reduction algorithm called sparsity preserving discriminant analysis (SPDA) based on SDA was developed. Experiments on AR, PIE and YaleB databases show that proposed method outperforms the SDA method.

Solution for single sample problem based on Fisherface method on generic dataset was presented in (Majumdar & Ward, 2008). The method was also extended to multiscale transform domains like wavelet, curvelet and contourlet. Results on Faces94 and the AT&T database show, that this approach outperforms SPCA and Eigenface Selection methods. Best results came from the Pseudo-fisherface method in the wavelet domain.

In (Gao et al., 2008), a method based on singular value decomposition (SVD) was used to evaluate the within-class scatter matrix so that the FLDA could be applied for face recognition with only one sample image in training set. The experiments on FERET, UMIST, ORL and Yale databases show, that the proposed method outperforms other state-of-the-art methods like E(PC)2A, SVD perturbation and different FLDA implementations.

A novel local appearance feature extraction method based on multi-resolution Dual Tree Complex Wavelet Transform (DT-CWT) was presented in (Priya & Rajesh, 2010). Experiments with ORL and Yale databases show, that this method and its block-based modification get very good results under illumination, perspective and expression variations conditions compared to PCA and global DT-CWT, while keeping low computational complexity.

In (Tan & Triggs, 2010) original LBP method used for face recognition was extended. More efficient preprocessing was proposed to eliminate illumination variances using LTP (local ternary patterns) – generalization and enhancement of the original LBP texture descriptor. By replacing the local histogram with a distance transform based similarity metrics the performance of the LBP/LTP face recognition was further improved. Experiments under difficult lighting conditions with Face Recognition Grand Challenge, Extended Yale-B, and CMU PIE databases provide results comparable to up to date methods.

Another extension of the LBP algorithm was presented in (Lei et al., 2008). The face image is first decomposed by multi-scale and multi-orientation Gabor filters. Local binary pattern analysis is then applied on the derived Gabor magnitude responses. Using FERET database with 1 image per person in the gallery, the method achieved results outperforming LBP, PCA and FLDA. To improve the recognition accuracy, it helps to add some synthetic samples of subject to the learning process. Standard procedures to create synthetic samples are the parallel deformation method (generate novel views of a single face image under different poses) (Tan et al., 2006), modification by noise or filtering original images. In (Xu & Yang, 2009) the feature extraction technique called Local Graph Embedding Discriminant Analysis(LGEDA) was proposed, where the imitated images were generated using a mean filter.

In (Su et al., 2010) an Adaptive Generic Learning (AGL) method was described. To better distinguish the persons with single face sample, a generic discriminant model was adopted. As a specific implementation of the AGL, a Coupled Linear Representation (CLR) algorithm was proposed to infer, based on the generic training set, the within-class scatter matrix and the class mean of each person given its single enrolled sample. Thus, the traditional Fisher's Linear Discriminant (FLD) can be applied to one sample problem task. Experiments are taken on images from FERET, XM2VTS, CAS-PEAL databases and a private passport database. The results show, that the Adaptive Gabor-FLD outperforms other methods like E(PC)2A, LBP and other FLD implementations. The proposed method is related to methods using virtual sample generation although it does not explicitly generate any virtual sample.

3. Face recognition methods

We use various methods in order to deeply explore the behavior of face recognition methods for single sample problem and to compare the methods using multiple face samples - both real-world samples and virtually generated samples. Used methods are briefly introduced in this subchapter.

3.1 Methods based on principal component analysis - PCA (PCA, 2D PCA and KPCA)

3.1.1 Principal component analysis - PCA

One of the most successful techniques used in face recognition is principal component analysis (PCA). The method based on PCA is named eigenface and was pioneered by Turk and Pentland (Turk & Pentland, 1991). In this method, each input image must be transformed into one dimensional image vector and set of these vectors forms input matrix. So the main idea behind PCA is that each *n*-dimensional face image can be represented as a linearly weighted sum of a set of orthonormal basis vectors.

This standard statistical method can be used for feature extraction. Principal component analysis reduces the dimension of input data by a linear projection that maximizes the scatter of all projected samples (Bishop, 1995).

For classification of projected samples Euclidean distance or other metrics can be used. Mahalanobis Cosine (MahCosine) is defined as the cosine of the angle between the image vectors that were projected into the PCA feature space and were further normalized by the variance estimates (Beveridge et al., 2003).

3.1.2 Two-dimensional PCA – 2D PCA

PCA is well-known feature extraction method mostly used as a baseline method for comparison purpose. Several extensions of PCA have been proposed. A major problem of using PCA lies in computation of covariance matrix what is computationally expensive. This computation can be significantly reduced by computing PCA features for columns (or rows) without previous matrix-to-vector conversion. This approach is also called two dimensional PCA (Yang et al., 2004). Main idea behind 2D PCA is the projection of image columns (rows) onto covariance matrix computed as the average of covariance matrices of each column for all training images. Let A be an m by n image matrix and average image $\overline{A}$ defined as $\overline{A} = 1/M \sum_k A_k$, where M is number of all *k* training images. Then covariance matrix can be calculated by

$$G = \frac{1}{M}\sum_{k=1}^{M}\sum_{i=1}^{m}(A_k^{(i)} - \overline{A}^{(i)})^T(A_k^{(i)} - \overline{A}^{(i)}) \tag{1}$$

Equation (1) reveals that the image covariance matrix can be obtained from the outer product of column (row) vectors of images, assuming the training images have zero mean.

For that reason, we claim that original 2D PCA works in the column direction of images. Result of feature extraction is then a matrix instead of a vector. Feature matrix has the same number of columns (rows) as width (height) of face image.

The extraction of image features is computationally more efficient using 2D PCA than PCA since the size of the image covariance matrix is quite small compared to the size of a covariance matrix in PCA (by using Turk & Pentlands optimization it depends on number of training images). 2D PCA is not only more efficient than PCA but it is possible to reach even higher recognition accuracy (Yang et al., 2004).

Despite its better efficiency, 2D PCA has also one disadvantage because it needs more coefficients for image representation than PCA. Because the size of the image covariance matrix for 2D PCA is equal to the width of images, which is quite small compared to the size of a covariance matrix in PCA, 2D PCA evaluates the image covariance matrix more accurately and computes the corresponding eigenvectors more efficiently than PCA.

3.1.3 Kernel PCA – KPCA

PCA is a linear algorithm that is not able to work with nonlinear data. Kernel PCA (Müller et al., 2001) is a method computing a nonlinear form of PCA. Instead of directly doing nonlinear PCA, it implicitly computes linear PCA in high-dimensional feature space that is in non-linear relation to input space.

3.2 Support vector machine - SVM

Support vector machines (SVM) (Asano, 2006; Hsu et al., 2003; Müller et al., 2001; Boser et al, 1992) are based on the concept of decision planes that define optimal boundaries. Its fundamental idea is very simple: the boundary is located to achieve the largest possible distance for the vectors of different sets. Example of this is shown in the Fig. 1. This figure illustrates linearly separable problem. In the case of linearly nonseparable problem, kernel methods are used. The concept of kernel method is a transformation of the vector space into a higher dimensional space.

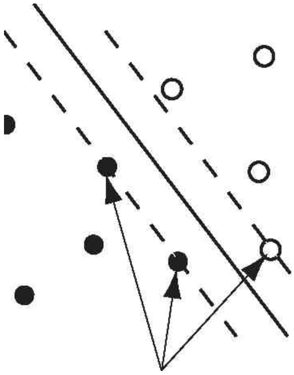


Fig. 1. Optimal boundary of support vector machine

The kernel function is defined as follows:

$$K(\mathbf{x},\mathbf{x}') = \Phi(\mathbf{x})^T \Phi(\mathbf{x}') \tag{2}$$

Kernel function is equivalent to the distance between $\mathbf{x}$ and $\mathbf{x}'$ measured in the higher dimensional space transformed by a nonlinear mapping Φ .

3.3 Methods based on neural networks (MLP, RBF network)

Neural network (Bishop, 1995; Haykin, 1994; Oravec et al., 1998) is a massive parallel processor that is inspired by biological nervous systems. Neural network is able to learn and to adapt its free parameters (connections between neurons known as synaptic weights are adjusted during the learning process).

3.3.1 Multilayer perceptron

Multilayer perceptron (MLP) (Bishop, 1995; Haykin, 1994; Oravec et al., 1998) is a layered feedforward network consisting of input, hidden and output layers.

Multilayer perceptron operates with functional and error signals. The functional signal propagates forward starting at the network input and ending at the network output as an output signal. The error signal originates at output neurons during the learning and propagates backward. MLP is trained by backpropagation algorithm.
MLP represents nested sigmoidal scheme (Haykin, 1994), its form for single output neuron is

$$F(\mathbf{x},\mathbf{w}) = \varphi\left(\sum_j w_{oj}\varphi\left(\sum_k w_{jk}\varphi\left(\ldots\varphi\left(\sum_i w_{li}x_i\right)\ldots\right)\right)\right) \quad (3)$$

where $\varphi(\cdot)$ is a sigmoidal activation function, w_{oj} is the synaptic weight from neuron j in the last hidden layer to the single output neuron o, and so on for the other synaptic weights, x_i is the i-th element of the input vector $\mathbf{x}$. The weight vector $\mathbf{w}$ denotes the entire set of synaptic weights ordered by layer, then neurons in a layer, and then number in a neuron.

3.3.2 Radial basis function network

Radial basis function network (RBF) (Oravec et al., 1998; Hlaváčková, 1993) is a feedforward network consisting of input, one hidden and output layer. Input layer distributes input vectors into the network, hidden layer represents RBFs h_i. Linear output neurons compute linear combinations of their inputs. RBF network topology is shown in Fig. 2.

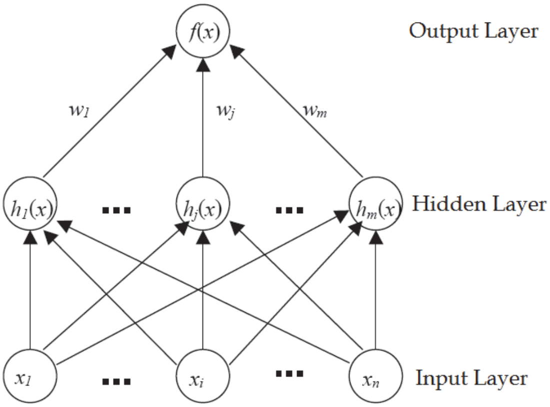


Fig. 2. RBF network topology

RBF network is trained in three steps:

1. Determination of centers of the hidden neurons
2. Computation of additional parameters of RBFs
3. Computation of output layer weights.

RBF network from Fig. 2 can be described as follows (Mark, 1996):

$$f(\mathbf{x}) = w_0 + \sum_{i=1}^{m} w_i h_i(\mathbf{x}) \quad (4)$$

where **x** is the input of RB activation function h_i and w_i are weights. Output of network is a linear combination of RBFs.

3.4 Local binary patterns – LBP

Local binary patterns (LBP) were first described in (Ojala et al., 1996). It is a computationally efficient descriptor to capture the micro-structural properties and was proposed for texture classification. The operator labels the pixels of an image by thresholding the 3x3-neighbourhood of each pixel with the center value and considering the result as a binary number. Later the LBP operator has been extended to use circle neighborhoods of different sizes - the pixel values are bilinearly interpolated (Fig. 3).

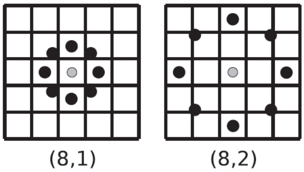


Fig. 3. The extended LBP operator with circular neighborhood

Another extension uses just uniform patterns. A local binary pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 00011110 and 10000011 are uniform patterns. Such patterns represent important features on the image like corners or edges. Uniform patterns account for most of the pattern in images (Ojala et al., 1996).

A system using LBP for face recognition is proposed in (Ahonen et al., 2004, 2006). Image is divided into non-overlapping regions. In each region a histogram of uniform LBP patterns is computed, the histograms are concatenated into one histogram (see Fig. 4 for illustration), which represents features extracted from the image in 3 levels (pixel, region and whole image).

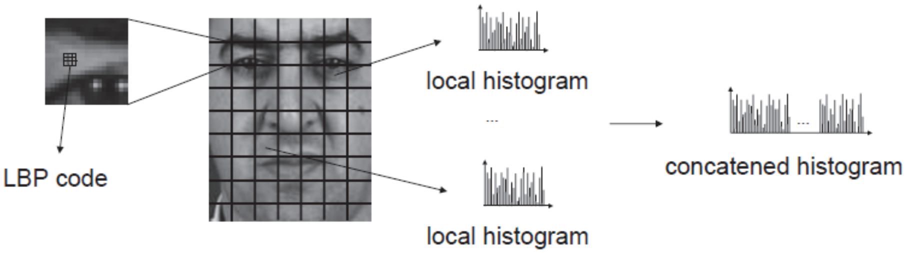


Fig. 4. Description of face using concatenated LBP histogram (image taken from (Marcel et al., 2007))

The χ2 metric is used as the distance metric for comparing the histograms:

$$\chi^2(\mathbf{S},\mathbf{M}) = \sum_{r,i} \frac{(S(i) - M(i))^2}{S(i) + M(i)} \tag{5}$$

where **S** and **M** are the histograms to be compared and *i* is the *i*-th bin of histogram.

4. Face database

We used images selected from FERET image database (Phillips et al., 1998). FERET face images database is de facto standard database in face recognition research. It is a complex and large database, which contains more than 14126 images of 1199 subjects of dimensions 256 x 384 pixels. Images differ in head position, lighting conditions, beard, glasses, hairstyle, expression and age of subjects.
We worked with grayscale images from Gray FERET (FERET Database, 2001). We selected image set containing total 665 images from 82 subjects. It consists of all available subjects from whole FERET database that have more than 4 frontal images containing also corresponding eyes coordinates (i.e. largest possible set fulfilling these conditions from FERET database was chosen). The used image sets are visualized in Fig. 5.

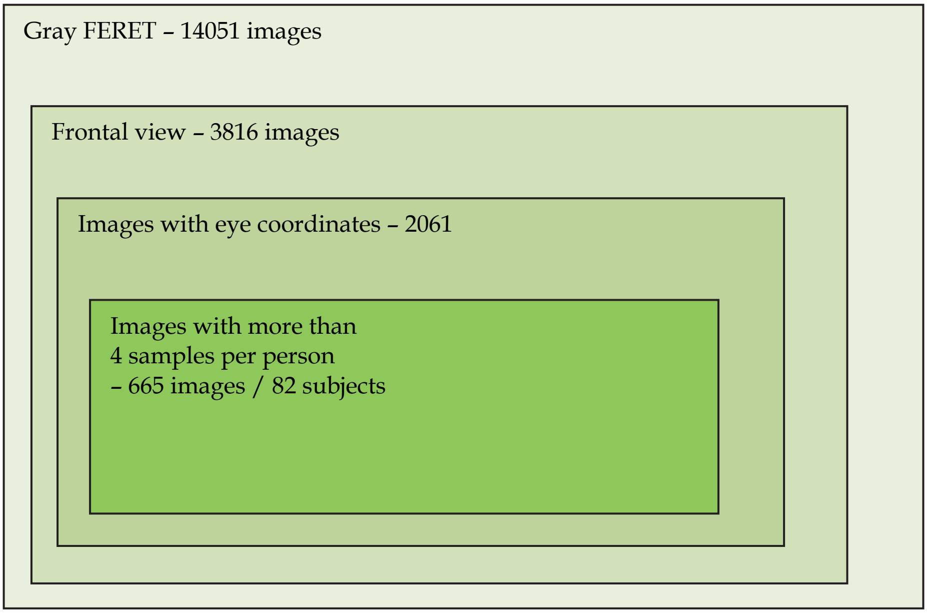

Gray FERET – 14051 images
Frontal view – 3816 images
Images with eye coordinates – 2061
Images with more than 4 samples per person – 665 images / 82 subjects

Fig. 5. Visualization of subset of images from FERET used in our experiments

The images were preprocessed. Our preprocessing consists of

- geometric normalization (aligning according to eye coordinates)
- histogram equalization
- masking (cropping an ellipse around the face)
- resizing to 65x75pix

Fig. 6 shows an example of the original image and the image after preprocessing.

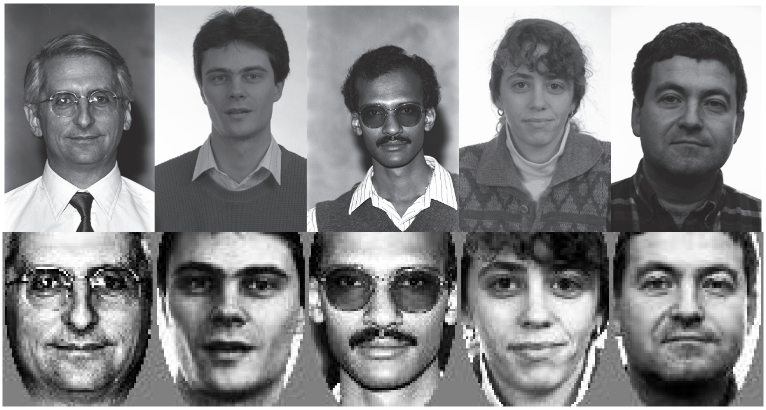


Fig. 6. Original images and corresponding images after preprocessing

5. Simulation results for 1 - 4 original training images per subject

In our experiments with original training images we compared the efficiency of several algorithms in scenario with 1 (single sample problem), 2, 3 and 4 images/subject. We carefully selected algorithms generally considered to play the major role in today face recognition research. Also standard PCA was included for comparison purposes. All these methods are briefly reviewed in subchapter 3. Face recognition methods.

	1_train	2_train	3_train	4_train
PCA	72.26	82.43	86.60	89.43
2D-PCA	73.41	81.90	86.36	89.02
KPCA	73.97	82.28	87.83	91.62
PCA+SVM	79.97	91.52	95.76	97.18
SVM	66.32	68.76	91.73	95.05
MLP	61.69	72.86	83.40	85.86
RBF	66.41	85.26	93.16	96.79
LBP-5x5	83.02	89.37	92.06	94.21
LBP 7x7	85.29	91.47	94.45	95.99
LBP 7x7w	85.81	92.91	95.05	96.59

Table 1. Results for different training sets (dependence of face recognition accuracy in % with regard to number of samples per subject in the training set)

In each test with different number of images in training set we made 4 runs with different selection of the images into the training set: original one with choosing the first images alphabetically by name and 3 additional training/testing collections with randomly shuffled images. The final test results are the average from these 4 values.

Our results are summarized in Table 1 and in Fig. 7. All figures and tables in this chapter contain values whose meaning is recognition accuracy in % achieved on test sets. The notation *n_train* means *n images (samples) per subject in training sets.*

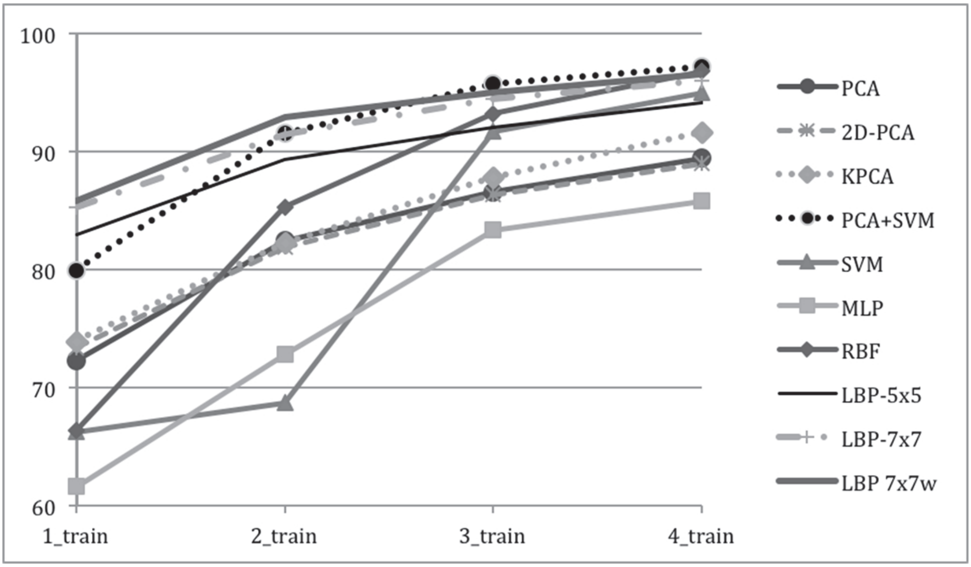


Fig. 7. Graphic comparison of the results for different training sets (dependence of face recognition accuracy in % with regard to number of samples per subject in the training set)

Presented results are summarized as follows:

Neural networks and SVM

For single sample per person training sets, methods based on neural networks (RBF network and MLP) and also SVM achieved less favorable results (below 70%). The extension of the training sets by second sample per person slightly increased face recognition test results for MLP and SVM methods. For RBF network, the second sample improved the result to the value above 85%. Impact of adding third sample per person into training sets caused a significant improvement of test results (for RBF and SVM above 90% accuracy was achieved). Adding more than four samples per person into training sets has only a minimal effect on increasing the face recognition results and has a negative impact on the computational and time complexity. The larger training sets the better recognition results were achieved.

PCA-based methods

PCA with Euclidean distance metric as a reference method shows that more images per subject in training set lead to more accurate recognition results, improving from 68% with 1 img./subj. to 89% with 4 img./subj. Although there was reported that 2D PCA can reach higher accuracy in term of precision, PCA slightly overcome 2D PCA in our experiments. However, 2D PCA still has big advantage in comparison to PCA which lies in faster training time due to using smaller covariance matrix. As it is shown in (Li-wei et al., 2005), 2D PCA is equal to block-based PCA and it means that it uses only several parts of covariance matrix used in PCA. In other words we lose information from rest of covariance matrix that can lead to worse recognition rates. KPCA achieved slightly better results compared to 2D PCA (KPCA is included for comparison purposes here and it will not be used further within this chapter).

PCA+SVM

PCA+SVM method is a two-stage setup including both feature extraction and classification. Features are first efficiently extracted by PCA with optimal truncating the vectors from the transform matrix. The parameters for the selection of the transformation vectors are based on our previous research (Oravec et al., 2010). The classification stage is performed by SVM. SVM model is created with the best parameters found using cross-validation on the training set. PCA+SVM has very good recognition rate even with 1 img./subj. and with 3 and 4 img./subj. it outperforms all other methods in our tests reaching 97% recognition rate with 4 img./subj.

LBP

In our experiments, we used local binary patterns method for face recognition in 3 different modifications. The image is divided into 5x5 or 7x7 blocks from which the concatenated histogram is computed. The "LBP 7x7w" modification adds also weighting of the histogram with different weights according to corresponding image regions. This weighting has been proposed in (Ahonen et al., 2004).

Results for all LBP methods are the best in our tests and were outperformed only slightly with PCA+SVM method with 3 and 4 img./subj. The main characteristic of LBP is that the recognition results are very good even for 1 img./subj.. From the graph in Fig. 7 we see that the recognition rates for the three LBP methods go parallel with each other. The LBP is starting with 83% reaching 94% accuracy with 4 img./subj. LBP 7x7 is approximately 1.5% better than the 5x5modification and the LBP 7x7w more than 2% better reaching almost 97% accuracy with 4 img./subj.

Within this chapter, we work with images of size 65x75 pixels after preprocessing. In Table 2, results for image size 130x150pix (FERET default standard) are shown for illustration. Generally, larger size of images can yield slightly better recognition rates.

Method	1_train
LBP-5x5	83.02
LBP 7x7	85.29
LBP 7x7w	85.81

Method	1_train
LBP 7x7 130x150pix	84,82
LBP 7x7w 130x150pix	86,66
LBP 10x10 130x150pix	87,48
LBP 10x10w 130x150pix	88,34

Table 2. Recognition rates for different LBP modifications

6. Simulation results for training sets enlarged by generating new samples

In the previous subchapter, we presented recognition results for methods trained by 1 img./subj. in training sets. We also presented the comparison to results for 2, 3 and 4 img./subj. in training sets, while 2nd, 3rd and 4th images were the *original* images, i.e. the images were *real*, taken from the *original* face database.

Herein we consider different situation: only 1 original sample is available and we try to enhance recognition accuracy by generating new samples to the training sets in artificial manner. Thus, we try to enlarge the training sets by generating new (virtual, artificial) samples. We propose to generate new samples by modifying single available original image in different ways – this is why we will use the term *image modification* (or *modified image*). Natural continuation of such approach leads to generating synthetic face images.

In our tests we use different modifications of available single per person images: adding noise, applying wavelet transform and performing geometric transformation.

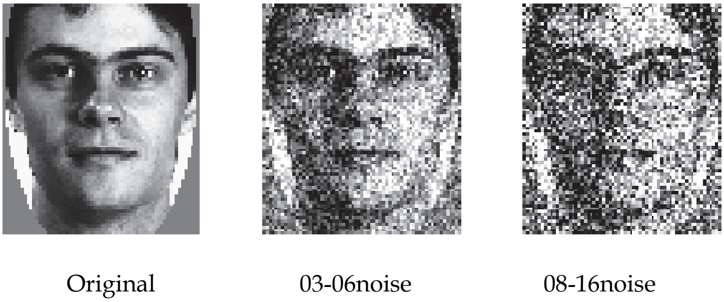
6.1 Modifications of face images by adding Gaussian noise

Noise in face images can seriously affect the performance of face recognition systems (Oravec et al., 2010). Each image capturing generates digital or analog noise of diverse intensity. The noise is also generated while transmitting and copying analog images. Noise generation is a natural property for image scanning systems. Herein we use noise for generating modified samples of original image. In our modifications, we use Gaussian (Truax, 1999) noise.

Gaussian noise was generated using Gaussian distribution function

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{6}$$

where μ is the mean value of the required distribution and σ^2 is a variance (Truax, 1999; Chiodo, 2006).



Original 03-06noise 08-16noise

Fig. 8. Examples of images modified by Gaussian noise

Gaussian noise was applied on each image with zero mean and in two random intervals of variance. Examples of images degraded by Gaussian noise can be seen in Fig. 8. The labels *03-06noise* and *08-16noise* mean that the variance of Gaussian noise is random between values 0.03 - 0.06 and 0.08 - 0.16, respectively. The same notation is used also in presented graphs and tables (Tab. 3a and 3b, Fig. 9a and 9b). Noise parameters settings for our simulations were determined empirically. Training sets were created by noise modification of samples added to the original one (*1+1noise*, *1+2noise* and *1+3noise*).

Presented results for noise modifications shown in Tab. 3a, 3b and Fig. 9a, 9b are summarized as follows:

Neural networks and SVM

The improvement for RBF, MLP and SVM is clearly visible. In both noise modifications (*03-06noise* and *08-16noise*), the most significant increase in accuracy of test results is achieved by RBF network (about 80% for 1+3 training sets). Similarly to the tests in subchapter 5, adding more samples into training sets has a constant effect on the recognition results.

	1_train	1+1_train 03-06noise	1+2_train 03-06noise	1+3_train 03-06noise
PCA	72.26	72.16	72.16	72.16
2D-PCA	73.41	73.18	73.24	73.24
PCA+SVM	79.97	78.73	79.03	78.43
SVM	66.32	67.44	74.76	74.84
MLP	61.69	65.99	68.39	71.48
RBF	66.41	79.19	79.73	79.87
LBP-5x5	83.02	83.02	83.02	83.02
LBP-7x7	85.29	85.29	85.29	85.29
LBP-7x7w	85.81	85.81	85.81	85.81

Table 3a. Results for generating new face samples (modifications of original face samples by Gaussian noise – lower variance)

	1_train	1+1_train 08-16noise	1+2_train 08-16noise	1+3_train 08-16noise
PCA	72.26	72.16	72.16	72.16
2D-PCA	73.41	73.24	73.13	73.07
PCA+SVM	79.97	78.90	77.83	77.79
SVM	66.32	74.84	74.87	74.81
MLP	61.69	64.83	65.57	65.63
RBF	66.41	77.1	79.3	80.07
LBP-5x5	83.02	83.02	83.02	83.02
LBP-7x7	85.29	85.29	85.29	85.29
LBP-7x7w	85.81	85.81	85.81	85.81

Table 3b. Results for generating new face samples (modifications of original face samples by Gaussian noise – higher variance)

PCA-based methods

The results of PCA and 2D PCA methods are only slightly affected when adding additional images with different amount of noise to the training set. The results with the noise images added are approximately 1% worse than the original recognition rate with 1 img./subj. Reason for this effect can be probably found in the fact that the transformation matrix computed from the training sample with added noise represents the variances in the space worse than after computing it from original images only. Adding samples to training set is also very uneconomical from the point of view of PCA methods since the time needed to compute the transform matrix grows.

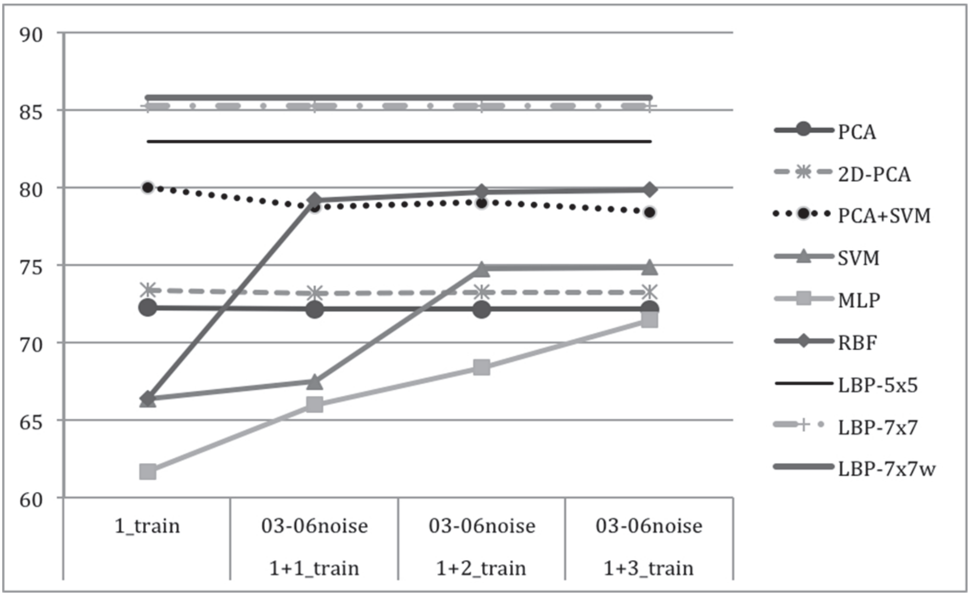


Fig. 9a. Graphic comparison of the results for generating new face samples (modifications of original face samples by Gaussian noise – lower variance)

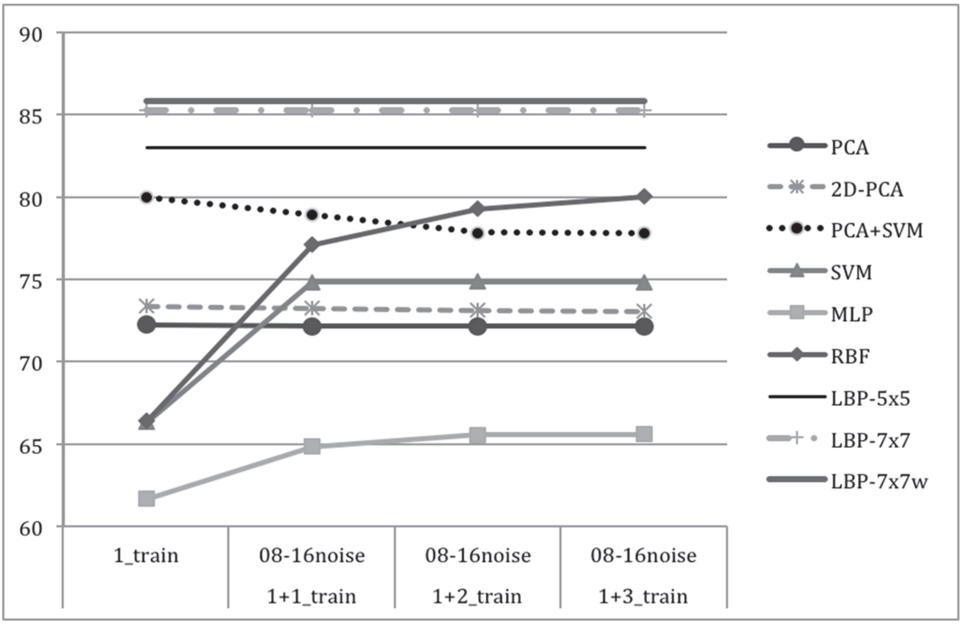


Fig. 9b. Graphic comparison of the results for generating new face samples (modifications of original face samples by Gaussian noise – higher variance)

PCA+SVM

The effect observed with PCA can be observed also with PCA+SVM method. Adding the noise images to the training set leads to worse results than with the original training set for about 1% for every scenario. The SVM classification model is influenced by the features extracted from noisy samples, but this accuracy drop is not dramatic.

LBP

The results of LBP methods are not influenced with the noisy samples at all. This has two reasons:

- By LBP method no model or transformation is calculated from the training images, so there cannot be such global effect to the recognition results as with PCA or SVM.
- The histograms of LBP patterns in noisy images change rapidly so the distance between the noisy image and the original image of the same person is higher than the distance between two original images of different persons. The consequence is that the minimal distances between the testing and training images do not change and the results are the same as without the noisy images in training set. See Table 4. for illustration of the distances between original and noisy images.

	Testimg1_subj1	Testimg1_subj2
Trainimg1_subj1	21.45	22.52
Trainimg1_subj1_noise	28.81	30.90

Table 4. Distances between the LBP 7x7 histograms for original and noisy train images compared with the same and different subject (see Fig. 10 for illustration of the images compared)

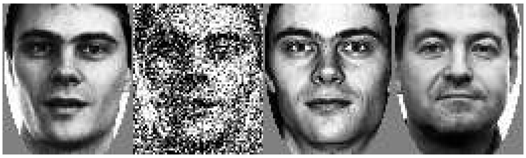


Fig. 10. Illustration of images used in comparison in Table 4

6.2 Modifications of face images based on wavelets

Discrete wavelet transform DWT (Puyati et al., 2006; Sluciak & Vargic 2008) (notation *wavelets* is used in our tables and charts) is defined as follows:

$$DWT(j,k) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(x)\psi\left(\frac{x}{2} - k\right)dx \qquad (7)$$

where *j* is the power of binary scaling, *k* is a constant of the filter and function ψ is a basic wavelet, *f(x)* is a function which is to be transformed.

Our modifications of face images were done by three steps:

1. Forward transform of image by DWT
2. Setting horizontal, diagonal and vertical details in frequency spectrum
3. Image reconstruction by inverse DWT

We used two types of wavelets: *Reverse biorthogonal 2.4* (Vargic & Procháska, 2005) and *Symlets 4* (Puyati et al., 2006) (Fig. 11.). These wavelets were chosen empirically – our aim was to produce slight change in the expression of a face. The training sets were created similarly to those with the noise modification (*1+1, 1+2* and *1+3*), see subchapter 6.1. An example of new samples is shown in Fig. 12.

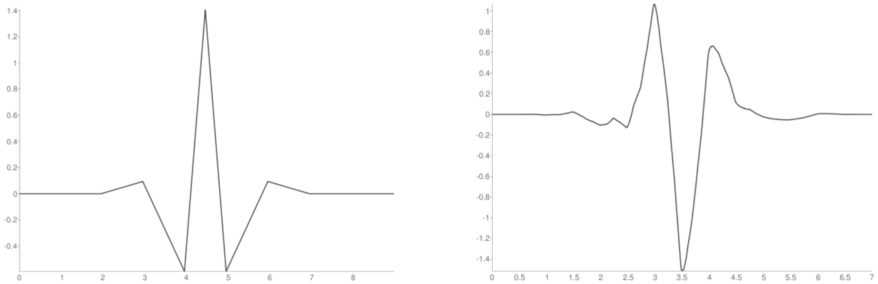


Fig. 11. Wavelet function ψ: Reverse biorthogonal 2.4, Symlets 4

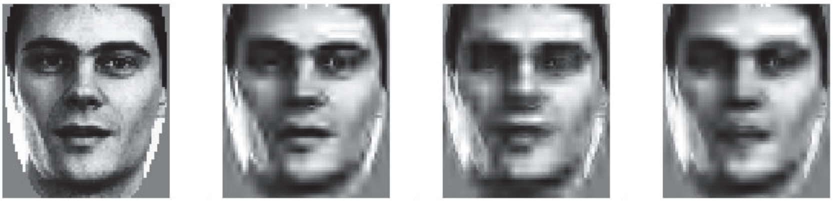


Fig. 12. Original image and three types of images modified by wavelet transform

		1+1_train	1+2_train	1+3_train
	1_train	wavelets	wavelets	wavelets
PCA	72.26	72.16	73.07	73.47
2D-PCA	73.41	73.30	73.18	73.36
PCA+SVM	79.97	78.43	65.27	50.47
SVM	66.32	28.04	24.71	23.61
MLP	61.69	68.92	73.65	74.73
RBF	66.41	77.49	77.70	79.12

Table 5. Results for generating new face samples (modifications of original face samples by wavelet transform)

Presented results for wavelet modifications (shown in Table 5 and in Fig. 13) are summarized as follows:

Neural networks and SVM

Experiment with wavelet transform demonstrated improvement of one sample per person face recognition using neural network methods - RBF network and MLP. These methods confirmed increase of recognition rate with extending the training sets with images modified by wavelet transform. Improvement above 10% was achieved for RBF network with adding three samples per person (*1+3_train*) into training sets. On the other hand, SVM method achieved very low face recognition accuracy.

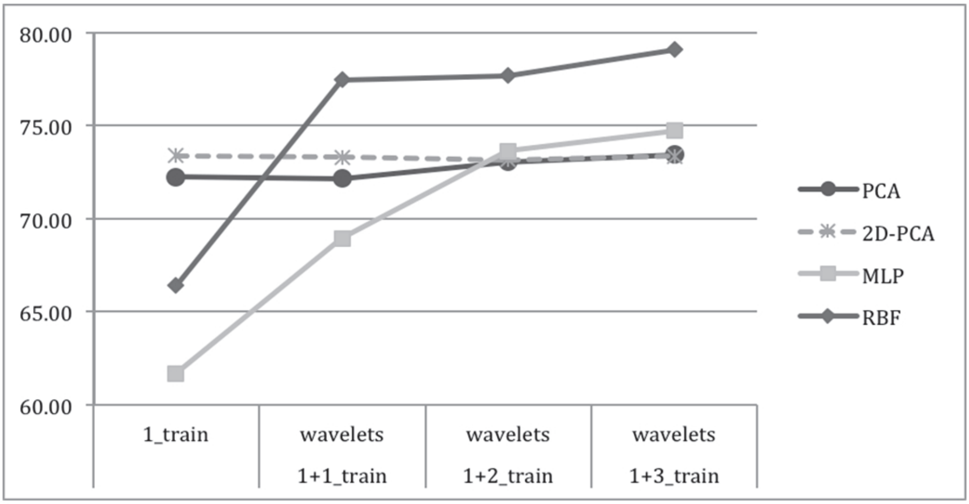


Fig. 13. Graphic comparison of the results for generating new face samples (modifications of original face samples by wavelet transform)

PCA-based methods

Experiments with extending the training set with images modified by wavelet transform show that there is only a small influence to the results of PCA and 2D PCA methods. The accuracy increases when adding the images with stronger wavelet modification. The accuracy of recognition results is only 1% higher than original 1img./subj. The modified images do not cause any significant change in recognition and so there is almost no gain by adding new sample.

PCA+SVM

In contrast to PCA, the effect of decreasing accuracy can be seen when also SVM is involved. When 3 images modified by wavelets are added to the training set, the recognition result is almost 30% worse than using the original image only. In this case, only 50% accuracy can be obtained.

LBP

The effect of the wavelet modifications to the LBP histogram is similar to that with the noise images, so the LBP results stay the same as with the original training set.

6.3 Modifications of face images based on geometry

One of the most successful approaches to samples generation is that based on geometric transformation. The idea is to learn some suitable manifolds and extend training set by new synthetic poses or expressions based on original image (Wen et al., 2003). Because generation of new samples is based on facial features and their position on the face, these features need to be localized at first.

After the all facial features are properly localized and represented by contour and middle points, the next step is to generate target expressions. Because the change of an expression involves moving detected feature points, there is a need to change texture information as well. Real expressions and direction of movements during the expression depends on strength of muscles contractions. We divided each face image into triangles according to direction of these contractions. Face features localization process and dividing into triangles (also called triangulation) is fully automated (unlike usual manual method described in (C.-kai Yang & Chiang, 2007)) using active shape models (Milborrow, 2008). Using active shape models produces very precise positions of facial features and facial boundaries. Result of triangulation is facial graph containing only triangles among detected points determining facial features.

Making use of rule based system, similar to system described in (Yang & Chiang, 2007), we generated different expressions from each training sample by moving location of points in the facial graph. Texture in each triangle containing moved points is then interpolated from original according new coordinates. This procedure with different rules creates new "smile" and "sad" expressions (Fig. 14) and represents more sophisticated approach to generating additional training samples.

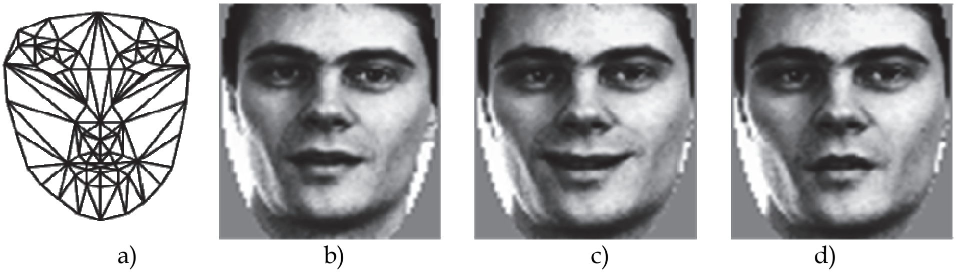


Fig. 14. Example of image modified by geometric transformation: a) example of triangular division of face, b) original face image, c) synthetic smile expression, d) synthetic sad expression

		1+1_train	1+1_train	1+2_train
	1_train	SMILE	SAD	SMILE+SAD
PCA	72.26	73.07	73.13	73.87
2D-PCA	73.41	73.58	73.24	73.87
PCA+SVM	79.97	80.19	80.32	75.09
SVM	66.32	48.26	47.63	28.82
MLP	61.69	71.62	69.60	75.61
RBF	66.41	76.50	75.69	72.96
LBP-5x5	83.02	82.98	83.10	83.23
LBP-7x7	85.29	85.33	85.33	85.51
LBP-7x7w	85.81	85.89	85.98	87.22

Table 6. Results for generating new face samples (modifications of original face samples by geometric transformation)

Simulation results for geometric modifications are summarized in Table 6 and Fig. 15. Only results for *SMILE* expression were included in the graph since it helps to improve recognition. It agrees with the fact that the face database contains more faces with smiles than sad faces. In this way it is also possible to present results consistent with other graphs – 1, 2 and 3 samples per face.
The results are summarized as follows:

Neural networks and SVM

Both RBF network and MLP achieved better recognition accuracy using *SMILE* face expression images (the increase compared with one sample per person about 10%). Tests with extending the training set by SMILE+SAD face expression images were most effective for MLP method (75.61%). For SVM method, these new samples caused the drop of recognition rate about 25%, similar to the wavelet transform.

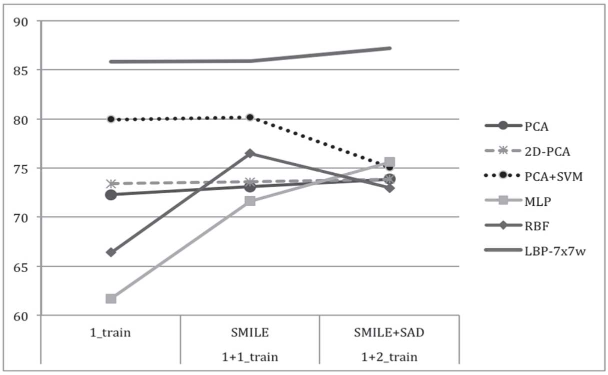


Fig. 15. Graphic comparison of the results for generating new face samples (modifications of original face samples by geometric transformation)

PCA-based methods

Geometric transformation results show comparable influence as those of PCA and 2D PCA using wavelet modifications. The accuracy increases when adding samples with *SMILE* expression. The accuracy of recognition results is only 1% higher than original 1 img./subj. The modified images do not cause any significant change in recognition. An improvement could be expected when more face expressions is taken in account.

PCA+SVM

Adding one image modified by geometry into the training set (either *SAD* or *SMILE* modification) improved the recognition rate for only about 0.2-0.3% (adding *SMILE* transformation helps slightly more). Surprisingly, when both transformed images were added to the training set, the recognition rate drops almost 5%.

LBP

As expected, adding transformed images with artificial change of expression (*SAD* and *SMILE* emotion) to the training set improves recognition. LBP method reaches better results because the system is more resistant against change in expression. Better results are reached when both transformed images (*SAD+SMILE*) are used. When also the images in the test set are transformed (for every sample also distances for *SAD* and *SMILE* transformation are computed), the results are even better, yielding 87.22% accuracy for LBP 7x7w method with 1 img./subj.

6.4 Comments and summary for methods that are influenced significantly by enlarging training sets by adding modified samples

This subchapter deals with methods for which extending the training set by modified images influences recognition results significantly (compared to recognition using multiple original images). The modifications of images described above (noise, wavelets and geometric tranformations) may be most helpful to neural networks. The comparison of recognition results for original training sets and extended training sets for RBF network and MLP is shown in Fig. 16. In Fig. 16 (and similarly in Fig. 17), the horizontal axis represents the number of images per person in training sets: the meaning for method using original images is 1, 2 ,3 and 4 original images in the training set; the meaning for modified images is 1 original image, 1 original plus 1, 2, or 3 modified images. For RBF network, above 10% improvement using modified images was achieved. For MLP, geometric transformation was the most successful modification of face images (75.61%).

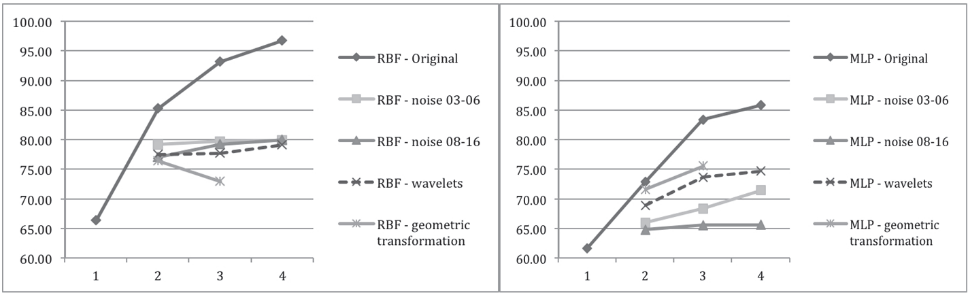


Fig. 16. Comparison of the results obtained using original and modified training images for RBF network and MLP (generated samples improve recognition)

Figure 17. shows the negative effects of adding newly generated samples into training sets. This effect is clearly visible for PCA+SVM and SVM, when training sets are extended by wavelet transform and geometric transformation.

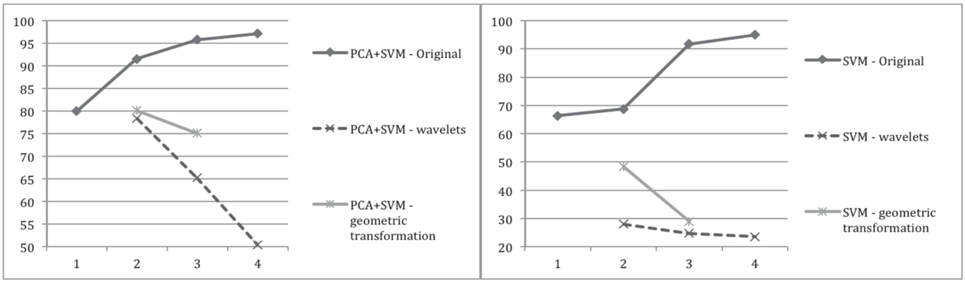


Fig. 17. Comparison of the results obtained using original and modified training images for PCA+SVM and SVM (generated samples degrade recognition)

7. Conclusion

In this chapter, we considered relevant issues related to one sample per person problem in the area of face recognition. We focused mainly on recognition efficiency of several methods working with single and multiple samples per subject. We researched techniques for enlargement of the training set by new (artificial, virtual or nearly synthetic) samples, in order to improve recognition accuracy. Such samples can be generated in many ways – we concentrated on modifications of the original samples by noise, wavelets and geometric transformation. We proposed methods for modifying expression of a subject by geometric transformation and by wavelet transform. We examined the impact of these extensions on various methods (PCA, 2D PCA, SVM, PCA+SVM, MLP, RBF and LBP variants).

Methods such as PCA+SVM or LBP achieved recognition results above 80% for single sample per person in the training set. For these methods, adding new samples (modified images) did not help significantly. On the other hand, the utilization of the extended training sets for neural networks (MLP and RBF network) always increased the face recognition rate. This confirms that an appropriate extension of the input data set enhances the learning process and the recognition accuracy. Adding more than three new samples per person into the training sets has almost no influence on the recognition rate and has a negative impact on the computational and time complexity. The SVM method improved recognition accuracy only for extension of the training set by noise modification of images.

Experimental results for PCA and 2D PCA show only negligible influence of adding modified samples. We can conclude that the use of modified samples for PCA and 2D PCA has no added value, especially when samples are modified by Gaussian noise only.

PCA+SVM (two-stage method with PCA for feature extraction and SVM for classification) achieved very good results even for 1 img./subj. Adding any modified images to the training set did not improve the recognition rates, but the results were still one of the best from the compared methods.

Our experiments show that LBP is one of the most efficient state-of-the-art methods in face recognition. Adding noise and wavelet modified images to the training set does not have any effect on the recognition rates of LBP – unlike other methods that use the training sample to compute models or transformation matrices. This is caused by the nature of the method, where the histogram of LBP patterns of the noisy image differs too much from the original images. This can be also a disadvantage, when the images in the test set are corrupted with noise. On the other hand, adding images with transformed face expression helps and the system is more resistant to expression change in the images.

LBP for face recognition has obvious advantages such as state-of-the-art recognition rates even with 1 img./subj. in the training set, no need to train models or transformation matrices and good computational efficiency. But there is still potential to improve the results by possible modifications and optimization, which can be researched further: selection of LBP patterns, different preprocessing or modifications of LBP operator. The geometric transformation of images (emotional expression or head pose) and generating synthetic samples seem to be good ways how to improve the results. Further research is needed, since a simple extension of the training set with modified images does not always help.

We are currently working on a more sophisticated geometric transformation to cover more facial expressions. Although the results in section 6.3 show only a small improvement (with the exception of MLP where the improvement was significant), we suppose there is great potential of using samples with synthetic expression. The triangular model of face enables to extend the generation algorithm by other possibilities like generation of samples with different poses and illumination conditions. In the future, we also plan to publish modules generating new samples (with different expressions, poses and illumination) for our universal biometric system BioSandbox[1] (used in our experiments).

Modification of images using wavelet transform has also large potential to generate new samples. One way to create new samples by wavelet transform is a fusion of two face images, where a new image is generated by applying the wavelet transform on two original images, followed by suitable manipulations of coefficients in a transformed space and finally merging images by inverse transform.

Using mean filter (Xu, J. & Yang, J., 2009) is another simple way of creating modified images. By using mean filter with different kernels (2x2, 3x3…15x15), we achieved results close to the modifications by wavelet transform.

Evaluating face recognition in single sample image per subject conditions reflects the real-world scenario. Also other effects such as various occlusions or lighting variation need to be taken into account when trying to reflect real conditions. We also need to test our methods using face databases that contain samples with these variations. Face databases such as ORL or AR could be used for this purpose.

For authentication and identification purposes, face recognition with 1 img./subj. only may not be enough, because its accuracy does not necessarily reach the required level. Therefore face recognition methods can be combined with different biometrics to form a multimodal system with much better characteristics than each of the biometrics itself (Ross & Jain, 2004).

8. Acknowledgment

Research described in this paper was done within the grants 1/0214/10 and 1/0961/11 of the Slovak Grant Agency VEGA. Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office. We would like to thank to our colleague Radoslav Vargic for valuable consultation regarding practical use of wavelets. We also thank to our student Ján Režnák for preparation of KPCA results.

9. References

Ahonen, T.; Hadid, A. & Pietikäinen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, TPAMI.2006.244

Ahonen, T.; Hadid, A. & Pietikäinen, M. (2004). Face Recognition with Local Binary Patterns, *Proceedings of Computer Vision – ECCV*, ISBN: 978-3-540-21981-1, Prague, Czech Republic, May, 2004

Akbari, R.; Bahaghighat, M. K. & Mohammadi, J. (2010). Legendre Moments for Face Identification Based on Single Image per Person, *2nd Int. Conference on Signal Processing Systems (ICSPS)*, ISBN: 978-1-4244-6892-8, Dalian, August 2010

Asano, A. (2006). Pattern Information Processing, (lecture 2006 Autumn semester), Hiroshima University, Japan, Available from <http://laskin.mis.hiroshima-u.ac.jp/Kougi/06a/ PIP/PIP12pr.pdf>

[1] Biosanbox project page – http://biosandbox.fei.stuba.sk

Beveridge, R.; Bolme, D.; Teixeira, M. & Draper, B. (2003). The CSU Face Identification Evaluation System *User's Guide, Version 5.0, Technical Report.*, Colorado State University, May 2003, Available from <http://www.cs.colostate.edu/evalfacerec/algorithms/version5/faceIdUsersGuide.pdf>

Bishop, C. M. (1995). Neural Networks for Pattern Recognition, Oxford University Press, Inc., ISBN 0 19 853864 2, New York

Boser, B.; Guyon, I. & Vapnik, V. (1992). A Training Algorithm for Optimal Margin Classifiers, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh, PA, USA, July 1992

Delac, K.; Grgic, M. & Bartlett, M., S. (Eds.) (2008). *Recent Advances in Face Recognition*, IN-TECH, Vienna, Retrieved from <http://intechweb.org/book.php?id=101>

Deng, W.; Hu, J.; Guo, J.; Cai, W. & Feng, D. (2010). Robust, Aaccurate and Efficient Face Recognition From a Single Training Image:A Uniform Pursuit Approach. *Pattern Recognition,* Vol. 43 Issue 5, May, 2010, pp. 1748–1762, ISSN:0031-3203

FERET Database (2001). Available from: <http://www.itl.nist.gov/iad/humanid/feret/, NIST>

Gao, Q.-X.; Zhang, L. & Zhang, D. (2008), Face Recognition Using FLDA With Single Training Image Per Person, *Applied Mathematics and Computation,* Volume 205, Issue 2, pp. 726–734, ISSN: 0096-3003

Haykin, S. (1994). *Neural Networks - A Comprehensive Foundation*, New York: Macmillan College Publishing Company, ISBN 0-02-352781-7

Hlaváčková, K. & Neruda, R. (1993). Radial Basis Function Networks. *Neural Network World*, Vol.3, No.1, pp. 93-102

Hsu, C.W.; Chang, C.C. & Lin, C.J. (2003). A Practical Guide to Support Vector Classification. Dept. of Computer Science and Inf. Engineering, National Taiwan University, April 15, 2010, Available from: <http://www.csie.ntu.edu.tw/~cjlin>

Chen, S. & Lovell, B. C. (2004). Illumination and Expression Invariant Face Recognition with One Sample Image, *17th Int. Conference on Pattern Recognition (ICPR' 04)* –Vol. 1, ISBN: 0-7695-2128-2, Cambridge UK, Aug. 2004

Chen, S. C., Zhang D.Q. & Zhou Z.-H. (2004), Enhanced (PC)2A For Face Recognition With One Training Image Per Person. *Pattern Recognition Letters*, Vol. 25, No.10, pp. 1173-1181, ISSN: 0167-8655

Chen, S.C.; Liu, J. & Zhou Z.-H. (2004). Making FLDA Applicable to Face Recognition with One Sample Per Person. *Pattern Recognition*, Vol. 37, (7), July 2004, pp. 1553-1555, ISSN:0031-3203

Chiodo, K. (2006). Normal Distribution. *NIST/SEMATECH e-Handbook of Statistical Methods*, Retrieved from <http://www.itl.nist.gov/div898/handbook/eda/section3/eda3661.htm>

Kanan, H. R. & Faez, K., (2010). Recognizing Faces Using Adaptively Weighted Sub-Gabor Array From a Single Sample Image Per Enrolled Subject. *Image and Vision Computing*, Vol. 28, No. 1, Jan. 2010, pp. 438–448, ISSN: 0262-8856

Lei, Z; Liao, S.; He, R.; Pietikäinen, M; & Li, S. Z. (2008). Gabor Volume Based Local Binary Pattern for Face Representation and Recognition, *8th IEEE Int. Conf. on Automatic Face & Gesture Recognition 2008 FG '08, Amsterdam,* ISBN: 978-1-4244-2153-4

Li, S. Z. & Jain, A. K. (Eds.) (2005). *Handbook of Face Recognition*, Springer, ISBN# 0-387-40595-x, New York

Li-Wei, W., Xiao, W., Ming, C., & Ju-Fu, F. (2005). Is Two-dimensional PCA a New Technique ? *Acta Automatica*, Vol.*31,*(5), pp. 782-787, ISSN: 0254-4156

Ma, W.J.; Li, S.; Yao, Y.F.; Lan, Ch.; Gao, S.Q.; Tang, H. & Jing, X.Y. (2009). Multi-modal Biometrics Pixel Level Fusion and KPCA-RBF Feature Classification for Single Sample Recognition Problem, *NSFC*, ISBN 978-1-4244-4131-0

Majumdar, A. & Ward, R. K. ,(2008). Pseudo-fisherface Method For Single Image Per Person Face Recognition, *ICASSP 2008*, Las Vegas, NV, ISBN: 978-1-4244-1483-3

Marcel, S. , Rodriguez, Y., & Heusch, G., (2007). On the Recent Use of Local Binary Patterns For Face Authentication. *International Journal on Image and Video Processing Special Issue on Facial Image Processing*, 2007. IDIAP-RR 06-34

Mark J. L. Orr, (1996). Introduction to Radial Basis Function Networks. Centre for Cognitive Science, University of Edinburgh, April 1996

Milborrow, S. (2008). Locating Facial Features With an Extended Active Shape Model. *Proc. of the 10th European Conf. on Computer Vision: Part IV,* Marseille, France, Springer-Verlag, Nov. 2010, Retrieved from from http://www.springerlink.com/index/5t8hjm7j02qx6184.pdf

Müller, K.R.; Mika, S.; Rätsch, G.; Tsuda, K. & Schölkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Trans. on Neural Networks*, Vol. 12, No. 2, March 200), pp. 181-201, ISSN: 1045-9227

Nguyen, H., & Bai, L. (2009). Local Gabor Binary Pattern Whitened PCA: A Novel Approach for Face Recognition from Single Image Per Person. *Advances in Biometrics*, Volume: 5558, Pages: 269-278, ISSN: 0302974, November 11, 2010, Retrieved from http://www.springerlink.com/index/t011q303142j7772.pdf

Ojala, T.; Pietikäinen, M. & Harwood, D. (1996). A Comparative Study of Texture Measures with Classification Based on Feature Distributions. *Pattern Recognition*, Vol. 29, pp. 51-59, ISSN:0031-3203

Oravec, M.; Mazanec, J.; Pavlovicova, J.; Eiben, P. & Lehocki, F. (2010). Face Recognition in Ideal and Noisy Conditions Using Support Vector Machines, PCA and LDA, In: *Face Recognition*, Milos Oravec (Ed.), ISBN: 978-953-307-060-5, INTECH, available from: http://sciyo.com/articles/show/title/face-recognition-in-ideal-and-noisy-conditions-using-support-vector-machines-pca-and-lda

Oravec, M.; Polec, J. & Marchevský, S. (1998). *Neural Networks for Digital Signal Processing* (in Slovak), Bratislava, Slovakia, ISBN 80-967503-9-9

Phillips, P.J. ; Wechsler, H.; Huang, J. & Rauss, P. (1998). The FERET Database and Evaluation Procedure For Face Recognition Algorithms, *Image and Vision Computing*, Vol. 16, No. 5, pp. 295-306, ISSN: 0262-8856

Priya, K. J & Rajesh, R.S., (2010). Dual Tree Complex Wavelet Transform Based Face Recognition with Single View, *The Int. Conference on Computing, Communications and Information Technology Applications (CCITA-2010)*, ISSN 1994-4608

Puyati, W.; Walairacht, S. & Walairacht, A. (2006). PCA in Wavelet Domain For Face Recognition, *The 8th International Conference Advanced Communication Technology*, pp. – 455, Phoenix Park, ISBN: 89-5519-129-4

Qiao, L.; Chen , S. and Tan, X., (2010). Sparsity *preserving discriminant analysis for single training image face recognition*. Pattern Recognition Letters Vol. 31, pp. 422–429, ISSN: 0167-8655

Que, D., Chen, B., Hu, Jin, & Ax, Y. (2008). A Novel Single Training Sample Face Recognition Algorithm Based on Modular Weighted $(2D)^2$ PCA, *9th Int. Conference on Signal Processing, 2008. ICSP 2008,* pp. 1552-1555, Beijing

Que, D.; Chen, B.; Hu, J. & Ax, Y. (2008). A Novel Single Training Sample Face Recognition Algorithm Based on Modular Weighted $(2D)^2$ PCA, 9th *Int. Conf. on Signal Processing, ICSP 2008,* pp. 1552-1555, ISBN: 978-1-4244-2178-7

Sluciak, O. & Vargic, R. (2008). An Audio Watermarking Method Based on Wavelet Patchwork Algorithm, *Proceedings of IWSSIP 2008*, June 2008, Bratislava, Slovak Republic, pp. 117-120, ISBN: 978-80-227-2856-0

Su,Y.; Shan, S.; Chen, X. & Gao, W., (2010). Adaptive Generic Learning for Face Recognition from a Single Sample Per Person, Proceedings of Int. Conference on Computer Vision and Pattern Recognition, CVPR2010, pp.2699~2706

Tan, X. & Triggs, B., (2010). Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *IEEE Transactions on Image Processing*, Vol.19(6), pp. 1635-1650, ISSN: 1057-7149

Tan, X., Chen, S., Zhou, Z.-H., & Zhang, F. (2006). Face recognition from a single image per person: A survey. *Pattern Recognition*, Vol. *39* (9), pp. 1725-1745. ISSN:0031-3203

Truax, B. (1999). (Ed.) Gaussian Noise In: *Handbook for Acoustic Ecology*, Available on: http://www.sfu.ca/sonic-studio/handbook/Gaussian_Noise.html Cambridge Street Publishing

Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, Win. 1991, pp. 71-86, Retrieved September 25, 2010, from http://portal.acm.org/citation.cfm?id=1326894#

Vargic, R. & Procháska J. (2005). An Adaptation of Shape Adaptive Wavelet Transform for Image Coding, *EURASIP2005,* Smolenice, Slovakia, June 2005

Wen, G., Shiguang, S., Xiujuan, C., & Xiaowei, F. (2003). Virtual Face Image Generation for Illumination and Pose Insensitive Face Recognition, *Proceedings of IEEE Int. Conference on Acoustics, Speech, and Signal Processing, ICASSP '03,* pp. IV-776-9, ISBN: 0-7803-7663-3

Wu, J. & Zhou, Z.-H., (2002) Face Recognition with One Training Image Per Person. *Pattern Recognition Letters*, Vol. 23(14) pp. 1711-1719, ISSN: 0167-8655

Xie, X. & Lam, K.-M., (2006). Gabor-Based Kernel PCA With Doubly Nonlinear Mapping for Face Recognition With a Single Face Image. *IEEE Trans. on Image Processing*, Vol. 15, No. 9, ISSN: 1057-7149

Xu, J. & Yang, J. (2009). Local Graph Embedding Discriminant Analysis for Face Recognition, School of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing 210094, China, 2009

Yang, C.-Kai, & Chiang, W.-ting. (2007). An Interactive Facial Expression Generation System. *Multimedia Tools and Applications*, Vol.40, No. 1, pp. 41-60

Yang, J., Zhang, D., Frangi, A. F., & Yang, J.-yu. (2004). Two-dimensional PCA: a New Approach to Appearance-Based Face Representation And Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*(1), pp. 131-137, ISSN:0162-8828

Zhan, C., Li, W., & Ogunbona, P. (2009). Face Recognition from Single Sample Based on Human Face Perception. *24th International Conference Image and Vision Computing* New Zealand, Wellington, pp. 56-61, ISBN: 978-1-4244-4698-8

Zhang, D.; Chen, S. & Zhou, Z.-H., (2005). A New Face Recognition Method Based on SVD Perturbation for Single Example Image Per Person. *Applied Mathematics and Computation,* Vol. 163, Issue 2, pp. 895-907

10

Constructing Kernel Machines in the Empirical Kernel Feature Space

Huilin Xiong[1] and Zhongli Jiang[2]
[1]*Shanghai Jiao Tong University, Shanghai*
[2]*Shanghai University of Political Science and Law, Shanghai*
China

1. Introduction

Over the last decade, kernel-based nonlinear learning machines, e.g., support vector machines (SVMs) Vapnik (1995), kernel principal component analysis (KPCA) Scholkopf (1998), and kernel Fisher discriminant analysis (KFDA) Mika (1999), attracted a lot of attentions in the fields of pattern recognition and machine learning, and have been successfully applied in many real-world applications Mika (1999); Yang (2002); Lu (2003); Yang (2004). Basically, the kernel-based learning methods work by mapping the input data space, $\mathcal{X}$, into a high dimensional space, $\mathcal{F}$, called the kernel feature space: $\Phi : \mathcal{X} \longrightarrow \mathcal{F}$, and then building linear machines in the kernel feature space to implement their nonlinear counterparts in the input space. This procedure is also known as a "kernelization", in which the so-called kernel trick is associated in such a way that the inner product of each pair of the mapped data in the kernel feature space is calculated by a kernel function, rather than explicitly using the nonlinear map, Φ.

The kernel trick provides an easy way to kernelize linear machines. However, in many cases, formulating a kernel machine via the kernel trick could be difficult and even impossible. For example, it is pretty tough to formulate the kernel version of the direct disciminant analysis algorithm (KDDA) Lu (2003) using the kernel trick. Moreover, for some recently developed linear discriminant analysis schemes, such as the uncorrelated linear discriminant analysis (ULDA) Ye (2004), and the orthogonal linear discriminant analysis (OLDA) Ye (2005), which have been shown to be efficient in many real-world applications Ye (2004), it is impossible to directly kernelize them via the kernel trick, since these schemes need first computing the singular value decomposition (SVD) of an interim matrix, namely, H_t (see Ye (2004)), which is generally of infinite column size in the case of the kernel feature space.

Theoretically, the kernel feature space is generally an infinite dimensional Hilbert space. However, given a training data set $\{x_i\}$ $(i = 1, 2, \ldots, n)$, the kernel machines we known perform actually in a subspace of the kernel feature space, $span\Phi(x_i)$ $(i = 1, 2, \ldots, n)$, which can be embedded into a finite-dimensional Euclidean space with all data's geometrical measurements, e.g., distance and angle, being preserved Xiong (2005). This finite-dimensional embedding space, called empirical kernel feature space, provides a unified framework for kernelizing all kinds of linear machines. With this framework, kernel machines can be “seamlessly” formulated from their linear counterparts without any difficulty: performing linear machines in the finite-dimensional empirical kernel feature space, the corresponding nonlinear kernel machines are then constructed in the input data space.

In this chapter, we propose to approach the kernelization from the empirical kernel feature space, that is, we formulate nonlinear kernel machines by directly performing their linear counterparts in the empirical kernel feature space. The kernel machines constructed, called empirical kernel machines, are usually different from the conventional kernel machines based on the kernel trick, and surprisingly, the empirical kernel machines are shown to be more efficient in many real-world applications, such as face recognition, facial expression recognition, and handwritten digit recognition, than the conventional nonlinear kernel machines and their linear counterparts.

The remainder of this chapter is organized as follows: In Section 2, we introduce the concepts and related notation concerning the empirical kernel feature space. Section 3 shows the difference in formulation between the conventional kernel principal component analysis (KPCA) and the empirical kernel principal component analysis (eKPCA), which is constructed by performing the linear principal component analysis (PCA) in the empirical kernel feature space. In Section 4, we formulate three other empirical kernel machines, namely, the empirical kernel direct discriminant analysis (eKDDA), the empirical kernel ULDA, denoted as eKUDA, and the empirical kernel OLDA, denoted as eKODA, via directly performing the DLDA Yu (2001), ULDA, and OLDA schemes in the empirical kernel feature space. Experiments for evaluating the performance of the empirical kernel machines in the real-world applications, e.g., face and facial expression recognition, are presented in Section 5.1. Finally, Section 6 concludes this chapter.

2. The empirical kernel feature space

Let $\{x_i, \xi_i\}_{i=1}^{n}$ be a d-dimensional training data with class labels $\{\xi_i\}$, the kernel matrix $K = [k_{ij}]_{n\times n}$, where $k_{ij} = \Phi(x_i)\cdot\Phi(x_j) = k(x_i, x_j)$, and $rank(K) = r, r \leq n$. Since K is a symmetrical positive semi-definite matrix, K can be decomposed as:

$$K_{n\times n} = P_{n\times r}\Lambda_{r\times r}P_{r\times n}^{T} \tag{1}$$

where Λ is a diagonal matrix only containing the r positive eigenvalues of K in decreasing order, and P consists of the eigenvectors corresponding to the positive eigenvalues. The map from the input data space to an r-dimensional Euclidean space Φ^e: $\mathcal{X} \longrightarrow \mathbf{R^r}$

$$x \longrightarrow \Lambda^{-\frac{1}{2}}P^T(k(x,x_1), k(x,x_2), \ldots, k(x,x_n))^T$$

is referred to the empirical kernel map in Xiong (2005); Scholkopf (1999). We call the subspace $span\{\Phi^e(x_i)\}$ the empirical kernel feature space, and denote it by $\mathcal{F}^e$. Obviously, we have $\overline{span\{\Phi^e(x_i)\}} \subset \overline{span\{\Phi^e(\mathcal{X})\}} \subset \mathbf{R}^r$. For the completion of the subspaces, it is easy to verify: $\overline{span\{\Phi^e(x_i)\}} = \overline{span\{\Phi^e(\mathcal{X})\}} = \mathbf{R}^r$.

It is well-known that various kernel machines, such as KPCA and SVM, perform only in a subspace of the kernel feature space: $span\{\Phi(x_i)\}$, which is actually isometric isomorphic with the empirical kernel feature space $span\{\Phi^e(x_i)\}$. In fact, let Y denote the data matrix with size $r \times n$ in the empirical kernel feature space, that is,

$$Y = (\Phi^e(x_1), \Phi^e(x_2), \ldots, \Phi^e(x_n)) = \Lambda^{-\frac{1}{2}} P^T K. \tag{2}$$

The dot product matrix of $\{\Phi^e(x_i)\}$ in the empirical kernel feature space can be calculated as

$$Y^T Y = KP\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}P^T K = K. \tag{3}$$

This is exactly the dot product matrix of $\{\Phi(x_i)\}$ in the feature space. Since the distances of the n vectors $\{\Phi(x_i)\}_1^n$ in the kernel feature space are uniquely determined by the dot product matrix, we can see the training data have the same distance matrix in both the empirical kernel feature space, $\mathcal{F}^e$, and the kernel feature space, $\mathcal{F}$, that is, as pointed out in Xiong (2005), $span\{\Phi(x_i)\}$ can be embedded into an r-dimensional Euclidean space with the distances between each pair of the training data being preserved. Note that the dimension of the samples in the empirical kernel feature space is always smaller than the sample size, $r \leq n$, which may help to some extent to alleviate the so-called “Small Sample Size” (SSS) problems Chen (2000); Yu (2001) in discriminant analysis.

3. Principal component analysis in the empirical kernel feature space

Principal component analysis (PCA) is a widely used subspace method in pattern recognition and dimension reduction. It gives the optimal representation of the pattern data with the minimum mean square error. The PCA transform (projection) matrix can be calculated from the eigendecomposition of the sample covariance matrix, or alternatively, from the eigendecomposition of the inner product matrix of samples in the case of high data dimensionality. Kernel principal component analysis (KPCA) is carried out by applying PCA in the kernel feature space. Using the kernel trick, the KPCA transform matrix can be computed from the eigendecomposition of the kernel matrix.

Let us perform the linear PCA in the empirical kernel feature space. The scheme obtained is called empirical kernel principal component analysis, denoted as eKPCA for short. Let K_c represent the centered kernel matrix, that is,

$$K_c = (I_{n\times n} - \frac{1}{n}1_{n\times n})K(I_{n\times n} - \frac{1}{n}1_{n\times n}),$$

where $I_{n\times n}$ is the $n \times n$ identity matrix, and $1_{n\times n}$ represents the $n \times n$ matrix with all entries being equal to unity. The centered kernel matrix can be decomposed as $K_c = Q\Sigma Q^T$, where Σ is a diagonal matrix containing the positive eigenvalues of K_c, and Q consists of the eigenvectors corresponding to the positive eigenvalues. Given a sample x, the conventional KPCA maps x to $\Sigma^{-\frac{1}{2}}Q^T(k(x,x_1),\ldots,k(x,x_n))^T$. However, when we perform the linear PCA in the empirical feature space, the x will be transformed to

$$\begin{aligned}
&\Sigma^{-\frac{1}{2}}Q^T Y^T \Phi^e(x)\\
&= \Sigma^{-\frac{1}{2}}Q^T Y^T \Lambda^{-\frac{1}{2}}P^T(k(x,x_1),\ldots,k(x,x_n))^T\\
&= \Sigma^{-\frac{1}{2}}Q^T P P^T(k(x,x_1),\ldots,k(x,x_n))^T.
\end{aligned}$$

This is our eKPCA formula. Note that $P^T P$ is the identity matrix of size $r \times r$, however, PP^T generally is not the identity matrix of size $n \times n$. If $Q^T PP^T = Q^T$, or equivalently, $PP^T Q = Q$

holds, our eKPCA scheme turns to be $\Sigma^{-\frac{1}{2}}Q^T(k(x,x_1),\ldots,k(x,x_n))^T$, which is actually the conventional KPCA. Many experiments (see the experiment section below) show that eKPCA and KPCA usually lead to the same results, which may suggest that the equation $PP^TQ = Q$ holds frequently in practices.

4. Discriminant analysis in the empirical kernel feature space

Currently, linear discriminant analysis (LDA) has become a classical statistical approach for pattern classification, feature extraction, and dimension reduction. It has been successfully applied in many real-world applications, e.g., face recognition Belhumeour (1997), information retrieval Berry (1995), and microarray gene expression data analysis Dudoit (2002). while PCA calculates the optimal projection for pattern representation, LDA projects data aiming to discriminate the labeled pattern data. LDA calculates the optimal projection directions by maximizing the ratio of the between-class scatter measure to the within-class scatter measure, and thus, achieves the maximum class discrimination. A big challenge facing the conventional LDA is that it requires the within-class scatter matrix (or the total scatter matrix) be nonsingular, which usually cannot be met in practices, specifically for the "SSS" problems Chen (2000); Yu (2001).

In recent years, we have witnessed a great development of the linear discriminant analysis (LDA) research in handling the problem caused by the singularity of the scatter matrices. A variety of linear schemes have been proposed, from the pseudo-inverse LDA Raudys (1998), the null space LDA Chen (2000), and the direct linear discriminant analysis (DLDA) Yu (2001), to the recently developed sophisticated schemes, the uncorrelated LDA (ULDA) Ye (2004) and orthogonal LDA (OLDA) Ye (2005).

In this section, we perform various linear discriminant analysis schemes in the r-dimensional empirical kernel feature space to formulate our kernel nonlinear discriminant analysis schemes. It needs to emphasis that, in the empirical kernel feature space, the data dimension and the scatter matrix size are always smaller than the sample size ($r \leq n$). However, even so, we still face the singularity problem of the scatter matrices. We choose to kernelize three LDA schemes, namely, the DLDA, ULDA, and OLDA schemes, which are three typical extensions of the classical LDA scheme in overcoming the singularity problem. With these examples, we want to highlight our point that performing linear LDA schemes in the empirical kernel feature space can seamlessly formulate the kernel versions of vaious linear discriminant analysis schemes.

Suppose the labeled training data $\{x_i, \xi_i\}_{i=1}^n$ are grouped into m class, and each class contains n_i samples, where $\sum_{i=1}^m n_i = n$. The data matrix of the training data in the empirical kernel feature space is Y, that is, $Y = \Lambda^{-\frac{1}{2}}P^TK$. Let us define three matrices H_b, H_w, and H_t as follows:

$$H_b = \frac{1}{\sqrt{n}}[\sqrt{n_1}(\overline{y}_1 - \overline{y}), \cdots, \sqrt{n_m}(\overline{y}_m - \overline{y})]$$

$$H_w = [\frac{1}{\sqrt{n}}(Y_1 - \overline{y}_1 1_{n_1}^T), \cdots, (Y_m - \overline{y}_m 1_{n_m}^T)]$$

$$H_t = \frac{1}{\sqrt{n}}(Y - \overline{y}1_n^T)$$

where Y_i and $\overline{y}_i$ respectively denote the data matrix and centroid of the i-th class in the empirical kernel feature space, $\overline{y}$ is the global centroid of the data in the empirical kernel feature space, and 1_{n_i} represents the n_i-dimensional vector with entries being unity. Then, the *between-class scatter matrix* S_b, the with-in class scatter matrix S_w, and the *total scatter matrix* S_t defined in Fukunaga (1990) can be represented as: $S_b = H_b H_b^T$, $S_w = H_w H_w^T$, and $S_t = H_t H_t^T$. It is easy to verify:

$$H_b = YE_b,\ H_w = YE_w, \text{ and } H_t = YE_t \tag{4}$$

and therefore, we have

$$S_b = YE_bY^T,\ S_w = YE_wY^T, \text{ and } S_t = YE_tY^T \tag{5}$$

where the three constant matrices, E_b, E_w, and E_t, are:

$$\begin{aligned} E_b &= D - \frac{1}{n}1_{n\times n} \\ E_w &= I_{n\times n} - D \\ E_t &= I_{n\times n} - \frac{1}{n}1_{n\times n} \end{aligned}$$

in which matrix D is:

$$\begin{pmatrix} \frac{1}{n_1}1_{n_1\times n_1} & & \\ & \ddots & \\ & & \frac{1}{n_m}1_{n_m\times n_m} \end{pmatrix},$$

$I_{n\times n}$ is the $n \times n$ identity matrix, and $1_{n_i\times n_i}$ represents the $n_i \times n_i$ matrix with all the entries being equal to unity.

4.1 Empirical kernel direct discriminant analysis

In discriminant analysis, it has been recognized that the null space of the within-class scatter matrix may contain significant discriminant information. The so-called "direct LDA", or DLDA in the literature, involves two schemes Chen (2000); Yu (2001) in extracting the discriminant information from the null space, and meanwhile addressing the singularity problem of the scatter matrix. Different from Chen *et.al.*'s scheme Chen (2000), Yu *et.al.*'s scheme Yu (2001) first projects the data into the range space of the between-class matrix, and then calculates the projection in the null space of the within-class scatter matrix. Yu *et.al.*'s scheme is more efficient in computation than Chen *et.al.*'s, and this scheme has been kernelized by Lu *et.al.* in Lu (2003). In this section, we formulate our kernel direct discriminant analysis by performing the Yu's DLDA scheme in the empirical kernel feature space. The obtained kernel direct discriminant analysis algorithm is called empirical kernel direct discriminant analysis, denoted as eKDDA in order to differentiate it from Lu's KDDA scheme:

- **Step 1**. Calculate the matrices Y, S_b, and S_w in Eq.(2) and Eq.(5).
- **Step 2**. Calculate the eigen decomposition of $S_b = YE_bY^T$ as $S_b = P_b\Lambda_b P_b^T$, where Λ_b is the diagonal matrix consisting of the r_b positive eigen values sorted in decreasing order, and $r_b = rank(S_b)$. Let $M_1 = P_b\Lambda_b^{-\frac{1}{2}}$.

- **Step 3**. Calculate $\widetilde{S}_w = M_1^T S_w M_1$, and decompose it as:

$$\widetilde{S}_w = \left(\widetilde{P}_w, \widetilde{N}_w\right) \begin{pmatrix} \widetilde{\Lambda}_w & \\ & 0 \end{pmatrix} \begin{pmatrix} \widetilde{P}_w^T \\ \widetilde{N}_w^T \end{pmatrix}$$

- **Step 4**. Suppose we need extracting q-dimensional feature vectors, where $q \leq m - 1$. Let $M = M_1 \widetilde{N}_w(:, 1 : q)$, then, for given $x \in \mathcal{X}$, eKDDA transform x to

$$G(k(x, x_1), k(x, x_2), \ldots, k(x, x_n))^T,$$

where $G = M^T \Lambda^{-\frac{1}{2}} P^T = \widetilde{N}_w^T \Lambda_b^{-\frac{1}{2}} P_b^T \Lambda^{-\frac{1}{2}} P^T$.

In the implementation of the eKDDA algorithm, to avoid possible numerical instability in step 2, we introduce an extra parameter, ε, to discard some tiny eigenvalues. The eigenvalue λ is considered to be zero if $\frac{\lambda}{\lambda_{max}} \leq \varepsilon$, where λ_{max} denotes the maximum eigenvalue. In the step 3, we only need calculate the eigen decomposition of the matrix $\widetilde{S}_w$, and sort the eigenvalues (or the absolute values of the eigenvalues) in ascend order. The $\widetilde{N}_w(:, 1 : q)$ is then composed of the q eigenvectors corresponding to the first p small eigenvalues.

4.2 Empirical kernel uncorrelated and orthogonal discriminant analysis

Uncorrelated linear discriminant analysis (ULDA) Ye (2004) and orthogonal linear discriminant analysis (OLDA) Ye (2005) are two recently developed LDA schemes, in which some sophisticated matrix techniques such as singular value decomposition (SVD) and QR-decomposition are used to address the singularity problem in the classical LDA scheme. In the ULDA and OLDA algorithms, we need first compute the SVD of the matrix H_t, which makes it difficult to kernelize ULDA and OLDA directly via the conventional kernel trick, since the dimension of the matrix H_t in the kernel feature space is infinite in general. In Ji (2008), an indirect kernelization scheme of ULDA and OLDA, refereed to as KUDA and KODA, respectively, is proposed. Essentially, in the scheme of Ji (2008), KUDA "is equivalent to applying ULDA to the kernel matrix, where each column is considered as an n-dimensional data point" Ji (2008). Since the geometrical structure, e.g., distance and angle, among the "column" data of the kernel matrix is different from that of the data in the kernel feature space, some discriminatory information may be changed or lost as we use the "column" data to replace the data in the kernel feature space. On the contrary, the empirical kernel feature space preserves the geometrical structure of the training data in the kernel feature space, therefore, there would be no information loss in performing LDA in the empirical kernel feature space instead of the kernel feature space. Furthermore, our experiments show (see the experiment section) that the kernel ULDA and OLDA formulated in the empirical kernel feature space perform substantially better than KUDA and KODA in most cases.

According to the schemes of ULDA and OLDA Ye (2004; 2005), we simply perform the ULDA and OLDA algorithms in the empirical kernel feature space to formulate our empirical kernel ULDA and OLDA, denoted as eKUDA and eKODA, respectively.

4.2.1 The eKUDA algorithm

- **Step 1**. Calculate the matrices Y, H_t, and H_b in Eq.(2) and (4).
- **Step 2**. Calculate the reduced SVD of H_t as $H_t = U_t \Sigma_t V_t^T$.

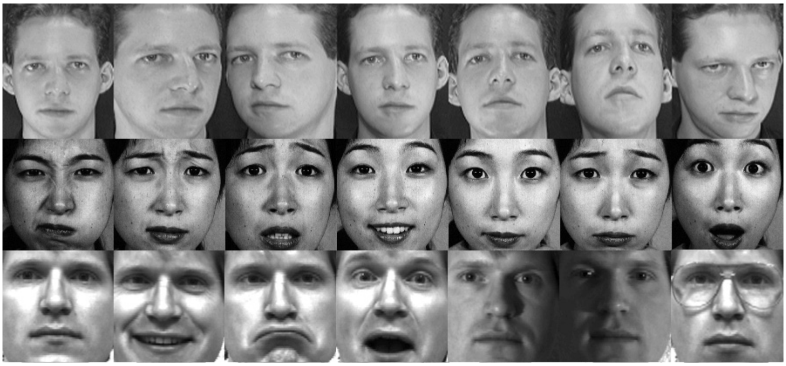


Fig. 1. Some sample images in the ORL, JAFFE, and Yale data sets.

- **Step 3**. Let $B = \Sigma_t^{-1} U_t^T H_b$, and $q = rank(B)$. Calculate the reduced SVD of B as $B = U_B \Sigma_B V_B^T$.
- **Step 4**. Let $X = U_t \Sigma_t^{-1} U_B$, $M = X(:, 1:q)$, then, for given $x \in \mathcal{X}$, eKUDA transform x to

$$G(k(x, x_1), k(x, x_2), \ldots, k(x, x_n))^T,$$

where $G = M^T \Lambda^{-\frac{1}{2}} P^T$.

4.2.2 The eKODA algorithm

- **Step 1**. Calculate the matrices Y, H_t, and H_b in Eq.(2) and (4).
- **Step 2**. Calculate the reduced SVD of H_t as $H_t = U_t \Sigma_t V_t^T$.
- **Step 3**. Let $B = \Sigma_t^{-1} U_t^T H_b$, and $q = rank(B)$. Calculate the reduced SVD of B as $B = U_B \Sigma_B V_B^T$.
- **Step4**. Let $X = U_t \Sigma_t^{-1} U_B$. Calculate the QR-decomposition of $X_q = X(:, 1:q)$ as $X_q = QR$, then, for a given sample $x \in \mathcal{X}$, eKODA transform x to

$$G(k(x, x_1), k(x, x_2), \ldots, k(x, x_n))^T,$$

where $G = Q^T \Lambda^{-\frac{1}{2}} P^T$.

5. Experiments

We conduct three types of experiments to investigate the efficiency of our empirical kernel machines in a wide range of real-world applications. We compare the performances of our empirical kernel machines, specifically, eKPCA, eKDDA, eKULDA ,and eKOLDA, with those of the kernel-trick-based machines, namely, KPCA, KDDA, KUDA, and KODA, and the linear machines, namely, PCA, ULDA, and OLDA, in the applications of face recognition, facial expression recognition, and handwritten digit recognition.

Four standard databases, including three face image data sets and one handwritten digit image data set, are used to evaluate the pattern classification algorithms

p	0.2	0.3	0.4	0.5	0.6
PCA	81.40±1.99	88.03±2.36	92.14±1.65	94.62±1.66	95.52±1.49
KPCA	81.44±1.97	88.16±2.36	92.26±1.62	94.90±1.65	95.94±1.35
eKPCA	81.44±1.97	88.15±2.36	92.26±1.62	94.90±1.65	95.94±1.35
KDDA	78.80±5.27	86.29±2.54	93.09±1.63	95.90±1.10	97.70±1.39
eKDDA	83.38±2.01	89.93±2.07	93.51±1.68	94.64±1.44	96.34±1.35
ULDA	80.84±2.57	86.46±2.01	90.18±1.91	92.05±2.26	93.33±1.49
KUDA	**85.96±2.06**	**91.78±1.88**	95.06±1.55	96.51±1.08	97.77±1.10
eKUDA	85.52±2.14	91.42±1.89	94.82±1.53	96.91±1.21	97.67±1.10
OLDA	84.96±2.18	90.86±2.09	94.18±1.47	96.01±1.25	97.25±1.35
KODA	85.07±2.44	91.41±1.95	95.08±1.75	96.57±1.16	97.84±1.33
eKODA	85.30±2.13	91.58±1.90	**95.37±1.35**	**96.95±1.10**	**98.09±1.11**

Table 1. Experimental results in terms of the average values and the standard deviations of the best recognition accuracy (%) on test data for the ORL data set

mentioned above. The three face image databases are ORL face images (available at http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html), Yale face images (available at http://cvc.yale.edu/projects/yalefaces/yalefaces.html), and JAFFE facial expression images Lyons (1998). The handwritten digit images, in size 16 × 16, are collected from the USPS database Hull (1994). Some samples of these image databases are shown in Fig.(1). Except the JAFFE images and Yale images, where the face part of each image is cropped, in size of 128 × 128 and 112 × 112, respectively, from the original images, no any other preprocessing is applied to the images. The ORL and Yale data are used to evaluate the algorithms for the task of face recognition, and the JAFFE face images are used for facial expression recognition.

We only consider the Gaussian kernel, $k(x,y) = exp(-\gamma\|x-y\|^2)$, in this chapter. There is no parameter need to be set in advance for the ULDA and OLDA schemes, and only one parameter, γ, need to set for the KUDA, eKUDA, KODA, and eKODA schemes. However, for the KPCA, eKPCA, KDDA, and eKDDA schemes, an extra parameter, ε, is introduced to avoid the numerical instability caused by the tiny eigenvalues. The tiny eigenvalue λ is considered to be zero, if $\frac{\lambda}{\lambda_{max}} \leq \varepsilon$, where λ_{max} denotes the maximum eigenvalue. We select the parameter γ from set {10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}, 10^{-9}, 10^{-10}}, and the parameter ε from set {10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 0}. For the KDDA and eKDDA schemes, the final projection dimension q, where $q \leq m-1$, still needs to be pre-specified. However, to avoid setting too many parameters, especially, for the KDDA scheme, we usually fix q at $m-2$. In the experiments, we implement the KDDA scheme using the Matlab code written by Lu, which is available for downloading at http://www.dsp.utoronto.ca/juwei/juwei_pubs.html). However, for the sake of fairness in the comparisons, the regularization constant, "*Eta_sw*", in Lu's KDDA code is set to zero, since no other scheme employs the regularization technique to further improve performance.

After data are mapped to the different projection spaces, the nearest neighbor (NN) classifier is employed to classify the sample images, and the classification accuracy on test samples are used to evaluated the performances of various learning machines.

p	0.2	0.3	0.4	0.5	0.6
PCA	57.59±4.91	64.62±3.01	67.26±4.32	68.92±4.11	72.12±4.85
KPCA	58.17±4.71	64.92±2.98	67.69±4.17	69.22±4.13	72.37±4.73
eKPCA	58.13±4.69	64.90±2.95	67.67±4.15	69.19±4.12	72.42±4.67
KDDA	49.46±6.14	69.15±3.67	74.21±4.35	76.69±4.04	81.00±4.43
eKDDA	63.33±3.87	74.94±3.79	77.45±3.59	82.53±3.68	86.46±3.60
ULDA	70.63±3.73	79.60±3.10	81.38±5.33	83.94±4.81	86.54±5.34
KUDA	68.74±6.57	79.69±2.93	76.29±15.86	74.08±21.16	72.88±24.09
eKUDA	**71.63±3.26**	**80.54±2.99**	**83.05±4.34**	85.44±4.36	88.58±4.46
OLDA	66.67±3.74	77.65±3.55	81.90±4.00	84.92±2.81	87.00±4.79
KODA	63.20±4.01	75.35±3.46	79.00±3.92	82.08±3.08	87.33±4.23
eKODA	67.19±3.75	78.21±3.44	82.55±3.95	**85.78±2.65**	**88.96±3.92**

Table 2. Experimental results in terms of the average values and the standard deviations of the best recognition accuracy (%) on test data for the Yale data set

p	0.5	0.6	0.7	0.8	0.9
PCA	58.48±4.48	64.52±4.15	69.25±5.69	73.04±5.69	78.33±8.62
KPCA	59.05±4.40	65.06±4.03	70.08±5.54	73.69±6.31	79.52±8.50
eKPCA	58.93±4.38	65.06±4.04	70.04±5.54	73.69±6.31	79.40±8.45
KDDA	61.57±5.32	65.51±4.36	68.33±5.20	70.77±7.15	73.33±8.68
eKDDA	69.48±5.00	73.87±4.51	77.06±4.97	79.46±4.96	86.55±7.54
ULDA	70.62±4.71	74.37±4.41	77.34±5.05	79.70±5.39	85.71±7.78
KUDA	71.69±4.50	75.71±4.12	79.25±5.00	82.74±5.08	88.07±6.82
eKUDA	71.83±4.48	75.95±4.23	79.44±5.01	83.04±4.80	88.10±7.23
OLDA	72.14±5.31	76.82±5.09	78.97±5.36	82.38±5.80	87.74±7.46
KODA	73.50±5.03	**78.42±5.09**	80.55±5.74	85.24±5.32	89.52±6.92
eKODA	**73.62±4.72**	78.07±5.01	**81.03±5.26**	**85.36±4.84**	**90.12±6.23**

Table 3. Experimental results in terms of the average values and the standard deviations of the best recognition accuracy (%) on test data for the JAFFE data set

5.1 Experiment on face recognition

In this experiment, we compare the empirical kernel machines with the kernel-trick-base kernel machines and the linear machines in the application of face recognition. The experiment is carried out on two face image database, the ORL and Yale database. The ORL data contain 40 persons, each having 10 different images of size 92 × 112 with the variation to a certain extent in pose and scaling, and the Yale data we used includes 15 individuals, each having 10 pictures (cropped to size 112 × 112) with different facial expressions and illuminations, wearing or without wearing glasses. The samples of each subject are randomly divided to two disjoint subsets, one is used as the training data, and the other the test data. The ratio of the training data number to the total sample number per class (individual), called training rate, is denoted by p.

We investigate the performances of different machines with different values of p. The best value of the recognition accuracy on the test data over different parameter settings is used to evaluate the performances of different algorithms. The experiment is repeated 40 times, and the experimental results in terms of the average values and the standard deviations of the recognition accuracy on test data are shown in Table 1, for the ORL data, and Table 2, for

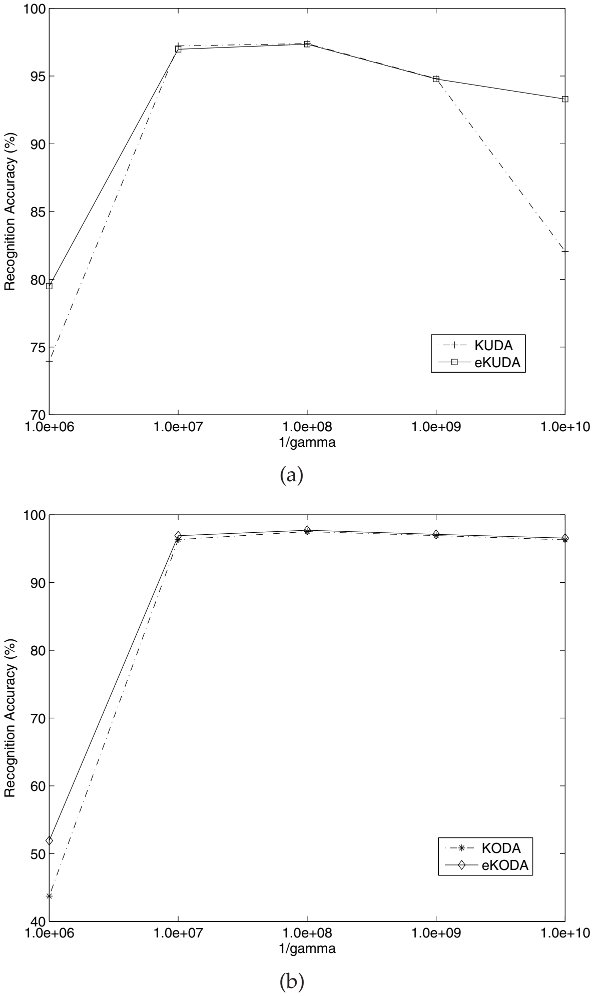
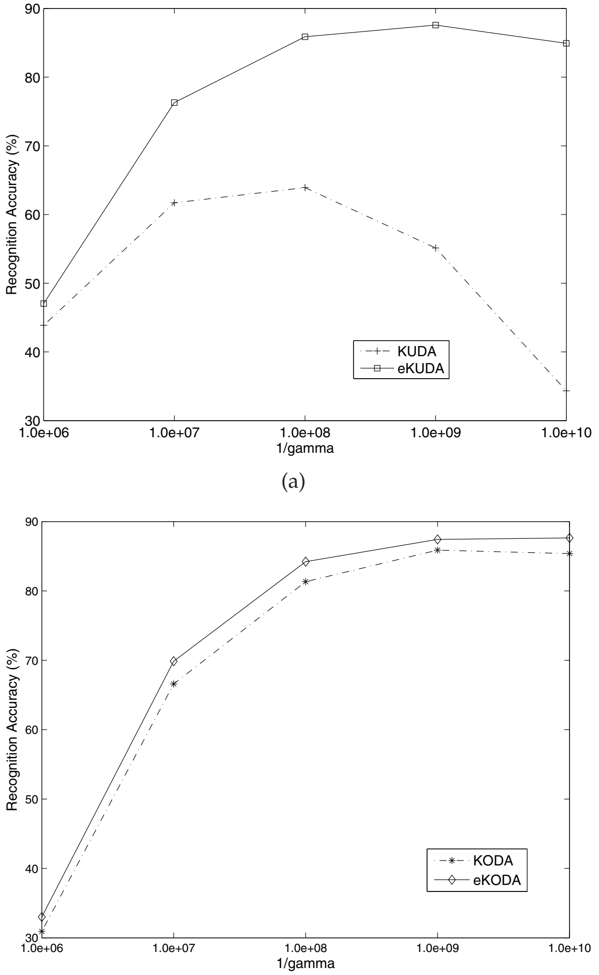


Fig. 2. Performance comparisons of (a) KUDA vs. eKUDA, and (b) KODA vs. eKODA on ORL data, with different parameter γ settings.

the Yale data. The best results under different training rates (p) are shown in boldface in the tables.

We also compare the performances of two pairs of kernel machines, namely, KUDA vs. eKUDA, and KODA vs. eKODA, when their unique parameter γ is set to different values. Fig.(2) (a) (b) illustrate the average test recognition accuracy (%) as a function of $1/\gamma$ on the ORL data set, where the training rate is set at $p = 0.6$. The corresponding result on the Yale data set is presented in Fig.(3)

The experimental results in Tables 1 and 2 and Figs.(2)(3) lead to following points:



(a)

(b)

Fig. 3. Performance comparisons of (a) KUDA vs. eKUDA, and (b) KODA vs. eKODA on Yale data, with different parameter γ settings.

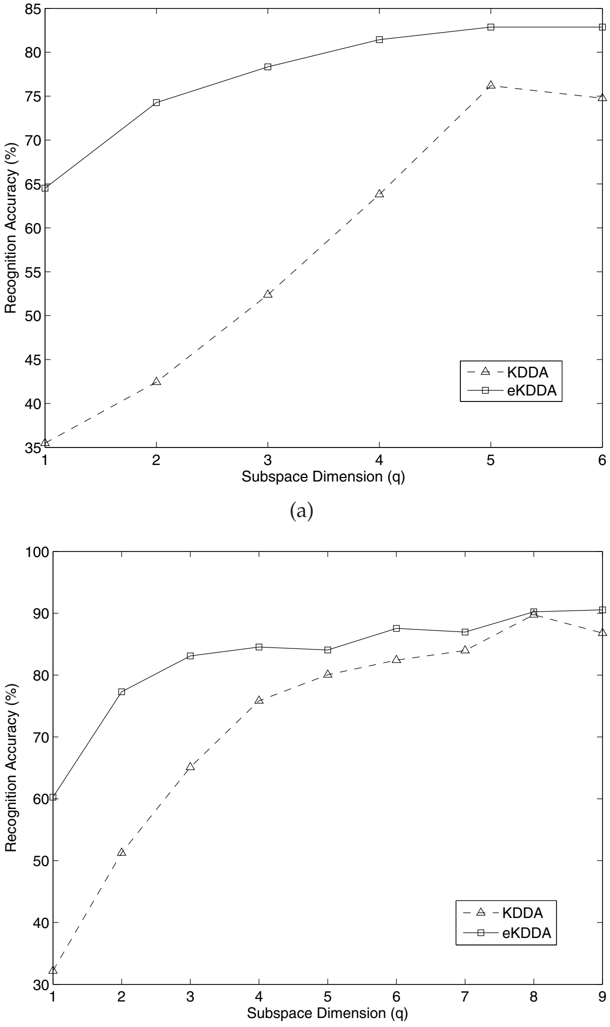
p	0.2	0.3	0.4	0.5	0.6
PCA	82.54±2.31	83.25±2.37	84.30±1.73	87.30±2.13	88.53±2.39
KPCA	83.33±2.21	84.25±1.99	85.24±1.67	87.97±1.86	89.41±1.89
eKPCA	83.33±2.21	84.25±1.99	85.22±1.66	88.00±1.85	89.41±1.89
KDDA	82.92±2.46	84.45±2.04	83.62±1.87	87.10±1.80	87.45±2.09
eKDDA	83.84±2.59	85.82±1.71	86.08±2.07	88.11±2.13	90.23±2.20
ULDA	60.80±4.00	52.75±4.95	46.60±3.60	40.99±4.11	29.18±3.04
KUDA	83.91±2.99	86.85±2.54	88.16±1.76	90.20±2.11	92.19±1.87
eKUDA	86.16±2.00	88.44±1.98	89.45±1.41	90.65±1.98	92.78±1.73
OLDA	67.32±3.50	59.85±3.52	57.60±3.06	50.84±3.81	37.05±3.63
KODA	83.22±2.62	85.93±2.22	86.98±1.66	88.45±2.13	89.37±2.24
eKODA	**86.74±2.20**	**89.12±1.97**	**89.75±1.33**	**91.09±1.66**	**92.91±1.90**

Table 4. Experimental results in terms of the average values and the standard deviations of the best recognition accuracy (%) on test data for the USPS data set

1. Empirical kernel machines achieve the best results in most cases.
2. Empirical kernel PCA performs almost the same as the conventional KPCA, which may suggests that the Eq.(2) holds or approximately holds in practices.
3. Lu's KDDA scheme works better than eKDDA in two cases on the ORL data. However, on the Yale data set, where the within-class scatter measure is much larger than that of the ORL data due to the variations of illumination, the eKDDA scheme performs much better than the KDDA scheme.
4. For the SVD-based discriminant analysis schemes, either ULDA, OLDA, or their kernel counterparts, they usually outperform the PCA schemes and the direct-LDA schemes. Moreover, while the KUDA and KODA schemes work better than their linear counterparts on the ORL data, their performances degenerate remarkably on the Yale data, especially for KUDA. However, in either case, our eKUDA and eKODA work well, and lead to most best results.

5.2 Experiment on facial expression recognition

We investigate the efficiency of our empirical kernel machines in the application of facial expression recognition, and compare their performances with those of the other pattern classification methods. Compared with face recognition, the facial expression recognition is a more challenging classification task, since the between-class discrimination among different facial expression patterns is much smaller than the within-class discrimination of the expression patterns. In this experiment, we use the JAFFE facial expression database to test and evaluate various algorithms. The JAFFE data set is a widely-used database for facial expression recognition. It contains ten Japanese women's face images with 7 typical facial expressions (angry, disgust, fear, happy, sad, surprise, and neutral), each expression having three different pictures, which are cropped to size 128 × 128. Since facial expression recognition is a difficult classification task, the training rate p is set to a relatively large value. The experimental results are shown in Table 3 in terms of the average best recognition accuracy on test data over 40 trails , corresponding to the training rate $p = 0.9, 0.8, 0.7, 0.6$, and 0.5, respectively. Furthermore, we also compare the performances of KDDA and eKDDA



(a)

(b)

Fig. 4. Comparison of the performances of the KDDA and eKDDA schemes on, (a) the JAFFE data, and (b) the USPS data, under different projection dimension q

with different projection dimension q (in the previous experiments, we fix q at $m - 2 = 5$), as the training rate p is at level 0.9. Fig.(4) (a) shows the results.

It can be seen that, for the facial expression recognition on the JAFFE data set, 1)the eKDDA scheme remarkably outperforms the KDDA scheme; 2)the orthogonal discriminant analysis schemes perform better than the uncorrelated disciminant analysis schemes, either in OLDA vs. ULDA, KODA vs. KUDA, or eKODA vs. eKUDA, and furthermore, the eKODA scheme achieves the best results in all cases except $p = 0.6$.

5.3 Experiment on handwritten digit recognition

To test our algorithms in a wide-range of applications, we conduct experiment for handwritten digit recognition using the USPS data. The USPS handwritten digit data set, available for downloading at http://www.csie.ntu.edu.tw/ ˜cjlin/libsvmtools/datasets/, is widely used as a benchmark for evaluating various learning methods. It contains more then 7 thousands training samples and two thousands test samples of handwritten digits from 0 to 9. Each sample is represented by an 16 × 16 image.

Since our goal in this experiment is focused at comparing different classification algorithms, to reduce the computational burden, we randomly select 800 samples, 80 samples per class, from the training set of the USPS data to form our experiment data set. Considering the data dimension in this experiment is much smaller than that of the data used in other experiments, we choose the value of the parameter γ from $\{10^0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$, and the parameter ε from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 0\}$. Table 4 gives the experimental results in terms of the average values of the best recognition accuracy on test data over 40 trails , corresponding to the training rate $p = 0.2, 0.3, 0.4, 0.5$, and 0.6, respectively. Furthermore, to compare the performances of the KDDA and eKDDA schemes under different projection dimension q (in the previous experiments, we always set $q = m - 2$), we illustrate the average test recognition accuracy (%) as a function of q in Fig.(4) (b), where the training rate is set at $p = 0.6$.

From Table 4 and Fig.(4), it is easy to see that the eKODA scheme achieves the best recognition results in all cases, and eKDDA performs substantially better than KDDA. Moreover, a big difference between Table 4 and other tables is that the linear versions of the SVD-based discriminant analysis, i.e., ULDA and OLDA, perform surprisingly worse than other methods this time. However, their kernel nonlinear versions still work well, especially, the empirical kernel versions.

6. Conclusion

We have presented a new way to “seamlessly” kernelize linear machines. The empirical kernel feature space, a finite-dimensional embedding space, in which the distances of the data in the kernel feature space are preserved, provides a unified framework for the kernelization. This method is different from the conventional kernel-trick based kernelization, and more importantly, the final empirical kernel machines performs more efficiently in many real-world applications, such as face recognition, facial expression recognition, and handwritten digit identification, than the kernel-trick based kernel machines.

7. Acknowledgements

We would thank Dr. Lu and Dr. Ye for providing Matlab codes for the implementation of the KDDA, ULDA, and OLDA algorithms. This work was supported by the National Natural Science Foundation of China (grant no. 60775008) and the National High Technology Research and Development Program (863 Program) of China (grant no. 2007AA01Z196).

8. References

V. Vapnik. *The Nature of Statistical Learning Theory*, Mew York:Spring, 1995.

B. Scholkopf, A.J. Smola, and K.-R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, vol.10, pp.1299–1319, 1998.

S. Mika, G. Rätsch, J. Weston, B. Schölkopf, and K.-R. Müller. Fisher discriminant analysis with kernel. *Proc. IEEE Int'l Workshop Neural Networks for Signal Processing IX*, pp.41-48, Aug. 1999.

S. Mika, G. Rätsch, B. Schölkopf, A. Smola, J. Weston, and K.-R. Müller. Invariant feature extracion and classification in kernel space. *Advances in Neural Information Processing Systems*, 12, Cambridge, Mass.:MIT Press, 1999.

M.H. Yang. Kernel eigenfaces vs. kernel Fisherfaces: Face recognition using kernel methods. *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp.215-220, May 2002.

J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans Neural Networks*, vol.14, no.1, 2003, pp.117-126.

J. Yang, A.F. Frangi, and J.-Y. Yang. A new kernel Fisher disciminant algorithm with application to face recognition. *Neurocomputing*, vol.56, pp.415-421, 2004.

J. Ye, T. Li, T. Xiong, and R. Janardan. Using uncorrelated discriminant analysis for tissue classification with gene expression data. *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol.1, no.4, 2004, pp.181-190.

J. Ye. Characterization of a family of algorithms for generalized discriminant analysis on undersampled problems. *Journal of Machine Learning Research*, vol.6, 2005, pp.483-503.

H. Xiong, M.N.S. Swamy, and M.O. Ahmad. Optimizing the kernel in the empirical feature space. *IEEE Trans. Neural Networks*, vol.16, no.2, 2005, pp.460-474.

B. Schölkopf, B. Mika, C.J.C. Burges, P. Knirsch, K.-R. Müler, G. Rätsch, and A.J. Smola. Input space versus feature space in kernel-based methods. *IEEE Trans. Neural Networks*, vol.10, no.5, 1999, pp.1000-1017.

L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, and G.J. Yu. A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition*, vol.33, 2000, pp.1713-1726.

H. Yu and J. Yang. A Direct LDA Algorithm for High-Dimensional Data–with Application to Face Recognition. *Pattern Recognition*, vol.34, no.10, pp.2067-2070, 2001.

P.N. brlhumeour, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific projection. *IEEE Trans. on Pattern Analysis and machine Intelligence*, vol.19, no.7, pp.711-720, 1997.

M.W. Berry, S.T. Dumaism, and G.W. O'Brie. Using linear algebra for intelligent information retrieval. *SIAM Review*, vol.37, pp.573-595, 1995.

S. Dudoit, J. Fridlyand, and T.P. Speed. Comparison of discriminant methods for the classification of tumors using gene expression data. *Journal of the American Statistical Association*, vol.97, pp.77-87, 2002.

S. Raudys and R.P.W. Duin. On expected classification error of the Fisher linear classifier with pseudo-inverse convariance matrix. *Pattern Recognition Letters*, vol.19, no.5-6, 1998, pp.385-392.

K. Fukunaga. *Introduction to Statistical Pattern Recognition*, second edition. Academic Press, 1990.

S. Ji and J. Ye. Kernel uncorrelated and regularized discriminant analysis: A theoretical and computational study. *IEEE Trans. on Knowledge and Data Engineering*, vol.20, no.10, pp.1311-1321, 2008.

J. J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.16, no.5, pp.550-554, 1994.

M.J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba. Coding Facial Expressions with Gabor Wavelets. *Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, pp. 200-205, April 14-16 1998.

Part 4

Robust Facial Localization & Recognition

11

Additive Noise Robustness of Phase-Input Joint Transform Correlators in Face Recognition

Alin Cristian Teusdea and Gianina Adela Gabor
University of Oradea
Romania

1. Introduction

This chapter introduces a new phase-input joint transform correlator (PiJTC) to accommodate the additive noise in face recognition. It begins with the *4f* Vander Lugt architecture in order to highlight the importance of spatial coordinate filtering in the Fourier plane. It continues with the introduction of non-zero order fringe-adjusted amplitude joint transform correlator (FA-AJTC) model. There are also presented the experimental optical and hybrid setups, with their benefits and drawbacks, for a better understanding of the correlation process.

The authors emphasize the presentation of this model because it represents the basis for the phase-input model. A better understanding of the (FA-AJTC) facilitates the use of the proposed sine modulated fringe-adjusted phase-input joint transform correlator (FA-sinePiJTC). All the steps that help the amplitude (FA-AJTC) model performances improve, are similar in the phase model with the benefits migration, too. A short step by step example of correlation process explains and reveals the pattern recognition performances improvements of the proposed phase-input JTC model.

It is pointed out that the sine modulation function comprises a degree of freedom through the limits of the modulation domain, $dfPRE = fPRE_2 - fPRE_1$. With these parameters one can adjust the pattern recognition performance of the correlation process due to the variations of faces in the captured images.

Next stage of this chapter analysis the face database recognition performances of the proposed (FA-sinePiJTC) model in additive noise conditions from 10% to 50% random noise. There are presented some computer simulation results of the correlation process over face images selected from a small test database with 21 individual classes x 3 face images/class. Due to the variations of the faces in the original images there were chosen different sine modulation domains to achieve better face recognition performances in additive noisy conditions for a larger database. Thus, there were selected 102 individual classes with 5 face images/class multiplied by 5 additive noise conditions. The goal was to build up 1,560,600 cross-correlations between all original face images and all the noisy ones. Where was necessary, the face recognition performances were analysed with the error rates: equal error rate (EER), genuine acceptance rate (GAR), false acceptance rate (FAR) and false rejection rate (FRR). The receiver operating characteristic (ROC) curve was also built up.

2. Amplitude joint transform correlators (AJTC)

Vander Lugt correlator (VLC) is based on *4f* setup involving two convergent lenses, L1 and L2, with the same focal length, f (figure 1) (Born & Wolf, 1970; Vlad et al., 1976). The two lenses represent the 2D Fourier operators and generate the Fourier transform in the output Fourier plane of the input image. Thus, there are two Fourier output planes in the VLC: the Fourier plane or the spatial frequencies plane, PF, of L1 lens and the output plane, Pe, of L2 lens. An important notice regarding the optical or hybrid correlators based on *4f* Vander Lugt principle is that they have two Fourier transforms related with the L1 and L2.

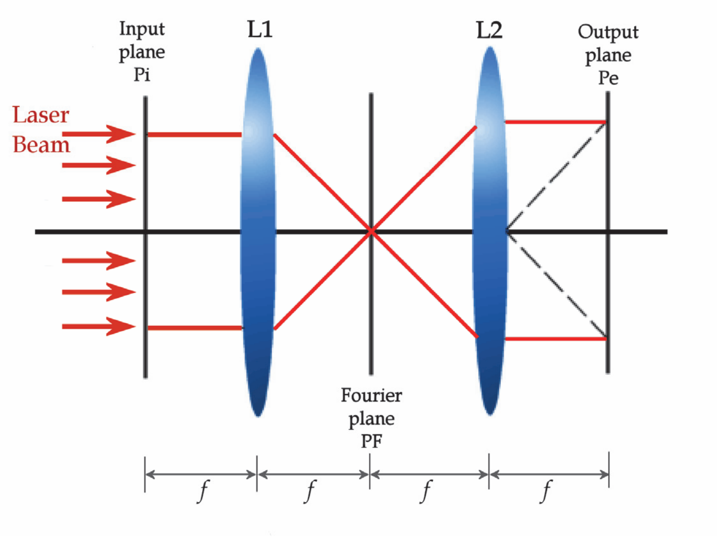


Fig. 1. Architecture of the *4f* VLC and JTC.

The mathematical model of (VLC) starts considering an input image, $scn(x,y)$, that contains the target image, $t(x,y)$, and clutter or the embedding noise, $n(x,y)$ (Born & Wolf, 1970; Vlad et al.,1976)

$$scn(x,y) = t(x-x_t, y-y_t) + n(x,y). \quad (1)$$

The L1 lens generates $Scn(u,v)$ as the Fourier transform of the input image, in the (VLC's) Fourier plane, PF. In this plane, the complex filter, $H(u,v)$, of the reference image that has to be recognized by the (VLC) is placed. In other words, the reference $ref(x+x_r, y+y_r)$ has to be discriminated in the input image, $scn(x,y)$, at the target image $t(x-x_t, y-y_t)$ location. The complex filter consists of complex conjugate Fourier of the reference, located at (x_r, y_r) (Born & Wolf, 1970; Vlad et al., 1976)

$$H(u,v) = Ref^*(u,v) \cdot \exp(2\pi i(x_r \cdot u + y_r \cdot v)), \quad (2)$$

where $Ref(u,v)$ is the Fourier transform of $ref(x,y)$, $u = \frac{2\pi}{\lambda f} \cdot x$ and $v = \frac{2\pi}{\lambda f} \cdot y$ are the spatial coordinates in the Fourier plane, f is the focal length of the lenses, λ is the wavelength of the coherent light used.

This filter is applied to

$$Scn(u,v) = T(u,v) \cdot \exp(-2\pi i(x_t \cdot u + y_t \cdot v)) + N(u,v) \tag{3}$$

and generates the filtered Fourier transform in the (VLC) Fourier plane (Vlad et al., 1976)

$$\begin{aligned} Scn'(u,v) = H(u,v) \cdot Scn(u,v) = \\ = Ref^*(u,v) \cdot T(u,v) \cdot \exp[2\pi i(x_r - x_t) \cdot u + 2\pi i(y_r - y_t) \cdot v] \\ + Ref^*(u,v) \cdot N(u,v) \cdot \exp(2\pi i(x_r \cdot u + y_r \cdot v)). \end{aligned} \tag{4}$$

The L2 lens generates the Fourier transform (second in number) of the $Scn'(u,v)$. The result is the optical cross-correlation 2D function or image (Vlad et al., 1976)

$$\begin{aligned} corr(x,y) = ref(x,y) \otimes t(u,v) \cdot \delta[(x_r - x_t),(y_r - y_t)] \\ + ref(u,v) \otimes n(u,v) \cdot \delta(x_r + y_r) \\ = T_1 + T_0. \end{aligned} \tag{5}$$

There are two output cross-correlation terms. The T_0 one is called the DC or zero-order term and represents the "useless" term of cross-correlation between the reference image and the embedding noise. The T_0 one is called the dc-term and represents the cross-correlation between the reference image and the target image. Now, if the target image consists in identical or distorted reference image, then this term is the auto-correlation term.

Pattern recognition (i.e. discrimination or detection) process fails, if the cross-correlation T_0 term presents higher correlation peak than the autocorrelation T_1. The L1 and L2 lenses optical axes alignment is the main disadvantage of the (VLC). If the optical axes of L1 and L2 lenses are not perfectly aligned then the spatial filtering in the (VLC) Fourier plane is missed and the correlation process doesn't generate accurate information. In this case, the pattern recognition process may fail.

Optical axis alignment, as being the major disadvantage of (VLC), generates the occurrence of the amplitude joint transform correlator (AJTC). This correlator is based on the *4f* Vander Lugt correlator, but uses a joint image to alleviate the optical axes alignment (Abookasis et al., 2001; Alam & Karim, 1993a). The joint image $jnt(x,y)$ presents the reference image $ref(x,y)$ and the scene image $scn(x,y)$ placed in two separate half (figure 2). The corresponding mathematical model of an optical (AJTC) consists in the following well known equations, starting with the joint image

$$jnt(x,y) = ref(x, y + y_r) + scn(x, y - y_t). \tag{6}$$

Here $ref(x, y + y_r)$ and $scn(x, y - y_t)$ are the reference image and the scene image (figure 2).

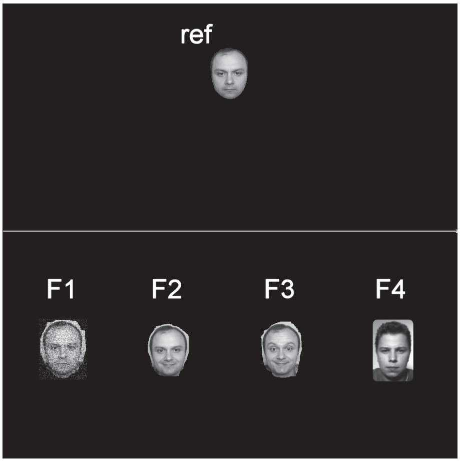


Fig. 2. Joint image: reference image is presented in the upper half and the scene image is presented in the lower half.

The next equations, after the L1 lens optical Fourier transform, are

$$\begin{aligned}JTPS(u,v) &= RPS(u,v) + SPS(u,v) \\ &+Ref(u,v)\cdot Scn^*(u,v)\exp(iv\cdot(y_r - y_t)) \\ &+Scn(u,v)\cdot Ref^*(u,v)\exp(-iv\cdot(y_r - y_t)).\end{aligned} \tag{7}$$

Here, $Ref(u,v)$ and $Scn(u,v)$ are the Fourier transforms of $ref(x,y)$ and $scn(x,y)$, respectively, and $u = \frac{2\pi}{\lambda f}\cdot x$, $v = \frac{2\pi}{\lambda f}\cdot y$ are the spatial frequency coordinates in the Fourier plane, f is the focal length of the lenses, λ is the wavelength of the light used, $JTPS(u,v)$, $SPS(u,v)$, $RPS(u,v)$, are the joint power spectra, scene power spectrum and reference power spectrum, respectively (Alam & Karim, 1993a; Huang, et al, 1997; Javidi & Kuom, 1988; Lu et al., 1997).

The L2 lens Fourier transform, of the $JTPS(u,v)$, generates the correlation result with the (AJTC) in the output correlation plane (Alam & Karim, 1993a; Javidi & Kuom, 1988)

$$\begin{aligned}corr(x,y) &= \left[ref(x,y)\otimes ref(x,y)\cdot\delta(x,y) + scn(x,y)\otimes scn(x,y)\cdot\delta(x,y)\right] \\ &+ref(x,y)\otimes scn(x,y)\cdot\delta[x,y-(y_r - y_t)] \\ &+scn(x,y)\otimes ref(x,y)\cdot\delta[x,y+(y_r - y_t)] \\ &= T_0 + T_1 + T_2.\end{aligned} \tag{8}$$

The correlation result presents the zero-order term, T_0, and two anti-parallel correlation "lines", as T_1 and T_2, with the same information.

Amplitude joint transform correlators (AJTC) are known as robust to embedding and additive noise, thus they are a good solution for the pattern recognition in these conditions. Also, the main disadvantage of the JTCs is that it has a very high intensity zero-order correlation peak (dc term), T_0. This term is useless for the discrimination (e.g. pattern recognition) process.

It was the past issue to resolve the JTCs. To remove the zero-order term and to accomplish the non-zero order JTC (NZAJTC), many methods have been proposed: for instance, phase-shifting technique (Su & Karim, 1998; Su & Karim, 1999), joint transform power spectrum (JTPS) subtraction strategy (Cheni & Wu, 2003; Cheni & Wu, 2005; Lu et al., 1997), Mach-Zehnder method.

The (NZAJTC) based on JTPS subtraction is more time consuming than the others, as a digital operation of non-zero order term subtraction is needed (Cheni & Wu, 2003; Cheni & Wu, 2005; Lu et al., 1997)

$$\begin{aligned} NZJTPS(u,v) &= JTPS(u,v) - RPS(u,v) - SPS(u,v) \\ &= Ref(u,v)\cdot Scn^{*}(u,v)\exp(iv\cdot(y_r - y_t)) \\ &\quad + Scn(u,v)\cdot Ref^{*}(u,v)\exp(-iv\cdot(y_r - y_t)) \end{aligned} \tag{9}$$

where $NZJTPS(u,v)$, $JTPS(u,v)$, $SPS(u,v)$, $RPS(u,v)$, are the non-zero joint power spectrum, joint power spectrum, scene power spectrum and reference power spectrum, respectively.

In order to achieve thin and higher correlation peaks with (NZAJTC), one must make a spatial frequency domain filtering that enhances the high order spatial frequencies which defines the reference object details.

One of the conventional methods to alleviate this issue is the use of amplitude fringe-adjusted joint transform correlator (FA-AJTC). The method applies to $NZJTPS(u,v)$, an amplitude fringe-adjusted filter (FAF) which consists in the inverse amplitude spectrum of the reference Fourier transform, $REF(u,v) = |Ref(u,v)|$, (Abookasis et al., 2001; Alam & Karim, 1993a; Alam & Karim, 1993b; Alam & Horache ,2004)

$$FAF(u,v) = \begin{cases} \dfrac{1}{\left[REF(u,v)\right]}, REF(u,v) > \varepsilon \\ \dfrac{1}{\left[REF(u,v) + Z(u,v)\right]}, REF(u,v) \le \varepsilon \end{cases} \tag{10}$$

where ε is the lowest positive real value that the computer recognizes and $Z(u,v)$ is a real non-zero function used to alleviate the poles of $REF(u,v)$.

The (FAF) creates better light diffraction efficiency because it works like inverse reference adaptive filtering. In other words, use of inverse adaptive filter increases the higher spatial frequencies and decreases the lower spatial frequencies of the $NZJTPS(u,v)$. The higher frequencies are responsible for the "details" of an object-image and the lower frequencies are responsible for "common" properties of the object-image. Thus, increasing the higher spatial frequencies of the $NZJTPS(u,v)$ with a reference inverse adaptive filter, it generates higher autocorrelation peaks and fades the cross-correlation peaks. As a consequence, a better discrimination of the reference in the scene image is achieved.

The final fringe-adjusted non-zero joint power spectrum in the Fourier plane is (Abookasis et al., 2001; Alam & Karim, 1993a; Alam & Karim, 1993b)

$$\begin{aligned}FA-JTPS(u,v) &= NZJTPS(u,v)\cdot FAF(u,v)\\ &= \frac{1}{REF(u,v)}\cdot\begin{bmatrix} Ref(u,v)\cdot Scn^{*}(u,v)\exp(iv\cdot(y_r-y_t))\\ +Scn(u,v)\cdot Ref^{*}(u,v)\exp(-iv\cdot(y_r-y_t))\end{bmatrix}\\ &= \frac{Ref(u,v)\cdot Scn^{*}(u,v)}{REF(u,v)}\cdot\exp(iv\cdot(y_r-y_t))\\ &+\frac{Ref^{*}(u,v)\cdot Scn(u,v)}{REF(u,v)}\cdot\exp(-iv\cdot(y_r-y_t)),\end{aligned}\tag{11}$$

and generates the correlation result

$$\begin{aligned}corr(x,y) &\propto ref(x,y)\otimes scn(x,y)\cdot\delta[x,y+(y_r-y_t)]\\ &+scn(x,y)\otimes ref(x,y)\cdot\delta[x,y-(y_r-y_t)]\\ &=T_1+T_2.\end{aligned}\tag{12}$$

According to this equation, the amplitude fringe-adjusted joint transform correlator (FA-AJTC) generates only the two anti-parallel cross-correlation “lines” ”, T_1 and T_2, with the usefull pattern recognition information (figure 5).

All optics experimental setup of the fringe-adjusted non zero-order amplitude joint transform correlator (FA-AJTC) invokes holographic (i.e. CD media type) or high resolution photographic media. With the help of the laser coherent light, these plates have to be “written” with the joint image in the input plane and the reference fringe-adjusted filter in the Fourier plane.

Alternatively a hybrid (optical-digital) (FA-AJTC) can be built up that is implemented with amplitude spatial modulator (ASLM) and a square law image capture device (CCD camera) (Demoli et al., 1997; Sharp et al., 1999). A coherent laser beam light is projected on the (ASLM1) which is addressed by a computer with the joint image in the input plane. After the Fourier transform with the L1 lens, the joint power spectrum, $JTPS(u,v)$, is captured by a CCD1 camera and stored in the computer. Here, the digital processing of $JTPS(u,v)$ generates the $NZJTPS(u,v)$, and then the application of FAF filter is accomplished. After these digital processing steps the resulting power spectrum is projected in the Fourier plane with a (ASLM2) on lens L2 to perform the inverse Fourier transform. In the output plane a CCD2 camera captures the correlation result as a digital image.

All the above steps describe the *4f* (FA-AJTC) architecture in the hybrid version. The *1/f* (FA-AJTC) architecture in the hybrid version can be done using just one amplitude spatial modulator (ASLM), one CCD camera and just one lens. The correlation peaks are generated by using two passes through this hybrid system.

The only drawback of the hybrid setup of (FA-AJTC) is the time consuming of the digital operations (i.e. digital addressing the (ASLM), digital image capturing with CCD camera and digital computer operations). The benefits of the hybrid experimental setup of (FA-AJTC) are the mathematical perfect filtering and processing of the digital $JTPS(u,v)$ - thus, there is no need of optical lenses axis alignment.

As follows, quantitative analysis should be done. Correlation performance criteria, that can be used to analyse the described joint transform correlators, need the definition of cross-correlation peak intensity, CPI, and the auto-correlation peak intensity, API. The ratio $DEC = API/CPI$ denotes the detection efficiency coefficient (Abookasis et al., 2001; Alam & Karim, 1993a; Alam & Karim, 1993b) and prescribes a pattern recognition failure for values less than the recommended threshold value of 1.2000 (and not 1.0000). Values greater than the threshold conclude in a successful pattern recognition process.

A better understanding of the correlation performances of the (JTC) is accomplished by a short example. Thus, a joint image (figure 2) with the scene image gathering a reference with additive noise (50% random type – indexed F1), two morphologically distorted reference images (indexed F2 and F3) and a non-reference image (with index F4) are considered.

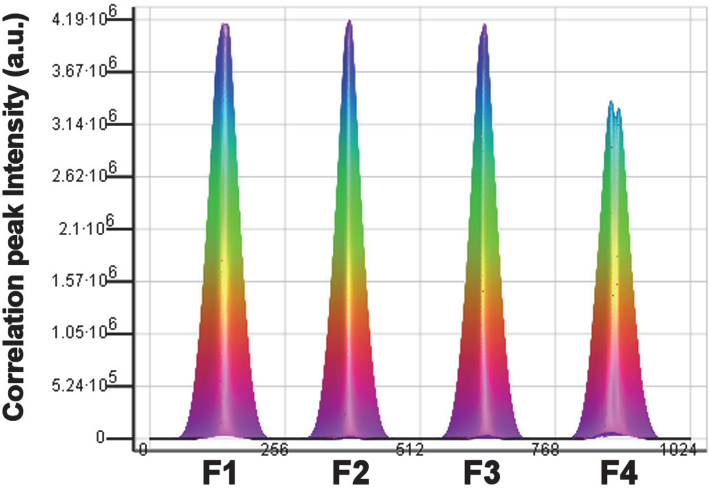


Fig. 3. Correlation performance of (NZAJTC) with the input image from figure 2.

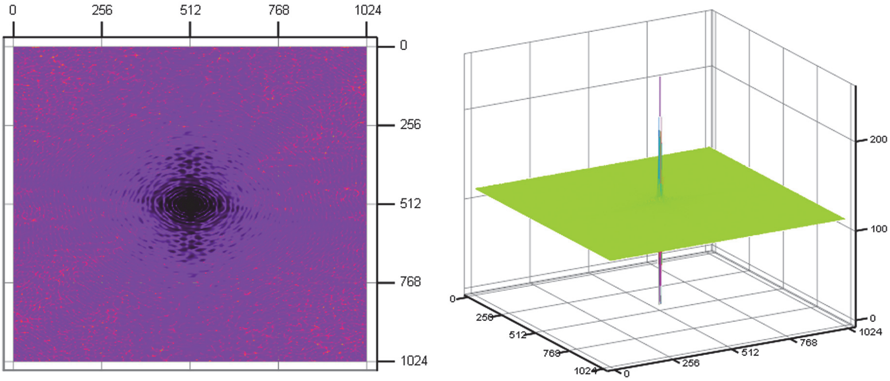


Fig. 4. Fringe-adjusted filter (FAF) (left image) and the corresponding filtered power spectrum, FA-JTPS (right image), in the case of the input image from figure 2.

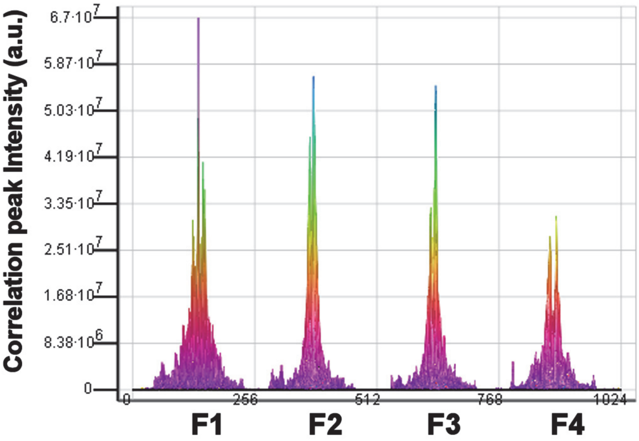


Fig. 5. Correlation performance of (FA-AJTC) with the input image from figure 2.

Correlation results, presented in figures 3 and 5, show the benefits of fringe-adjusted frequency domain filtering (figure 4). After this filtering, the output correlation peaks are sharper and the minimum value of detection efficiency coefficients calculated with CPI for F4 and APIs for F1, F2, F3, for (FA-AJTC) (DEC = 1.7501) is greater than for (NZAJTC) (DEC = 1.1944). There can be noticed that the (NZAJTC) fails to discriminate between the reference-type image, F3, and the non-reference image, F4. The F3 image generates the lowest API from the reference-image class and the DEC value is less sensible than the threshold.

3. Phase-input joint transform correlators (PiJTC)

Amplitude encoding joint transform correlators need (ASLM) with large bandwidth and thus, generate wider correlation peaks and lower light efficiency. In the holographic research, there is stated that the phase encoding needs small bandwidth of the recording media to generate higher light efficiency. These reasons have made the transition from amplitude JTC to phase-input JTC, denoted (PiJTC), worthy.

Since the appearance of phase spatial light modulators (PSLM) there were several approaches to encode the joint image in the phase domain. In early attempts direct phase encoding domain was restricted to $\left[0;\pi\right]$. Nowadays, (PSLM) direct phase encoding domains cover the full range of $\left[0;2\pi\right]$ (Cohn & Liang, 1996; Labastida et al., 1994; Takahashi & Ishii, 2010; Takahashi & Ishii, 2006). Here it must be mentioned that all optics phase-input correlators can be built with holographic approach, but not a versatile one.

The (PiJTC) model is the same as the amplitude one, but it uses the phase transformation of the reference, scene and joint input image. This method, assumes that the amplitude image $AmplitudeImage(x,y)$ is somehow transformed from intensity gray levels (usually from 0 to 255) in phase levels with a domain usually of $dfPSLM \equiv \pi - 0$ or $dfPSLM \equiv 2\pi - 0$. Phase transformation function, $PhT[\cdot]$, invoked to obtain a phase image $PhaseImage(x,y)$, is mathematically described by (Lu & Yu, 1996; Sharp et al., 1998)

$$\begin{aligned} PhaseImage(x,y) &= PhT\left[AmplitudeImage(x,y)\right] \\ &= \exp\left(i \cdot ScalAmplitudeImage(x,y)\right), \end{aligned} \tag{13}$$

$$ScalAmplideImage(x,y) = \left(\frac{AmpliudeImage(x,y) - Min}{Max - Min}\right) \cdot dfPSLM + fPSLM_1 , \tag{14}$$

$$dfPSLM = fPSLM_2 - fPSLM_1 , \tag{15}$$

where $dfPSLM$ is the phase depth, Max, Min, are the maximum and the minimum gray levels of the amplitude encoded image.

Reference, scene and joint phase encoded images are optically Fourier transformed to obtain the reference $RPS(u,v)$, scene $SPS(u,v)$ and joint $JPS(u,v)$ power spectrum in the Fourier plane (Chang & Chen, 2006; Nomura, 1998; Su & Karim, 1998; Su & Karim, 1999). These power spectrums can be processed as in the amplitude non-zero order fringe-adjusted joint transform correlator (FA-AJTC) to provide the correlation peaks in the output plane. The second Fourier transform generated by the L2 lenses uses the amplitude coded $NZJTPS(u,v)$ image projection and does not need just the amplitude one, (ASLM).

These steps describe the non-zero order fringe adjusted phase-input joint transform correlates (FA-PiJTC). The "phase-input" terminology comes from the fact that in the input plane of the (FA-PiJTC) a (PSLM) is used to make a direct phase projection of an amplitude input image.

In optical holography, when the object is a phase encoded one, then the whitened holographic method must be applied in order to get the best contrast and light diffraction efficiency. Similarly, in the (FA-PiJTC) the input direct projection phase encoded image must have a non-black background to achieve the theoretically predicted performances. One solution is to apply a gray level grating background at the (PSLM) pixel level. But this processing is due to the amplitude gray domain range of the input image and the phase encoding domain of the (PSLM), $dfPSLM$. If this step is done manually, then real time applications are not possible. Thus, for speed reasons and better pattern recognition performances, the authors proposed a new (FA-PiJTC) which can automatically adjust the amplitude gray domain range of the input image and the phase encoding domain of the (PSLM) in the same time denoted as (FA-sinePiJTC).

Sine modulated fringe-adjusted non-zero order phase-input joint transform correlator, (FA-sinePiJTC) presents the amplitude domain pre-processing step, other than the scalar transform of the input amplitude image. In this step, as in the previous one, a scalar transformation of the input amplitude is involved. This is required by a pre-processing trigonometric function, chosen by the authors as the *sine* function

$$\begin{aligned} sinePhaseImage(x,y) &= PhT\left[TrigPreProccAmplitudeImage(x,y)\right] \\ &= \left[\sin\left[ScalAmplitudeImage(x,y)\right]\right] \\ &= \left[\exp\left(i \cdot \left[\sin\left[ScalAmplitudeImage(x,y)\right]\right]\right)\right]. \end{aligned} \tag{16}$$

$$TrigPreProccAmplideImage(x,y) = \sin\left[\left(\frac{AmpliudeImage(x,y) - Min}{Max - Min}\right) \cdot dfPRE + fPRE_1\right] \quad (17)$$

$$dfPRE = fPRE_2 - fPRE_1\,, \quad (18)$$

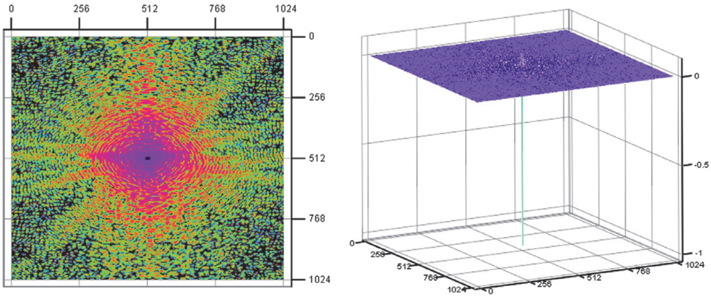


Fig. 6. Fringe-adjusted filter (FAF) (left image) and the corresponding filtered power spectrum, *FA-PiJTPS* (right image).

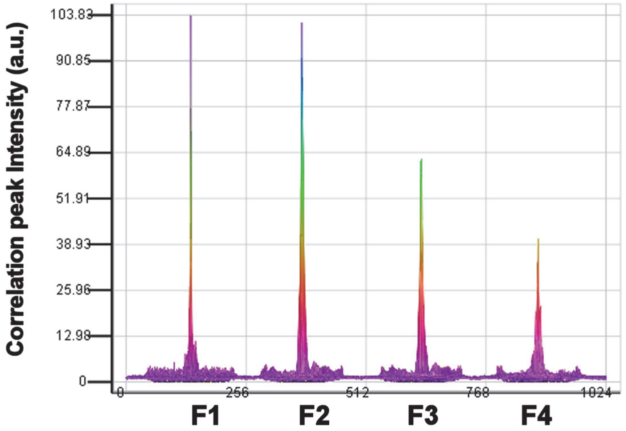


Fig. 7. Correlation performance of FA-PiJTC.

The pre-processing amplitude sine function needs a definition domain within the amplitude image to be scaled in, *dfPRE* . The limits of this definition domain represent the parameters of the (FA-sinePiJTC) and they can be considered as degree of freedom. With these parameters, one can make the adjustments for better correlation performance. One of the aims of this paragraph is to analyse the correlation performance adjustments possibilities with these two parameters.

A better understanding of the correlation performances of the (FA-PiJTC) and (FA-sinePiJTC) is accomplished by a short example which consider the same joint image used previously for the amplitude models.

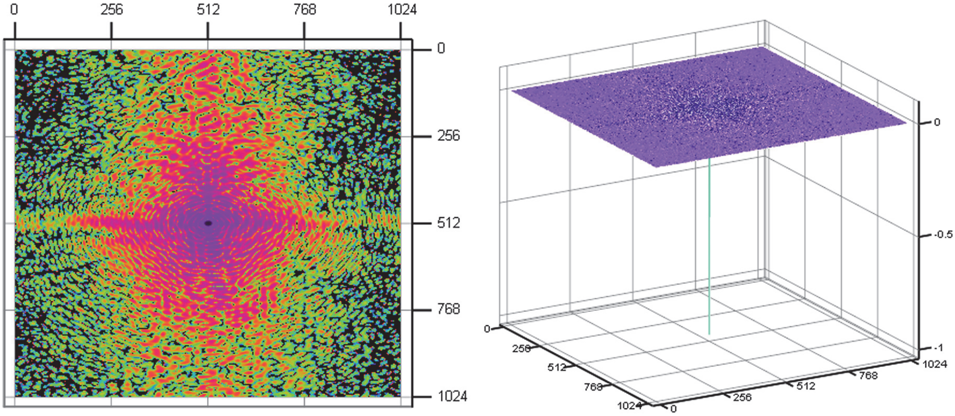


Fig. 8. Fringe-adjusted filter (FAF) (left image) and the corresponding filtered power spectrum, (FA-sinePiJTPS) (right image).

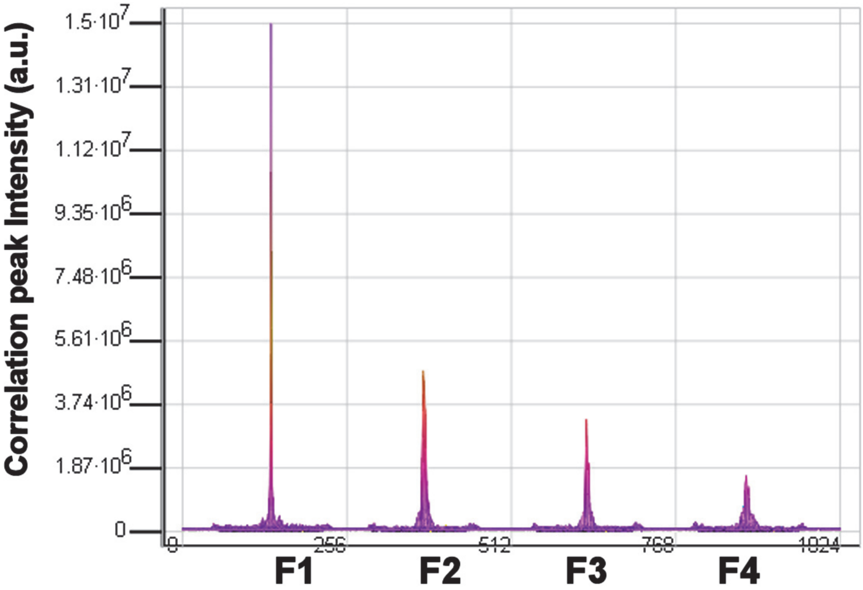


Fig. 9. Correlation performance of (FA-sinePiJTC).

In the same way, from figure 7 and 9, it can be noticed that (FA-PiJTC) has sharper correlation peaks than (FA-AJTC) and the minimum detection efficiency coefficient value is DEC=1.5547, with $\left[0;\ 2\pi\right]$ the PSLM definition domain; the (FA-sinePiJTC) has the sharpest correlation peaks and the best pattern recognition performance DEC=2.0036, with $\left[-0.8\pi;\ 0.8\pi\right]$ the sine definition domain and $\left[0;\ 2\pi\right]$ the (PSLM) definition domain.

These results conclude that (FA-sinePiJTC) can be chosen for database (i.e. inter-class and intra-class) face recognition with and without additive noise (that consists in an image distortion).

4. Experimental and computer simulation aspects of phase-input fringe-adjustment joint transform correlators (PiJTC)

Previous paragraphs present the mathematical models of (FA-AJTC) and (FA-sinePiJTC) and the experimental setup for (FA-AJTC) in all optics and hybrid versions.

The following presentation emphasizes the experimental hybrid setup, the computer simulation conditions and results for (FA-sinePiJTC) model. In the *4f* architecture the (FA-sinePiJTC) can be done more accurate. The reason is that in the first stage the joint image projection is done in the phase domain with a (PSLM) and in the second stage the $FA-JTPS(u,v)$ is projected in the amplitude domain with a (ASLM). Thus for a phase-input joint transform correlator, two spatial light modulators are imminently needed: one in phase domain and one in the amplitude domain. This is the fact that generates the synonym "phase-amplitude" architecture correlator for the *4f* phase-input joint transform correlator. From this point of view, some researchers proved that "phase-phase" architecture does not bring any improvements to pattern recognition performances in comparison with "amplitude-amplitude" one.

The drawbacks and benefits of the *4f* hybrid architecture of (FA-AJTC) model are the same as for the (FA-sinePiJTC) model. The (FA-sinePiJTC) model has a degree of freedom, the sine modulation domain limits, that can adjust the performances of the correlation process.

The computer simulations were done on a face database with the sine modulated fringe-adjusted phase-input joint transform correlator, (FA-sinePiJTC). The (FA-sinePiJTC) model has a degree of freedom, the sine modulation domain limits, that can adjust the performances of the correlation process. In the computer simulations two sine modulation domains: $[-0.5\pi;\ 0.5\pi]$ and $[-0.75\pi;\ 0.75\pi]$ are used.

The face databases used are two parts from the "General Purpose Recognition and Face Recognition" section from Computer Vision Research Projects supervised by Dr. Libor Spacek from Department of Computer Science University of Essex, Colchester, UK (*http://cswww.essex.ac.uk/mv/allfaces/index.html*). As it is described by the database owner, the face images are held in four directories (Faces94, Faces95, Faces96, Grimace), in order of increasing the difficulty degree. Faces96 and Grimace are the most difficult, for two different reasons - variation of background and scale, versus extreme variation of expressions.

The presented computer simulations use the "male" and "malestaff" directories from Faces94 (updated Friday, 16-Feb-2007 15:52:52 GMT). The subjects sit at fixed distance from the camera and are asked to speak, whilst a sequence of images is taken. The speech is used to introduce facial expression variation (see table 1).

Joint images, built up to perform the correlation process, have 512x512 pixels. The reference image is located in the upper half and two target images are located in the lower half composing the scene image.

	“malestaff”	“male”
Total number of individuals	21	102
Number of images per individual	20	20
Total number of images	420	2040
Gender	contains images of male subjects	
Race	contains images of people of various racial origins	
Age	Range older individuals are also present; the images are mainly of first year undergraduate students, so the majority of individuals are between 18-20 years old	
Glasses	Yes	
Beards	Yes	
Head Scale	none	
Head turn, tilt and slant	significant variation in these attributes	
Position of face in image	minor changes	
Expression variation	considerable expression changes	
Additional comment	there is no individual hairstyle variation as the images were taken in a single session	
Backgrounds	is plain green	
Lighting	artificial, mixture of tungsten and fluorescent overhead – no variation	
Image format	180x200 (WxH pixels); 24bit colour JPEG	
Camera used	S-VHS camcorder	

Table 1. Face image database description.

5. Correlation performances of phase-input fringe-adjustment joint transform correlator (FA-PiJTC) in face recognition with additive noise

Computer simulations of correlation process over a database can be very time consuming. In order to avoid time loss, a step by step strategy is chosen. In this strategy a small database is tested and after that, due to the results, the decisions are applied to a larger database in order to draw final conclusions. The smaller database from directory ”malestaff” the larger the one from directory “male”.

The test face database used is organized in individual classes Figure 10. One individual class has a total of 20 images of the same person (i.e. individual). For the test database there are selected only 3 face images (figure 11) from each class of the 21 classes (figure 12).

Five additive noise conditions - with 10%, 20%, 30%, 40%, 50% random noise levels - are presented in figure 11. References, just the face parts, were subtracted from the original images to prevent the background false correlation. CIELab colour filter was used to extract the predominant green background.

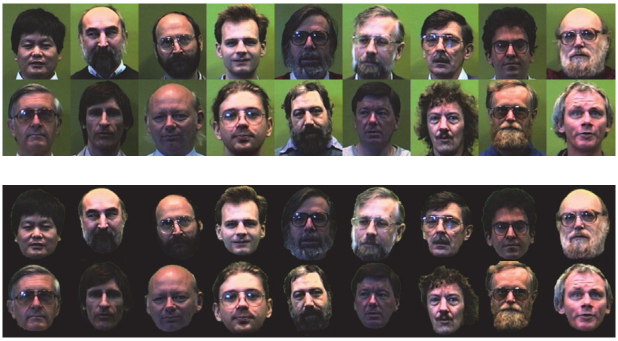


Fig. 10. A selection of test database original face images (upper line); reference images filtered from the original ones (lower line).

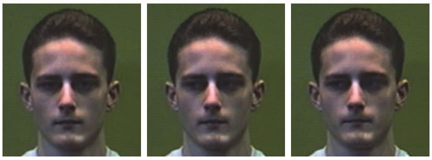


Fig. 11. Example of one individual class (3 face images) from the test database.

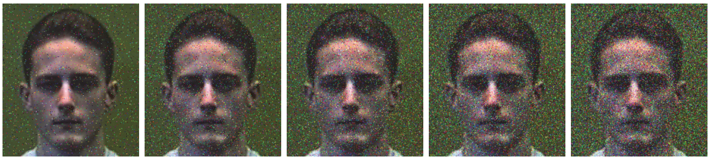


Fig. 12. Five additive noise conditions of a face image (target images as reference-type images), with 10%, 20%, 30%, 40%, 50% random noise levels, respectively.

Correlation performance analyse over the face database without additive noise should be done. If the pattern recognition process fails or has poor detection efficiency for the situation without additive noise, then the additive noise condition will collapse the process. At the start, autocorrelations over all class reference images were done: 21 classes x ((20x20) reference images/class). For each class only the overall (20x20=400 values) minimum value represented with diamond line - figure 13 as intra-class correlation peak intensity (ICPI) was retained.

Large domain values of ICPI can be noticed from figure 13. The main reason is that the face images were captured during a monologue. At the start of session, in the first images, the individuals take care about their posing. After a few seconds they have head motions that generate significant variations in the face images. As a consequence, the ICPI decreases very fast for 8 individual classes. In order to analyze the additive noise correlator performances, as a counter measure, the authors decide to drop the number of reference images for the test database from 20 to only 3 and the target images to just one (the first in the class).

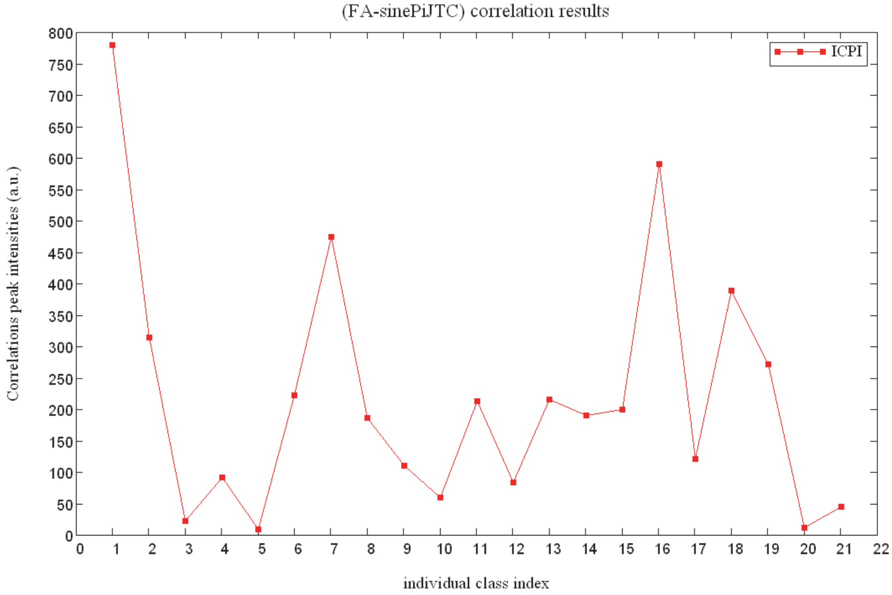


Fig. 13. Autocorrelation results for all the classes reference images of the test database.

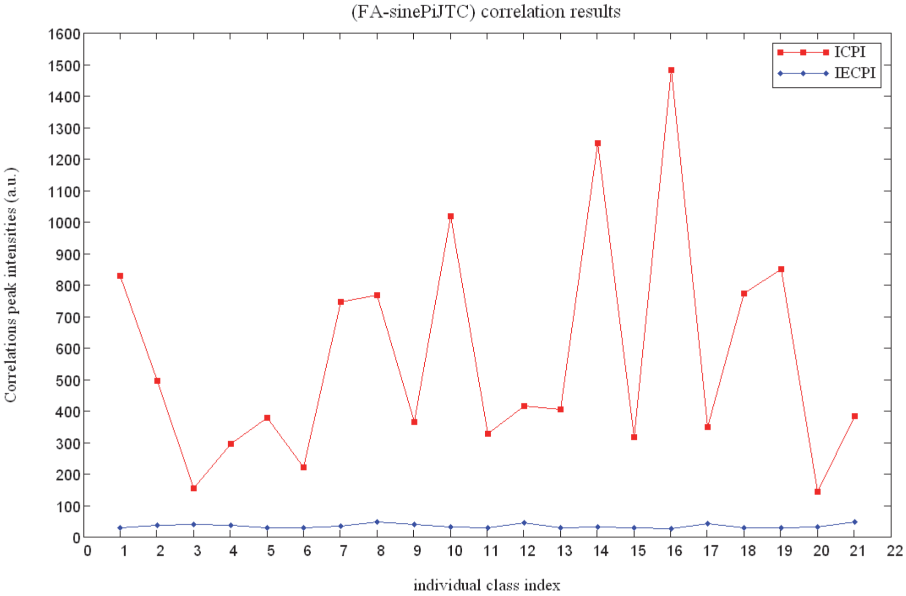


Fig. 14. Correlation performances of FA-sinePiJTC for interclass (diamond line) and intra-class (boxed line) cases from face database without additive noise.

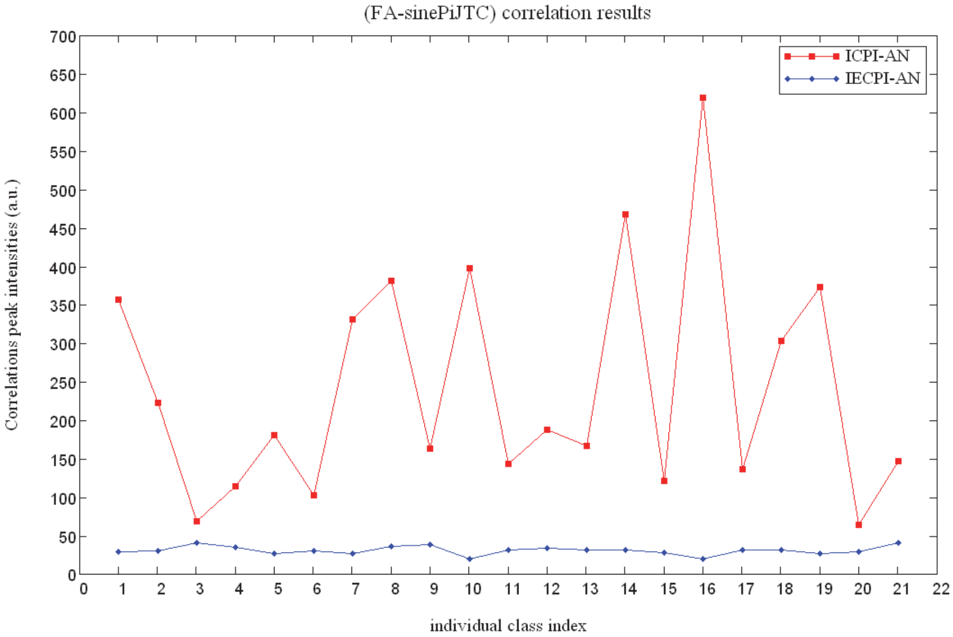


Fig. 15. Correlation performances of FA-sinePiJTC for interclass (diamond line) and intra-class (boxed line) cases from face database with additive noise.

Autocorrelations between 63 reference images (21 classes x 3 reference images) were done. The diagonal (3x3) block from the cross-correlation matrix represents the intra-class results (ICPI). For each class was retained only the overall (3x3=9 values) minimum value represented with boxed line - figure 12 - as intra-class correlation peak intensity line (ICPI). The rest of the values consist of the inter-class cross-correlations values. For each class, only the maximum values of the inter-class correlation values were retained, as the inter-class correlation peak intensity (IECPI) – represented in figure 14 with diamond line.

In the same manner, the results for cross-correlations between 63 reference images (21 classes x 3 reference images) and 105 additive noisy images (21 classes x (1 original image x 5 noise conditions)) were generated. Graphical results are presented in figure 15 as the ICPI-AN and IECPI-AN (AN from Additive Noise) correlations peak intensity lines.

Figure 14 shows that the ICPI line does not intersect with IECPI line; figure 15 shows that the ICPI-AN line does not intersect with IECPI-AN line. If the lines intersect then there is an image from an individual class that is similar with an image from different class. But, these facts prove that the (FA-sinePiJTC) is robust to random additive noise.

Figure 14/15 present gaps between the ICPI/ICPI-AN values and IECPI/IECPI-AN values along all classes. The significance of these gaps is a very good discrimination (e.g. pattern recognition) performance of the (FA-sinePiJTC). Thus, the (FA-sinePiJTC) can discriminate with high accuracy over all classes (figure 16), between faces of different classes and do the registration of the faces from the same class successfully. The accuracy (gap magnitude) decreases in additive noise conditions (figure 16) at 22.95 (a.u.), from 96.05 (a.u in the situation without additive noise.

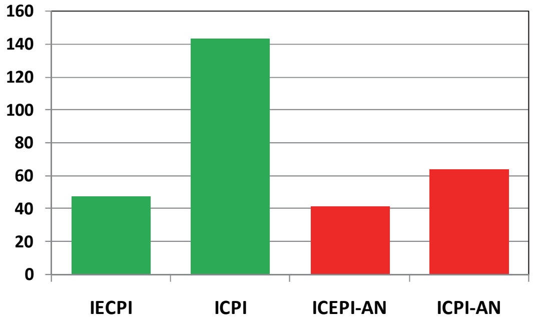


Fig. 16. Overall correlation performances comparison for FA-sinePiJTC on face database without and with additive noise.

The algorithm revealed by correlation results with the (FA-sinePiJTC) over the test face image database, can be now applied on a larger face images database. First, the (21 classes x (20x20) reference images) autocorrelations are done. The results presented in figure 17 show small values of ICPI for 10% of the classes. As consequence, only 5 reference images per individual class were considered.

With these conditions, the autocorrelations between 510 reference images (102 classes x 5 reference images) were done. The results are presented in figure 18 as the ICPI and IECPI correlations peak intensities lines.

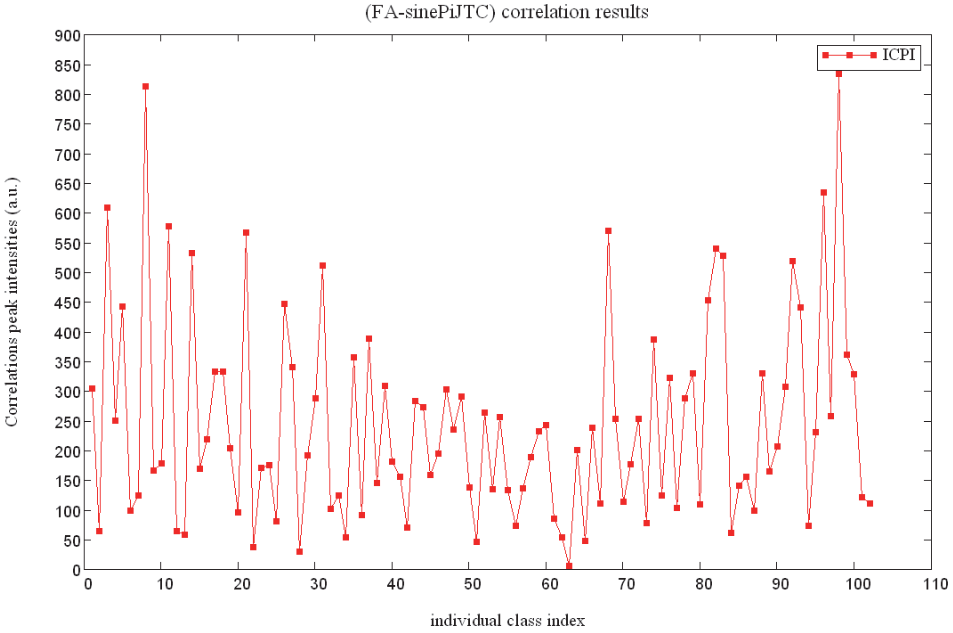


Fig. 17. Autocorrlation results for the all classes reference images of the large database.

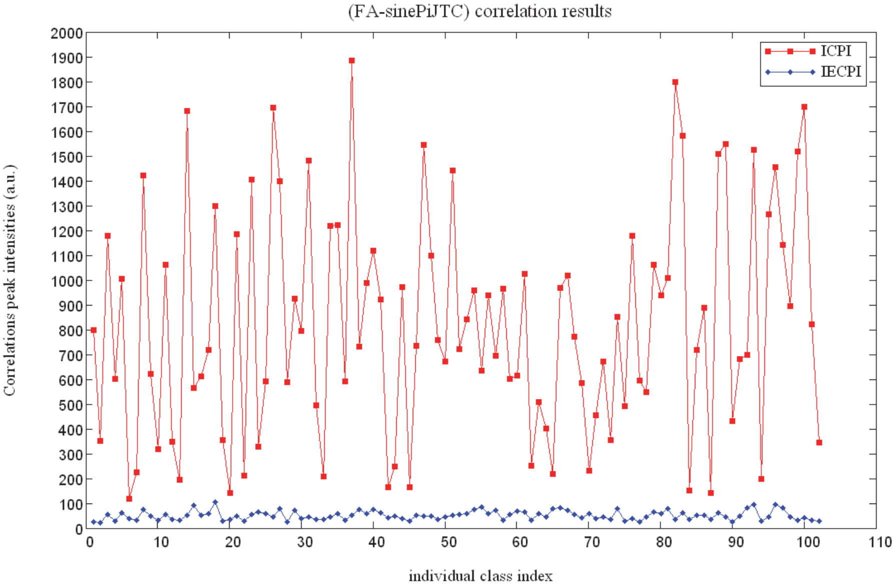


Fig. 18. Correlation performances of FA-sinePiJTC for interclass (diamond line) and intra-class (boxed line) cases from face database without additive noise.

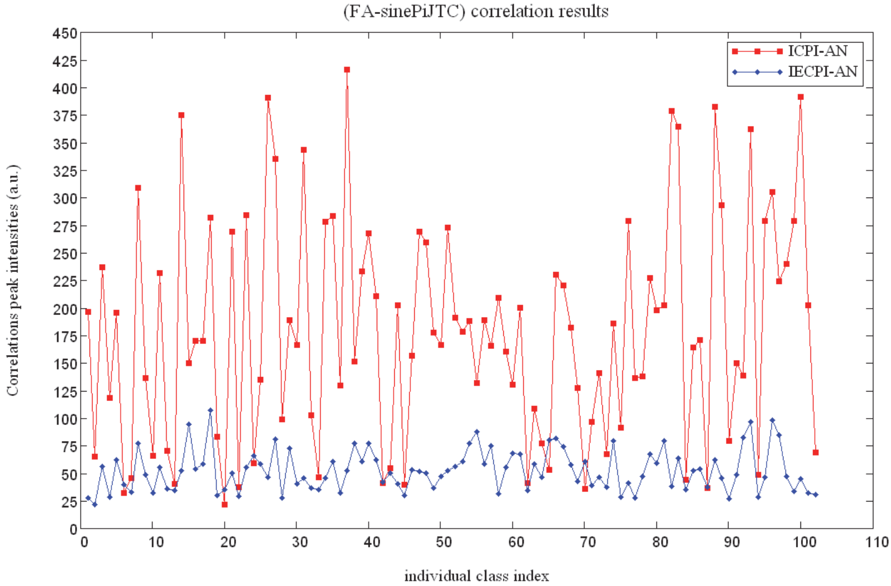


Fig. 19. Correlation performances of (FA-sinePiJTC) for interclass (diamond line) and intra-class (boxed line) cases from large face database with additive noise.

Correlation lines, ICPI and IECPI, in figure 18 do not intersect, but the gap between them is very small. As mentioned before, the main reason is the significant variation of the faces due to the motion during the monologue.

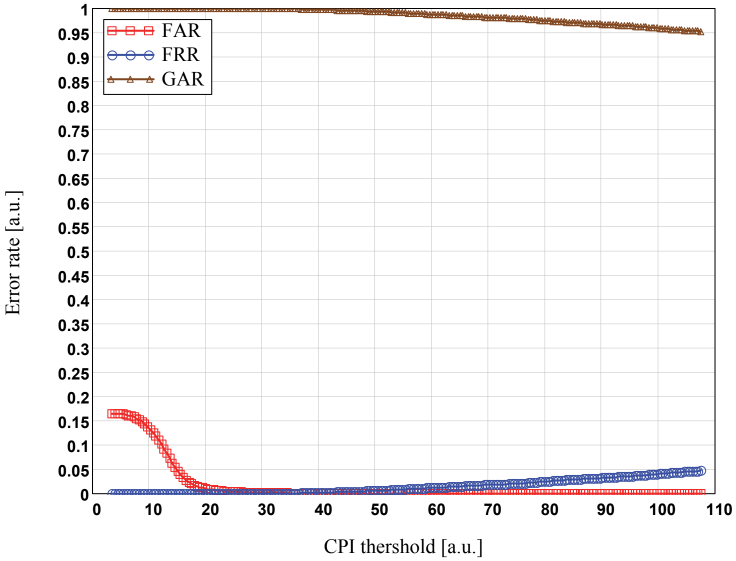


Fig. 20. Error rates curves.

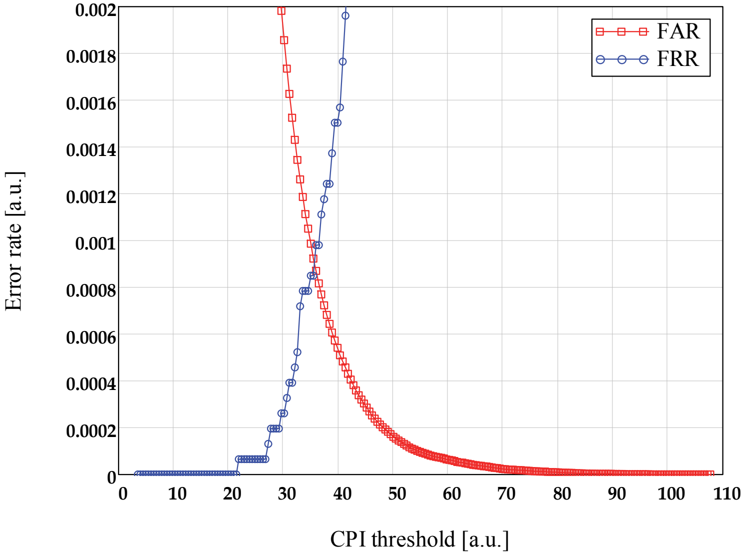


Fig. 21. Error rates curves – emphasize part in the EER point neighbor.

In analogue conditions, the cross-correlations between 510 reference images (102 classes x 5 reference images) and 2550 additive noisy images (102 classes x (5 original image x 5 noise conditions)) were done. Graphical results are presented in figure 19 as ICPI-AN and IECPI-AN correlations peak intensity lines. Correlation lines ICPI-AN and IECPI-AN, intersect because no gap between them is present. This fact conclude that some individuals appear to be similar using the (FA-sinePiJTC) in the mentioned conditions.

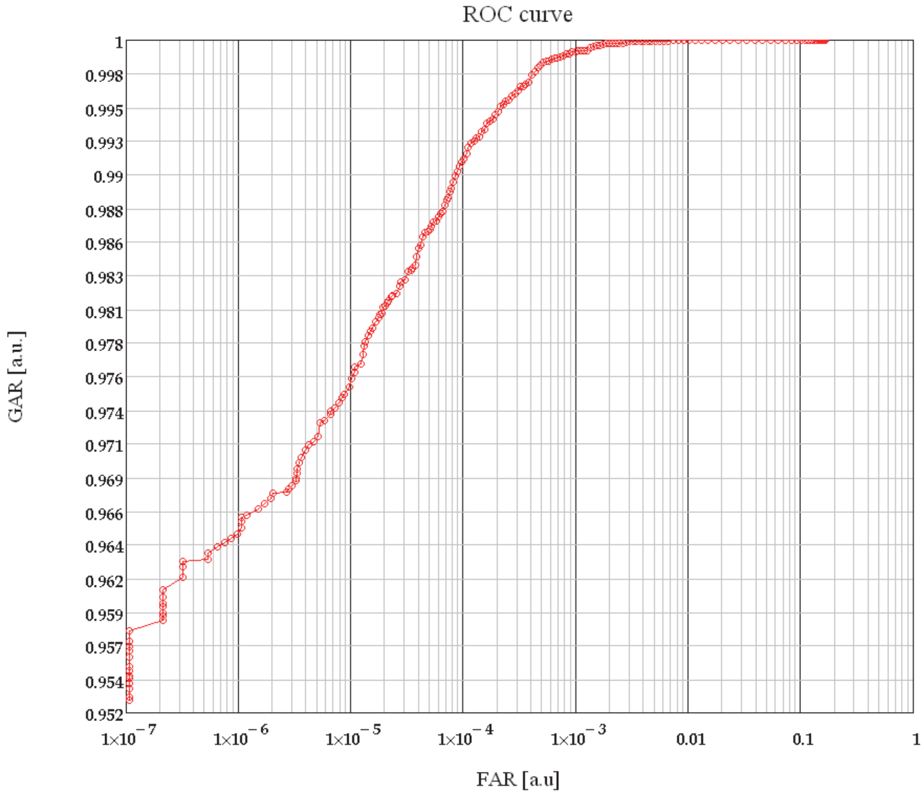


Fig. 22. Receiver operating characteristic (ROC) curve.

The pattern recognition performances in this case are measured by equal error rate (EER), genuine acceptance rate (GAR), false acceptance rate (FAR) and false rejection rate (FRR) (figures 20-22). Also there must be built up the receiver operating characteristic (ROC) curve (figure 22).

6. Conclusion

This chapter presents additive noise robustness of sine modulated fringe-adjusted phase-input joint transform correlator (FA-sinePiJTC), tested over a human face images database. There were inspected two databases: a small – test database for preliminary results, and then a large database for the final results and conclusions. The architecture for the proposed

JTC is the *4f*, with non-zero order algorithm and a fringe-adjusted filter, in order to achieve better pattern recognition efficiency.

The sine modulation function gives the correlation process a degree of freedom through the limits of the modulation domain, $dfPRE = fPRE_2 - fPRE_1$. These parameters can adjust the pattern recognition performance of the correlation process due to the variations of faces in the captured images. These variations are in plane and in axis head rotations, known as very altering for pattern discrimination performances. To ensure better performances the correlation processes were made with $dfPRE = [-0.5\pi;0.5\pi]$, for the test database, and with $dfPRE = [-0.75\pi;0.75\pi]$ for the large database. There has to be pointed out that the conjunction of: the sine modulation function, the phase-input and the fringe-adjusted filter, provides best pattern recognition performances over the inspected face databases.

Presented results of the proposed joint transform correlator algorithm were achieved with 512x512 pixels input images. The joint image has two target images in the scene section, so it works twice faster than the Vander Lugt correlator (VLC), or other single pair matching correlators. Computer simulations were done on Intel i7-870 processor at 2.93 MHz with 3 GB of RAM with an operation speed of 0.185 seconds/single reference-target correlation. The application was built up to be used as a single process (CPU thread) or as multiple processes run simultaneously in maximum 5 CPU threads. The limited number of the CPU threads used is due to RAM memory and not to the processor. In these conditions, the single reference-target correlation time drops to only 37 milliseconds and make possible the achievement of 1,560,600 single pair correlations from the large database. In this way, the statistical significance of the face recognition results is very high.

The (FA-sinePiJTC) face recognition for the test database reveals that choosing only 3 out of 20 face images at the start of capturing process provides zero ERR and FAR error rates. This is stated because of the gap between the maximum IECPI and the minimum ICPI over the entire database process, without and with random additive noise. We can conclude that the (FA-sinePiJTC) is robust to additive noise up to 50% for database face recognition.

The selection of more than 3 initial face images, to 5 face images in the individual classes, widens the variations of faces. In this situation non-zero error rates. The equal error rate appear value is EER=0.0886% for CPI value of 35.4007, and the genuine acceptance rate is GAR=99.9020% at a false acceptance rate of FAR=0.0869%. The error rates values are very low revealing high accuracy and high robustness of the (FA-sinePiJTC) at most 50% additive random noise in face recognition.

Low resolution used (200x180 pixels) in the above performance analysis validate video security systems applications.

Authors experience with the same (FA-sinePiJTC) confirm the good pattern recognition results in correlation applications involving human fingerprint and irises, mostly based on the sine modulation function, the phase-input and the fringe-adjusted filter, previously emphasized.

Future work can involve 3D face recognition systems based on (FA-sinePiJTC) correlation algorithm.

7. References

Abookasis, D.; Arazi, O.; Rosen J. & Javidi, B. (2001). Security optical systems based on a joint transform correlator with significant output images, Opt. Eng., 40(8), pp.1584–1589 (August 2001), ISSN 0091-3286/2001

Alam, M. S. & Horache, E. H. (2004). Optoelectronic implementation of fringe-adjusted joint transforms correlator, Journal of Optics Communications, Vol. 236, No.1, (January 2004), pp. 59–67, ISSN 0030-4018

Alam, M.S. & Karim, M.A. (1993a). Fringe-adjusted joint transform correlation, Journal of Applied Optics, Vol. 32, No. 23, (August 1993) pp. 4344 - 4350, ISSN 1559-128X (print), 2155-3165 (on-line)

Alam, M.S. & Karim, M.A. (1993b). Joint-trasform correlation under varying illumination, Journal of Applied Optics, Vol. 32, No.23, (August 1993), pp. 4351, ISSN 1559-128X (print), 2155-3165 (on-line)

Born, M. & Wolf, E. (1970). Principles of Optics, Pergamon Press, ISBN 0 521 642221, New York, USA

Chang, H.T. & Chen, C.C. (2006). Fully-phase asymmetric-image verification system based on joint transform correlator, Optics Express, Vol. 14, No. 4 (20 February 2006), pp. 1458-1467, ISSN 1094-4087

Cheni, C. & Wu, C.S. (2003). Polychromatic pattern recognition using the non-zero-order joint transform correlator with cross-correlation peak optimization, Journal of Modern Optics, Vol. 50, No. 9, (2003), pp. 1353-1364, ISSN 095M340 print/lSSN 1362-3044 online

Cheni, C. & Wu, C.S. (2005). Modification of training images for the optimisation of the nonzero order joint transform correlator, Journal of Modern Optics, Vol. 52, No. 1, (2005), pp. 21-32, ISSN 095M340 print/lSSN 1362-3044 online

Cohn, R.W. and Liang, M. (1996). Pseudorandom phase–only encoding of real–time spatial light modulators, Journal of Applied Optics, Vol. 35, No. 14, (May 1996), pp. 2488 – 2498, ISSN 1559-128X (print), 2155-3165 (on-line)

Demoli, N.; Dajms, U.; Gruber, H. & Wernicke, G. (1997). Influence of flatness distorsion on the output of a liquid – crystal – television – based joint transform correlator system, Journal of Applied Optics, Vol. 36, No. 32, (November 1997), pp. 8417 – 8426, ISSN 1559-128X (print), 2155-3165 (on-line)

Huang, X.; Lai, H. & Gao, Z. (1997). Multiple–target detection with use of a modified amplitude–modulated joint transform correlator, Journal of Applied Optics, Vol. 36, No. 35, (December 1997), pp. 9198 – 9204, ISSN 1559-128X (print), 2155-3165 (on-line)

Javidi, B. & Kuo, C.J. (1988). Joint transform image correlation using a binary spatial light modulator at the Fourier plane, Journal of Applied Optics, Vol. 27, No. 4, (February 1988), pp. 663 – 665, ISSN 1559-128X (print), 2155-3165 (on-line)

Javidi, B. & Tang, Q. (1994). Chirp–encoded joint transform correlators with a single input plane, Journal of Applied Optics, Vol. 33, No. 2, (January 1994), pp. 227 – 230, ISSN 1559-128X (print), 2155-3165 (on-line)

Jutamulia, S. (1994). Phase–only Fourier transform of an optical transparency, Journal of Applied Optics, Vol. 33, No. 2, (January 1994), pp. 280 – 282, ISSN 1559-128X (print), 2155-3165 (on-line)

Labastida, I.; Carnicer, A.; Martin-Badosa, E.; Santiago, V. and Ignacio, J. (2000). Optical correlation by use of partial phase-only modulation with VGA liquid –crystal displays, Journal of Apllied Optics, Vol.39, No.5, (February 2000), pp. 766-769, ISSN 1559-128X (print), 2155-3165 (on-line)

Lu, G. & Yu, F.T.S. (1996). Performance of a phase-transformed input joint transform correlator, Journal of Applied Optics, Vol.35, No.2, (January 1996), pp. 304-313, ISSN 1559-128X (print), 2155-3165 (on-line)

Lu, G.; Zhang, Z.; Wu, S. & Yu, F.T.S (1997). Implementation of a non-zero-order joint-transform correlator by use of phase-shifting techniques, Journal of Applied Optics, Vol.36, No.2, (January 1997), pp. 470-483, ISSN 1559-128X (print), 2155-3165 (on-line)

Nomura, T. (1998). Phase-encoded joint transform correlator to reduce the influence of extraneous signals, Journal of Applied Optics, Vol.37, No.17, (June 1998), pp. 3651-3655, ISSN 1559-128X (print), 2155-3165 (on-line)

Rosen, J. (1998). Three – dimensional joint transform correlator, Journal of Applied Optics, Vol. 37, No. 32, (November 1998), pp. 7538 – 7544, ISSN 1559-128X (print), 2155-3165 (on-line)

Sharp, J.H.; Budgett, D.M.; Slack, T.G. and Scott, B.F. (1998). Compact phase–conjugating correlator: simulation and experimental analysis, Journal of Applied Optics, Vol. 37, No. 20, (July 1998), pp. 4380 – 4388, ISSN 1559-128X (print), 2155-3165 (on-line)

Sharp, J.K.; Mackay, N.E.; Tang, P.C.; Watson, I.A.; Scott, B.F.; Budgett, D.M.; Chatwin, C.R.; Young, R.C.D.; Touda, S.; Huignard, J.P.; Slack, T.G.; Collings, N.; Pourzand, A.R.; Duelli, M.; Grattarola, A. & Braccini, C. (1999). Experimental systems implementation of a hybrid optical – digital correlator, Journal of Applied Optics, Vol. 38, No. 29, (October 1999), pp. 6116 – 6127, ISSN 1559-128X (print), 2155-3165 (on-line)

Su, H.J. & Karim, M.A. (1998). Performance improvement of a phase-shifting joint transform correlator by use of phase-iterative techniques, Journal of Applied Optics, Vol.37, No.17, (June 1998), pp. 3639-3642, ISSN 1559-128X (print), 2155-3165 (on-line)

Su, J.H.; & Karim, M.A. (1999). Phase–shifting joint transform correlation with phase – iterative algorithm: effect of the dynamic range limit, Journal of Applied Optics, Vol. 38, No. 26, (December 1999), pp. 5556 – 5559, ISSN 1559-128X (print), 2155-3165 (on-line)

Takahashi, T & ISHII, Y. (2006). Laser-Diode Phase-Shifting Joint-Transform Correlator with Multiple-Object Recognition, Optical Review, Vol. 13, No. 2 (2006), pp.53–63, ISSN 1340-6000 (printed version), ISSN 1349-9432 (electronic version)

Takahashi, T & ISHII, Y. (2010). All Optical Phase-Only Filtering Correlation with Binarized Inputs by a Ferroelectric Liquid-Crystal Spatial Light Modulator, Optical Review, Vol. 17, No. 3 (2010), pp.195–203, ISSN 1340-6000 (printed version), ISSN 1349-9432 (electronic version)

Vlad,V.I.; Zaciu, R.; Miron, N.; Maurer, J. & Sporea D. (1976). Prelucrarea optică a informației-Optical Information Processing (in romanian language), Editura Academiei Republicii Socialiste Romania, Bucuresti, Romania

Willett, P.; Javidi, B. & Lops, M. (1998). Analysis of image detection based on Fourier plane nonlinear filtering in a joint transform correlator, Journal of Applied Optics, Vol. 37, No. 8, (March 1998), pp. 1329 – 1341, ISSN 1559-128X (print), 2155-3165 (on-line)

Zhang, S. & Karim, M.A. (1999). Illumination-invariant pattern recognition with joint-transform-correlator-based-morphological correlation, Journal of Applied Optics, Vol.38, No.35, (Decembre 1999), pp. 7228-7237, ISSN 1559-128X (print), 2155-3165 (on-line)

Zhong, S.; Jiuxing, J.; Liu, S. & Li, C. (1997). Binary joint transform correlator based on differential processing of the joint transform power spectrum, Journal of Applied Optics, Vol.36, No.8, (March 1997), pp. 1776-1780, ISSN 1559-128X (print), 2155-3165 (on-line)

12

Robust Face Detection through Eyes Localization using Dynamic Time Warping Algorithm

Somaya Adwan
Factuly of engineering, department of electrical engineering, University of Malaya
Kuala Lumpur
Malaysia

1. Introduction

Detection of faces and facial patterns in static or video images is an important but challenging problem in computer vision, as faces may present in different scales, orientations, positions and poses in an uncontrolled background. To date, numerous approaches have been implemented for face and facial pattern detection [1-4]. These approaches have been grouped into four broad categories [2]. These are knowledge-based, feature based, appearance based and template based methods.

In this chapter, we present a simple yet robust algorithm for template matching method for face detection through eyes localization by using Dynamic Time Warping (DTW) algorithm with different poses and variations. By using DTW, we overcome some of the main limitations of Classical template matching.

We elaborate through discussion and experiments' results that how our image processing strategy applied to DTW algorithm effectively increase the performance and detection rate. Using such an approach does not require translation of human knowledge about the face and facial patterns to computer representation, training data set and training mechanisms or face geometry, and does not require large numbers of model templates to handle different pose and variation changes shown by facial features. In the coming section, short study of Dynamic Time Warping algorithm (DTW) is presented. Proceeding to the next section we discuss the necessary steps for our image processing strategy. This is followed by presenting our approach of image partitioning, and in the next section we present dynamic time warping algorithm application in our work. Towards the end of this paper, results and discussion are put forward and conclusion is presented.

2. Visiting Dynamic Time Warping review

Dynamic Time Warping (DTW) is a fast and efficient algorithm for measuring similarity between two sequences. It is used to find the optimal alignment between two time series, if one time series may be warped non-linearly along its time axis. This warping between the two time series can be used to determine the similarity which may vary in time or speed. Similarity is measured by aligning two sequences and computing a distance between them

[5]. This technique has made it possible to make a matching between two curves that are "subject not only to alteration by the usual additive random error but also to variations in speed from one portion to another" [5]. Dynamic Time Warping (DTW) was initially introduced to recognize spoken words [7]. Since then it has been used and proved useful in different applications such as: handwriting recognition [6, 8], signature verification [9-12], fingers print verification [13], face recognition [14, 15], speech and gesture recognition [16], protein structure [17], gene expression [18, 19], chromosome classification [20], object detection and classification [21, 22], curve matching [23, 24] control system [25], image and shape matching [24,26] database and data mining [27- 29, 31, 32, 34-37), brain activity classification [38], voice control [39] and so on.

The DTW algorithm uses a dynamic programming technique to align time series with a given template so that a total distance measure is minimized [25]. To align these sequences (the reference/template vector X of length M and the test vector Y of length N), a local cost matrix is defined where each cell (i,j) represents the pair wise distances between the i^{th} component of X and the corresponding component j^{th} of vector Y. Generally the Euclidean distance is suggested as a distance cost function in DTW algorithm to build a local cost matrix [30], the cost (distance) function has a small value when sequences are similar and large value if they are different. After computing the local cost matrix between two 1-dimenasional features' vectors X and Y, a warping path is constructed by building a cumulative/global cost matrix D(i,j) [6,7]. The cumulative distances matrix D(i,j) is defined as the sum of the local distance d(i,j) found in the current cell and the minimum of the cumulative distances of the adjacent cells. A warping path is a concatenation of cells starting from (1,1) to the cell (M,N). The objective is to search an optimal path for which a least cost is associated. This path provides the low cost areas between the two feature-vectors (see section 3.5 for more details of building the local and global matrices).

In DTW, contsriants are applied for constructing optimal warping path [7, 27]. Constraints are widely used to speed up DTW [7, 27, 30, 33] by reducing the number of paths to be considered during the computation. "These constraints serve to reduce the search space- the space of possible warping paths. Searching through all possible warping paths is computationally expensive. Therefore, out of concern for efficiency, "it is important to restrict the space of possible warping paths" [27]. Several well-known constraints have been applied to the problem to restrict the moves that can be made from any point in the path. Sakoe and Chiba applied a pattern matching algorithm with a nonlinear time-normalization effect using dynamic programming to spoken word recognition [7]. They introduced several kinds of reasonable constraints as outlined below [27]:

- Monotonicity: The alignment path doesn't go back in time index, $i_{k-1} \leq i_k$ and $j_{k-1} \leq j_k$.

This guarantees that features are not repeated in the alignment.

- Continuity: The alignment doesn't jump in time index, i_k - $i_{k-1} \leq 1$ and j_{k-} $j_{k-1} \leq 1$. This guarantees that important features are not omitted.
- Boundary: The alignment starts at the top-left and ends at the bottom-right. This guarantees that the sequences are not considered only partially.
- Warping window: A good alignment path is unlikely to wander too far from the diagonal. This guarantees that the alignment doesn't try to skip different features or get stuck at similar features, and allowable points must fall within a given window, $|i_k - i_{k-1}| \leq w$; where w is a positive integer window width (threshold).

A new restriction called a slope constraint was introduced in their work and it was claimed that symmetric form and slope constraint are effective, which resulted in optimum performance in comparison to several dynamic programming algorithms [7]. Slope constraint was also used later by Wang et al. [23], to prevent unrealistic warping.

Berndt and Clifford [27] applied the constraints introduced in Sakoe and Chiba work [7] using symmetric form to find patterns in time series for archiving purpose. They used tracing backward technique to find the optimal warping path. Daniel et al. [33] proposed the threshold DTW (TDTW) algorithm based on the DTW technique, for two dimensional spatial activity recognition, by introducing a user defined threshold to the diagonal matching condition of the DTW. Their work showed that the TDTW is less affected to noise and they accomplished higher classification accuracy than DTW. David Clifford et al. [36] used DTW for aligning chromatogram signals. A variable penalty was introduced into the DTW that was added to the distance metric to reduce the number of non-diagonal paths. The percentage of non-diagonal moves taken during the usual DTW ranges from 52% to 70%. After adding the variable penalty to DTW, This range was reduced and it ranges from 1% to 5%. It was shown in the experiments that penalized DTW is highly significant and achieves good alignment of peaks in chromatograms.

These constraints are best visualized in [27, 23, 40]. Using constraints helps speed up DTW execution time by a constant factor but the DTW algorithm time complexity remains $O(N^2V)$ [40].

Some of the principal advantages of using Dynamic Time Warping include its capability to handle different scale and translation, its ability to compare two sequences of signals with different elongations. It does not require complex mathematical models.And “it is more robust against noise and provides scaling along the time axis. This ability allows DTW to identify similarities far more accurately and so enhances the functionality of the applications that use it” [31]. Furthermore, a simple pattern representation template is enough for similarity matching and detecting purposes with DTW [28]. However, DTW algorithm poses limitations in terms of time and space complexity [28, 40]. The time and space complexity of DTW is approximately $O\ (NMV)$, where N and M are sequences lengths ($O\ (N^2V)$ if the two sequences is equal in length), and V is the number of templates to be considered. This means that practical applications may be quite slow. To overcome these limitations, researchers proposed different methods to make DTW faster and to reduce the complexity of time and calculations required for alignment process.

Another significant issue with DTW is to achieve the higher accuracy for classification problems. This issue is dealt on application by application or domain by domain basis. Research efforts are being put forward to address this issue [48,49].

2.1 Dynamic Time Warping and face detection

Dynamic time warping gained significant attention in the last decade as a classification algorithm and similarity measure technique for different applications in different domains. Researchers have also attempted to use DTW with different existing algorithm within the domain of face localization and face recognition.

Luis et *al.* [43] introduced a face localization algorithm using DTW supplemented with skin color filter in RGB space to reduce the computational complexity. The template used in DTW is an average face created by manually extracting and aligning 40 different human faces with a constant size of 80 × 83 from 66 still images. The averaged face is quantized and smoothed using median filter prior to extracting the feature vectors. The feature vectors, in their work, consist of vertical projection of the template and two horizontal vectors one for the eyebrows and eye region and the later for the nose and mouth region. Prior to DTW classification algorithm, they had segmented the image using a morphological skin color filter in RGB space. Then they had extracted the large regions, which contain the skin color as a candidate face region and restricted the DTW classification algorithm to the extracted regions that had reduced the total computational complexity of their approach. Finally the DTW is used to calculate the Pearson correlation coefficient as a similarity distance measure for every extracted feature vector in stepwise. Alexander et *al.* [15] proposed DTW and LSTM ANN algorithms for face recognition. For DTW, they used the Euclidian distance as similarity measure. The templates consists of 50 face images of the size of 19 × 19 pixels converted into 1-D 361 element feature vector. The test set consists of 50 face images for the same persons in the templates and the DTW classification algorithm is applied to calculate the similarity measure for every test image with all the templates. To validate their approach, they repeated the experiments in three scenarios:

1. The templates and the test images are free of noise.
2. The templates are free of noise and the test images are degraded with Gaussian noise (σ^2=0.1).
3. The templates and the test images are degraded with Gaussian noise (σ^2=0.1).

The first scenario showed a 100% detection rate and the path sizes matrix showed a perfect alignment of the templates and the test faces. The performance of the algorithm was reduced to 94% when applied to the second scenario; however, it shows the robustness of the algorithm in the presence of noise. The last scenario the performance is further reduced to 70%. All the scenarios are repeated using MSE as similarity measure and the performances were reported as 92%, 72%, and 34% respectively. For LSTM neural network, they used 50 faces converted to 1×361 feature vector as training set to represent the main classes. Eigen faces were extracted using PCA to reduce the dimensionality of the input space. The experiment is performed in two scenarios:

1. Using 10 PCA, and
2. Using 20 PCA.

Selecting the learning rate to be 0.02 and training the LSTM ANN for 50000 epochs in both scenarios to reach 0.05 mean square error. In the first scenario the detection rate was 96% and 88% for the second one. Computational complexity of their approach was not investigated. The proposed LSTM ANN exhibited high performance and robustness.

3. Method for face detection

The eyes region has the most edge information in a facial image [42], and hence, it can be considered as a vital signature of a human face. Thus, the eyes region was chosen in this study as a facial pattern to apply the proposed method.

To have an eye region as a facial pattern template, a face image of a person from a data set (reference) has been cropped to eyes. Significant information (features) is then extracted from this particular template and transformed into a 1-Dimendioanl (1-D) series sequence. In addition, this facial pattern is used as a reference to detect eyes in the input images from the data set. On the other hand, an input face image from the image data set is gone through a partitioning process (see sub-section 3.2.3). This process produces virtual partitions of an input image. Each partition is equivalent to the template image in size. For each partitioned region, significant information is extracted and converted into a 1-D vector sequence to be aligned with the reference sequence (from template image) by DTW. The partition which belongs to a sequence that best matches the reference sequence is selected as the eyes/face region of the facial image. The proposed image processing strategy consists of the following steps:

i. Edge Detection,
ii. Histogram Equalization,
iii. Image Partitioning, and
iv. Features-Vectors Extraction.

Figures 1 and 2 present the functional flow of the image processing strategy and the image processing strategy respectively. All the steps stated above are described in the following sections.

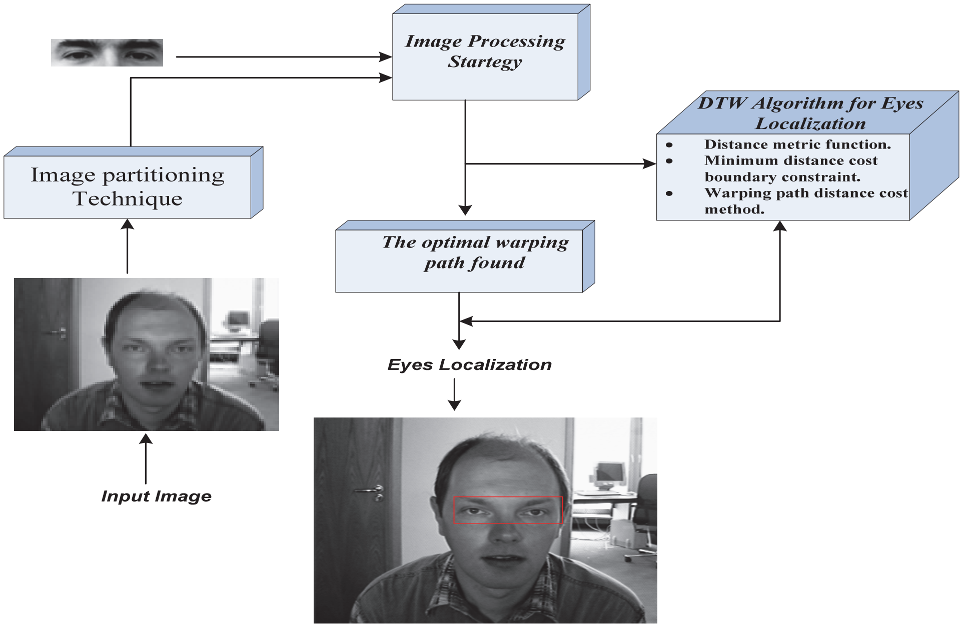


Image Processing Startegy
DTW Algorithm for Eyes Localization
• Distance metric function.
• Minimum distance cost boundary constraint.
• Warping path distance cost method.
Image partitioning Technique
The optimal warping path found
Eyes Localization
Input Image

Fig. 1. Functional flow of Eye region detection process

3.1 Edge detection and cropping rules

The first step of proposed strategy is edge detection. An input image is processed for edge detection to find the region boundaries. Standard Sobel operators are used with 3X3 filters for edge detection. Once the boundaries of the image are determined, a cropping step is applied.

Cropping an image is important in reducing the size of the data and iterations required to scan full image by DTW. For this purpose, a cropping rule is set instead of the manual cropping. An image is cropped till the first boundary region determined in the previous step is obtained. Figure 3 shows the original input image and the edge image for the same image, as well as the result of the cropping rules.

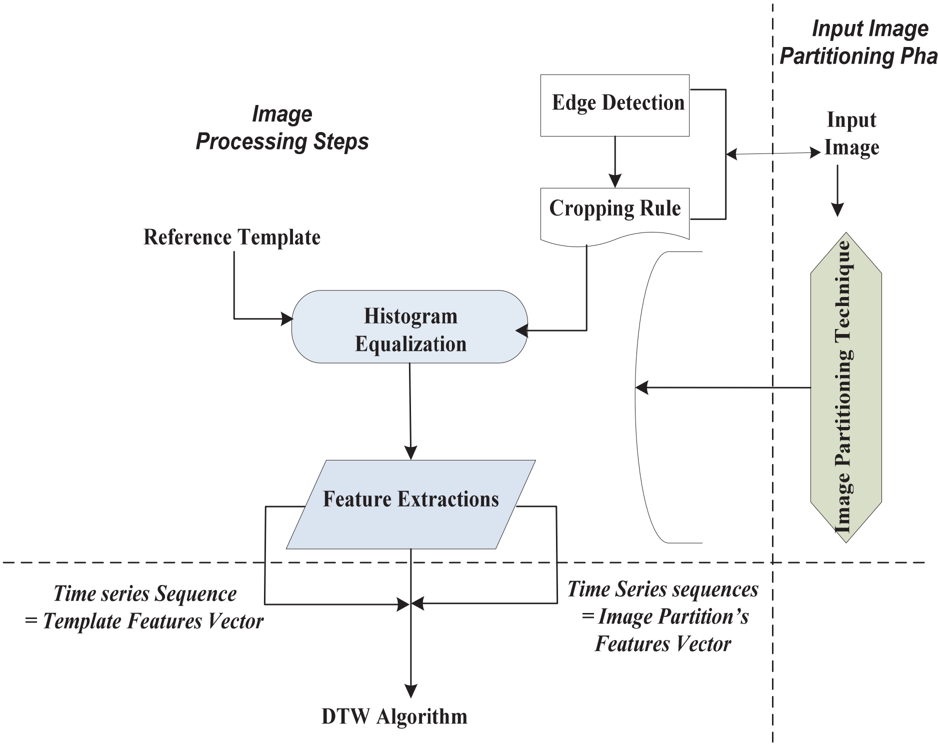
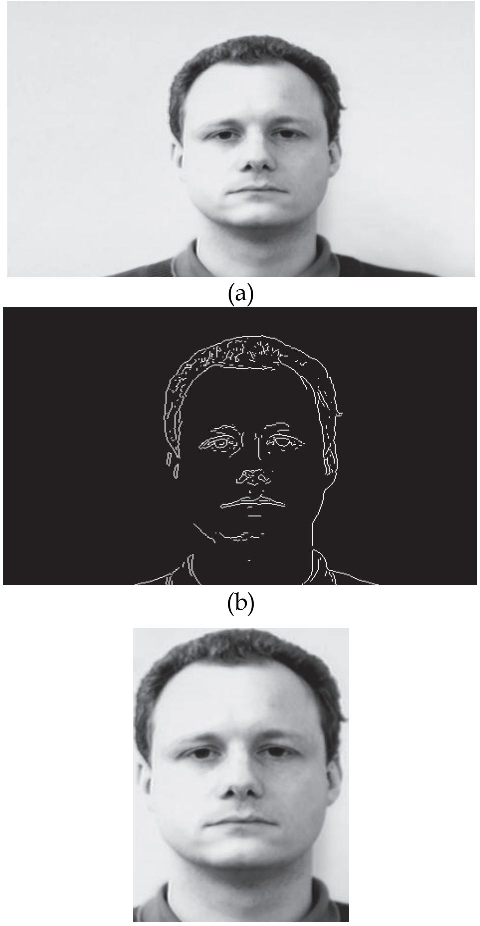

Input Image Partitioning Phase
Image Processing Steps
Edge Detection
Input Image
Cropping Rule
Reference Template
Histogram Equalization
Image Partitioning Technique
Feature Extractions
Time series Sequence = Template Features Vector
Time Series sequences = Image Partition's Features Vector
DTW Algorithm

Fig. 2. Image processing strategy



(a)

(b)

(c)

Fig. 3. (a) The original image, (b) the image after the application of Sobel operator, (c) the cropped image.

3.2 Histogram equalization

Intensities in the images are highly sensitive to external factors, such as illuminations, pose, rotations of the images, etc. These external factors affect the distribution of intensities in the histogram of the images [44], which in turn affects the detection rate to a great extent. In order to make the intensities relatively insensitive to a particular contrast, and brightness of the original image, etc., a histogram equalization is applied with a flat envelop on the image to re-distribute the intensities throughout the range. Figure 4 shows the histogram of an image before applying the histogram equalization step. In comparison to the histogram shown in Figure 4, the histogram of the same image is shown in Figure 5 after applying the histogram equalization. It can be seen in Figure 5 that the intensities of the image have been distributed throughout the range of the histogram.

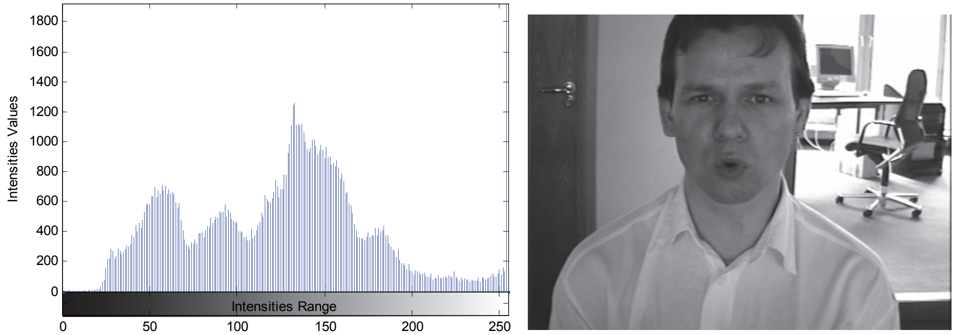


Fig. 4. A histogram distribution of an image before histogram equalization

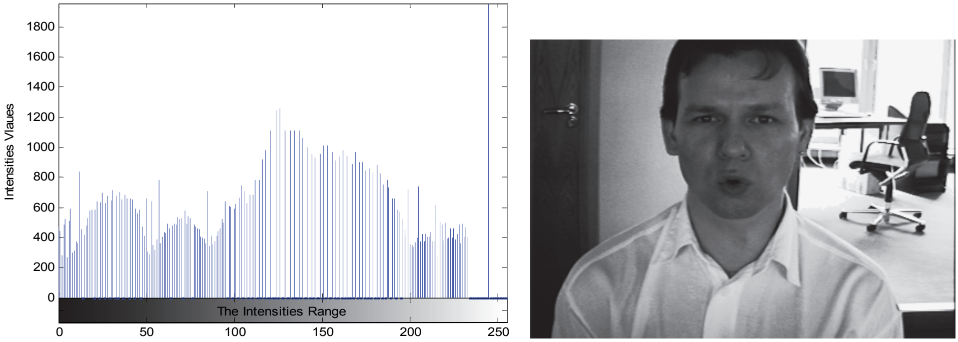


Fig. 5. A histogram distribution of an image after histogram equalization

3.3 Image partitioning

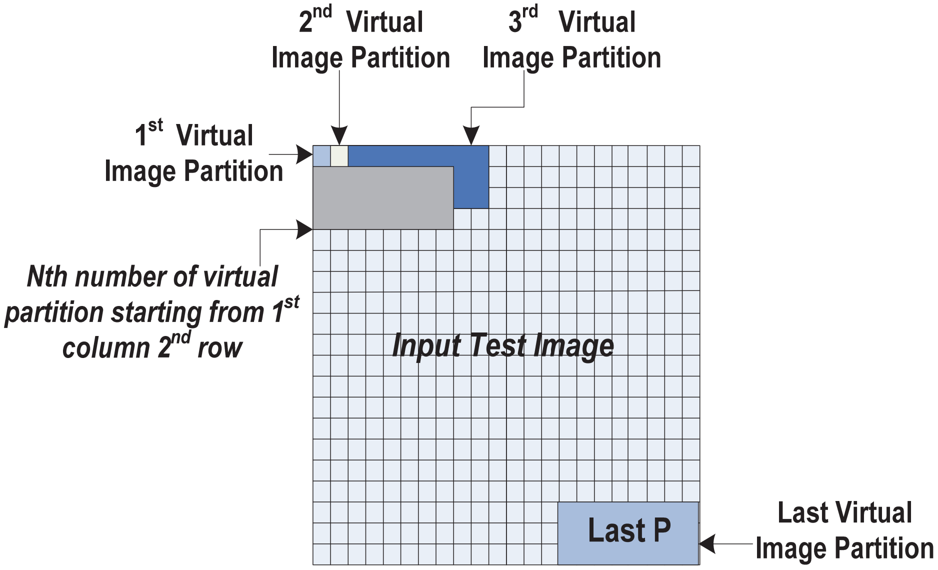
Image partitioning is an important step of the proposed strategy for face detection using DTW. In this step, virtual image partitions are extracted from an image. It is applied only on the test (input) images, which have been cropped through the steps of the image processing strategy. The image partitioning performed by our heuristic is given in Figure 6 below.

All partitions are equal in the size of the template image. Partitioning of an image is performed from the beginning of the first row and column and these are continued till the last row and column. All image partitions are further processed through the steps of the proposed image processing strategy, as shown in Figure 1. These image partitions are transformed into 1-D feature vectors (see section 3.4). It is important to note that at this point, DTW algorithm is used to measure the similarity between the template features vector and the vectors of the partition features to detect facial patterns and faces.

Partitioning process has been elaborated with the example given in Figure 7. In this figure, the first input image partition is virtually extracted (equivalent in the size of the template) from the first row and the first column, as shown by the sky blue colour. The second partition is extracted from the first row and the second column, which is shown in white (see Figure 7). In this way, the partitioning of an input test image goes through each column and row in a test image to find the optimal path between the two sequences. Thus, horizontally, a column and vertically a row becomes a *step size*.

```
//initialization
T= template image;
I= test image;
k=0;

[Trow,Tcol]=size(T);
[Irow,Icol]=size(I);

imax=Irow-Trow;
jmax=Icol-Tcol;

//iterations
for every i from 1 to imax
for every j from 1 to jmax
increment k
partition   I →Ipart=I(i:i+Trow,j:j+Tcol);

repeat for i
repeat for j
```

Fig. 6. Heuristic for image partitions



2nd Virtual Image Partition
3rd Virtual Image Partition
1st Virtual Image Partition
Nth number of virtual partition starting from 1st column 2nd row
Input Test Image
Last P
Last Virtual Image Partition

Fig. 7. Image Partitioning Process

3.4 Features vector extraction

To extract a particular feature from the image using 1-D DTW algorithm, horizontal and vertical projections are computed by integrating the columns and rows to produce 1-D vectors. The horizontal and vertical integral projections are presented in Equations 1 and 2, respectively.

$$Horizontal\,relief\ (y - \text{relief}) = \left[\sum_{j=1}^{N} I_{1j} \quad \sum_{j=1}^{N} I_{2j} \cdots \sum_{j=1}^{N} I_{Mj}\right] \tag{1}$$

$$Vertical\,relief\ (x - \text{relief}) = [\sum_{i=1}^{M} I_{i1} \quad \sum_{i=1}^{M} I_{i2} \cdots \sum_{i=1}^{M} I_{iN}] \tag{2}$$

In Equations 1 and 2, N is the width of the template/input partition image (the total number of columns available), M is the height of the template/input partition image (total number of rows). Figure 8 and 9 show the horizontal and vertical relief of the image (eyes region) respectively.

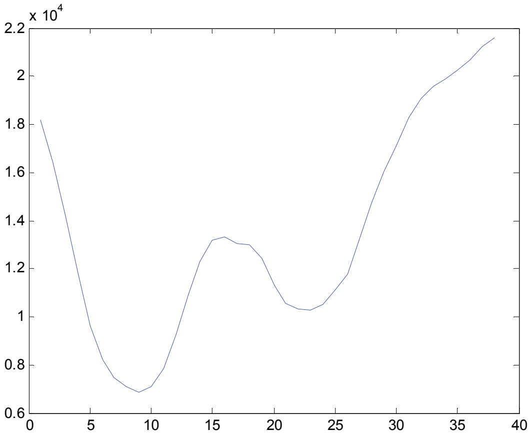


Fig. 8. Horizontal projections for image template

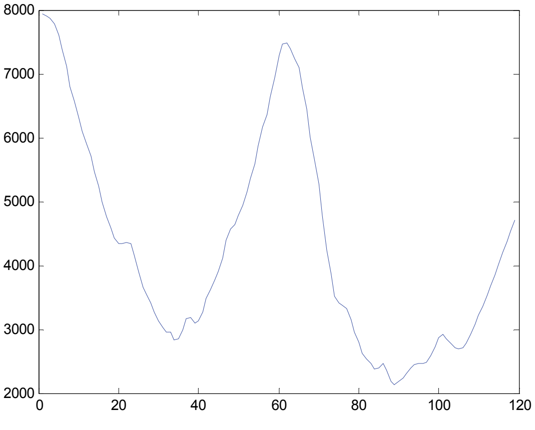


Fig. 9. Vertical projections for image template

The feature vector is formed by concatenating the y-relief and x-relief vectors. From Equations 1 and 2, the feature vector f can be obtained, as follows:

$$feature\ vector\ f = x - \text{relief} \oplus y - \text{relief} = [f_1 \quad f_2 \cdots f_k] \tag{3}$$

Where ⊕ denotes vector concatenation and k=N+M.

Equation 3 represents the feature vector extracted from a 2-D image. In the present work, the last element of the row sequence was joined to the first element of the column sequence. It was found through the experiment that an advantage of concatanating the horizontal and vertical projections in the features would include the cross correlation between the values in the rows and columns in the resulted accomulative distance matrix (in DTW algorithm). This row-column cross correlation has been found to make the classification less sensitive to the variation in the pose and orientation of the images. Figure 10 shows the result of cocatnation the vertical and horisontal vectors of the image template (eyes in this case).

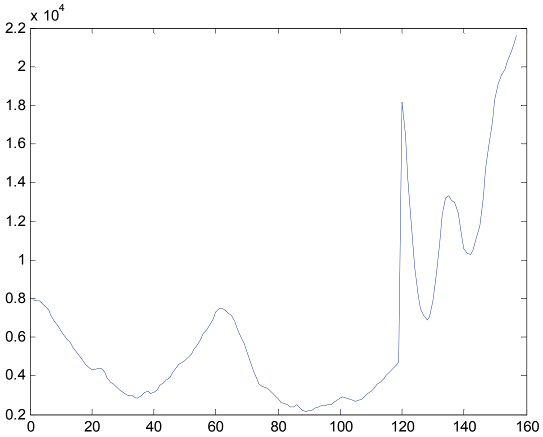


Fig. 10. Concatenated projections for image template

3.5 Dynamic Time Warping algorithm

Assume that there are two image vectors, I (the input image partition) and T (representing template image) each having noise n_I and n_T, the vectors are:

$$I_n = I + n_I = \left[I_1 + n_{I_1}\ I_2 + n_{I_2} \dots\ I_m + n_{I_m}\right] \tag{4}$$

$$T_n = T + n_T = \left[T_1 + n_{T_1}\ T_2 + n_{T_2} \dots\ T_k + n_{T_k}\right] \tag{5}$$

In Equations 4 and 5, m denotes the number of columns (the length) in the input image's partition I, and k denotes the number of column in the template image T (both vectors I and T have same length in the proposed method), whereas I_n and T_n are the input image partition and template image vectors with noise, respectively.

To align the image sequences, first a grid, I × T, local cost matrix is defined, where each cell represents the distance between the corresponding indices of the two feature-vectors. Normally, Euclidean distance is chosen to calculate the distance between the two elements in the matrix, d. The local distance matrix d can be computed as follows:

$$d(I,T) = |I - T| \tag{6}$$

The d matrix is of the size of m×k, presenting the cost values between each element of vector I and T, as shown in Equation 3.7 below:

$$d = \begin{bmatrix} I_1 + n_{I_1} - T_1 - n_{T_1} & \cdots & I_1 + n_{I_1} - T_k - n_{T_k} \\ \vdots & \ddots & \vdots \\ I_m + n_{I_m} - T_1 - n_{T_1} & \cdots & I_m + n_{I_m} - T_k - n_{T_k} \end{bmatrix} \tag{7}$$

Rearranging the terms in each element in Equation 7, the local distance matrix can be written as shown in Equation 8:

$$d(I_n, T_n) = d(I,T) + d(n_I, n_T) \tag{8}$$

After computing and building the local distance cost matrix (d) between two 1-D features' vectors I and T, a warping path (WP) is constructed by building a cumultaive/global cost matrix D(i,j). As previously mentioned (see section 2), the cumulative distance matrix D(i,j) is defined as the sum of the local distance d(i,j) found in the current cell and the minimum of the cumulative distances of the adjacent cells. A warping path is a concatenation of the cells starting from (1,1) to the cell (m,k). Let's define the minimum operator as follows:

$$M_{ij} = min\left[D_{i-1\,,j}, D_{i,j-1}, D_{i-1,j-1}\right] \tag{9}$$

Where M_{ij} is the minimum of the three cells in the global cost matrix D, from where the cell can be reached (depicted in Figure 11).

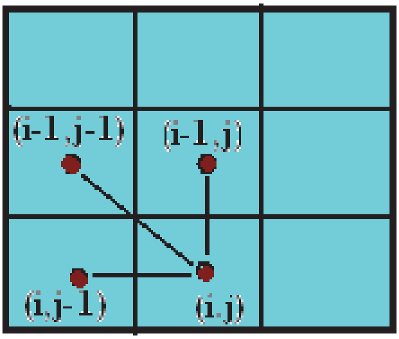


Fig. 11. The cell (i, j) and the three cells surrounding it.

Therefore, the global distance matrix (or the accumulative distance) can be formed from *d* (*I*, *T*) in Equation 7, as presented in Equation 10.

$$D = \begin{bmatrix} 0 & \infty \ldots & \infty \\ \infty & d_{1,1} + M_{2,2} \ldots & d_{1,k} + M_{2,k+1} \\ \vdots & \vdots & \vdots \\ \infty & d_{m,1} + M_{m+1,2} \ldots & d_{m,k} + M_{m+1,k+1} \end{bmatrix} \tag{10}$$

Applying the minimum operator and the steps to create the global distance matrix will result in building matrix D. Equations 11-21 represent the process of building the global cost matrix D.

The first element in the D matrix is equivalent to zero.
The entire first row and first column of the global matrix D is filled by infinity to ensure that the first element D (1,1) is the minimum. The second element of D, D (2, 2) is filled as follows:

$$D(2,2) = d(1,1) + M_{2,2} \tag{11}$$

$$D(2,2) = d(1,1) + minimum\left(D_{1,2}\ , D_{2,1}, D_{1,1}\right) \tag{12}$$

$$D(2,2) = d(1,1) + minimum(inf, inf, 0) \tag{13}$$

$$D(2,2) = I_1 - T_1 + 0 = I_1 - T_1 \tag{14}$$

Proceeding in the same manner for the second column of D and the second row of D:

$$D(3,2) = d(2,1) + M_{3,2} \tag{15}$$

$$D(3,2) = d(2,1) + minimum\left(D_{2,2}, D_{3,1}, D_{2,1}\right) \tag{16}$$

$$D(3,2) = \ d(2,1) + minimum(A_1 - B_1, inf, inf) \tag{17}$$

$$D(3,2) = I_2 - T_1 + I_1 - T_1 = \sum_{l=1}^{2} I_l - 2T_1 \tag{18}$$

In the same manner, the contents of the second column can be written as:

$$D(m,2) = \sum_{l=1}^{m-1} I_l - (m-1)T_1 \tag{19}$$

Once again, the second row elements of the matrix D can be expressed in the same manner, as:

$$D(2,k) = (k-1)I_1 - \sum_{w=1}^{k-1} T_w \tag{20}$$

The rest of D elements can be computed on the same manner, D will be:

$$D = \begin{bmatrix} 0 & \infty \ldots & \infty \\ \infty & I_1 - T_1 \ldots & (k-1)I_1 - \sum_{w=1}^{k-1} T_w \\ \vdots & \vdots & \vdots \\ \infty & \sum_{l=1}^{m-1} I_l - (m-1)T_1 \ldots & d_{m,k} + M_{m+1,k+1} \end{bmatrix} \tag{21}$$

Constraints are applied for constructing the optimal warping path (Sakoe & Chiba 1978), (see section 2 for more details).

- Monotonicity: The alignment path does not go back in the time index, $i_{l-1} \leq i_l$ and $j_{w-1} \leq j_w$.
- This guarantees that the features are not repeated in the alignment (see Figure 3.12).
- Continuity: The alignment does not jump in the time index, i_l – $i_{l-1} \leq 1$ and j_w- $j_{w-1} \leq 1$. This guarantees that the important features are not omitted. Continuity condition is shown in Figure 12, along with the monotonicity constraint.
- Boundary: The alignment starts at the beginning of the matrix and ends at the bottom-right. This guarantees that the sequences are not considered only partially. The blue boxes in Figure 12 show that the path starts at D (1,1) and ends at D (m,k).
- Warping window: A good alignment path is unlikely to wander too far from the diagonal. This guarantees that the alignment does not try to skip different features or

get stuck at similar features, and allowable points must fall within a given window, $|i_l - i_{l-1}| \leq w$; where w is a positive integer window width (threshold). Windowing constraint appears in Figure 12, as the two dark red lines surrounded the purple diagonal line.

Figure 12 illustrates the global cost matrix together with the warping path and the constraints applied to the paths.

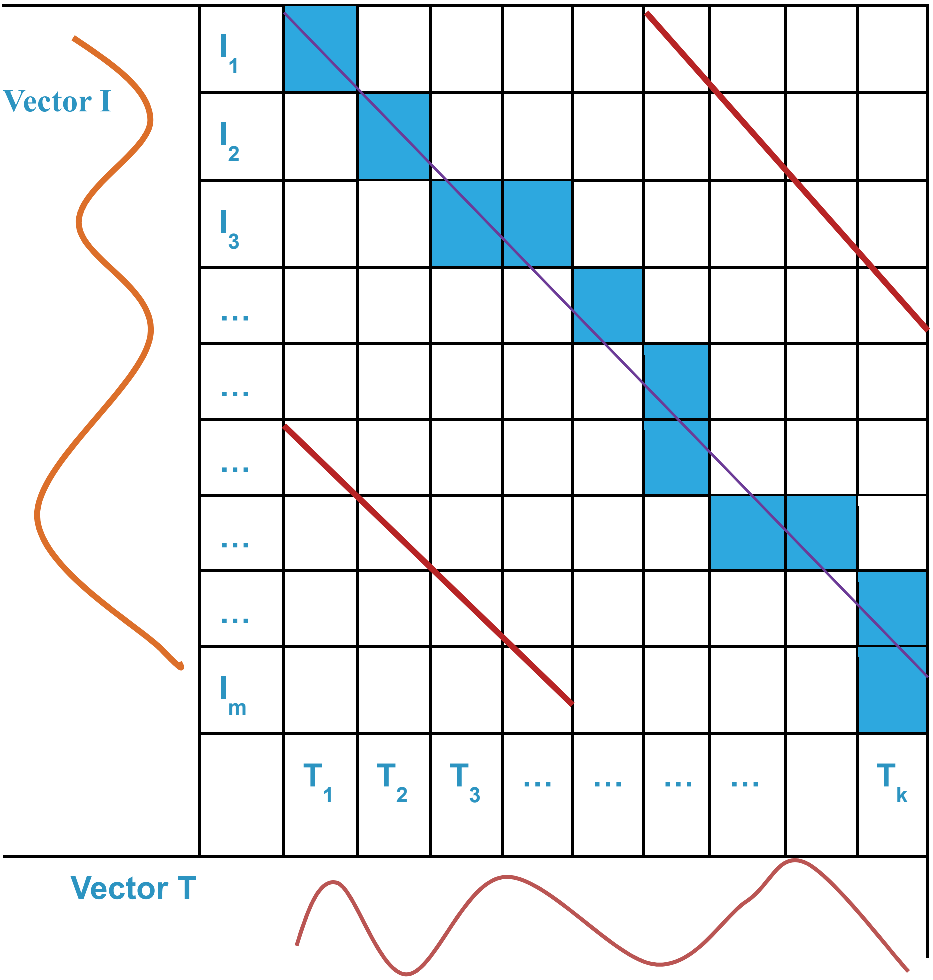


Fig. 12. Alignment between the two image vectors I and T

4. Working example of the Dynamic Time Warping algorithm for face detection

The proposed algorithm is implemented in MatLab 9 running on a computer with 2.0GHz dual core OPTERON processor and 1.87GB RAM. To evaluate the proposed techniques and method, two publicly available databases were used in this study. One database, containing face images with a resolution of 512 x 342 pixels under controlled lighting conditions and different rotations, is available from University of Bern, Switzerland [47]. Additionally, face images have changes in their facial expressions and immediate changes in head poses (right, left, up down and straight). The second one is the BioID face images database with the resolution of 384 x 286 pixels [46]. This database has face images with a complex background, along with different expressions and various illumination conditions, poses, and rotations. We conducted the experiments in different settings to evaluate the efficacy of our approach. In first setting, we applied the technique for eyes localization using DTW without our image processing strategy; the detection rate was very low. In second setting, we applied our image processing strategy with histogram equalization and DTW algorithm on the same dataset, the detection rate improved within 35%. Figure 13 shows the arbitrary eyes template chosen for our experiment.

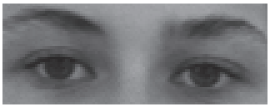
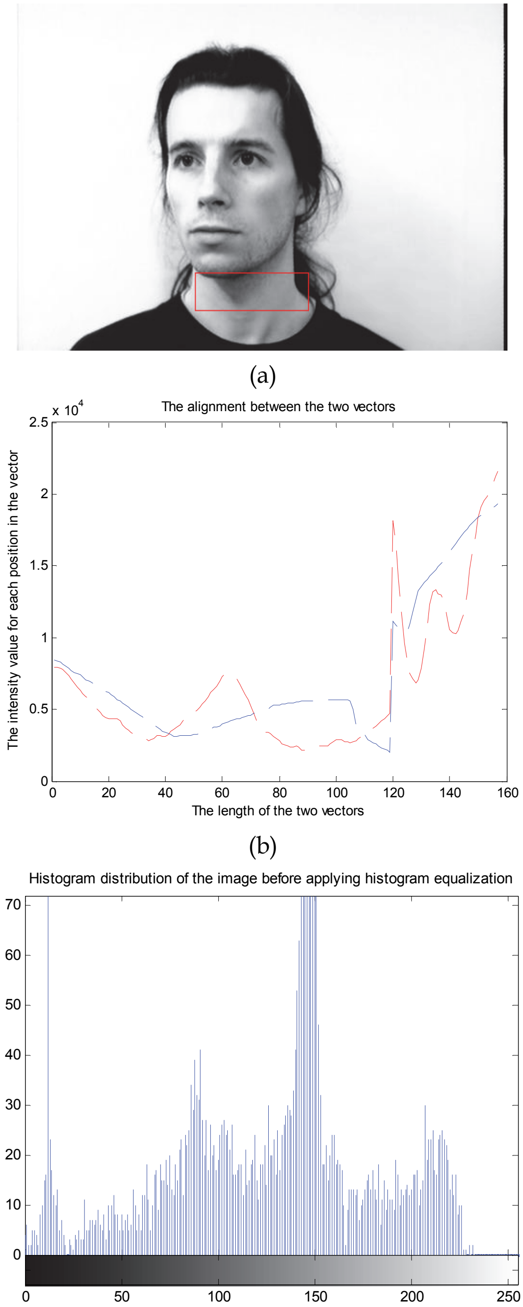


Fig. 13. The arbitrary template of eyes

With our experimental analysis and observation, histogram equalization reduced the effect of light on intensity distribution.

Comparing to other template matching methods [45] our system is simple fast and easy to implement with no training data set and shape model is needed, and template of eyes is manually cropped. The proposed method is showing acceptable detection rate and less time complexity comparing to other classifier and distance measurement systems.

Using the proposed image processing strategy, features vector were extracted (see section 3.4). With these extracted features, histogram equalization was applied with a flat envelop for both the template image and input test image(s) so as to re-distribute the intensities throughout the range. Histogram equalization distributes the intensities of an image throughout a range of histogram. The better distribution of the intensities has helped to detect facial patterns and face in bright background with existing illumination and lighting conditions. For example, the face image shown in Figure 14(a) was processed in Experiment setting 1 for the eye detection. Figure 14(a) shows that the eyes were not detected before histogram equalization was applied. Figure 14(b) illustrates the alignment between both the template and input image template vectors. Before that figure 14(c) depicts the histogram distribution for the image shown in Figure 14(a). It is obvious that both vector sequences are not well aligned. With our experimental analysis, histogram equalization was applied to reduce the effect of light on intensity distribution. The same facial image is shown in Figure 15(a) after applying histogram equalization. From the figure, it can be seen that the eyes have been detected correctly. Furthermore, both the vectors have been well aligned, as shown Figure 15(b) below:

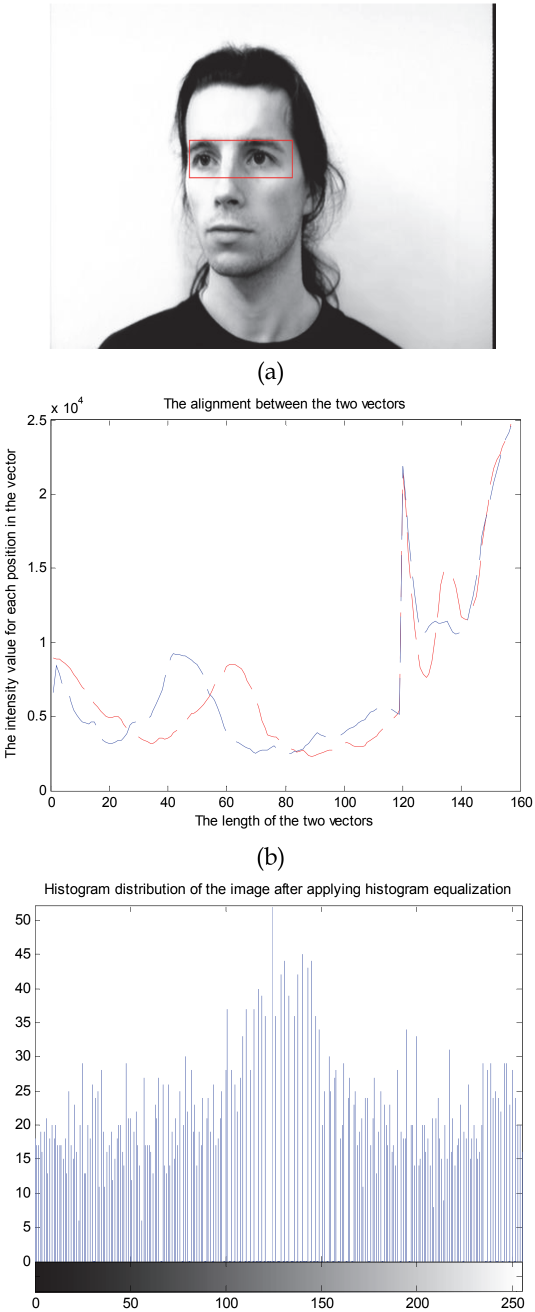


(a)

(b)

(c)

Fig. 14. Before applying image processing strategy



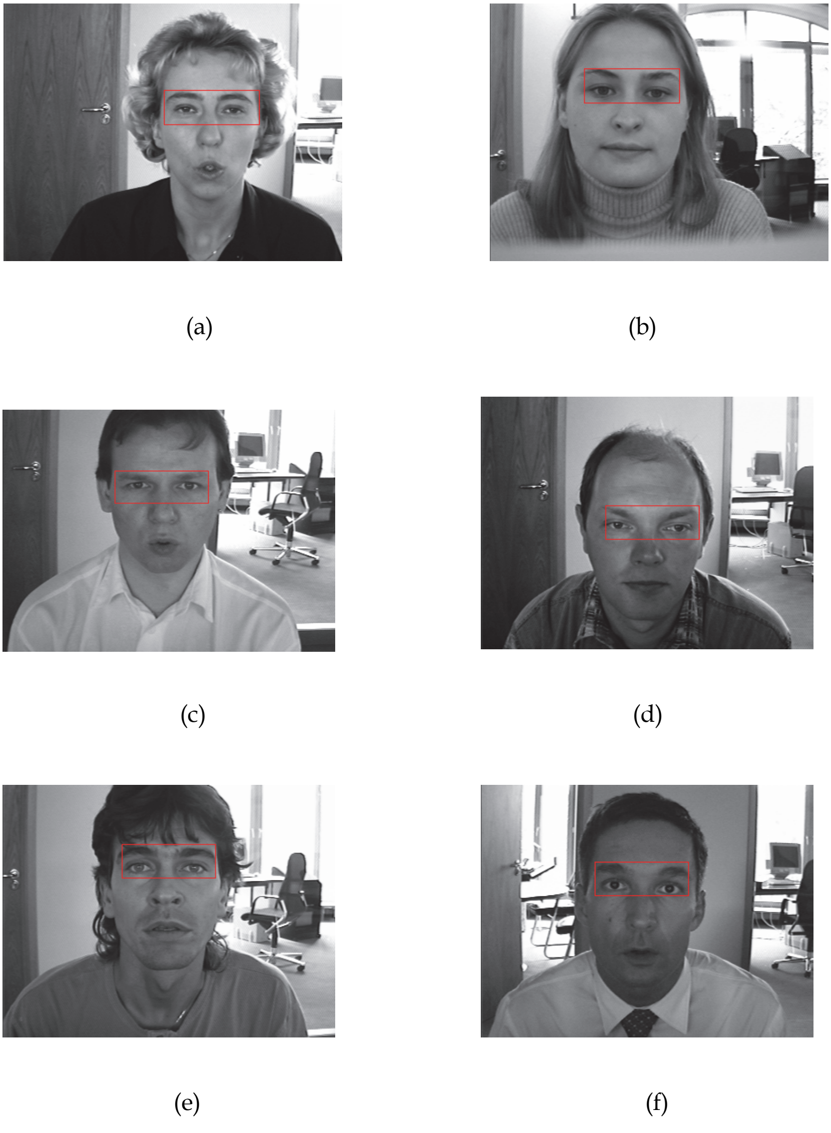
(a)

(b)

(c)

Fig. 15. Correct detection of the eyes after applying image processing strategy

Figure 16 and 17 show some of the image faces in different poses and lighting conditions, which are detected correctly and the alignment between the template image and the test images corresponding to these images, respectively.



(a) (b)

(c) (d)

(e) (f)

Fig. 16. (a-f) Face Images with positive eyes localized. (BioID images database [22])

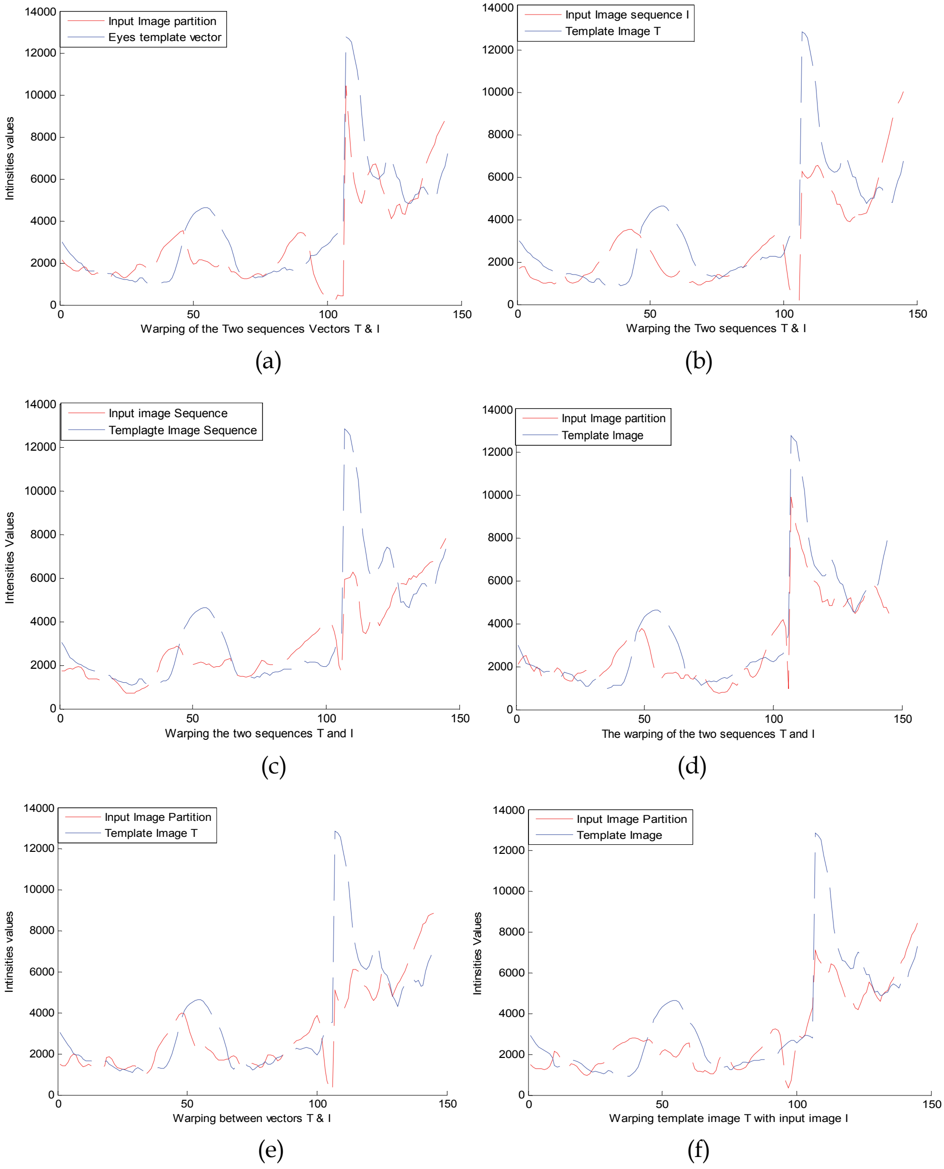


Fig. 17. (a-f) the corresponding graph of the warping path for the same image presented in Figure 16.

5. Conclusions

We have presented our approach of template-based matching method and Dynamic time warpping to efficiently enhance the detection and localization of eyes in complex background with face images under differnt illumination, light conditions, expressions and poses. This method overcomes some of the limitations of tempalte matching for face detection. With more developed DTW technique this method helps in getting high accuracy in matching process. Results of our approach show that supplementing DTW with proposed image processing strategy enhances the detection of facial feature patterns and face detection by overcoming intensities and face pose variations. Future work involves the improvement of DTW algorithm and Dynamic Programming constraints to further increase the detection rate.

6. References

[1] Md. Al-Amin B., Vuthichai A., Shin-yo M. H. U., "Face Detection and Facial Feature Localization for Human-machine Interface", NII Journal 5(3), (2003): 25-39.

[2] Ming-Hsuan Y., Kriegman D. J. and Ahuja N., "Detecting faces in images: a survey," IEEE Trans on PAMI 24(2002): 34-58.

[3] Hjelmas E. L., Boon K. "Face detection: A survey", Computer Vision and Image Understanding 83(3) (2001): 236-274.

[4] Zhao W., Chellappa R., Phillips P. J., Rosenfeld A., "Face Recognition: A Literature Survey", ACM Computing Surveys 35(4) (2003): 399–458.

[5] Rath T., Manmatha R., "Word image matching using dynamic time warping", (paper presented at the IEEE computer society conference on computer vision and pattern recognition, vol. 2. Los Alamitos, CA, USA: IEEE Computer Society. pp II-521-527, 2003).

[6] Ralph N., "Dynamic Time Warping An Intuitive Way of Hand writing Recognition", (paper presented at in Faculty Of Social Sciences Department Of Artificial Intelligence / Cognitive Science,, vol. Master: Radboud University Nijmegen, 2004).

[7] Sakoe H., Chib S. "Dynamic Programming Algorithm Optimization For Spoken Word Recognition", (paper presented at IEEE Trans. Acoustics, Speech, and Signal Proc, 1978, ASSP-26).

[8] Carlo Di B., Ralph N., Anneloes O., Gabriel L., Wouter H., "Dynamic time warping: a new method in the study of poor handwriting", Journal of Human Movement Science. 27(2) , (2008) : 242-255.

[9] Faundez-Zanuy M., "On-Line Signature Recognition Based on VQ-DTW", *Pattern Recognition*, 40(2007): 981-992.

[10] Efrat A., Fan Q., and Venkatasubramanian S., "Curve matching, time warping, and light Fields: New algorithms for computing similarity between curves", Journal of Mathematical Imaging Vis. 27(3), (2007): 203-216.

[11] Olaf H., Sascha M., "Effects of Time Normalization on the Accuracy of Dynamic Time Warping", (First IEEE International Conference on Biometrics: Theory, Applications, and Systems. BTAS 2007, pp.1-6, 27-29 Sept. 2007).

[12] Jayadevan R., Satish R., Pradeep M., "Dynamic Time Warping Based Static Hand Printed Signature Verification", Journal Of Pattern Recognition Research, 4(2009):52-65.

[13] Zsolt M., KovaÂ C.-V., "A Fingerprint Verification System Based on Triangular Matching and Dynamic Time Warping", IEEE Transactions on Pattern Analysis And Machine Intelligence, 22(11)(2000): 1266 – 1276.

[14] Hichem S., and Nozha B., "Robust Face Recognition Using Dynamic Space Warping", Lecture Notes in Computer Science 2359(2002): 121-132.

[15] Alexandre L. M. L., Débora C. C., Denis H. P. S., José H. S. and Nelson D. A. M., "Novel Approaches for Face Recognition: Template-Matching using Dynamic Time Warping and LSTM Neural Network Supervised Classification", (Presented in the *15th International Conference on Systems, Signals and Image Processing, 2008. IWSSIP 2008.*, pp.241-244, 25-28 June 2008).

[16] Scott M., George L., "Elastic matching in linear time and constant space", (DAS '10, June 9-11, 2010, Boston, MA, USA).

[17] Han-Wen H.,Wen-H., Cheng-K. and Jeffrey J. P. T., "A Novel Feature with Dynamic Time Warping and Least Squares Adjustment for Protein Structure Alignment", Asian Journal of Health and Information Sciences 1(3)(2006): 261-275.

[18] Aach J, Church GM, "Aligning gene expression time series with time warping algorithms", Bioinformatics. 17(6)(2001):495-508.

[19] Jo C. and Elena T., "Gene Time E*x*pression Warper: a tool for alignment, template matching and visualization of gene expression time series", Bioinformatics Applications Note 22 (2) (2006): 251–252.

[20] Benoıˆt L., Chang C.S. , Ong S.H., Soek-Ying N., Nallasivam P., "Chromosome classification using dynamic time warping", Journal of Pattern Recognition Letters 29(2008): 215–222.

[21] Aparna L. R., Grimson W. E. L., And William M. W. I., "Object Detection and Localization by Dynamic Template Warping", *International Journal Of Computer Vision* 36(2)(2000)131-147.

[22] Santosh K.C, "Use of Dynamic Time Warping for Object Shape Classification through Signature", Kathmandu University Journal of Science, Engineering And Technology 6(I) (2010): 33-49.

[23] Kongming W. and Theo G., "Alignment of curves by dynamic time warping", Annals of Statistics 25(3) (1997): 1251-1276.

[24] Longin J. L., Vasileios M., Qiang W., Deguang Y., "An elastic partial shape matching technique." Pattern Recognition 40(2007): 3069-3080.

[25] Joan C., Joaquim M., Fco. I. G., "Pattern Recognition Based On Episodes and DTW, Application To Diagnosis Of A Level Control System", (Presented at the 16th International Workshop on Qualitative Reasoning, 2002).

[26] Hansheng L., Govindaraju, "Direct Image Matching by Dynamic Warping", (Presented in the Computer Vision and Pattern Recognition Workshop CVPRW '04: 76- 76, 27-02 June 2004).

[27] Berndt D. and Clifford J., "Using dynamic time warping to find patterns in time series", (Presented at AAAI-94 Workshop on Knowledge Discovery in Databases, 1994).

[28] Keogh E., "Exact indexing of dynamic time warping", *Knowledge and Information Systems* 7 (3) (2005): 358-386.

[29] Keogh E.and Pazzani M., "Derivative Dynamic Time Warping", (Presented at the SIAM International Conference on Data Mining, Chicago, 2001).

[30] Ratanamahatana A. and Keogh E., "Making Time-series Classification More Accurate Using Learned Constraints", (Presented in the SIAM Intl. Conf. on Data Mining, pp. 11-22, Lake Buena Vista, Florida, 2004).

[31] Yasushi S., Masatoshi Y., and Chrisos F., "FTW: Fast Similarity Search under the Time Warping Distance". (Presented In *PODS*, pp. 326–337, 2005).

[32] Yutao S., Nikos M., and David W. C., "Fast and Exact Warping of Time Series Using Adaptive Segmental Approximations", Machine Learning 58 (2005): 231–267.

[33] Daniel E. R., Svetha V., and Wanquan L., "Threshold dynamic time warping for spatial activity recognition", *International Journal of Information and Systems Sciences* (3) (2007): 392-405.

[34] Daniel L., "Faster Sequential Search with a Two-Pass Dynamic-Time-Warping Lower Bound", arXiv:0807.1734v4 [cs.DB], Oct 2008.

[35] Johannes A., Thomas B., Hans-Peter K., Peer K., and Matthias R., "Periodic Pattern Analysis in Time Series Databases. In Database Systems For Advanced Applications", Lecture Notes In Computer Science, 5463 (2009): 354-368.

[36] Clifford D., Stone G., Montoliu I., Rezzi S., Martin F., Guy P., Bruce S., And Kochhar S., "Alignment Using Variable Penalty Dynamic Time Warping", *Anal. Chem* 81(2009): 1000-1007.

[37] Xavier A., Robert M., Nuria O., "Partial Sequence Matching Using an Unbounded Dynamic Time warping Algorithm", (ICASSP 2010).

[38] Chaovalitwongsea W. A. and Pardalos P. M., "On the Time Series Support Vector Machine Using Dynamic Time Warping Kernel For Brain Activity Classification", Cybernetics and Systems Analysis 44(1) (2008): 125-138.

[39] Marek I., Agnieszka M., "Voice control", (Presented in the XXIV Symposium *Vibrations in Physical Systems,* Poznan – Bedlewo, May 12-15, 2010).

[40] Salvador S., Chan P., "FastDTW: Toward Accurate Dynamic Time Warping In Linear Time and Space", (Presented in the *3rd KDD workshop on mining temporal and sequential data*, pp.70-80, 2004).

[41] Cha Z. and Zhengyou Z., "A Survey of Recent Advances in Face Detection", (Microsoft Research, Microsoft Corporation. June 2010).

[42] Kawaguchi T., Rizon M., "Iris Detection Using Intensity And Edge Information", Journal of Pattern Recognition 36(2) (2007): 549 –562.

[43] Luis, E., Raul, P., Jonathan V., "Face Localization In Color Images Using Dynamic Time Warping and Integral Projections", (Presented in the International Joint Conference On Neural Networks, August 12-17, 2007, Orlando, Florida, USA).

[44] Schwenker F., Andreas, S., Palm G., Kestler H. A., "Orientation Histograms for Face Recognition", Artificial Neural Networks in Pattern Recognition 4087(2006): 253 – 259.

[45] Qiong W., Jingyu Y., "Eye Detection in Facial Images with Unconstrained Background", Journal of Pattern Recognition Research 1(2006): 55-62.

[46] Jesorsky O., Kirchberg K., Frischholz R. The BioID face database. *HumanScan,* 2001,[Online]. Available: http://www.bioid.com/downloads/facedb/index.php.

[47] B. Achermann, The face database of University of Bern
[http://iamwww.unibe.ch/fkiwww/staH/achermann.html],
Institute of Computer Science and Applied Mathematics, University of Bern, Switzerland, 1995.

[48] Somya A., Hamzah A., "Dynamic Time Warping Approach Toward Face Patterns Detection", International Journal Of Academic Research 3(1)(2011): 89-95.

[49] Somya A., Hamzah A., "A New Approach for an Efficient DTW in Face Detection through Eyes Localization", Journal of Electronics and Electrical Engineering 2(108) (2011): 103-108.

Part 5

Face Recognition in Video

13

Video-Based Face Recognition Using Spatio-Temporal Representations

John See[1], Chikkannan Eswaran[1] and Mohammad Faizal Ahmad Fauzi[2]
[1]*Faculty of Information Technology, Multimedia University*
[2]*Faculty of Engineering, Multimedia University*
Malaysia

1. Introduction

Face recognition has seen tremendous interest and development in pattern recognition and biometrics research in the past few decades. A wide variety of established methods proposed through the years have become core algorithms in the area of face recognition today, and they have been proven successful in achieving good recognition rates primarily in still image-based scenarios (Zhao et al., 2003). However, these conventional methods tend to perform less effectively under uncontrolled environments where significant face variability in the form of complex face poses, 3-D head orientations and various facial expressions are inevitable circumstances. In recent years, the rapid advancement in media technology has presented image data in the form of *videos* or *video sequences*, which can be simply viewed as a temporally ordered collection of images. This abundance and ubiquitous nature of video data has presented a new fast-growing area of research in video-based face recognition (VFR).

Recent psychological and neural studies (O'Toole et al., 2002) have shown that facial movement supports the face recognition process. Facial dynamic information is found to contribute greatly to recognition under degraded viewing conditionsand also when a viewer's experience with the same face increases. Biologically, the media temporal cortex of a human brain performs motion processing, which aids the recognition of dynamic facial signatures. Inspired by these findings, researchers in computer vision and pattern recognition have attempted to improve machine recognition of faces by utilizing video sequences, where temporal dynamics is an inherent property.

In VFR, temporal dynamics can be exploited in various ways within the recognition process (Zhou, 2004). Some methods focused on directly modeling temporal dynamics by learning transitions between different face appearances in a video. In this case, the sequential ordering of face images is essential for characterizing temporal continuity. While its elegance in modeling dynamic facial motion "signatures" and its feasibility for simultaneous tracking and recognition are obvious, classification can be unstable under real-world conditions where demanding face variations can caused over-generalization of the transition models learned earlier.

In a more general scenario, VFR can also be performed by means of image sets, comprising of independent unordered frames of a video sequence. A majority of these methods characterize the face manifold of a video using two different representations – (1) face subspaces[1], and (2)

[1] Sometimes termed as *local sub-manifolds* or *local models* in different works.

face exemplars, or representative images that summarizes face appearances in a video. While these methods are attractive due to their straightforward representation and computational ease, they are typically dependent on the effectiveness of extracting meaningful subspaces or exemplars that can accurately represent the videos. Certain works have incorporated temporal dynamics into the final classification step to some degree of success.

This chapter presents a new framework for video-based face recognition using spatio-temporal representations at various levels of the task. Using the exemplar-based approach, our spatio-temporal framework utilizes additional temporal information between video frames to partition training video data into local clusters. Face exemplars are then extracted as representatives of each local cluster. In the feature extraction step, meaningful spatial features are extracted from both training and test data using the newly proposed Neighborhood Discriminative Manifold Projection (NDMP) method. Finally, in order to facilitate video-to-video recognition, a new exemplar-driven Bayesian network classifier which promotes temporal continuity between successive video frames is proposed to identify the subject in test video sequences. In the next section, some related works in literature will be discussed.

2. Related work

By categorizing based on feature representation, recent methods in video-based face recognition (VFR) can be loosely organized into three categories: (1) direct modeling of temporal dynamics, (2) subspace-based representation, and (3) exemplar-based representation.

In video sequences, continuity is observed in both face movement and change in appearances. Successful modeling of temporal continuity can provide an additional dimension into the representation of face appearances. As such, the smoothness of face movement can also be used for face tracking. Simultaneous tracking and recognition by Zhou and Chellappa is the first approach that systematically incorporate temporal dynamics in video-based face recognition (Zhou et al., 2003). A joint probability distribution of identity and head motion using sequential importance sampling (SIS) was modelled. In another tracking-and-recognition work (Lee et al., 2005), a nonlinear appearance manifold representing each training video was approximated as a set of linear sub-manifolds, and transition probabilities were learned to model the connectivity between sub-manifolds. Temporal dynamics within a video sequence can also be modeled over time using Hidden Markov Models (HMM) (Liu & Chen, 2003). Likelihood scores provided by the HMMs are then compared, and the identity of a test video is determined by its highest score. Due to the nature of these representations, many of these methods lack discriminating power due to disjointed person-specific learning. Moreover, the learning of temporal dynamics during both training and recognition tasks can be very time-consuming.

Subspace-based methods represent entire sets of images as subspaces or manifolds, and are largely parametric in nature. Typically, these methods represent image sets using parametric distribution functions (PDF) followed by measuring the similarity between distributions. Both the Mutual Subspace Method (MSM) (Yamaguchi et al., 1998) and probabilistic modeling approaches (Shakhnarovich et al., 2002) utilize a single Gaussian distribution in face space while Arandjelovic et al. (Arandjelovic et al., 2005) extended this further using Gaussian mixture models. While it is known that these methods suffer from the difficulty of parameter estimation, their simplistic modeling of densities is also highly sensitive to conditions where training and test sets have weak statistical relationships. In a specific work on image sets, Kim et al. (Kim et al., 2007) bypass the need of using PDFs by computing similarity between subspaces using canonical correlations.

Exemplar-based methods offer an alternative model-free method of representing image sets. This non-parametric approach has become increasingly popular in recent VFR literature. Krüeger and Zhou (Krüeger & Zhou, 2002) first proposed a method of selecting exemplars from face videos using radial basis function network. There are some comprehensive works (Fan & Yeung, 2006; Hadid & Peitikäinen, 2004) that proposed view-based schemes by applying clustering techniques to extract view-specific clusters in dimensionality-reduced space. Cluster centers are then selected as exemplars and a probabilistic voting strategy is used to classify new video sequences. Later exemplar-based works such as (Fan et al., 2005; Liu et al., 2006) performed classification using various Bayesian learning models to exploit the temporal continuity within video sequences. Liu et al. (Liu et al., 2006) also introduced a spatio-temporal embedding that learns temporally clustered *keyframes* (or exemplars) which are then spatially embedded using nonparametric discriminant embedding. While all these methods have good strengths, none of these classification methods consider the varying influence of different exemplars with respect to their parent clusters.

Our proposed exemplar-based spatio-temporal framework for video-based face recognition can be summarized by the following contributions at each step of the recognition process:

1. **Clustering and Exemplar Extraction**: Motivated by the advantages of using hierarchical agglomerative clustering (HAC) (Fan & Yeung, 2006), a new spatio-temporal hierarchical agglomerative clustering (STHAC) method is proposed to exploit temporal continuity in video sequences. For each training video, the nearest faces to the cluster means are selected as exemplars.
2. **Feature Representation**: The newly proposed Neighborhood Discriminative Manifold Projection (NDMP) (See & Ahmad Fauzi, 2011) is applied to extract meaningful features that are both discriminative and neighborhood preserving, from both training and test data.
3. **Classification**: A new exemplar-driven Bayesian network classifier which promotes temporal continuity between successive video frames is proposed to identify the subject in test video sequences. Our classifier accumulates evidences from previous frames to decide on the subject identity. In addition, causal relationships between each exemplar and its parent class are captured by the Bayesian network.

Figure 1 illustrates the steps taken in our proposed framework, and where spatial and temporal information has been utilized.

3. Exemplar-based spatio-temporal framework

In this section, we elaborate in detail our proposal of an exemplar-based spatio-temporal framework for video-based face recognition.

3.1 Problem setting

The abundance of images in video poses a methodological challenge. While conventional still image-based face recognition is a straightforward matching of a test image to a gallery of training images *i.e.* an *image-to-image*[2] recognition task, it is an ill-posed problem for video

[2] In the abbreviated recognition settings, the first and third words denote data representation for each subject in the training/gallery and test/probe sets respectively.

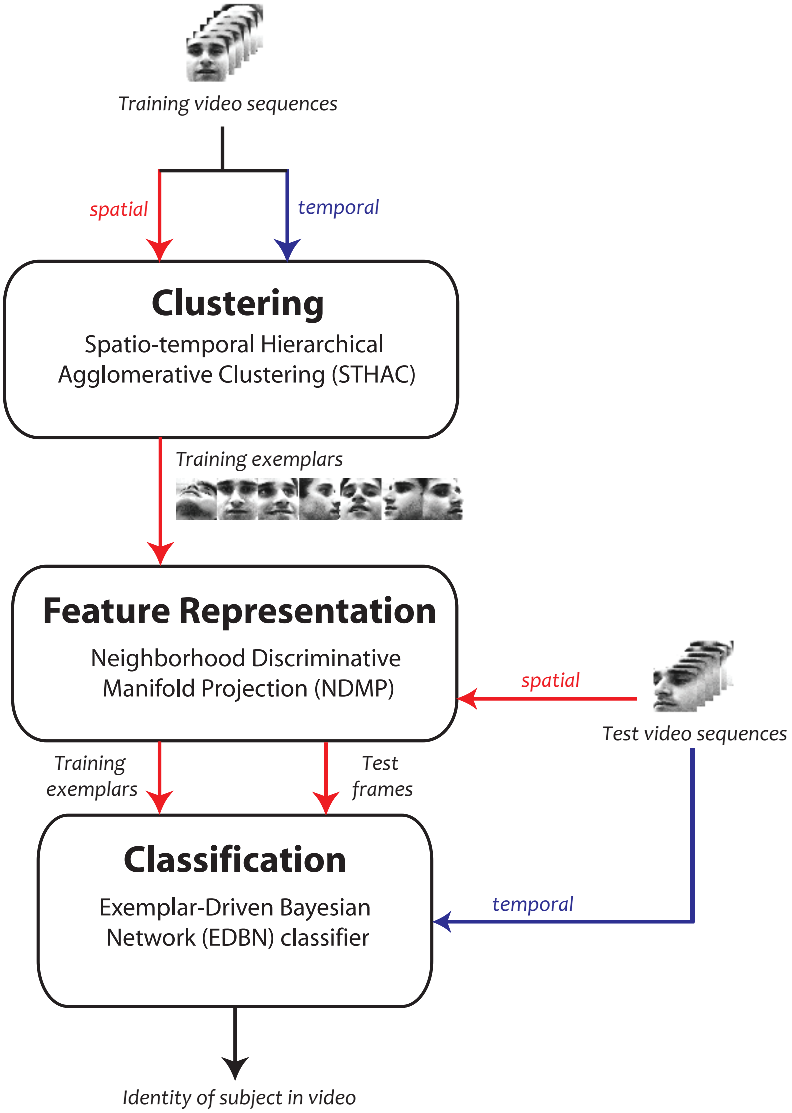

Training video sequences
spatial
temporal
Clustering
Spatio-temporal Hierarchical Agglomerative Clustering (STHAC)
Training exemplars
Feature Representation
Neighborhood Discriminative Manifold Projection (NDMP)
spatial
Test video sequences
Training exemplars
Test frames
Classification
Exemplar-Driven Bayesian Network (EDBN) classifier
temporal
Identity of subject in video

Fig. 1. The proposed exemplar-based spatio-temporal framework for video-to-video face recognition. The usage of spatial and temporal information are indicated in red and blue respectively in this diagram. The STHAC method in the clustering step and EDBN classifier in the classification step utilizes both spatio-temporal dynamics. Feature representation (NDMP method) takes only spatial information since the extracted training exemplars are used here, unlike subspace-based and temporal modeling-based methods (as discussed in Section 2).

sequences. Which image from the training video is to be matched with images from the test video?

To accomplish a complete *video-to-video* setting, one possible configuration used by exemplar-based approaches is to simplify it to a *image-to-video* recognition task, whereby each training video is represented by a set of exemplars (Fan & Yeung, 2006; Hadid & Peitikäinen, 2004). The availability of multiple image frames in the test video provides a good platform for utilizing temporal information between successive frames. For general notation, given a sequence of face images extracted from a video,

$$X_c = \{x_1^c, x_2^c, \ldots, x_{N_c}^c\} \ , \tag{1}$$

where N_c is the number of face images in the video. Assuming that each video contains the faces of the same person and c is the subject identity of a C-class problem, $c \in \{1, 2, \ldots, C\}$, its associated exemplar set

$$E_c = \{e_1^c, e_2^c, \ldots, e_M^c\} \ , \tag{2}$$

where $E_c \subseteq X_c$ and the number of exemplars extracted, $M \ll N_c$. Thus, the overall exemplar-class set can be succintly summarized as

$$E = \{e_{c,m} | c = 1, \ldots, C; m = 1, \ldots, M\} \ , \tag{3}$$

in which there are a total of $C \times M$ unique exemplar-classes. In cases where more than one training video of a particular class is used, image frames from all similar-class videos are aggregated to extract M exemplars.

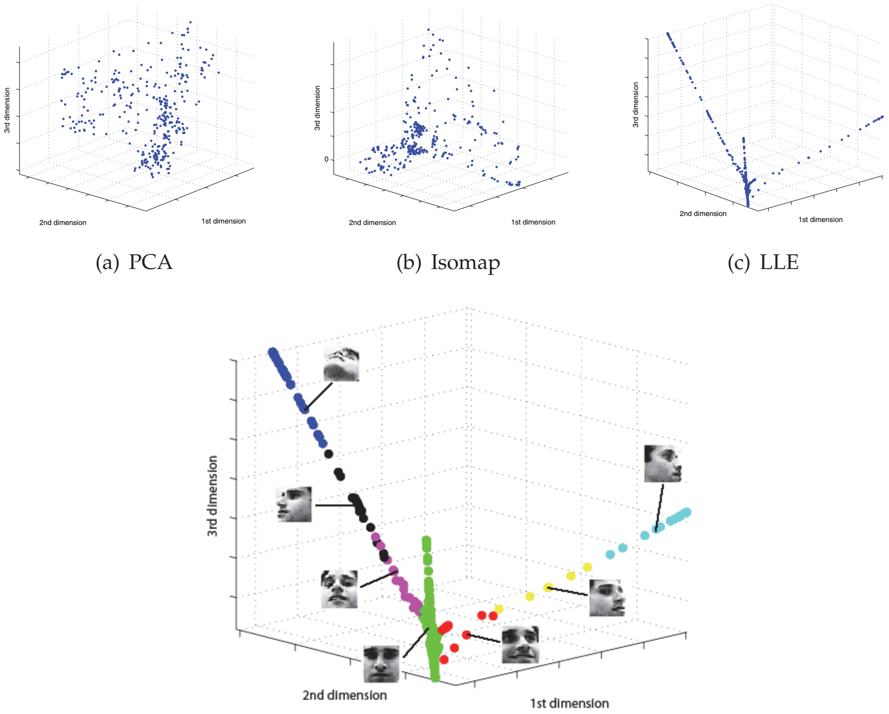
3.2 Embedding face variations

Considering the large amount of face variations in each training video sequence, a suitable dimensionality reduction method is necessary to uncover the intrinsic structure of the data manifold which may originally lie on a complex hyperplane. Recent nonlinear dimensionality reduction techniques such as Locally Linear Embedding (LLE) (Roweis & Saul, 2000) and Isomap (Tenenbaum et al., 2000) have been proven effective at seeking a low-dimensional embedding of a data manifold. LLE in particular, is known for its capability in modeling the global intrinsic structure of the manifold while preserving local neighborhood structures to better capture various face variations such as pose, expression and illumination. Conventional unsupervised methods such as Principal Component Analaysis (PCA) (Turk & Pentland, 1991) and Multidimensional Scaling (MDS) (Cox & Cox, 2001) have the tendency of overestimating the intrinsic dimensionality of high-variability data.

For each training video in our method, we apply the LLE algorithm to embed the face variations in a low-dimensional space, in preparation for the subsequent clustering step. Fig. 2(a), 2(b) and 2(c) shows the embedding of image data from a single video in PCA, Isomap and LLE spaces respectively. From Fig. 2(d), we can clearly see that the LLE method is able to uncover linear patches in its embedded space after performing spectral clustering, where each cluster represents a particular face appearance.

3.3 Spatio-temporal clustering for exemplar extraction

In the next step, clustering is performed on the LLE-space by extracting M number of clusters that group together faces of similar appearances. In many previous works (Fan et al., 2005; Hadid & Peitikäinen, 2004), k-means clustering is the primary choice for assigning data into different clusters due to its straightforward implementation. However, it has some obvious limitations – firstly, it is sensitive to the initial seeds used, which can differ in every run, and secondly, it produces suboptimal results due to its inability to find global minima.



(a) PCA

(b) Isomap

(c) LLE

(d) Linear cluster patches (in different colors) in LLE-space.

Fig. 2. The embedding plots of data taken from a sample video sequence, embedded using (a) PCA, (b) Isomap, and (c) LLE dimensionality reduction methods. Each data point represents an image in the embedded space. In the case of LLE, the formation of linear patches enables the spreading of different face appearances across its spectral projection (d).

3.3.1 Hierarchical Agglomerative Clustering (HAC)

Hierarchical Agglomerative Clustering (HAC) is a hierarchical method of partitioning data points by constructing a nested set of partitions represented by a cluster tree, or *dendrogram* (Duda et al., 2000). The agglomerative approach works from "bottom-up" by grouping smaller clusters into larger ones, as described by the following procedure:

1. Initialize each data point (of all N_c points) as a singleton cluster.
2. Find the nearest pairs of clusters, according to a certain distance measure $d(\Phi_i, \Phi_j)$ between clusters Φ_i and Φ_j. Commonly used measures are such as single-link, complete-link, group-average-link and Ward's distance criterion. Merge the two nearest clusters to form a new cluster.
3. Continue distance computation and merging (repeat Step 2), and terminate when all points belong to a single cluster. The required number of clusters, M is selected by partitioning at the appropriate level of the dendrogram.

3.3.2 Spatio-Temporal Hierarchical Agglomerative Clustering (STHAC)

Our proposed Spatio-Temporal Hierarchical Agglomerative Clustering (STHAC) differs from the standard HAC in terms of the computation of the nearest pair of clusters (Step 2). A *spatio-temporal distance measure* is proposed by fusing both spatial and temporal distances to influence the location of data points in time-space dimension. Since clustering procedures generally only utilize the distances between points, perturbations can be applied to the original spatial distances without cumbersome modeling of points in time-space dimension.

While spatial distance is measured by simple Euclidean distance between points, temporal distance is measured by the time spanned between two frame occurrences (x_i and v_j) in a video sequence,

$$d_T(x_i, x_j) = \left| t_{x_i} - t_{x_j} \right| \tag{4}$$

where t is a discretized unit time. This formulation is intuitive enough to quantify temporal relationships across sequentially ordered data points. The matrices containing pairwise spatial Euclidean distances $D_S(x_i, x_j)$ and temporal distances $D_T(x_i, x_j)$ between all data points, are computed and normalized. We present two varieties of fusion schemes, one which functions at the global structural level, and another at the local neighborhood level.

The Global Fusion (STHAC-GF) scheme blends the contribution of spatial and temporal distances using a temporal tuning parameter, α. The tuning parameter adjusts the perturbation factor defined by its upper and lower bounds, p_{max} and p_{min} respectively, which acts to increase or reduce the original distances. GSTF defines the spatio-temporal distance as

$$D_{ST,global} = (p_{max} - \alpha)D_S + (\alpha + p_{min})D_T, \qquad 0 \leq \alpha \leq 1\ . \tag{5}$$

The Local Perturbation (STHAC-LP) scheme perturbs spatial and temporal distances based on local spatio-temporal neighborhood relationships between a point and its neighbors. For each point x_i, a temporal window segment,

$$S_{x_i} = \{x_{i-w}, \ldots, x_i, \ldots, x_{i+w}\}\ , \tag{6}$$

of length $(2w + 1)$ is defined as its temporal neighborhood. The spatial neighborhood of point x_i,

$$Q_{x_i} = \{x_1, x_2, \ldots, x_k\} \tag{7}$$

is simply a set containing k-nearest neighbors of x_i computed by Euclidean distance. A point x_j is identified as a *common spatio-temporal neighbor* (CSTN) of point x_i if it belongs to both spatial and temporal neighborhood point sets, hence the criterion,

$$CSTN_{x_i} = S_{x_i} \cap Q_{x_i}\ . \tag{8}$$

With that, we introduce a perturbation affinity matrix containing multipliers that represent the degree of attraction and repulsion betwee

$$P_{ij} = \begin{cases} 1 - \lambda_{sim}, & \text{if } x_j \in CSTN_{x_i} \\ 1 + \lambda_{dis}, & \text{if } x_j \in (S_{x_i}\ CSTN_{x_i}) \\ 1, & \text{otherwise} \end{cases} \tag{9}$$

where λ_{sim} and λ_{dis} are the similarity and dissimilarity perturbation constants respectively, taking appropriate values of $0 < \{\lambda_{sim}, \lambda_{dis}\} < d(x_i, x_j)$. To simplify parameter tuning, we use a single perturbation constant, that is $\lambda = \lambda_{sim} = \lambda_{dis}$. In short P_{ij} seeks to accentuate the similarities and dissimilarities between data samples by artificially reducing and increasing spatial and temporal distances between samples. By matrix multiplication, STHAC-LP defines the spatio-temporal distance as

$$D_{ST,local} = P_{ij}(D_S + D_T) \ . \tag{10}$$

The linkage criterion chosen in our work for merging clusters is Ward's distance criterion,

$$d(\Phi_i, \Phi_j) = \frac{n_i n_j}{n_i + n_j} \left\| m_i - m_j \right\|^2 . \tag{11}$$

where m_i and m_j are means of cluster i and j respectively, while n_i and n_j are the cardinality of the clusters. Ward's criterion (Webb, 2002) is chosen due to its attractive sample-dependent mean squared error term, which is a good heuristic for approximating the suitable number of clusters via finding the "elbow" of the residual error curve (see Fig. 3).

After clustering the face data in each training video, face images that are nearest to each cluster mean are chosen as exemplars.

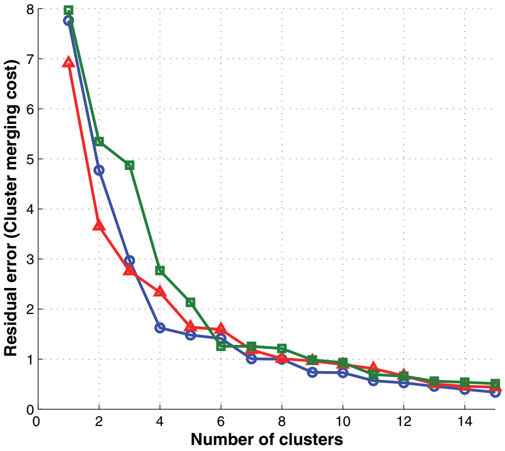


Fig. 3. Residual error curve of three different training videos (plotted with different colors) that were partitioned into different number of clusters. The "elbow" of the curve is approximately at 5 – 9 clusters.

3.4 Feature representation

Traditional linear subspace projection methods such as Principal Component Analysis (PCA) (Turk & Pentland, 1991) and Linear Discriminant Analysis (LDA) (Belhumeur et al., 1997) have been widely used to great effect in characterizing data in smooth and well-sampled manifolds. Recently, there has been a flurry of manifold learning methods such as Locality Preserving Projections (LPP) (He & Niyogi, 2003), Marginal Fisher Analysis (MFA) (Yan et al., 2007) and Neighborhood Preserving Embedding (NPE) (He et al., 2005), that are able to effectively derive optimal linear approximations to a low-dimensional embedding of complex manifolds. NPE in particular, has an attractive neighborhood preserving property due to its formulation based on LLE.

For improved extraction of features for face recognition, we apply the Neighborhood Discriminative Manifold Projection (NDMP) method (See & Ahmad Fauzi, 2011), which is our earlier work on feature representation. NDMP is a supervised discriminative variant of the NPE which seeks to learn an optimal low-dimensional projection by distinguishing between intra-class and inter-class neighborhood reconstruction. Global structural and local neighborhood constraints are imposed in a constrained optimization problem, which can be solved as a generalized eigenvalue problem:

$$(\mathbf{X}\mathbf{M}_{\mathbf{intra}}\mathbf{X}^{\mathbf{T}})\mathbf{A} = \Lambda(\mathbf{X}\mathbf{M}_{\mathbf{inter}}\mathbf{X}^{\mathbf{T}} + \mathbf{X}\mathbf{X}^{\mathbf{T}})\mathbf{A}, \tag{12}$$

where $\mathbf{X}$ denotes the face exemplar set in $\Re^D$, while $\mathbf{M}_{\mathbf{intra}}$ and $\mathbf{M}_{\mathbf{inter}}$ are the intra-class and inter-class orthogonal weight matrices respectively.

A new test sample $\mathbf{X}'$ can be projected to the embedded space in $\Re^{D'}$ by the linear transformation

$$\mathbf{Y}' = \mathbf{A}^{\mathbf{T}}\mathbf{X}' \tag{13}$$

where $D' \ll D$. More details on the theoretical formulation of the NDMP method can be found in (See & Ahmad Fauzi, 2011).

Fig. 4 shows the plots of face exemplar points in LDA, LPP and NDMP feature space. Note that among the three supervised methods, the NDMP dimensionality reduction method provides the best discrimination between classes. The insufficient amount of separation between different-class points (for LPP) and attraction within same-class points (for LDA and LPP) can potentially contribute towards erroneous classification in the next task.

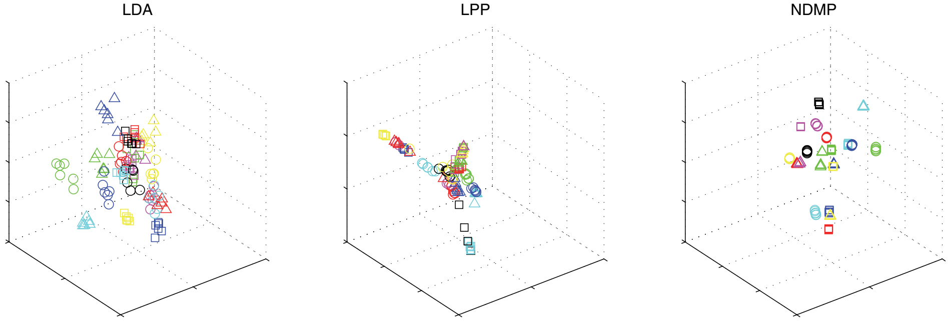


Fig. 4. Data points of a 20-class exemplar set plotted in LDA *(left)*, LPP *(center)*, and NDMP *(right)* feature space. Exemplars of each class are indicated by a different shape-color point.

3.5 Exemplar-driven Bayesian network classification

Many conventional still-image face recognition systems evaluate the performance of an algorithm by measuring recognition accuracies or error rates, which can simply be computed based on the number of correctly or incorrectly identified test images in a test set. In video-based evaluation, this is usually extended to a voting scheme, where the face in every video frame is identified independently, and then a voting method (typically majority vote or confidence-based methods such as sum rule and product rule) is performed to decide on the overall identity of the person in the sequence.

In this work, we propose a new classification method for video sequences using an exemplar-driven Bayesian network (EDBN) classifier, which introduces a joint probability function that is estimated by *maximum a posteriori* (MAP) decision rule within a Bayesian inference model. In comparison to other Bayesian methods (Fan et al., 2005; Liu et al., 2006), the EDBN classifier incorporates temporal continuity between successive video frames with consideration for the causal relationships between exemplars and their parent classes.

In a Bayesian inference model (Duda et al., 2000), the subject identity of a test video X can be found by estimating the MAP probability decision rule,

$$c^* = arg \max_C P(c|x_{1,\ldots,N_c}), \tag{14}$$

where the subscript notation of x succintly represents a sequence of N images.

In a typical Naive Bayes classifier, estimation based on MAP decision rule can be expressed as

$$P(c|x_{1,\ldots,N_c}) = \frac{P(c)P(x_{1,\ldots,N_c}|c)}{P(x_{1,\ldots,N_c})} = \frac{P(c)P(x_{1,\ldots,N_c}|c)}{\sum_c P(x_{1,\ldots,N_c}|c)P(c)}, \tag{15}$$

where $P(c)$ is the prior probability of each class, $P(x_{1,\ldots,N_c}|c)$ is the likelihood probability of x given class c and the denominator is a normalization factor to ensure that the sum of the likelihoods over all possible classes equals to 1.

Within an embedding space, the extracted exemplars can be irregularly located due to varying degree or magnitude of appearances. Thus, they should be weighted according to their influence or contribution in the subspace. Intuitively, contribution of exemplars that lie farther from the within-class exemplar subspace (more limited or uncommon) should be emphasized while exemplars that are near the within-class exemplar subspace (more likely found) should contribute at a lesser degree. To introduce causal relationship between exemplars and their parent classes, we formulate a joint probability function,

$$P(c, E, X) = P(X|c, E)P(E|c)P(c), \tag{16}$$

which includes the exemplar-class set E as a new latent variable. The graphical model of the EDBN classifier is shown in Fig. 5. Thus, the MAP classifier is redefined by maximizing the *joint posterior probability* of the class c and exemplar-class E given observation X:

$$\begin{aligned}\max_C P(c, E|X) &= \max_C \frac{P(c, E, X)}{P(X)} \\ &= \max_C \sum_{j=1}^{M} \frac{P(X|c, e_{c,j})P(e_{c,j}|c)P(c)}{P(X)} \\ &= \max_C \sum_{j=1}^{M} \prod_{i=1}^{N_c} \frac{P(x_i|c, e_{c,j})P(e_{c,j}|c)P(c)}{P(x_i)}.\end{aligned} \tag{17}$$

The marginal probability $P(x_i)$ does not depend on both c and E, thus functioning only as a normalizing constant. Since the class prior probability $P(c)$ is assumed to be non-informative at the start of observation sequence X, using uniform priors is a good estimation. The conditional probability $P(e_{c,j}|c)$ represents the *exemplar prominence probabilitiy* while $P(x_i|c, e_{c,j})$ is the class likelihood probability.

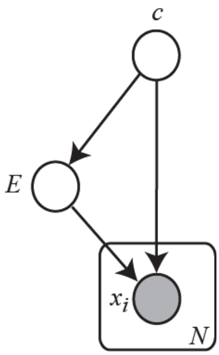


Fig. 5. Graphical model of the exemplar-driven Bayesian network (EDBN) classifier

3.5.1 Computation of class likelihood probability

Conventional Bayesian classifiers typically estimate the distribution of data using a multivariate Gaussian density function. However, accurate estimation of distribution can be challenging with the limited sample size in our problem setting. Alternatively, we can perform non-parametric density estimation by applying distance measures (or a kernel density estimator with uniform kernel), which are computationally inexpensive.

We define Frame Similarity Score (FSS) as the reciprocal of the ℓ^2-norm distance between the observed face image x_i and the j-th exemplar-class $e_{c,j}$,

$$S_i^{FSS}(x_i, e_{c,j}) = \frac{1}{d_{\ell^2}(x_i, e_{c,j})}. \tag{18}$$

The likelihood probability of the test face image x_i given the class c and exemplar-class e is determined by the ratio of FSS for exemplar-class $e_{c,j}$ to the total sum of FSS across all $C \times M$ exemplar-classes,

$$P(x_i|c, e_{c,j}) = \frac{S_i^{FSS}(x_i, e_{c,j})}{\sum_{k=1}^{C}\sum_{m=1}^{M} S_i^{FSS}(x_i, e_{k,m})}. \tag{19}$$

3.5.2 Computation of exemplar prominence

Causal relationship between exemplars and their parent classes can be represented by the exemplar prominence probability $P(e_{c,j}|c)$. Similar to the computation of likelihood probability, we avoid estimating density functions by representing the influence of an exemplar in its own class subspace using a normalized Hausdorff distance metric,

$$d_h(e_{c,j}, E_c) = \frac{1}{\kappa} \min_{e' \in E_c} \left\| e_{c,j} - e' \right\| , \tag{20}$$

where E_c is the exemplar set of class c and κ is a normalizing factor to ensure that distances are relatively scaled for each class.

By defining Exemplar Prominence Score (EPS) as the reciprocal of the distance metric,

$$S_{c,j}^{EPS}(E_c, e_{c,j}) = 1/d_h(e_{c,j}, E_c), \tag{21}$$

the exemplar prominence probability can be formulated as

$$P(e_{c,j}|c) = \frac{S_{c,j}^{EPS}(E_c, e_{c,j})}{\sum_{m=1}^{M} S_{c,j}^{EPS}(E_c, e_{c,m})}\,, \tag{22}$$

which can be pre-computed since it does not depend on input observation X.

4. Experimental setup

Unlike VFR, still image-based face recognition uses very standardized evaluation procedures and there exist many benchmark datasets up-to-date (Gross, 2004). Due to the different evaluation settings used for video-based face recognition, it is generally difficult to make direct comparisons between results reported in different works in the literature. Also, a large variety of datasets have been used or customized for the purpose of video-based face recognition.

4.1 Datasets

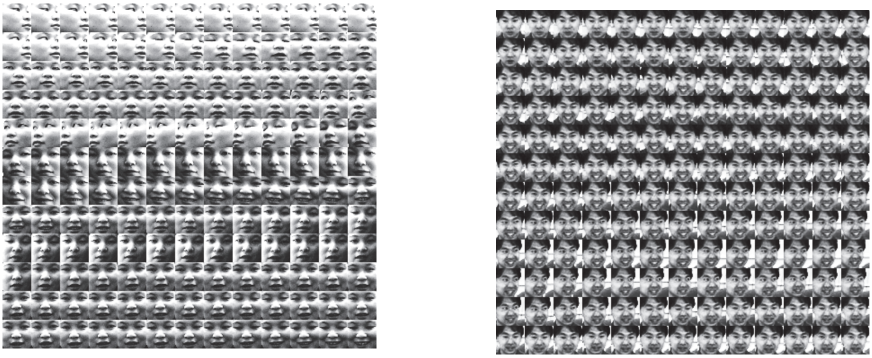


Fig. 6. Sample face images of a video sequence from the Honda/UCSD *(left)*, and CMU MoBo *(right)* datasets

In order to ensure a comprehensive evaluation was conducted, we use two standard video face datasets: Honda/UCSD Face Video Database (Lee et al., 2005) and CMU Motion of Body (MoBo) (Gross & Shi, 2001).

The first dataset, Honda/UCSD, which was collected for the purpose of VFR, consists of 59 video sequences of 20 different people (each person has at least 2 videos). Each video contains about 300-600 frames, comprising of large pose and expression variations and significant 3-D head rotations. The second dataset, CMU MoBo is another commonly used benchmark dataset customized for VFR, although it was originally intended for human motion analysis. It consists of 96 sequences of 24 different subjects (each person has 4 videos). Each video contains about 300 frames.

For both datasets, faces were extracted using the Viola-Jones face detector (Viola & Jones, 2001) and IVT object tracker (Ross et al., 2008) (which is very robust towards out-of-plane head movements), to ensure all frames with the presence of a face are successfully processed. The extracted faces are resized to grayscale images of 32 × 32 pixels, followed by histogram equalization to normalize lighting effects. Sample face images of a video sequence from both datasets are shown in Fig. 6.

4.2 VFR evaluation settings

For each subject, one video sequence is used for training, and the remaining video sequences for testing. To ensure extensive evaluation on the datasets, we construct our test set by randomly sampling 20 sub-sequences consisting of 100 frames from each test video sequence. We use 7 exemplars for Honda/UCSD and 6 exemplars for CMU MoBo[3]. Tuning parameters for STHAC-GF and STHAC-LP were set at $\alpha = 0.75$ and $\lambda = 0.2$ respectively.

5. Evaluation results and discussions

To evaluate our proposed exemplar-based spatio-temporal framework for video-based face recognition, we conduct a collection of experiments to test the effectiveness of novel spatio-temporal representations at various levels of the VFR task. Finally, some rank-based identification results are reported.

5.1 Exemplar extraction/clustering methods

Focusing our attention first to exemplar extraction/clustering methods, we perform experiments on the following exemplar-based methods:

- Random exemplar selection
- LLE+*k*-means clustering (used in (Hadid & Peitikäinen, 2004))
- Geodesic distances + HAC (used in (Fan & Yeung, 2006))
- LLE + HAC
- LLE + STHAC-GF
- LLE + STHAC-LP

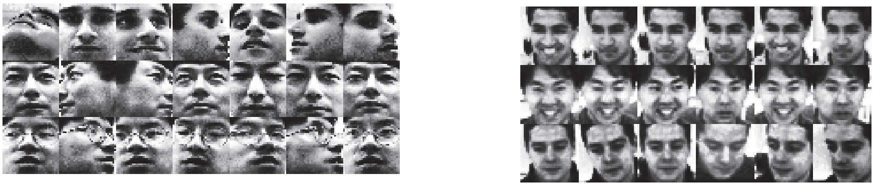


Fig. 7. Sample extracted exemplar sets of three different training videos (one in each row) from the Honda/UCSD *(left)*, and CMU MoBo *(right)* datasets.

To narrow the scope of our experiments, we only conduct this experiment on the Honda/UCSD dataset since it possesses much larger and more complex face variations compared to the CMU MoBo. Test sequences are evaluated using the proposed

[3] The number of exemplars selected from each video is heuristically determined using the "elbow" rule of thumb from the residual error curve of Ward’s distance criterion.

Exemplar-driven Bayesian Network (EDBN) classifier. Some sample extracted exemplars from both datasets are shown in Fig. 7.

Methods\Feature	PCA	LDA	NPE	NDMP
Random selection	63.68	64.81	65.68	66.09
LLE + *k*-means	68.54	70.43	65.36	73.66
Geodesic + HAC	73.69	71.30	66.07	76.75
LLE + HAC	66.18	71.20	71.70	86.90
LLE + STHAC-GF	74.89	76.94	80.68	95.33
LLE + STHAC-LP	81.91	87.21	90.84	94.52

Table 1. Average recognition rates (%) of different exemplar extraction methods on the Honda/UCSD

Table 1 demonstrates the superiority of our proposed spatio-temporal hierarchical agglomerative clustering (STHAC) technique in partitioning data into relevant clusters for exemplar extraction. The Global Fusion variant (STHAC-GF) performs slightly better than the Local Perturbation variant (STHAC-LP) when NDMP features are used. However, the STHAC-LP is notably more effective using the other tested features. This substantiates our hypothesis that the usage of the temporal dimension plays a vital role in processing continuous stream data such as video. Conventional spatial techniques such as HAC and *k*-means clustering do not utilize the temporal continuity to its good potential.

5.2 Feature representation methods

In our second evaluation, we examine the effectiveness of different feature representation methods in extracting meaningful features that can produce the best possible recognition results. Experiments were conducted on both Honda/UCSD and CMU MoBo datasets, with six different feature extraction techniques: PCA, LDA, LPP, MFA, NPE and our recently-proposed NDMP method. Training exemplars are selected using the STHAC-GF clustering method. Test sequences are evaluated using the proposed Exemplar-driven Bayesian Network (EDBN) classifier (EDBN).

Feature\Datasets	Honda/UCSD	CMU MoBo
PCA	74.89	85.32
LDA	78.04	89.79
LPP	75.00	91.41
MFA	58.88	80.68
NPE	81.62	91.80
NDMP	95.33	95.55

Table 2. Average recognition rates (%) of different feature representation methods on the evaluated datasets

The NDMP method is able to produce the best results among the tested state-of-art methods by extracting features that are discriminative, structure-constraining and neighborhood-preserving, as shown in Table 2. It is worth noticing that classic methods for still image-based recognition such as PCA and LDA struggle to find meaningful features within a complex face manifold. Our method also performed better than recent neighborhood-preserving methods such as MFA and NPE.

Overall, it can also be observed that evaluation on the CMU MoBo dataset yielded better recognition rates than the Honda/UCSD dataset. This is an expected outcome as the Honda/UCSD dataset contains much more challenging pose variations, head movements and illumination changes.

5.3 Classification methods

In our third evaluation, we fixed the methods used for the exemplar extraction/clustering and feature representation steps to our best options (STHAC-GF and NDMP respectively) while testing the recognition capability of an array of classification schemes. Our experiment implements the following schemes:

1. **Majority voting** (with nearest neighbor matching), where a vote is taken in each frame and the class with the majority vote is classified as the subject.
2. **Probabilistic voting**, where the likelihood probabilities of each frame (from Eq. 19) are combined cumulatively by simple sum rule. The class with the highest cumulative likelihoods is classified as the subject. The class with the largest sum of likelihoods is classified as the subject.
3. **Naive Bayes classifier** (from Eq. 15)
4. **Exemplar-driven Bayesian Network (EDBN) classifier** (from Eq. 17)

Classifier\Datasets	Honda/UCSD	CMU MoBo
Majority voting	78.68	92.67
Probabilistic voting	83.95	93.10
Naive Bayes	90.67	94.04
EDBN	95.33	95.55

Table 3. Average recognition rates (%) of different classification methods on the evaluated datasets

From the results in Table 3, it is clear that Bayesian classifiers performed better in the video-based setting where rapidly changing face pose and expression can easily cause recognition failure. It is no surprise that the superiority of the EDBN classifier is more pronounced on the Honda/UCSD dataset, where facial appearances change quickly. In Fig. 8, the posterior plot of a sample test sequence from the Honda/UCSD dataset demonstrates that the EDBN classifier is capable of arriving at the correct identity, even if initial frames were incorrectly classified.

Unlike conventional voting schemes, the major advantage of our framework is the ability to incorporate temporal dependencies between successive frames. Also, the establishment of causality between exemplars and their parent classes in our proposed Bayesian network slightly enhances recognition accuracy since exemplars that are more prominent are given more influence in the model and *vice versa*.

Our proposed framework can achieved the top recognition rate of more than 95% on both Honda/UCSD and CMU MoBo datasets using the best combination of spatio-temporal representations – exemplar extraction by clustering with STHAC-GF (Global Fusion scheme), representating the exemplar features with Neighborhood Discriminative Manifold Projection

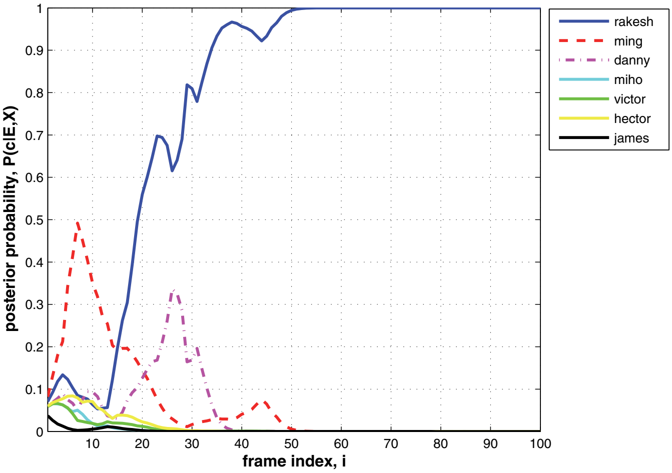


Fig. 8. Plot of posterior $P(c|E, X)$ versus frame index i of a sample 100-frame *rakesh* video subsequence from Honda/UCSD dataset. Posterior probabilities of the seven most probable subjects are shown in different colors. The subject *rakesh* (in blue line) is correctly identified for this test subsequence.

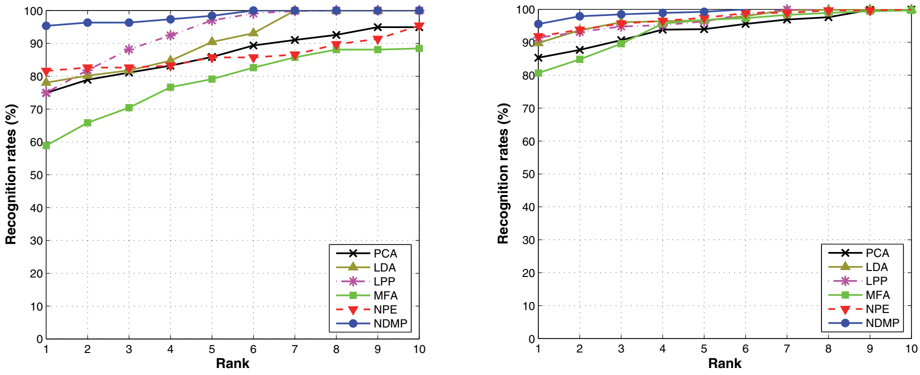
(NDMP) dimensionality reduction method, and classification with the Exemplar-Driven Bayesian Network (EDBN) classifier.

5.4 Rank-based identification

We further evaluate the reliability of the proposed spatio-temporal framework in a rank-based identification setting, by presenting its performance using a cumulative match curve (CMC). To accomodate this setting, we alter the MAP decision rule in Eq. 14 to take the top-n matches (insted of the maximum) based on the posterior probabilities. The subject in the test sequence is identified if it matches any class from the top-n matches. We gauge the rank-based performance of our proposed framework by comparing the effect of using different feature representation and clustering methods through their respective CMCs.

Figs. 9(a) and 9(b) shows the CMC of using different feature representation methods (PCA, LDA, LPP, MFA, NPE, NDMP) on both Honda/UCSD ad CMU MoBo datasets. At any given rank, the NDMP feature always yields better recognition rates compared to the other tested features. The NDMP feature is able to achieve a perfect recognition score of 100% as early as rank-6 on both datasets.

In the comparison of clustering methods in Fig. 10, the spatio-temporal HAC with Global Fusion (STHAC-GF) method is able to outperform other spatial methods such as k-means clustering and HAC in the rank-based identification setting. For both our methods, dimensionality reduction is performed using LLE. Interestingly, the Local Perturbation variant (STHAC-LP) is unable to improve its recognition rate even when the rank increases.



(a) (b)

Fig. 9. Cumulative match curves of different feature representation methods on the (a) Honda/UCSD dataset, and (b) CMU MoBo dataset. On both datasets, it is clear that using NDMP features to represent exemplars consistently yield better recognition rates than using other features at any given rank. Assume STHAC-GF and EDBN classifier are used for clustering and classification respectively.

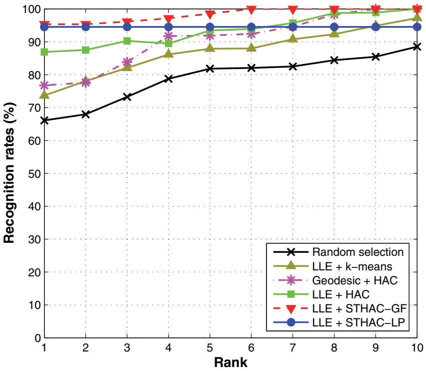


Fig. 10. Cumulative match curves of different clustering methods for exemplar extraction on the Honda/UCSD dataset. It can be observed that the proposed spatio-temporal HAC method with Global Fusion (STHAC-GF) is clearly superior to spatial clustering methods such as *k*-means and HAC. Assume NDMP and EDBN classifier are used for feature representation and classification respectively.

6. Future directions

While our proposed exemplar-based spatio-temporal framework has demonstrated promising results in our video-based face recognition evaluation, there are many avenues for future improvements in the following areas:

Clustering and Exemplar Extraction: While the straightforward STHAC-GF method is able produce good clusters for exemplar extraction, the STHAC-LP has shown unpredictable results especially in the rank-based identification setting in Section 5.4. It is possible that the optimum values for the λ_{sim} and λ_{dis} parameters in Eq. 9 have not been found, thus limiting its full potential for better performance. Future work can also extend beyond the exemplar-based approach by utilizing the already-extracted clusters as "local clusters" to create a dual exemplar-and-image set representation, with works by (Kim et al., 2007) and (Wang et al., 2008) being the primary motivations of this idea.

Feature Representation: The superior performance of the NDMP method compared to other existing state-of-art dimensionality reduction methods shows its potential usage for real-world applications such as biometric authentication, and video surveillance. Theoretically, the NDMP method can also be further generalized to a nonlinear form in kernel feature space, which may increase its robustness towards nonlinear manifolds.

Classification: The elegance of using a Bayesian network lies in its ability to encode causality between latent variables involved in the decision-making. While our EDBN model is the first step towards utilizing additional information to enhance classification, we can further formulate conditional probabilities between clusters, which can be easily simplified by computing canonical correlations.

Evaluation: With the luxury of temporal information in video, more comprehensive experiments can be further conducted to test the robustness of our proposed recognition framework in real-world scenarios, such as (1) having multiple identities in a same video sequence, (2) variable-length video sequences, and (3) degraded or low-quality video.

7. Conclusion

In this chapter, a novel exemplar-based spatio-temporal framework for video-based face recognition is presented. The paradigm of this framework involves identifying feasible spatio-temporal representations at various levels of a video-to-video recognition task. There are major contributions in all three stages of an exemplar-based approach – clustering for exemplar extraction, feature representation and classification. In the training stage, a new spatio-temporal hierarchical agglomerative clustering (STHAC) partitions data from each training video into clusters by exploiting additional temporal information between video frames. Face exemplars are then extracted as representatives of each cluster. In feature representation step, meaningful spatial features are extracted from both training and test data using the Neighborhood Discriminative Manifold Projection (NDMP) method. In the final classification task, a new exemplar-driven Bayesian network (EDBN) classifier which promotes temporal continuity between successive video frames is proposed to identify the subject in test video sequences. Extensive experiments conducted on two large video datasets demonstrated the effectiveness of the proposed framework and its associated novel methods compared to other existing state-of-art techniques. Furthermore, the reported results show promising potential for other real-world pattern recognition applications.

8. References

Arandjelovic, O., Shakhnarovich, G., Fisher, J., Cipolla, R. & Darrell, T. (2005). Face recognition with image sets using manifold density divergence, *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 581–588.

Belhumeur, P., Hespanha, J. & Kriegman, D. (1997). Probabilistic recognition of human faces from video, *IEEE Trans on PAMI* 19: 711–720.

Cox, T. & Cox, M. (2001). *Multidimensional Scaling, 2nd ed.*, Chapman and Hall.

Duda, R., Hart, P. & Stork, D. (2000). *Pattern Classification*, John Wiley.

Fan, W., Wang, Y. & Tan, T. (2005). Video-based face recognition using bayesian inference model, *Proceedings of AVBPA*, Springer-Verlag, pp. 122–130.

Fan, W. & Yeung, D.-Y. (2006). Face recognition with image sets using hierarchically extracted exemplars from appearance manifolds, *Proceedings of IEEE Automatic Face and Gesture Recognition*, pp. 177–182.

Gross, R. (2004). Face databases, *in* S. Li & A. Jain (eds), *Handbook of Face Recognition*, Springer-Verlag, Berlin.

Gross, R. & Shi, J. (2001). The cmu motion of body (mobo) database, *Technical Report CMU CMU-RI-TR-01-18*, Robotics Institute, CMU.

Hadid, A. & Peitikäinen, M. (2004). From still image to video-based face recognition: An experimental analysis, *Proceedings of IEEE Automatic Face and Gesture Recognition*, pp. 813–818.

He, X., Cai, D., Yan, S. & H.J., Z. (2005). Neighborhood preserving embedding, *Proceedings of IEEE Int. Conf. on Computer Vision*, pp. 1208–1213.

He, X. & Niyogi, P. (2003). Locality preserving projections, *Proceedings of NIPS* 16: 153–160.

Kim, T., Kittler, J. & Cipolla, R. (2007). Discriminative learning and recognition of image set classes using canonical correlations, *IEEE Trans. PAMI* 29(6): 1005–1018.

Krüeger, V. & Zhou, S. (2002). Exemplar-based face recognition from video, *Proceedings of European Conference on Computer Vision*, pp. 732–746.

Lee, K., Ho, J., Yang, M. & Kriegman, D. (2005). Visual tracking and recognition using probabilistic appearance manifolds, *Computer Vision and Image Understanding* 99(3): 303–331.

Liu, W., Li, Z. & Tang, X. (2006). Spatio-temporal embedding for statistical face recognition from video, *Proceedings of Eur. Conf. on Computer Vision*, Springer-Verlag, pp. 374–388.

Liu, X. & Chen, T. (2003). Video-based face recognition using adaptive hidden markov models, *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 340–345.

O'Toole, A., Roark, D. & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis, *Trends in Cognitive Sciences* 6(6): 261–266.

Ross, D., Lim, J., Lin, R.-S. & Yang, M.-H. (2008). Incremental learning for robust visual tracking, *Int. Journal of Computer Vision* 77(1): 125–141.

Roweis, S. & Saul, L. (2000). Nonlinear dimensionality reduction by locally linear embedding, *Science* 290: 2323–2326.

See, J. & Ahmad Fauzi, M. (2011). Neighborhood discriminative manifold projection for face recognition in video, *Proceedings of Int. Conf. on Pattern Analysis and Intelligent Robotics (ICPAIR)*, to appear.

Shakhnarovich, G., Fisher, J. & Darrell, T. (2002). Face recognition from long-term observations, *Proceedings of European Conf. on Computer Vision*, pp. 851–868.

Tenenbaum, J., de Silva, V. & Langford, J. (2000). A global geometric framework for nonlinear dimensionality reduction, *Science* 290: 2319–2323.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3(1): 71–86.

Viola, P. & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features, *Proceedings of IEEE Computer Vision and Pattern Recognition*, pp. 511–518.

Wang, R., Shan, S., Chen, X. & Gao, W. (2008). Manifold-manifold distance with application to face recognition based on image set, *Proceedings of IEEE Computer Vision and Pattern Recognition*.

Webb, A. (2002). *Statistical Pattern Recognition, 2nd ed.*, John Wiley.

Yamaguchi, O., Fukui, K. & Maeda, K. (1998). Face recognition using temporal image sequence, *Proceedings of IEEE Automatic Face and Gesture Recognition*, pp. 318–323.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q. & Lin, S. (2007). Locality preserving projections, *IEEE Trans. PAMI* 29(1): 40–51.

Zhao, W., Chellappa, R., Phillips, P. & Rosenfeld, A. (2003). Face recognition: A literature survey, *ACM Computing Surveys* 35(4): 399–485.

Zhou, S. (2004). Face recognition using more than one still image: What is more, *Proceedings of 5th Chinese Conference on Biometric Recognition, (SINOBIOMETRICS)*, pp. 225–232.

Zhou, S., Krüeger, V. & Chellappa, R. (2003). Probabilistic recognition of human faces from video, *Computer Vision and Image Understanding* 91(1–2): 214–245.

14

Real-Time Multi-Face Recognition and Tracking Techniques Used for the Interaction between Humans and Robots[1]

Chin-Shyurng Fahn and Chih-Hsin Wang[2]
National Taiwan University of Science and Technology, Taipei
Taiwan, R. O. C.

1. Introduction

The technology of biometric recognition systems for personal identification commonly manipulate the input data acquired from irises, voiceprints, fingerprints, signatures, human faces, and so on. The recognition of irises, voiceprints, fingerprints, and signatures belongs to the passive methods that require the camera with a high resolution or capture people's biometric information at a short range. These methods are not suitable for our person following robot to be developed, because they cannot provide convenience for users. Face recognition belongs to one of the active methods that users need keep away from a camera at a certain distance only. In this chapter, face recognition is regarded as a kind of human computer interfaces (HCIs) that are applied to the interaction between humans and robots. Thus, we attempt to develop an automatic real-time multiple faces recognition and tracking system equipped on the robot that can detect human faces and confirm a target in an image sequence captured from a PTZ camera, and keep tracking the target that has been identified as a master or stranger, then employ a laser range finder to measure a proper distance between target's owner and the robot.

Generally, three main procedures: face detection, face recognition, and face tracking are implemented on such a system. In literature, a large number of face localization techniques had been proposed. According to the literature (Hjelmås & Low, 2001), the methods for face detection can be basically grouped into feature-based and image-based approaches. The development of the feature-based approaches can be further divided into three areas: active shape models, feature analysis, and low-level analysis for edges, grey-levels, skin colours, motions, and so forth. On the other hand, the image-based approaches can be categorized into linear subspace methods, neural networks, and statistical approaches.

The development of face recognition is more and more advanced in the past twenty years (Zhao et al., 2003). Each system has its own solution. At present, all researches about face recognition take the features that are robust enough to represent different human faces. A

[1] This work was supported in part by the National Science Council of Taiwan (R. O. C.) under Grant NSC98-2221-E-011-119.

[2] C. S. Fahn and C. H. Wang are with the Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei 10607, Taiwan, R. O. C.

novel classification method, called the nearest feature line (NFL), for face recognition was proposed (Li et al., 1999). The derived FL can capture more variations of face images than the original feature points do, and it thus expands the capacity of an available face image database. However, if there are a lot of face images in the database, the recognition accuracy will reduce. Two simple but general strategies for a common face image database are compared and developed two new algorithms (Brunelli & Poggio, 1993); the first one is based on the computation of a set of geometrical features, such as nose width and length, mouth position, and chin shape, and the second one is based on almost-grey-level template matching. Nevertheless, under the different lighting conditions, the characteristic of geometry will change. In the literature (Er et al., 2002), the researchers combined the techniques of principal component analysis (PCA), Fisher's linear discriminant, and radial basis function to increase the correct rate of face recognition. However, their approach is unable to carry out in real time.

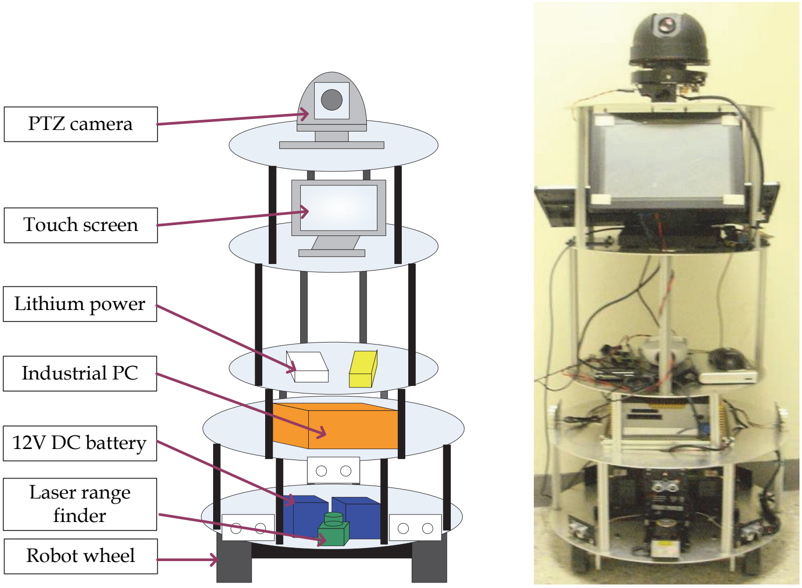
On the whole, the methods of tracking objects can be categorized into three ways: match tracking, predictive tracking, and energy functions (Fan et al., 2002). Match tracking has to detect moving objects in the entire image. Although the accuracy of match tracking is considerable, it is very time-consuming and we can not improve the performance effectively. Energy functions are often adopted in a snake model. They provide a good help for contour tracking. In general, there are two methods of predictive tracking: the Kalman filter and a particle filter. The Kalman filter has great effects when the objects move in linear paths, but it's not appropriate for the non-linear and non-Gaussian movements of objects. On the contrary, the particle filter performs well for non-linear and non-Gaussian problems. A research team completed a face tracking system which exploits a simple linear Kalman filter (An et al., 2005). They used a small number of critical rectangle features selected and trained by an AdaBoost learning algorithm, and then detected the initial position, size, and incline angle of a human face correctly. Once a human face is reliably detected, they extract the colour distributions of the face and the upper body from the detected facial regions and the upper body regions for creating respective robust colour modelling by virtue of *k*-means clustering and multiple Gaussian models. Then fast and efficient multi-view face tracking is executed using several critical features and a simple linear Kalman filter. However, two critical problems, lighting condition change and the number of clusters in the *k*-means clustering, are not solved well yet. In addition to the Kalman filter, some real-time face tracking systems based on particle filtering techniques were proposed (Fahn et al., 2009; Fahn & Lin, 2010; Fahn & Lo, 2010). The researchers utilized a particle filter to localize human faces in image sequences. Since they have considered the hair colour information of a human head, it will keep tracking even if the person is back to the sight of a camera.

In this chapter, an automatic real-time multiple faces recognition and tracking system installed on a person following robot is presented, which is inclusive of face detection, face recognition, and face tracking procedures. To identify human faces quickly and accurately, an AdaBoost algorithm is used for training a strong classifier for face detection (Viola & Jones, 2001). As to face recognition, we modify the discriminative common vectors (DCVs) algorithm (Cevikalp & Wilkes, 2004; Cevikalp et al., 2005; Gulmezoglu et al., 2001) and employ the minimum Euclidean distance to measure the similarity of a detected face image and a candidate one. To raise the confidence level, the most likely person is determined by the majority voting of ten successive recognition results from a face image sequence. In the sequel, the results of recognition will be assigned into two classes: "master" and "stranger."

In our system, the robot will track the master unceasingly except that the power is turned off. In the face tracking procedure, we propose an improved particle filter to dynamically locate multiple faces. According to the position of the target in an image, we issue a series of commands (moving forward, turning left or turning right) to drive the motors of wheels on the robot, and evaluate the distance between a master and the robot by means of a laser range finder to issue a set of commands (stop or turn backward) until the robot follows to a suitable distance in front of the master.

2. Hardware description

The following describes the hardware system of our experimental robot whose frame size is 40 cm long, 40 cm wide, and 130 cm high as Fig. 1 shows. The robot has three wheels; the motors of two front wheels driving the robot to move forward/backward and turn left/right, whereas one rear wheel without dynamic power is used for supporting the robot only. By controlling the turning directions of the two front wheels, we can change the moving direction of the robot. In order to help the robot to balance, we settle one ball caster in the rear.



(a) (b)

Fig. 1. Illustration of our experimental robot: (a) the robot frame; (b) the concrete robot.

Fig. 2 illustrates the hardware architecture of our experimental robot. We can receive the distance data of the laser range finder from the Industrial PC via RS-232. In the camera system, the images are acquired through the USB 2.0 video grabber. After processing the captured images and receiving the distance data from the laser range finder, the program will send commands through RS-232 to the motor system. The H8/3694F microprocessor can link the micro serial servo controller, and then the left and right motors of the front wheels are driven.

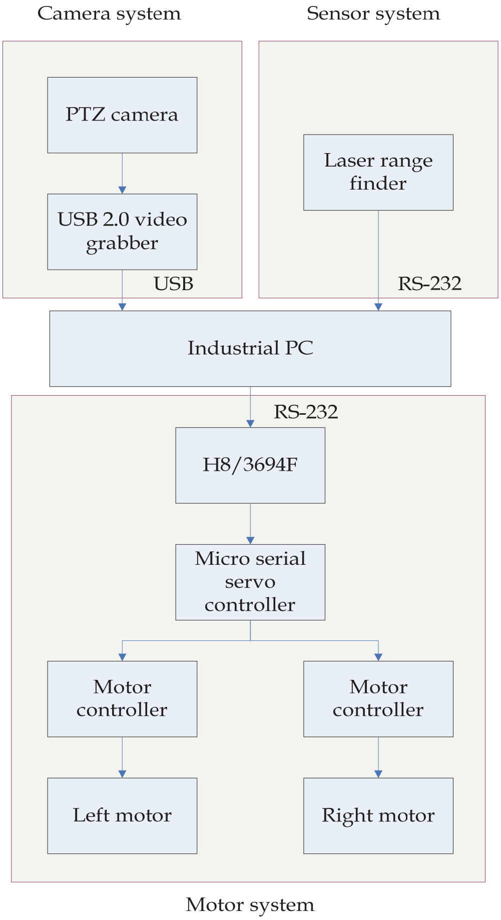


Fig. 2. Hardware architecture of our experimental robot.

3. Face detection

The face detection is a crucial step of the human face recognition and tracking system. For enabling the system to extract facial features more efficiently and perfectly, we must narrow the range of face detection first, and the performance of the face detection method can't be too low. In order to detect human faces quickly and accurately, we take advantage of an AdaBoost (Adaptive Boosting) algorithm to build a strong classifier with simple rectangle features involved in an integral image (Viola & Jones, 2001). Therefore, no matter what the size of a rectangle feature we use, the execution time is always constant. Some rectangle features are fed to the strong classifier that can distinguish positive and negative images.

Although the AdaBoost algorithm spends a lot of time on training the strong classifier, the face detection result attains high performance in the main. Fig. 3 depicts the boosting learning algorithm that we adopt for face detection.

Given: n training images (x_1,y_1), (x_2,y_2), …, and (x_n,y_n), where y_i = 0, 1 stand for negative and positive examples, respectively.

Initialize weights:

$$\begin{cases} w_{1,i} = \dfrac{1}{2m} & \text{if } m \text{ is the number of negative examples} \\ w_{1,i} = \dfrac{1}{2l} & \text{if } l \text{ is the number of positive examples.} \end{cases}$$

For t **= 1**, 2, …, T**:**

1. Normalize the weights:

$$w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}.$$

2. Compute the weighted error of each weak classifier:

$$\varepsilon_t = \min_{f,p,\theta} \sum_i w_i \left| h(x_i,f,p,\theta) - y_i \right|.$$

3. Choose the weak classifier that has the minimum error:

$$h_t(x) = h(x, f_t, p_t, \theta_t).$$

4. Update the weights:

$$\begin{cases} w_{t+1,i} = w_{t,i}\beta_t & \text{if } x \text{ is classified correctly} \\ w_{t+1,i} = w_{t,i} & \text{otherwise,} \end{cases}$$

where $\beta_t = \dfrac{\varepsilon_t}{1-\varepsilon_t}$.

The final strong classifier is

$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise,} \end{cases}$$

where $\alpha_t = \log \dfrac{1}{\beta_t}$.

Fig. 3. The boosting learning algorithm used in face detection.

The cascaded structure of weak classifiers using the AdaBoost algorithm is constructed one by one, and each weak classifier has its own order. A positive result will be processed by the next weak classifier, whereas a negative result at any currently processed point leads to the immediate rejection for the corresponding sub-window. First, we input the rectangle features via different sub-windows into many weak classifiers, and the detection error rate is the minimum in the first weak classifier that can delete a lot of negative examples, then the remaining images are more difficult to be removed, which are further processed by the successive weak classifiers. Such an operation is repeatedly executed until the last weak classifier is performed, and the remainders are face images.

In order to take advantage of the integral image, we adopt the concept like processing a pyramid of images. That is, a sub-window constituting the detector scans the input image on many scales. For example, the first detector is composed of 20×20 pixels, and an image of 320×240 pixels is scanned by the detector. After that, the current image is smaller than the previous one by 1.25 times. A fixed scale detector is then employed to scan each of these images. When the detector scans an image, subsequent locations are acquired from shifting the window by some number of pixels, $s \times \Delta$, where s is the current scale. Note that the choice of Δ affects both the speed of the detector and the accuracy of the detection. Detection results are clustered into a class to determine the representative position of the target via comparing the distance between the centres of a detection result and the class with a given threshold. The threshold changes with the current scale like shifting pixels mentioned above. If more than one detection result overlaps with each other, the final region is computed from the average of their positions of each overlapping part.

4. Face recognition

Through the face detection procedure, we can take face images that are then fed to the face recognition procedure. To begin with, we execute the image normalization to make the sizes of face images be the same. The intensity of the images will be also adjusted for reducing the lighting effect. After the size normalization and intensity adjustment, we subsequently perform the feature extraction process to obtain the feature vector of a face image. The idea of common vectors was originally introduced for isolated word recognition problems in the case that the number of samples in each class is less than or equal to the dimensionality of the sample space. The approaches to solving these problems extract the common properties of classes in the training set by eliminating the differences of the samples in each class. A common vector for each individual class is obtained from removing all the features that are in the directions of the eigenvectors corresponding to the nonzero eigenvalues of the scatter matrix of its own class. The common vectors are then used for pattern recognition. In our case, instead of employing a given class's own scatter matrix, we exploit the within-class scatter matrix of all classes to get the common vectors. We also present an alternative algorithm based on the subspace method and the Gram-Schmidt orthogonalization procedure to acquire the common vectors. Therefore, a new set of vectors called the discriminative common vectors (DCVs) will be used for classification, which results from the common vectors. What follows elaborates the algorithms for obtaining the common vectors and the discriminative common vectors (Gulmezoglu et al., 2001)

4.1 The DCV algorithm

Let the training set be composed of C classes, where each class contains M samples, and let x_m^c be a d-dimensional column vector which denotes the m-th sample from the c-th class. There will be a total of $N = MC$ samples in the training set. Supposing that $d > M - C$, three scatter matrices S_W, S_B, and S_T are respectively defined below:

$$S_W = \sum_{c=1}^{C} \sum_{m=1}^{M} (x_m^c - \overline{x}^c)(x_m^c - \overline{x}^c)^T, \tag{1}$$

$$S_B = \sum_{c=1}^{C} (\overline{x}^c - \overline{x})(\overline{x}^c - \overline{x})^T, \tag{2}$$

and
$$S_T = \sum_{c=1}^{C} \sum_{m=1}^{M} (x_m^c - \overline{x})(x_m^c - \overline{x})^T = S_W + S_B, \tag{3}$$

where $\overline{x}$ is the mean of all samples and $\overline{x}^c$ is the mean of samples in the c-th class.

In the special case, $w^T S_W w = 0$ and $w^T S_B w \neq 0$ for all $w \in R^d - 0_d$, modified Fisher's linear discriminant criterion attains a maximum. However, a projection vector w, satisfying the above conditions, does not necessarily maximize the between-class scatter matrix. The following is a better criterion (Bing et al., 2002; Turk & Pentland, 1991):

$$F(W_{out}) = \underset{\left|W^T S_W W\right|=0}{\arg\max} \left|W^T S_B W\right| = \underset{\left|W^T S_W W\right|=0}{\arg\max} \left|W^T S_T W\right|. \tag{4}$$

To find the optimal projection vectors w in the null space of S_W, we project the face samples onto the null space of S_W and then obtain the projection vectors by performing principal component analysis (PCA). To accomplish this, the vectors that span the null space of S_W must first be computed. Nevertheless, this task is computationally intractable since the dimension of this null space is probably very large. A more efficient way of realizing this task is resorted to the orthogonal complement of the null space of S_W, which significantly becomes a lower-dimensional space.

Let R^d be the original sample space, V be the range space of S_W, and $V^{\perp}$ be the null space of S_W. Equivalently,

$$V = span\left\{\alpha_k \middle| S_W \alpha_k \neq 0,\ k = 1, 2, \ldots, r\right\} \tag{5}$$

and
$$V^{\perp} = span\left\{\alpha_k \middle| S_W \alpha_k = 0,\ k = r+1, r+2, \ldots, d\right\}, \tag{6}$$

where $r < d$ is the rank of S_W, $\{\alpha_1, \alpha_2, \cdots, \alpha_d\}$ is an orthonormal set, and $\{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ is the set of orthonormal eigenvectors corresponding to the nonzero eigenvalues of S_W.

Consider the matrices $Q = \left[\alpha_1\ \alpha_2 \ldots \alpha_r\right]$ and $\overline{Q} = \left[\alpha_{r+1}\ \alpha_{r+2} \ldots \alpha_d\right]$. Since $R^d = V \oplus V^{\perp}$, every face image $x_m^c \in R^d$ has a unique decomposition of the following form

$$x_m^c = y_m^c + z_m^c, \tag{7}$$

where $y_m^c = Px_m^c = QQ^T x_m^c \in V, z_m^c = \overline{P}x_m^c = \overline{Q}\overline{Q}^T x_m^c \in V^{\perp}$, and P and $\overline{P}$ are the orthogonal projection operators onto V and $V^{\perp}$, respectively. Our goal is to compute

$$z_m^c = x_m^c - y_m^c = x_m^c - Px_m^c. \tag{8}$$

To do this, we need to find a basis in V, which can be accomplished by an eigenanalysis in S_W. In particular, the normalized eigenvectors α_k corresponding to the nonzero eigenvalues of S_W will be an orthonormal basis in V. The eigenvectors can be obtained by calculating the eigenvectors of the smaller $N \times N$ matrix defined as $S_W = A^T A$ where A is a $d \times N$ matrix of the form depicted below:

$$A = \left[x_1^1 - \overline{x}^1 \cdots x_M^1 - \overline{x}^1 \;\; x_1^2 - \overline{x}^2 \cdots x_M^C - \overline{x}^C \right]. \tag{9}$$

Let λ_k and v_k be the k-th nonzero eigenvalue and the corresponding eigenvector of $A^T A$, where $k < N - C$. Then $\alpha_k = Av_k$ will be the orthonormal eigenvector that corresponds to the k-th nonzero eigenvalue of S_W. The sought-for projection onto $V^{\perp}$ is achieved using Eq. (8). In this manner, it turns out that we obtain the same unique vector for all samples of a class as follows:

$$x_{com}^c = x_m^c - QQ^T x_m^c = \overline{Q}\overline{Q}^T x_m^c, \;\; m = 1,2,\ldots,M, \;\; c = 1,2,\ldots,C. \tag{10}$$

That is, the vector on the right-hand side of Eq. (10) is independent of the sample indexed m. We refer to the vectors x_{com}^c as the common vectors.

The theorem states that it is enough to project a single sample from each class. This will greatly reduce the computational load of the calculations. After acquiring the common vectors x_{com}^c, the optimal projection vectors will be those that maximize the total scatter of the common vectors:

$$\begin{aligned} F(W_{out}) &= \underset{|W^T S_W W| = 0}{\arg\max} \left| W^T S_B W \right| = \underset{|W^T S_W W| = 0}{\arg\max} \left| W^T S_T W \right| \\ &= \underset{W}{\arg\max} \left| W^T S_{com} W \right|, \end{aligned} \tag{11}$$

where W is a matrix whose columns are the orthonormal optimal projection vectors w_k, and S_{com} is the scatter matrix of the common vectors:

$$S_{com} = \sum_{c=1}^{C} (x_{com}^c - \overline{x}^{com})(x_{com}^c - \overline{x}^{com})^T, \;\; c = 1,2,\ldots,C, \tag{12}$$

where $\overline{x}^{com}$ is the mean of all common vectors:

$$\overline{x}^{com} = \frac{1}{C}\sum_{c=1}^{C} x_{com}^{c}. \tag{13}$$

In this case, the optimal projection vectors w_k can be found by an eigenanalysis in S_{com}. Particularly, all eigenvectors corresponding to the nonzero eigenvalues of S_{com} will be the optimal projection vectors. S_{com} is typically a large $d \times d$ matrix. Thus, we can use the smaller $C \times C$ matrix $A_{com}^{T} A_{com}$ to find nonzero eigenvalues and the corresponding eigenvectors of $S_{com} = A_{com} A_{com}^{T}$, where A_{com} is the $d \times C$ matrix of the form expressed as

$$A_{com} = \left[x_{com}^{1} - \overline{x}_{com} \cdots x_{com}^{C} - \overline{x}_{com} \right]. \tag{14}$$

There will be $C-1$ optimal projection vectors since the rank of S_{com} is $C-1$ if all common vectors are linearly independent. If two common vectors are identical, then the two classes, which are represented by this vector, cannot be distinguished from each other. Since the optimal projection vectors w_k belong to the null space of S_W, it follows that when the image samples x_m^c of the i-th class are projected onto the linear span of the projection vectors w_k, the feature vector $\Omega_c = \left[\left\langle x_m^c, w_1 \right\rangle \cdots \left\langle x_m^c, w_{C-1} \right\rangle \right]^T$ of the projection coefficients $\left\langle x_m^c, w_k \right\rangle$ will also be independent of the sample indexed m. Therefore, we have

$$\Omega_c = W^T x_m^c, \;\; m = 1,2,\ldots,M, \; c = 1,2,\ldots,C . \tag{15}$$

We call the feature vectors Ω_c discriminative common vectors, and they will be used for the classification of face images. To recognize a test image x_{test}, its feature vector is found by

$$\Omega_{test} = W^T x_{test}, \tag{16}$$

which is then compared with the discriminative common vector Ω_c of each class using the Euclidean distance. The discriminative common vector found to be the closest to Ω_{test} is adopted to identify the test image.

Since Ω_{test} is only compared to a single vector for each class, the classification is very efficient for real-time face recognition tasks. In the Eigenface, Fisherface, and Direct-LDA methods, the test sample feature vector Ω_{test} is typically compared to all feature vectors of samples in the training set. It makes these methods be impractical for real-time applications for large training sets. To overcome this, they should solve the small sample size problem alternatively (Chen et al., 2000; Huang et al., 2002).

The above method can be summarized as follows:

Step 1. Compute nonzero eigenvalues and their corresponding eigenvectors of s_w using the matrix $A^T A$, where $S_W = AA^T$ and A is given by Eq. (9). Set matrix $Q = \left[\alpha_1 \; \alpha_2 \ldots \alpha_r \right]$, where $\alpha_k, k = 1,2,\ldots,r$ are the orthonormal eigenvectors of S_W with rank r.

Step 2. Choose any sample from each class and project it onto the null space of S_W to obtain the common vectors:

$$x_{com}^c = x_m^c - QQ^T x_m^c, \;\; m = 1,2,\ldots,M, \;\; c = 1,2,\ldots,C. \tag{17}$$

Step 3. Compute the eigenvectors w_k of S_{com}, associated with the nonzero eigenvalues, by use of the matrix $A_{com}^T A_{com}$, where $S_{com} = A_{com} A_{com}^T$ and A_{com} is given in Eq. (14). There are at most $C-1$ eigenvectors corresponding to the nonzero eigenvalues to constitute the projection matrix $W = \left[w_1 \; w_2 \ldots w_{C-1} \right]$, which will be employed to obtain feature vectors as shown in Eqs. (15) and (16).

4.2 Similarity measurement

There exist a lot of methods about similarity measurement. In order to implement face recognition on the robot in real time, we select the Euclidean distance to measure the similarity of a face image to those in the face database after feature extraction with the aid of the DCV transformations.

Let the training set be composed of C classes, where each class contains M samples, and let y_m^c be a d-dimensional column vector which denotes the m-th sample from the c-th class. If y is the feature vector of a test image, the similarity of the test image and a given sample is defined as

$$d(y, y_m^c) = \left\| y - y_m^c \right\|. \tag{18}$$

And the Euclidean distance between the feature vectors of the test image y and class c is

$$d(y,c) = \min\left\{ d(y,y_1^c), d(y,y_2^c), \ldots, d(y,y_m^c) \right\}. \tag{19}$$

After the test image compared with the feature vectors in all training models, we can find out the class whose Euclidean distance is minimum. In other words, the test image is identified the most possible person, namely Id_y. Note that a range of the threshold T_c is prescribed for every class to eliminate the person who does not belong to the face database as Eq. (20) shows:

$$Id_y = \begin{cases} \arg\min\limits_{c} d(y,c) & \text{if } \min\left\{ d(y,c) \le T_c, \;\; c = 1,2,\ldots,C \right\} \\ \text{NULL} & \text{otherwise.} \end{cases} \tag{20}$$

5. Face tracking

Our detection method can effectively find face regions, and the recognition method can know the master for people in front of the robot. Then the robot will track the master using our face tracking system. On the other hand, we can exchange the roles of the master with a stranger, and the robot will resume its tracking to follow the stranger. Until now, many tracking algorithms have been proposed by researchers. Of them, the Kalman filter and particle filter are applied extensively. The former is often used to predict the state sequence of a target for linear predictions, and it is not suitable for the non-linear and non-Gaussian movement of objects. However, the latter is based on the Monte Carlo integration method and suit to nonlinear predictions. Considering that the postures of a human body are nonlinear motions, our system chooses a particle filter to realize the tracker (Foresti et al., 2003; Montemerlo et al., 2002; Nummiaro et al., 2003). The key idea of this technique is to represent probability densities by sets of samples. Its structure is divided into four major parts: probability distribution, dynamic model, measurement, and factored sampling.

5.1 Particle filter

Each interested target will be defined before the face tracking procedure operates. The feature vector of the i-th interested face in an elliptical shape in the t-th frame is denoted as follows:

$$F_{t,i} = \left\{x_{t,i}, y_{t,i}, w_{t,i}, h_{t,i}, Id_{t,i}\right\}, \tag{21}$$

where $(x_{t,i}, y_{t,i})$ is the centre of the i-th elliptical window in the t-th frame (i.e., at time step t), $w_{t,i}$ and $h_{t,i}$ symbolize the minor and major axes of the ellipse, respectively, and $Id_{t,i}$ is person’s identity assigned from the face recognition result.

According to the above target parameters, the tracking system builds some face models and their associated candidate models, each of which is represented by a particle. Herein, we select the Bhattacharyya coefficient to measure the similarity of two discrete distributions resulting from the respective cues of a face model and its candidate models existing in two consecutive frames.

Three kinds of cues are considered in the observation model; that is, the colour cue, edge cue, and motion cue. The colour cue employed in the observation model is still generated from the practice presented in the above literatures, instead of modelling the target by a colour histogram. The raw image of each frame is first converted to a colour probability distribution image via the face colour model, and then the samples are measured on the colour probability distribution image. In this manner, the face colour model is adapted to find interested regions by the face detection procedure, next handle the possible changes of the face regions due to variable illumination frame-by-frame, and the Bhattacharyya coefficient is used to calculate the similarity defined as

$$M_{C_{t,i}} = BC\left(C_{face_{t-1,i}}(l,m), C_{face_{t,i}}(l,m)\right), \tag{22}$$

where $C_{face_{t-1,i}}(l,m)$ and $C_{face_{t,i}}(l,m)$ mean the discrete distributions of the colour cues of the i-th face model and its candidate model at time steps $t-1$ and t, respectively.

In comparison to our earlier proposed methods (Fahn et al., 2009; Fahn & Lin, 2010; Fahn & Lo, 2010), the edge cue is a newly used clue. Edges are a basic feature in images, and they are not more influenced than the colour model suffering from illumination. First, the Sobel filter is applied to obtain all the edges in each frame, and the Bhattacharyya coefficient is employed to represent the similarity measurement just like the colour model:

$$M_{E_{t,i}} = BC\left(E_{face_{t-1,i}}(l,m), E_{face_{t,i}}(l,m)\right), \tag{23}$$

where $E_{face_{t-1,i}}(l,m)$ and $E_{face_{t,i}}(l,m)$ stand for the discrete distributions of the edge cues of the i-th face model and its candidate model at time steps $t-1$ and t, respectively.

As for the motion feature, our tracking system records the centre positions of interested objects in the previous num frames continued, say twenty. Both the distance and direction are then computed in the face tracking procedure for each interested object acting as a particle. At time step t, the average moving distance $Adis_{t,i}$ of the i-th particle (i.e., a candidate model) referring to its average centre positions $Ax_{t,i}$ and $Ay_{t,i}$ in the previous num frames are expressed below:

$$Ax_{t,i} = \frac{\sum_{k=1}^{num} x_{t-k,i}}{num}, \tag{24}$$

$$Ay_{t,i} = \frac{\sum_{k=1}^{num} y_{t-k,i}}{num}, \tag{25}$$

and
$$Adis_{t,i} = \frac{\sum_{k=1}^{num} \sqrt{(x_{t,i} - x_{t-k,i})^2 + (y_{t,i} - y_{t-k,i})^2}}{num}. \tag{26}$$

There are two states of the motion, including the accelerated velocity and decelerated velocity. If the interested face moves with the accelerated velocity, the distance between the current centre position and the average centre position will be larger than the average moving distance. If the interested face moves with the decelerated velocity, conversely, the distance between the current centre position and the average centre position will be less than the average moving distance.

In addition to the distance of face motion, the direction of face motion is another important factor. The angle $\theta_{t,i}$ symbolizing the direction of the i-th particle at time step t is expressed as

$$\theta_{t,i} = \frac{\left(\tan^{-1} \frac{y_{t,i} - Ay_{t,i}}{x_{t,i} - Ax_{t,i}}\right) \times 180}{\pi}. \tag{27}$$

It also depends on the average centre positions of the i-th particle in the previous num frames. All the scope of angles ranged from -90° to $+90^{\circ}$ is partitioned into four orientations, each of which covers the angle of 45°. In order to get rid of an irregular face motion trajectory, we consider the most possibility of the direction by virtue of the majority voting on the number of times of the corresponding orientation of $\theta_{t-k,i}$, $k = 1, 2, \ldots, num$. Thus, the likelihood of face motion at time step t is determined by the following equation:

$$M_{v_{t,i}} = \begin{cases} \dfrac{dis_{t,i}}{Adis_{t,i}} \times 0.25 + vote_{t,i} \times 0.2 + 0.6 & \text{if } dis_{t,i} \le Adis_{t,i} \\ \dfrac{dis_{t,i}}{Adis_{t,i}} \times 0.35 + vote_{t,i} \times 0.2 + 0.2 & \text{otherwise,} \end{cases} \tag{28}$$

where $dis_{t,i}$ is the distance between the centre position in the current frame and the average centre position in the previous num frames, $Adis_{t,i}$ is referred to Eq. (26), and $vote_{t,i}$ is the maximum frequency of the four orientations appearing in the previous num frames. Therefore, the likelihood of the motion feature combines both the measurements of the moving distance and moving direction.

When the occlusion of human faces happens, the motion cue is obvious and becomes more reliable. On the contrary, when the occlusion does not happen, we can weigh particles mainly by the observation of the colour and edge cues. As a result, we take a linear combination of colour, edge, and motion cues as follows:

$$M_{t,i} = (1-\alpha_t)\beta_t M_{c_{t,i}} + \alpha_t \beta_t M_{E_{t,i}} + (1-\beta_t) M_{v_{t,i}}, \tag{29}$$

where $0 \le \alpha_t, \beta_t \le 1$, and $\alpha_t + \beta_t = 1$. Fig. 4 shows the iteration procedure for tracking the i-th interested face using our improved particle filter.

Given: a particle set $S_{t-1,i} = \left\{ \left(s^{(j)}_{t-1,i}, \pi^{(j)}_{t-1,i} \right) \middle| j = 1,2,...,N \right\}$, a target $(s^{\otimes}_{t-1,i}, \pi^{\otimes}_{t-1,i}) \in S_{t-1,i}$, and its centre positions $x_{t-k,i}$ and $y_{t-k,i}$, $k = 1, 2, \ldots, num$ at time step $t-1$, perform the following steps:

Propagation: fluctuate each particle $s^{(j)}_{t-1,i}$ with the weight $\pi^{(j)}_{t-1,i}$ by a dynamic model to obtain the particle set $s_{t,i}$.

Observation: weigh the particles using colour, edge, and motion cues as:

1. Calculate the distance between each particle and the target by
   $Dis^{(j)}_{t,i} = \sqrt{1 - M^{(j)}_{t,i}}$, where $M^{(j)}_{t,i}$ is a linear combination of the three cues.
2. Weigh each particle with
   $\pi^{(j)}_{t,i} = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{Dis^{(j)2}_{t,i}}{2\sigma^2}}$, where σ is the standard deviation of a Gaussian distribution.

Selection: check the weights of particles. If the particle $s^{(j)}_{t,i}$ with $\pi^{(j)}_{t,i}$ exceeds or equals a threshold, then duplicate $K^{(j)}_{t,i}$ even-weighted particles from it; otherwise, ignore the particle.

Estimate: acquire the new target $s^{\otimes}_{t,i}$ with the largest weight as:

$\pi^{\otimes}_{t,i} = \pi^{*(k)}_{t,i}$ for $k = \arg \max\limits_{k} \pi^{*(k)}_{t,i}$ and its centre positions $x^{\otimes}_{t,i}$ and $y^{\otimes}_{t,i}$.

Fig. 4. The iteration procedure for tracking the i-th interested face by the improved particle filter.

After the current filtering process is finished, we execute the next filtering process again until a local search is achieved. Such a method will track faces steadily even if humans move very fast. In the first filtering process, we establish thirty particles to track faces. But, in the subsequent filtering processes, only ten particles are set up to facilitate the face tracking operation.

5.2 Occlusion handling

Occlusion handling is a major problem in visual surveillance (Foresti et al., 2003). The occlusion problem usually occurs in a multi-target system. When multiple moving objects occlude each other, both the colour and edge information are more unreliable than the velocities and directions of moving objects. In order to solve the occlusion problem, we have improved the strategies of the filtering process and modified the weights of the particles depicted in the previous subsection. Compared with the earlier proposed methods (Fahn et al., 2009; Fahn & Lin, 2010; Fahn & Lo, 2010), the major improvement is that the directions of moving objects in a number of the past frames are incorporated into the constituents of the motion cue. We will check whether the occlusion ceases to exist. If only one human face is detected in the face detection procedure, then the occlusion will not occur in this condition. Nevertheless, if two and more faces are detected, the occlusion will occur when moving faces are near to each other, and the system will go to the occlusion handling mode.

5.3 Robot control

We can predict the newest possible position of a human face by means of the particle filtering technique. According to the possible position, we issue a control command to make the robot track the i-th face continuously. Such robot control comprises two parts: PTZ camera control and wheel motor control. The following is the command set of robot control described at time step *t*, where the first two commands belong to the PTZ camera control and the remaining ones used for the wheel motor control:

$$S_{Command_{t,i}} = \{up_{t,i}, down_{t,i}, forward_{t,i}, stop_{t,i}, backward_{t,i}, left_{t,i}, right_{t,i}\}. \tag{30}$$

5.3.1 PTZ camera control

The principal task of the PTZ camera control is to keep the i-th face appearing in the centre of the screen. It prevents the target from situating on the upper or lower side of the scope of the screen. If $y_{t,i}$ is lower than or equal to 50, then we assign the command $up_{t,i}$ to the PTZ camera at time step *t*; if $y_{t,i}$ is greater than or equal to 190, then we assign the command $down_{t,i}$ to the PTZ camera at time step *t*, as shown in Eq. (31):

$$Command_{t,i} = \begin{cases} up_{t,i} & \text{if } y_{t,i} \le 50 \\ down_{t,i} & \text{if } y_{t,i} \ge 190 \\ \text{no-operation} & \text{otherwise,} \end{cases} \tag{31}$$

where $y_{t,i}$ is the centre of the i-th elliptic window in the *y*-direction at time step *t*.

5.3.2 Wheel motor control

The main task of the wheel motor control is to direct the robot to track the i-th face continuously. We utilize two kinds of information to determine which command is issued. The first information is the position of the face appearing in the screen. According to this, we assign the commands to activate the wheel motors that make the robot move appropriately. If $x_{t,i}$ is lower than or equal to 100, then we assign the command $left_{t,i}$ to the wheel motors at time step *t*; if $x_{t,i}$ is greater than or equal to 220, then we assign the command $right_{t,i}$ to the wheel motors at time step *t* as Eq. (32) shows:

$$Command_{t,i} = \begin{cases} left_{t,i} & \text{if } x_{t,i} \leq 100 \\ right_{t,i} & \text{if } x_{t,i} \geq 220 \\ \text{no-operation} & \text{otherwise,} \end{cases} \tag{32}$$

where $x_{t,i}$ is the centre of the i-th elliptic window in the *x*-direction at time step *t*.

The second information is the laser range data. We equip a laser range finder on the front of the robot to measure the distance $D_{t,i}$ between the robot and a target which is allowed to have the incline angle ranged from -10° to $+10^{\circ}$. In accordance with the distance $D_{t,i}$, we can assign the commands to the wheel motors to control the movement of the robot. If $D_{t,i}$ is lower than or equal to 60, then we assign the command $backward_{t,i}$ to the wheel motors at time step *t*; if $D_{t,i}$ is greater than or equal to 120, then we assign the command $forward_{t,i}$ to the wheel motors at time step *t*; otherwise, we assign the command $stop_{t,i}$ to the wheel motors at time step *t*, as stated in Eq. (33):

$$Command_{t,i} = \begin{cases} backward_{t,i} & \text{if } D_{t,i} \leq 60 \\ forward_{t,i} & \text{if } D_{t,i} \geq 120 \\ stop_{t,i} & \text{otherwise.} \end{cases} \tag{33}$$

Through controlling the robot, we can increase its functionality of interacting with people. If the robot can recognize humans' identities and follow him/her abidingly, then the relation between the robot and human being is definitely closer to each other, and it can be more extensive to use. So, we control the movement of the robot on the condition of state changes between the face tracking and face recognition modes. After face detection and target confirmation, the system will switch to the recognition mode and start the robot to execute a certain task.

6. Experimental results

Table 1 list both the hardware and software that are used for the development of the real-time face tracking and recognition system installed on our robot. All the experimental results presented in this section were obtained from taking an average of the outcome data for different 10 times to demonstrate the effectiveness of the proposed methods.

Hardware	Software
Host computer: Intel Core 2 Duo CPU T2300 2.0GHz with 1.0 GB RAM PTZ camera: Sony product FCB-EX480C Microprocessor: H8/3694F	Operating system: Microsoft Windows XP SP3 equipped with DirectX 9.0 Developing tool: Borland C++ Builder 6.0 H8/3694F Editor Software

Table 1. The Developing Environment of Our Face Tracking and Recognition System Installed on the Robot

6.1 Face detection

The training database for face detection consists of 861 face images segmented by hand and 4,320 non-face images labelled by a random process from a set of photographs taken in outdoors. The sizes of these face and non-face images are all 20×20 pixels. Fig. 5 illustrates some examples of positive and negative images used for training. Notice that each positive image contains a face region beneath the eyebrows and beyond the half-way between the mouth and the chin.

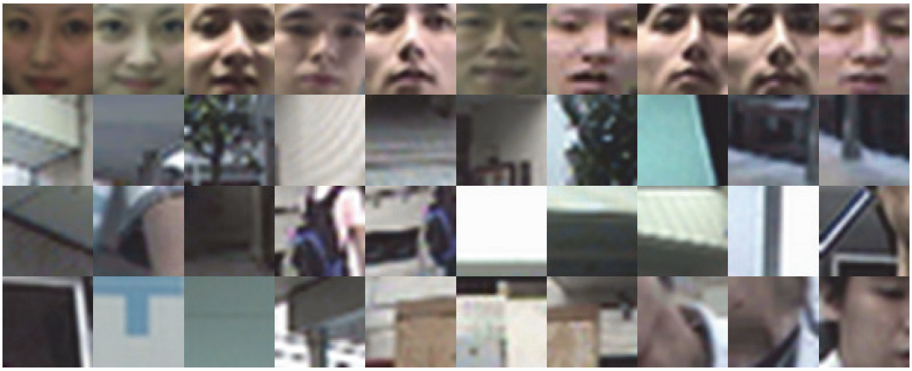


Fig. 5. Some examples of positive and negative images used for training. The first row is face images and the others are non-face images.

We perform the experiments on face detection in two different kinds of image sequences: ordinary and jumble. The image sequence is regarded as a jumble if the background in this sample video is cluttered; that is, the environment is possessed of many pixels like skin colours and/or the illumination is not appropriate. Then we define the “error rate” as a probability of detecting human faces incorrectly; for example, regarding an inhuman face as a human face. And the “miss rate” is a probability which a target appears in the frame but not detected by the system. Table 2 shows the face detection rates for the above experiments. From the experimental results, we can observe that the ordinary image sequences have better correct rates, because they encounter less interference, and the error rates are lower in this situation. The performance on the jumble image sequences is inferior to that on the ordinary ones. The main influence is due to the difference of luminance, but the effect of the colour factor is comparatively slight, particularly for many skin-like regions existing in the background. It is noted that the error rate of the jumble image sequences is larger than that of the ordinary ones, because our database has a small number of negative examples. That will cause the error rate raise. Fig. 6 demonstrates some face detection results from using our real-time face detection system in different environments.

Measurement / Experiment	The number of faces	The number of detected faces	Detection rate	Error rate	Miss rate
Ordinary image sequence	50	48	96.00%	2.00%	2.00%
	100	94	94.00%	2.00%	4.00%
	150	142	94.67%	0.67%	4.67%
	200	187	93.50%	1.00%	5.50%
Jumble image sequence	50	45	90.00%	6.00%	4.00%
	100	91	91.00%	5.00%	4.00%
	150	134	89.33%	6.00%	4.67%
	200	181	90.50%	6.50%	3.00%

Table 2. The Face Detection Rates of Two Different Kinds of Image Sequences

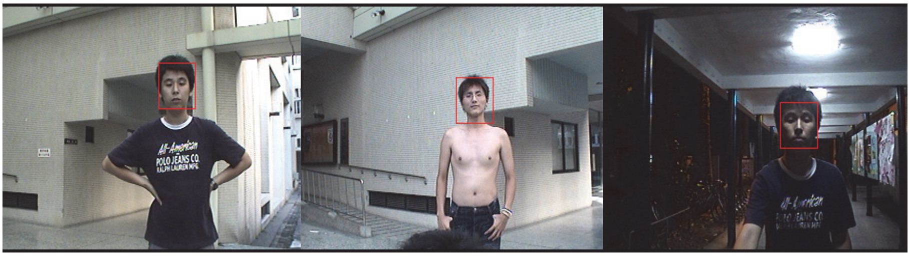


Fig. 6. Some face detection results from using our real-time face detection system.

6.2 Face recognition

Our face image database comprises 5 subjects to be recognized, and each of them has 20 samples. The nicknames are Craig, Eric, Keng-Li, Jane, and Eva (3 males and 2 females) respectively. The facial expressions (open or closed eyes and smiling or non-smiling) are also varied. The images of one subject included in the database are shown in Fig. 7.

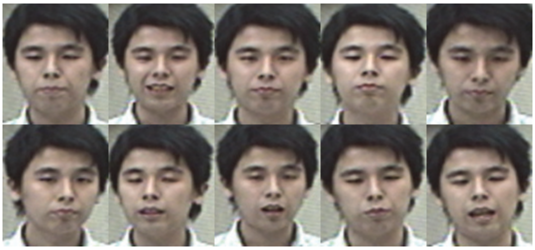
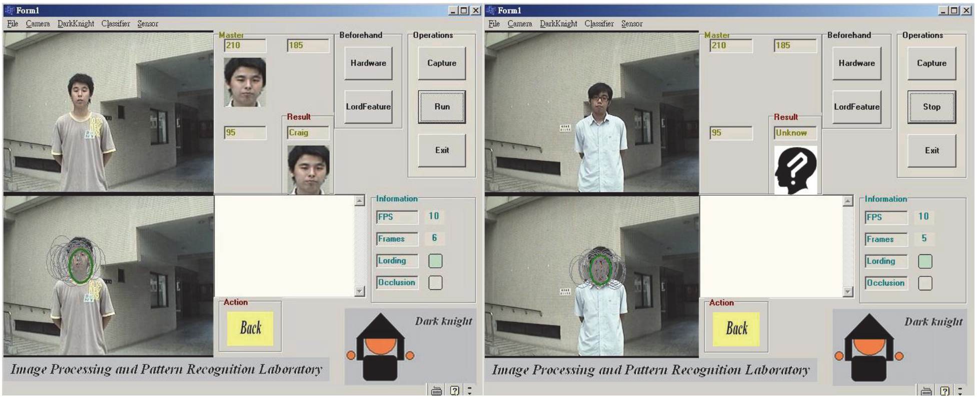


Fig. 7. Images of one subject included in our database.

Person \ Subject	Craig	Eric	Keng-Li	Jane	Eva
Craig	0~210	374~461	420~513	207~351	238~491
Eric	381~437	0~185	483~526	525~637	448~664
Keng-Li	372~450	348~493	0~95	366~513	418~622
Jane	215~429	524~618	342~572	0~226	259~373
Eva	225~411	459~586	409~597	276~356	0~100

Table 3. The Range of Thresholds for Each Class to Eliminate the Corresponding Person from the Possible Subjects

We have our robot execute face recognition in real time using the DCV method. After comparing the feature vector of the captured and then normalized face image with those of all training models, we can find out the class whose Euclidean distance is the minimum. In other words, it is identified as the most possible person. Table 3 presents the range of thresholds for each class to eliminate the corresponding person from the possible subjects included in the database. If the feature similarity of a captured and normalized face image is lower than or equal to the threshold value for every class, the corresponding person is assigned to one of classes. For every 10 frames per a period, we rank the frequency of assignments for each class to determine the recognition result. Fig. 8 graphically shows some face recognition results from using our real-time face recognition system installed on the robot.



Master: 210, 185, 95 — Result: Craig — FPS: 10 — Frames: 6

Master: 210, 185, 95 — Result: Unknow — FPS: 10 — Frames: 5

Fig. 8. Some face recognition results from using our real-time face recognition system installed on the robot.

6.3 Face tracking

Several different types of accuracy measures are required to evaluate the performance of tracking results. In this subsection, we utilize track cardinality measures for estimating the tracking rates. The reference data should be first determined in order to obtain any of the accuracy measures. Let A be the total number of actual tracks as the reference data, R be the total number of reported tracks, A' be the number of actual tracks that have corresponding reported tracks, and R' be the number of reported tracks that do not correspond to true tracks. The measures are determined by m_c, the fraction of true tracks, f_c, the fraction of reported tracks that are false alarms where they do not correspond to true tracks, and the miss rate is m_d. The following defines these variables:

$$m_c = \frac{A'}{A}, \tag{34}$$

$$f_c = \frac{R'}{A}, \tag{35}$$

and

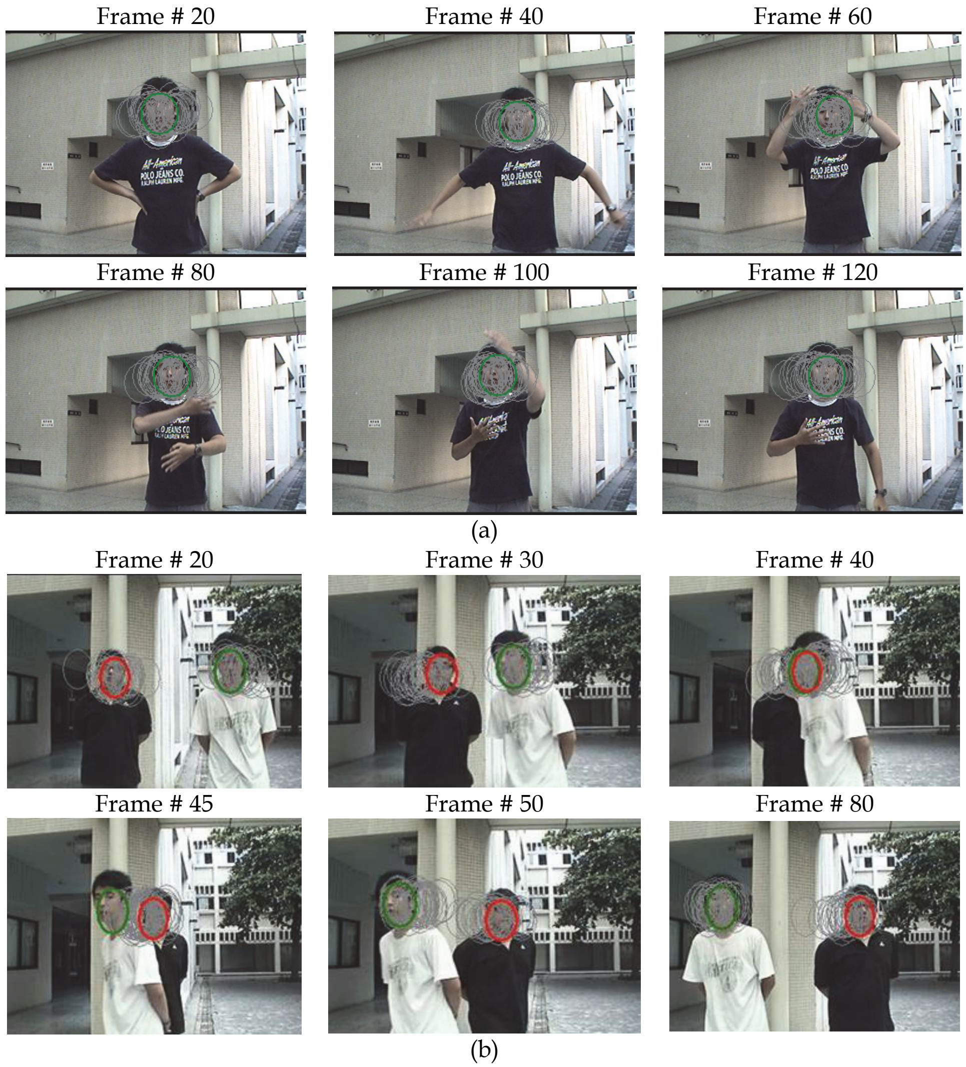
$$m_d = 1 - \frac{R}{A}. \tag{36}$$

The types of our testing image sequences are roughly classified into three kinds: a single face (Type A), multiple faces without occlusion (Type B), and multiple faces with occlusion (Type C). Each type has three different image sequences, and we will perform five tests on each of them. In these experiments, we use 30 particles to track human faces in each frame for the first filtering process, followed by 10 particles. Then we analyze the performance of face tracking on the three kinds of testing image sequences according to the aforementioned evaluation methods. Table 4 shows the face tracking performance for the above experiments.

Estimate Experiment	A	R	A'	R'	f_c (%)	m_d (%)	m_c (%)
Type A	100	99	98	1	1.00	1.00	98.00
	150	147	146	1	0.67	2.00	97.33
	200	198	196	2	1.00	1.00	98.00
	250	248	246	2	0.80	0.80	98.40
Type B	100	99	98	1	1.00	1.00	98.00
	150	148	147	1	0.67	1.33	98.00
	200	198	195	3	1.50	1.00	97.50
	250	247	245	2	0.80	1.20	98.00
Type C	100	98	83	15	15.00	2.00	83.00
	150	146	124	23	15.33	2.67	82.00
	200	194	167	27	13.50	3.00	83.50
	250	241	205	36	14.40	3.60	82.00

Table 4. The Face Tracking Performance of Three Different Kinds of Image Sequences

From the experimental results, the tracking accuracy of Type C is lower than those of the other two types. This is because when the occlusion happens, the particles are predicted on the other faces that are also the colour regions. On the other hand, the correct rates of face tracking for Types A and B are similar to each other; that is, the tracking performance for a single face is almost the same to that for multiple faces without occlusion. When the robot and the target move simultaneously, some objects whose colours are close to the skin colour may exist in the background. In this situation, we must consider the speed of the target and assume that the target disappearing on the left side of the image does not appear on the right side immediately. Through simple distance limits, the robot can track the target continuously. Moreover, at regular intervals we will either abandon the currently tracked targets then intend to detect the face regions once more, or keep tracking on the wrong targets until the head moves near the resampling range of the particle filter. Attempting to enlarge the resampling range can slightly conquer this problem. However, the wider the resampling range is, the sparser the particles are. It has good outcomes for the robot to track faces, but sometimes it will reduce the tracking correctness. Fig. 9 shows some face tracking results from using our real-time face tracking system equipped on the robot.



(a)

(b)

Fig. 9. Some face tracking results from using our real-time face tracking system for: (a) a single human; (b) multiple humans within the field of view of the PTZ camera equipped on the robot.

7. Conclusion and feature works

In this chapter, we present a completely automatic real-time multiple faces recognition and tracking system installed on a robot that can capture an image sequence from a PTZ camera, then use the face detection technique to locate face positions, and identify the detected faces as the master or strangers, subsequently track a target and guide the robot near to the target continuously. Such a system not only allows robots to interact with human being adequately, but also can make robots react more like mankind.

Some future works are worth investigating to attain better performance. In the face recognition procedure, if the background is too cluttered to capture a clear foreground, the recognition rate will decrease. Because most of our previous training samples were captured in a simple environment, sometimes static objects in the uncomplicated background are identified as the foreground. We can increase some special training samples in a cluttered background to lower the miss rate during the face detection. Of course, it will raise the face recognition accuracy, but need a lot of experiments to collect special and proper training samples.

8. References

K. H. An, D. H. Yoo, S. U. Jung, and M. J. Chung, "Robust multi-view face tracking," in Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Alberta, Canada, pp. 1905-1910, 2005.

Y. Bing, J. Lianfu, and C. Ping, "A new LDA-based method for face recognition," in Proceedings of the International Conference on Pattern Recognition, Quebec, Canada, vol. 1, pp. 168-171, 2002.

R. Brunelli and T. Poggio, "Face recognition: features versus templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, pp. 1042-1052, 1993.

H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana, "Discriminative common vectors for face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 1, pp. 4-13, 2005.

H. Cevikalp and M. Wilkes, "Face recognition by using discriminative common vectors," in Proceedings of the International Conference on Pattern Recognition, Cambridge, United Kingdom, vol. 1, pp. 326-329, 2004.

L. F. Chen, H. Y. M. Liao, M. T. Ko, J. C. Lin, and G. J. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," Pattern Recognition, vol. 33, no. 10, pp. 1713-1726, 2000.

M. J. Er, S. Wu, J. Lu, and H. L. Toh, "Face recognition with radial basis function (RBF) neural networks," IEEE Transactions on Neural Networks, vol. 13, no. 3, pp. 697-710, 2002.

C. S. Fahn, M. J. Kuo, and K. Y. Wang, "Real-time face tracking and recognition based on particle filtering and AdaBoosting techniques," in Proceedings of the 13th International Conference on Human-Computer Interaction, LNCS 5611, San Diego, California, pp.198-207, 2009.

C. S. Fahn and Y. T. Lin, "Real-time face detection and tracking techniques used for the interaction between humans and robots," in Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications, Taichung, Taiwan, 2010.

C. S. Fahn and C. S. Lo, "A high-definition human face tracking system using the fusion of omnidirectional and PTZ cameras mounted on a mobile robot," in Proceedings of the 5th IEEE Conference on Industrial Electronics and Applications , Taichung, Taiwan, 2010.

K. C. Fan, Y. K. Wang, and B. F. Chen, "Introduction of tracking algorithms," Image and Recognition, vol. 8, no. 4, pp. 17-30, 2002.

G. L. Foresti, C. Micheloni, L. Snidaro, and C. Marchiol, "Face detection for visual surveillance," in Proceedings of the 12th IEEE International Conference on Image Analysis and Processing, Mantova, Italy, pp. 115-120, 2003.

M. B. Gulmezoglu, V. Dzhafarov, and A. Barkana, "The common vector approach and its relation to principal component analysis," IEEE Transactions on Speech and Audio Processing, vol. 9, no. 6, pp. 655-662, 2001.

E. Hjelmås and B. K. Low, "Face detection: a survey," Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274, 2001.

R. Huang, Q. Liu, H. Lu, and S. Ma, "Solving the small size problem of LDA," in Proceedings of the International Conference on Pattern Recognition, Quebec, Canada, vol. 3, pp. 29-32, 2002.

S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," IEEE Transactions on Neural Networks, vol. 10, no. 2, pp. 439-443, 1999.

M. Montemerlo, S. Thrun, and W. Whittaker, "Conditional particle filters for simultaneous mobile robot localization and people-tracking," in Proceedings of the IEEE International Conference on Robotics and Automation, Washington, D.C., vol. 1, pp. 695-701, 2002.

K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," Image and Vision Computing, vol. 21, no. 1, pp. 99-110, 2003.

M. Turk and A. Pentland, "Face recognition using eigenfaces," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, pp. 586-591, 1991.

P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, vol. 1, pp. 511-518, 2001.

W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399-458, 2003.

Part 6

Perceptual Face Recognition in Humans

15

Face Recognition without Identification

Anne M. Cleary
Colorado State University
USA

1. Introduction

Most people have had the experience of recognizing a person's face as familiar despite failing to identify who the person is or where the person was seen before. Many domains of research have aimed to study this phenomenon; thus, there exist many research paradigms for attempting to tap it in laboratory settings. Among these paradigms are dual-process recognition memory paradigms within the recognition memory literature (e.g., Yonelinas, 2002), feeling-of-knowing paradigms within the metacognition literature (Koriat, 1995), and face identification paradigms within the more general face recognition literature (e.g., Burton, Bruce & Hancock, 1999).

2. Dual process methods of studying recognition memory

2.1 Familiarity-based recognition

Recognition memory is the type of memory that enables people to determine that they have experienced something previously. Dual process theories of recognition memory hold that it can be based on either of two processes: Recollection or familiarity (see Diana et al., 2006, or Yonelinas, 2002, for reviews). Recollection-based recognition occurs when one recognizes having experienced something before based on the retrieval of specifics about the prior occurrence. For example, one might pass someone on the sidewalk and recognize that the person has been seen before by calling to mind the specific instance in which the person was seen before: This person was the receptionist at the dentist the other day. In contrast, familiarity-based recognition occurs when one recognizes having experienced something before based only on a gut feeling or sense about the situation. For example, one might pass someone on the sidewalk and only recognize that the person is familiar without recalling where that person was seen before. The person simply *seems* familiar.

From a dual-process perspective, studying recognition that is familiarity-based requires teasing apart instances of familiarity-based recognition and instances of recollection-based recognition. Over the years, researchers have developed many methods of doing so within list-learning paradigms (see Yonelinas, 2002, or Mandler, 2008, for reviews). Some existing methods are: The process dissociation procedure (e.g., Jacoby, Toth, & Yonelinas, 1993), the tasks procedure (e.g., Cleary & Greene, 2001; Yonelinas, 1997), analyses of receiver operating characteristics (ROCs, Yonelinas, 1994, 1997), the signal-lag procedure (e.g., Hintzman & Curran, 1994) and the remember-know procedure (e.g., Rajaram, 1993).

By separating familiarity from recollection in studies of recognition, presumably the characteristics of familiarity can be studied. Indeed, much has been learned about familiarity from dual-process methods. Yonelinas (1994) combined the process-dissociation procedure with the analysis of receiver operating characteristics (ROCs) and found that although overall recognition memory tends to lead to a z-ROC slope of significantly less than 1.0, when the contribution of familiarity is isolated, the slope of the z-ROC approximates 1.0, as would be predicted by simple signal detection theory; this suggests that familiarity may be well-described by simple single detection theory when it is isolated. Rajaram (1993) used the remember-know procedure (whereby subjects simply indicate whether the basis for each "yes" response on a recognition test was recollection or familiarity) to show that familiarity, but not recollection, is affected by manipulations of perceptual fluency, such as the rapid presentation of a test stimulus prior to presenting it for the recognition decision. Jacoby, et al. (1993) found similar results using the process-dissociation procedure. Rajaram and Geraci (2000) used the remember-know technique to show that familiarity is affected by manipulations of conceptual fluency, such as the presentation of a semantically related word prior to presenting the recognition test stimulus. Unlike recollection, familiarity is unaffected by divided attention at encoding, as has been shown using the process-dissociation procedure (Jacoby, et al., 1993) and the remember-know paradigm (Gardiner & Parkin, 1990).

Research using the signal-lag procedure, in which subjects are given varying response deadlines across recognition test trials, has shown that familiarity-based old-new discrimination emerges sooner in the processing stream than recollection-based old-new discrimination. This has been shown with such tasks as the plurality task, in which subjects must discriminate between words that remain the same from study to test (e.g., frog) and words that changed plurality from study to test (e.g., frogs). Subjects can discriminate which root words were studied versus unstudied significantly earlier than they can discriminate between correct and incorrect pluralities, suggesting that familiarity becomes available earlier on in processing than recollection (Hintzman & Curran, 1994). This finding is consistent with studies of event-related potentials (ERPs) during recognition testing (e.g., Curran, 2000; Curran & Cleary, 2003), which have suggested that the brain electrophysiological correlate to familiarity occurs earlier (300-500 ms) than that of recollection (e.g., 500-800 ms).

Mathematical models often describe familiarity in terms of features (e.g., Clark & Gronlund, 1996). Memory traces for encoded items each exist as a set of the separable features that composed the item itself. At the time of the recognition test, the features in the test item are matched, on a feature by feature basis, with all of the features that have been stored in memory. From this perspective, features should play a critical role in familiarity-based recognition. Indeed, many studies have shown that there are various features that, when isolated, can produce familiarity-based recognition. Among the features that have been shown to play a role in familiarity are: Letters of words (Cleary & Greene, 2000, 2001), phonemes (Cleary, Winfield & Kostic, 2007), geometric shapes within pictures (Cleary, Langley & Seiler, 2004), song notes (Kostic & Cleary, 2009) and song rhythm (Kostic & Cleary, 2009).

Finally, familiarity appears to be left intact with certain forms of memory impairment and with aging. Many amnesics have been shown to be impaired on recollection with familiarity relatively spared (e.g., Aggleton & Brown, 1999; Vann et al., 2009), suggesting that whereas recollection involves the hippocampus proper, familiarity may involve other regions of the medial temporal lobe (MTL) that are often spared in amnesics. Functional neuroimaging studies have generally converged on this idea (e.g., Cohn et al., 2009; see Diana et al., 2007 and Eichenbaum et al., 2007, for reviews).

With regard to familiarity and aging, Mantyla (1993) used the remember-know paradigm and showed that "know" responses (indicating familiarity) were unaffected by aging, while "remember" responses (indicating recollection) declined with aging. In another study, Parkin and Walter (1992) found that "know" responses actually increased with age, while "remember" responses decreased. Other more recent studies have converged on the idea that recollection tends to be more impaired by aging than familiarity (e.g., Jacoby, 1999; Jacoby & Rhodes, 2006; McCabe, Roediger, McDaniel, & Balota, 2009; Rhodes, Castel, & Jacoby, 2008).

2.2 Familiarity-based face recognition

In the dual-process recognition literature, the most commonly-used example of familiarity-based recognition is that of recognizing a face as familiar without recollecting any specifics about the person. For example, in their dual-process study of recognition memory, Curran and Cleary (2003, p. 191) state, "We have all had the experience of knowing a face is familiar despite an inability to recollect details such as the person's name," and in his review of dual-process theory, Yonelinas (2002, p. 441) states, "The distinction is illustrated by the common experience of recognizing a person as familiar but not being able to recollect who the person is or where they were previously encountered." Finally, in her dual-process study, Rajaram (1993, p. 90) states, "There are times when we meet someone on the street whom we met at a party a few days ago. Although we know that we met this person at the party, we may not remember actually meeting the person, or his/her name." Although most dual-process studies use face recognition as an anecdotal example of familiarity-based recognition, most such studies use stimuli other than faces to isolate and study familiarity; in most cases, the stimuli are words. This section examines dual-process studies that have used faces as stimuli in trying to isolate familiarity-based recognition of faces.

Yonelinas, Kroll, Dobbins and Soltani (1999) examined recognition memory for faces. These authors were following up on prior work that had suggested that whereas item recognition (i.e., recognizing a single item as having been studied on an earlier list) can be based on either familiarity or recollection, associative recognition (i.e., recognizing which items were paired together in an earlier studied list and which were re-paired from study to test) appears to require recollection. Specifically, a number of studies have suggested that when subjects study pairs of words (e.g., apple-pond, rock-cat, desk-bottle) and are later tested on their ability to discriminate old from new words, this overall old-new discrimination involves a combination of both familiarity and recollection. However, when participants are instead later tested on their ability to discriminate intact (e.g., apple-pond) from rearranged (e.g., rock-bottle or desk-cat) pairs, recollection is required to make the discrimination (Hintzman, Caulton & Levitin, 1998; Yonelinas, 1997); familiarity alone is thought to lead subjects to false alarm to rearranged pairs.

Yonelinas et al. (1999) examined whether the same principle would apply to faces. To create intact and rearranged faces, these researchers manipulated the features of the faces so as to present some of the faces as rearranged versions of studied faces. Analogously to the rearranged word pairs mentioned above, the rearranged faces were each a combination of two different studied faces' features. Thus, subjects had to discriminate actually studied (intact) faces from faces that were actually recombined versions of studied faces. In this case, unlike with rearranged word pairs, familiarity was found to contribute to the ability to discriminate intact from rearranged faces, as suggested by the shape of the ROC curve. Yonelinas et al. suggested that the reason familiarity can contribute to this type of associative recognition with faces is because faces tend to be processed holistically, rather than decomposed into features (e.g., Searcy & Bartlett, 1996).

Prior work has suggested that while faces tend to be processed holistically when presented upright, when presented upside-down, they tend to instead be decomposed into features (e.g., Searcy & Bartlett, 1996). Accordingly, Yonelinas et al. (1999) found that familiarity-based discrimination between intact studied faces and faces comprised of recombined, studied features occurred only when the faces were presented upright. Recollection was required for such discrimination when the faces were presented upside-down. Thus, holistic processing of faces indeed appears to contribute to the ability to use familiarity alone to discriminate actually studied, intact faces from highly familiar, feature-rearranged faces.

Aly, Knight and Yonelinas (2010) investigated whether faces may be more likely to drive familiarity-based recognition than other types of stimuli. These researchers noted that many studies of amnesic patients (i.e., patients with severe memory impairment due to damage to the medial temporal lobe region) demonstrated impaired recognition memory for such stimuli as words or scenes, but relatively spared recognition memory for faces (e.g., Bird & Burgess, 2008; Carlesimo et al., 2001; Taylor, Henson & Graham, 2007). Aly et al. found that, indeed, overall word recognition was more impaired than overall face recognition in their amnesic patients. However, ROC analysis revealed that the amnesics were impaired in recollection for both words and faces. Furthermore, the type of medial temporal lobe damage made a difference; all patients showed intact familiarity for faces, but some of the patients showed impaired familiarity for words. From the full pattern of results, Aly et al. argued that the reason why amnesic patients may often appear less impaired on face recognition may be because 1) face recognition relies more heavily on familiarity than other types of stimuli and 2) face familiarity remains relatively spared in many cases of amnesia.

The research presented thus far suggests that faces may be somewhat unique within recognition memory. First, the evidence suggests that faces tend to be processed holistically rather than decomposed into features, and as such, familiarity can serve as a basis for discriminating similar faces from actually studied faces, or rearranged faces from intact faces. Second, people may rely more heavily on familiarity in face recognition than in the recognition of other types of stimuli. Finally, face familiarity tends to be relatively spared during impairment to other types of familiarity and to recollection.

2.3 Face recognition without identification

A relatively unique laboratory approach to studying familiarity-based recognition within the dual-process framework is that which is used to induce what has been termed *recognition without identification* (e.g., Cleary & Greene, 2000, 2001; Cleary et al., 2004; Peynircioglu, 1990). In this method, one examines recognition memory in situations where participants fail to identify the experimental reason for the feeling of recognition. For example, after listening to a list of words spoken through a set of computer speakers, subjects may receive a recognition test containing fragments of spoken words, such that only certain spliced phonemes of a given word are presented through the speakers (Cleary et al., 2007). Some of these phoneme fragments come from studied words and some come from unstudied words.

For each such fragment presented, subjects attempt to identify the word to which it corresponds. They are also asked to rate the likelihood that the fragment came from a studied word. Recognition without identification is the finding that among unidentified test items (as when the word from which a phoneme fragment came cannot be identified), people give higher recognition ratings to studied than to unstudied items. In short, people can recognize a test item as familiar despite an inability to identify the experimental source of that familiarity; the source in this case is the particular study episode that led to the familiarity (i.e., the particular spoken studied word corresponding to given unidentifiable phoneme fragment).

Cleary and Specker (2007) attempted to apply the recognition without identification paradigm to face recognition. They gave subjects celebrity names at study (e.g., Adrien Brody, Jennifer Connelly). At test, they gave the subjects pictures of celebrity faces, half of which were of celebrities whose names were studied, and half of which were of celebrities whose names were not studied. For each face presented on the test, subjects first attempted to identify the person by typing the person's name. Then, regardless of whether the face could be identified, subjects also rated the likelihood that the person's name was studied. Among celebrity faces that went unidentified on the test, subjects discriminated between those of celebrities whose names were studied and those of celebrities whose names were not. In this case, the unidentifiable experimental source of the familiarity was the person's name. Thus, subjects demonstrated some ability to recognize faces as familiar within the context of the experiment, yet were unable to identify the experimental source of that familiarity. Cleary and Specker termed this finding *recognition without face identification*. The finding suggests that recognition without identification of faces can be based on semantic information, as this effect required a pre-existing link in memory between the celebrity names and their corresponding faces.

Cleary and Specker (2007) also linked their recognition without face identification effect to the tip-of-the-tongue (TOT) phenomenon, which occurs when a person feels as if a word's retrieval is imminent, on the verge of being retrieved, yet remains inaccessible at the moment. Specifically, Cleary and Specker added an additional question to the test phase of a second experiment; after giving a recognition rating to the face, subjects were asked to indicate if they were experiencing a TOT state for the name or not. The results suggested a relationship between the recognition without face identification effect (i.e., higher recognition ratings for unidentified faces of celebrities whose names were studied than for unidentified faces of celebrities whose names were not studied) and the TOT phenomenon. Specifically, when the recognition ratings were broken down into those given during reported TOT states and those given during reported non-TOT states, the recognition without face identification effect was only found when subjects reported being in a TOT state; it was not present when subjects reported not being in a TOT state.

This finding suggests that the feeling of being able to recognize a face without being able to identify who the person is may be related to the more general TOT phenomenon. Indeed, some have used the example of face recognition without identification to illustrate the TOT phenomenon itself. For instance, Yarmey's (1973) article is entitled, "I recognize your face but I cannot remember your name: Further evidence on the tip-of-the-tongue phenomenon," and Schwartz (2002, p. 114) gives the following example in his review of the TOT experience: "You see an acquaintance approaching. Instantly, you are hit with a TOT. You cannot retrieve the person's name, although you are sure that you know it." Cleary and

Specker (2007), Cleary and Reyes (2009), and Cleary, Konkel, Nomi and McCabe (2010) suggest that the feeling of recognizing something as familiar, such as a face, may subjectively resemble the feeling of being in a TOT state. They based this assertion on the additional finding that subjects consistently give higher familiarity ratings overall when in a TOT state than when not in a TOT state (e.g., Cleary et al., 2010; Cleary & Reyes, 2009; Cleary & Specker, 2007).

3. Feelings of knowing

3.1 The feeling of knowing phenomenon

Feelings of knowing (FOKs) are judgments that people make for momentarily unretrievable information about the likelihood that they would recognize that information if presented with it in the future. Koriat (1995, p. 311) used the example of person recognition without identification to illustrate the FOK phenomenon: "The FOK phenomenon is best illustrated by the many everyday situations in which people try to recall the name of a person but fail to find it. These situations are sometimes accompanied by the subjective conviction that one knows the name and that one is likely to recall it given sufficient time and effort." However, as with the dual-process recognition literature, most FOK studies use stimuli other than faces and their corresponding names, even though people's faces and names are often used to illustrate the real-world phenomenon under investigation.

In one of the first FOK studies, Hart (1965) gave subjects general knowledge questions (e.g., What is the largest planet in the solar system?). When subjects failed at retrieving an answer, they rated the likelihood that they would be able to recognize the answer in a future forced-choice recognition test. In comparing subjects' predictions with their actual performance on the later forced-choice test, Hart found that subjects could predict at above-chance levels which of the then-unretrievable answers would be recognized on the later test. Since Hart's study, the FOK phenomenon has been the subject of a fairly large literature (see Koriat, 2007, for a review).

Many theories of the FOK phenomenon have been proposed over the years (e.g., Koriat & Levy-Sadot, 2001; Nelson, Gerler & Narens, 1984; Yaniv & Meyer, 1987). One of the most widely-held theoretical frameworks is that of Koriat and Levy-Sadot. This framework combines two different theoretical accounts of the FOK phenomenon into a single two-stage account. The first of the two stages is cue familiarity. Cue familiarity refers to the familiarity of the test probe or test question itself, and has been shown to be a basis for FOKs (e.g., Metcalfe, Schwartz & Joaquim, 1993). The second of the two stages is accessibility (Koriat, 1993, 1995, 2007). Accessibility refers to the amount of information that is retrievable in response to the cue, and perhaps even the ease with which it is accessed. In the second stage, subjects attribute any retrieved information and the ease with which it was accessed, whether correct or incorrect, to the likelihood that they will recognize the target if presented with it later (Koriat, 1993, 1995; Koriat & Levy-Sadot, 2001).

According to Koriat and Levy-Sadot (2001), subjects first assess the familiarity of the cue itself (i.e., the question or probe). If it seems familiar, this familiarity prompts them to proceed to the accessibility stage, at which point they search memory for any accessible information that can be retrieved in response to the cue. Benjamin (2005) found support for this idea by showing that, when subjects had to make an FOK judgment in a time-constrained manner, cue-familiarity had a larger influence than accessibility. In short, accessibility played a larger role in subjects' FOK judgments when they had enough time to proceed to that stage.

3.2 Feelings of knowing with faces

Hosey, Peynircioglu, and Rabinovitz (2009) examined subjects' FOKs for pictures of people's faces whose names failed to be retrieved. Hosey et al. also required subjects to indicate the bases of their FOK judgments. These researchers were particularly interested in whether subjects reported relying on cue familiarity or accessibility more often. Though few FOK studies had actually examined FOKs for names in response to faces, at least one study had indicated that cue familiarity with the face itself has an influence on FOKs for faces (Hanley & Cowell, 1988). In line with this idea, Hosey et al. found that subjects indicated relying on cue familiarity as a basis of their FOK judgments more often than they indicated relying on accessibility. These authors assert that this finding is consistent with a claim made by Schwartz, Benjamin and Bjork (1997) that feelings of knowing a person's name in response to a face may actually be driven largely by the familiarity of the face itself. As Schwartz et al. (p. 136) state, ". . .if you feel that a passerby's name is on the 'tip of your tongue,' it is not because you know the person's name, although it is likely that you do, but because the person's face is familiar."

If indeed such day-to-day feelings of knowing about people are driven largely by familiarity with people's faces themselves, then studying familiarity from the perspective of dual-process theory within recognition memory paradigms may be a complimentary experimental approach toward attempting to understand such day-to-day phenomena, as dual-process paradigms attempt to understand the cue familiarity process itself. Indeed, in Metcalfe et al.'s (1993) demonstration that cue familiarity can drive FOK judgments, cues were familiarized through earlier presentation in the experiment, similarly to how familiarity is manipulated in standard list-learning approaches to dual-process theory in recognition memory.

An interesting future direction for research on feelings of knowing with faces would be to determine how reliance on cue familiarity differs when feelings of knowing with faces are compared to feelings of knowing with other types of stimuli, such as verbal materials. Given the findings by Aly et al. (2010) that faces may tend to elicit a greater reliance on familiarity than other types of stimuli, it may be the case that subjects rely more heavily on cue familiarity when giving FOKs to faces than when giving FOKs to other types of stimuli.

4. Modeling the processes of face recognition

Burton, Bruce and Hancock (1999) developed a model of face recognition that includes mechanisms for explaining instances where face recognition occurs, but the person's name cannot be retrieved. This model stems neither from the dual-process recognition memory literature nor the FOK literature, but rather from a more general literature on identifying faces. This model is called the Interactive Activation and Competition (IAC) model (Burton et al., 1999; Burton, Bruce & Johnston, 1990). The model contains multiple levels of units that contribute to face recognition: Face recognition units (FRUs), person identification nodes (PINs), and semantic information units (SIUs), which carry general semantic information about a person including name information. The model also contains lexical output units for identifying the person's face. In this model, people's names are more difficult to retrieve than other types of semantic information about a person (e.g., occupation) because names are more distinctive than other types of semantic information, and distinctive information can often be difficult to retrieve.

In the IAC model, different pieces of information become available at different points in time. Face familiarity occurs at the earliest stage, at the level of the PINs. A face is recognized as familiar if a PIN's activation exceeds a determined threshold; this mechanism allows for a face to be recognized as familiar even though no information about the identity of the face may be recalled. Access to semantic information about the person's face becomes available next and occurs at the level of the SIUs. Activation at this level may allow a person to access semantic information associated with the face (e.g., the person's occupation) even though the person's name may still be unretrievable. Again, because they are distinctive, people's names tend to be more difficult to access than general semantic information; thus, semantic information becomes available earlier. At the latest stage of processing, retrieval of the person's name may finally occur.

The stages of processing proposed in the IAC model are supported by a number of empirical studies on general face recognition and identification processes. First, Johnston and Bruce (1990) have shown that subjects are able to determine that a face is familiar earlier than they are able to retrieve semantic information about the person. This type of finding is analogous to the signal-lag and ERP studies of dual-process theory discussed above, which have suggested that familiarity becomes available earlier on in processing than recollection (e.g., Hintzman & Curran, 1994; Curran, 2000; Curran & Cleary, 2003). Second, there is a lot of evidence to suggest that people commonly retrieve semantic information about a person without being able to retrieve the person's name, yet almost no evidence suggests that people can retrieve a person's name in the absence of any semantic information about the person (e.g., Hay, Young & Ellis, 1991; Young, Hay & Ellis, 1985). Third, a substantial literature suggests that people have greater difficulty recalling people's names than recalling general semantic information about people (e.g., Bredart & Valentine, 1998; Cohen, 1990; Stanhope & Cohen, 1993).

5. Summary, conclusions, and future directions

This chapter is concerned with the common experience of recognizing a person's face as familiar, despite an inability to identify who the person is, or very often, anything specific at all about the person. As illustrated here, many different research approaches to this phenomenon have been taken. Many dual-process recognition paradigms aim to study the process of familiarity-based recognition, which is thought to underlie, or at least contribute to, the real-life experience of recognizing without identifying a person. Feeling-of-knowing (FOK) paradigms aim to study the experience of feeling as if one knows something that cannot currently be accessed from memory, and the feeling that one would recognize that information as the sought-after information if later presented with it. Finally, the Interactive Activiation and Competition (IAC) model aims to simulate the processes involved in the day-to-day experience of recognizing without identifying a person, as when one looks at a picture of person, recognizes the face as familiar, yet cannot identify the person.

All three of these different approaches aim to tap the same real-world phenomenon, as illustrated by the fact that face recognition without identification is the most common example of the phenomenon under investigation in all of these approaches. However, the extent to which these different approaches all actually tap the same phenomenon remains to be determined. That said, many commonalities exist between the three approaches. First, the relative timeline for when different types of information become available is very similar across the three paradigms. In dual-process theory, familiarity is thought to become available earlier than recollection (Hintzman & Curran, 1994; Curran, 2000; Curran & Cleary, 2003). In FOK theory, cue familiarity is thought to become available before partial information becomes accessible (Benjamin, 2005). Finally, in the IAC model, face familiarity becomes available before semantic information about the person can be accessed, which in turn becomes available before the name itself can be accessed.

Second, as mentioned, the cue familiarity thought to contribute to FOK judgments (e.g., Metcalfe et al., 1993) may actually be the same type of familiarity that drives familiarity-based recognition in dual-process recognition paradigms. Thus, the findings from different paradigms that 1) subjects in FOK paradigms may rely more heavily on face familiarity than on accessibility of partial information when giving FOK judgments to faces (Hosey et al., 2009) and 2) that faces may elicit a greater reliance on familiarity-based recognition than other types of stimuli in dual-process paradigms (Aly et al., 2010), may indicate a convergence on the idea that face familiarity itself largely drives real-world cases of face recognition without identification. As mentioned, the IAC model contains a mechanism for this: The activation of Person Identification Nodes (PINs) allows a face to be recognized as familiar even when no other information can be accessed. In short, the three methods discussed here may converge on the same general ideas regarding face recognition without identification. Future research should aim to further determine how well they actually do converge.

6. References

Aggleton, J. P. & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences, 22*, 425-490.

Aly, M., Knight, R. T., & Yonelinas, A. P. (2010). Faces are special but not too special: Spared face recognition in amnesia is based on familiarity. *Neuropsychologia, 48*, 3941-3948.

Benjamin, A.S. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review*, 12 (5), 874-879.

Bird, C. M., & Burgess, N. (2008). The hippocampus supports recognition memory for familiar words but not unfamiliar faces, *Current Biology, 18*, 1932-1936.

Bredart, S., & Valentine, T. (1998). Descriptiveness and proper name retrieval. *Memory, 6*, 199-206.

Bruce, V. & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British Journal of Psychology, 76*, 363-383.

Burton, A.M., Bruce, V. & Hancock, P.J.B. (1999). From pixels to people: a model of familiar face recognition. *Cognitive Science, 23*, 1-31.

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology, 81,* 361-380.

Carlesimo, G.A., Fadda, L., Turriziani, P., Tomaiuolo, F., & Caltagirone, C. (2001). Selective sparing of face learning in a global amnesic patient. *Journal of Neurology, Neurosurgery, and Psychiatry, 71*, 340-346.

Cleary, A. M., & Greene, R. L. (2000). Recognition without identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1063-1069.

Cleary, A. M., & Greene, R. L. (2001). Memory for unidentified items: Evidence for the use of letter information in familiarity processes. *Memory & Cognition, 29*, 540-545.

Cleary, A. M., Konkel, K.E., Nomi, J. N., McCabe, D. P. (2010). Odor recognition without identification. *Memory & Cognition, 38*, 452-460.

Cleary, A.M., Langley, M.M., & Seiler, K.R. (2004). Recognition without picture identification: Geons as components of the pictorial memory trace. *Psychonomic Bulletin & Review*, 11 (5), 903-908.

Cleary, A.M., & Specker, L.E. (2007). Recognition without face identification. *Memory & Cognition*, 35, 1610-1619.

Cleary, A.M., & Reyes, N.L. (2009). Scene recognition without identification. *Acta Psychologia*, 131 (1), 53-62.

Cleary, A.M., Winfield, M.M., & Kostic, B. (2007). Auditory recognition without identification. *Memory & Cognition*, 35 (8), 1869-1877.

Cohen, G. (1990). Why is it difficult to put names to faces? *British Journal of Psychology, 81,* 287-297.

Cohn, M., Moscovitch, M., Lahat, A. & McAndrews, M. P. (2009). Recollection versus strength as the primary determinant of hippocampal engagement at retrieval. *Proceedings of the National Academy of Sciences, 106,* 22451-22455.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28,* 923-938.

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. Cognitive Brain Research, 15, 191-205.

Diana, R.A., Reder, L.M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin and Review, 13*(1), 1-21.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: A three-component model. *Trends in Cognitive Sciences, 11,* 379-386.

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience, 30,* 123-152.

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*, 579-583.

Hart, J.T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational psychology,* 56 (4), 208-216.

Hay, D. C., Young, A.W., & Ellis, A. W. (1991). Routes through the face recognition system. *Quarterly Journal of Experimental Psychology, 43A*, 761-791.

Hanley, J.R. & Cowell, E. S. (1988). The effects of different types of retrieval cues on the recall of names of famous faces. *Memory and Cognition, 16*, 545-555.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition, 26*, 449-462.

Hintzman, D. L. & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1-18.

Hosey, L. A., Peynircioglu, Z. F. & Rabinovitz, B. E. (2009). Feeling of knowing for names in response to faces. *Acta Psychologica, 130*, 214-224.

Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *25*, 3-22.

Jacoby, L. L., & Rhodes, M. G. (2006). False remembering in the aged. *Current Directions in Psychological Science*, *15*, 49-53.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139-154.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*, 609-639.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: further evidence for the accessibility model. *Journal of Experimental Psychology: General*, *124*, 311-333.

Koriat, A. (2007). *Metacognition and Consciousness*. In: Cambridge handbook of consciousness. Cambridge University Press, New York, USA.

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27,* 34-53.

Kostic, B.K., & Cleary A.M. (2009). Song recognition without identification: when people cannot"name that tune" but can recognize it as familiar. *Journal of Experimental Psychology: General, 138*, 146-159.

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science, 3*, 390-400.

Mantyla, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition, 21*, 379-388.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging decreases veridical remembering but increases false remembering: Neuropsychological test correlates of remember/know judgments. *Neuropsychologia, 41*, 2164-2173.

Metcalfe, J., Schwartz, B.L., & Joaquim, S.G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology, Learning, Memory, & Cognition*, 19, 851-861.

Nelson, T.O., Gerler, D., & Narens, L. (1984). Accuracy of feeling of knowing judgments for predicting perceptual identification and relearning. *Journal of Experimental Psychology: General*, 113, 282-300.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7*, 290-298.

Peynircioglu, Z.F. (1990). A feeling-of-recognition without identification. *Journal of Memory and Language,* 29, 493-500.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89-102.

Rajaram, S. & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1070-1074.

Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging*, *23*, 239-249.

Schwartz, B. L. (2002). *Tip-of-the-tongue states: Phenomenology, mechanism, and lexical retrieval*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6,* 132-137.

Searcy, J. H. & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relation information in faces. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 904-915.

Stanhope, N. & Cohen, G. (1993). Retrieval of proper names: Testing the models. *British Journal of Psychology, 84,* 51-65.

Taylor, K. J., Henson, R.N.A., & Graham, K. S. (2007). Recognition memory for faces and scenes in amnesia: Dissociable roles of medial temporal lobe structures. *Neuropychologia, 45*, 2428-2438.

Vann, S. D., Tsivilis, D., Denby, C. E., Quamme, J. R., Yonelinas, A. P., Aggleton, J. P., Montaldi, D., Mayes, A. R. (2009). Impaired recollection but spared familiarity in patients with extended hippocampal system damage revealed by 3 convergent methods. *Proceedings of the National Academy of Sciences, 106,* 5442-5447.

Yaniv, I., & Meyer, D.E. (1987). Activation and metacognition of inaccessible stored information: potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 187-205.

Yarmey, A. D. (1973). I recognize your face but I can't remember your name: Further evidence on the tip-of-the-tongue phenomenon. *Memory and Cognition, 1*, 287-290.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1341-1354.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*, 747-763.

Yonelinas, AP (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517.

Yonelinas, A. P. Kroll, N. E. A., Dobbins, I. G., & Soltani, M. (1999). Recognition memory for faces: When familiarity supports associative recognition judgments. *Psychonomic Bulletin & Review, 6*, 654-661.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognizing people. *British Journal of Psychology, 76,* 495-523.