Winter 2007

Neighborhood Defined Feature Selection Strategy for Improved Face Recognition in Different Sensor Modalitie

Satyanadh Gundimada
Old Dominion University

NEIGHBORHOOD DEFINED FEATURE SELECTION STRATEGY FOR IMPROVED FACE RECOGNITION IN DIFFERENT SENSOR MODALITIES

by

Satyanadh Gundimada
M.S (E.C.E) Dec 2002, Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ELECTRICAL ENGINEERING

OLD DOMINION UNIVERSITY
December 2007

Approved by:

K. Vijayan Asari (Director)

Frederic D. McKenzie (Member)

Jiang Li (Member)

Jessica Crouch (Member)

ABSTRACT

NEIGHBORHOOD DEFINED FEATURE SELECTION STRATEGY FOR IMPROVED FACE RECOGNITION IN DIFFERENT SENSOR MODALITIES

Satyanadh Gundimada
Old Dominion University, 2007
Director: Dr. K. Vijayan Asari

A novel feature selection strategy for improved face recognition in images with variations due to illumination conditions, facial expressions, and partial occlusions is presented in this dissertation. A hybrid face recognition system that uses feature maps of phase congruency and modular kernel spaces is developed. Phase congruency provides a measure that is independent of the overall magnitude of a signal, making it invariant to variations in image illumination and contrast. A novel modular kernel spaces approach is developed and implemented on the phase congruency feature maps. Smaller sub-regions from a predefined neighborhood within the phase congruency images of the training samples are merged to obtain a large set of features. These features are then projected into higher dimensional spaces using kernel methods. The unique modularization procedure developed in this research takes into consideration that the facial variations in a real world scenario are confined to local regions. The additional pixel dependencies that are considered based on their importance help in providing additional information for classification. This procedure also helps in robust localization of the variations, further improving classification accuracy. The effectiveness of the new feature selection strategy has been demonstrated by employing it in two specific applications via face authentication in low resolution cameras and face recognition using multiple sensors

(visible and infrared).

The face authentication system uses low quality images captured by a web camera. The optical sensor of the web camera is very sensitive to environmental illumination variations. It is observed that the feature selection policy overcomes the facial and environmental variations. A methodology based on multiple training images and clustering is also incorporated to overcome the additional challenges of computational efficiency and the subject's non involvement. A multi-sensor image fusion based face recognition methodology that uses the proposed feature selection technique is presented in this dissertation. Research studies have indicated that complementary information from different sensors helps in improving the recognition accuracy compared to individual modalities. A decision level fusion methodology is also developed which provides better performance compared to individual as well as data level fusion modalities. The new decision level fusion technique is also robust to registration discrepancies, which is a very important factor in operational scenarios.

Research work is progressing to use the new face recognition technique in multi-view images by employing independent systems for separate views and integrating the results with an appropriate voting procedure.

ACKNOWLEDGMENTS

While working towards my Ph.D. degree I have been very fortunate to be surrounded by a group of wonderful people, whose contribution to this work I would like to acknowledge. I would first like to express my deepest gratitude to my doctoral advisor, Vijayan Asari, for his extraordinary support and guidance during my graduate years at Old Dominion University. From his course that introduced me to the fascinating world of pattern recognition and computer vision, to the insightful suggestions that have always pointed me to the right direction in my research, to the final advice on refining my dissertation, he has continuously been a source of inspiration, energy and invaluable knowledge. For these, and for many other reasons that helped make this dissertation and the completion of my Ph.D. studies a reality, I will be always indebted to him. Heartfelt thanks to the other members of my Qualifying and Dissertation Committee: Min Song, Rick Mckenzie, and Youzong shen, for their thorough and valuable comments that helped shape the final version of my dissertation. I would like to thank my lab mates in the Computer Vision group for the friendly and stimulating atmosphere that made the past years in the lab a wonderful experience. Special thanks to Hau Ngo, Praveen Sankaran and Saibabu, for always being there to help, for the great conversations, and for the numerous weekends and nights spent together in the lab before deadlines. I would like to thank my great friends Viswanadham and Samyukta for their constant motivation and support throughout my Ph.D course. I have saved the last of my acknowledgments for the people whom I can never hope to adequately thank – my family. My love and gratitude to my parents, Jaganmohan Rao and Jayakumari and my sister Sunitha and brother in law

Saikiran, and my niece Aneesha for their support and for the many sacrifices they made so that I can achieve the best in my life. I dedicate this dissertation to them, with all my love.

TABLE OF CONTENTS

Chapter **Page**

Chapter **Page**

LIST OF FIGURES

Figure **Page**

Figure **Page**

Figure **Page**

Figure **Page**

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Automatic human face recognition has received substantial attention from researchers in computer vision, pattern recognition, biometrics, machine learning and computer graphics. This common interest among researchers working in diverse fields is motivated by a human's remarkable ability to recognize faces and also by the fact that this human activity is one of the primary concerns in everyday life. There are several commercial, security, and forensic applications that require the use of face recognition technologies. These applications include automated crowd surveillance, access control, mug shot identification, face reconstruction, design of human computer interface, multimedia communications and content based image database management. There are several commercial face recognition systems available, including, 2D systems from Cognitec Systems [1], Viisage [2] and 3D systems from A4Vision [3], Geometrix [4], and Genex technologies [5]. Most of these systems work under controlled lighting and environmental conditions.

A biometric system deals with automatic recognition of people based on their distinctive anatomical characteristics such as face, fingerprint, iris, retina, hand geometry, vein, voice, etc and behavioral characteristics such as signature and gait. Face is an effective biometric attribute. Different biometric indicators are suited for different kinds of identification applications due to their performance with regard to intrusiveness, accuracy, cost, and ease of sensing [6]. The face biometric provides good non-intrusiveness although the current state of accuracy is low under operational scenarios.

The reference model for this work is *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Global biometric revenues were $719 million in 2003. They are expected to reach $4.6 billon by 2008, driven by large scale public sector biometric departments, the emergence of transactional revenue models, and the adoption of standardized biometric infrastructures and data formats [6]. Among the emerging biometric technologies, facial biometrics is projected to reach annual revenues of $802 million in 2008. Face recognition scenarios can be classified into two types: (i) face verification (or authentication) and (ii) face identification (or recognition).

1.1 Face Verification

Face verification is a one-to-one match that compares a query face image against a template face image whose identity is being claimed.

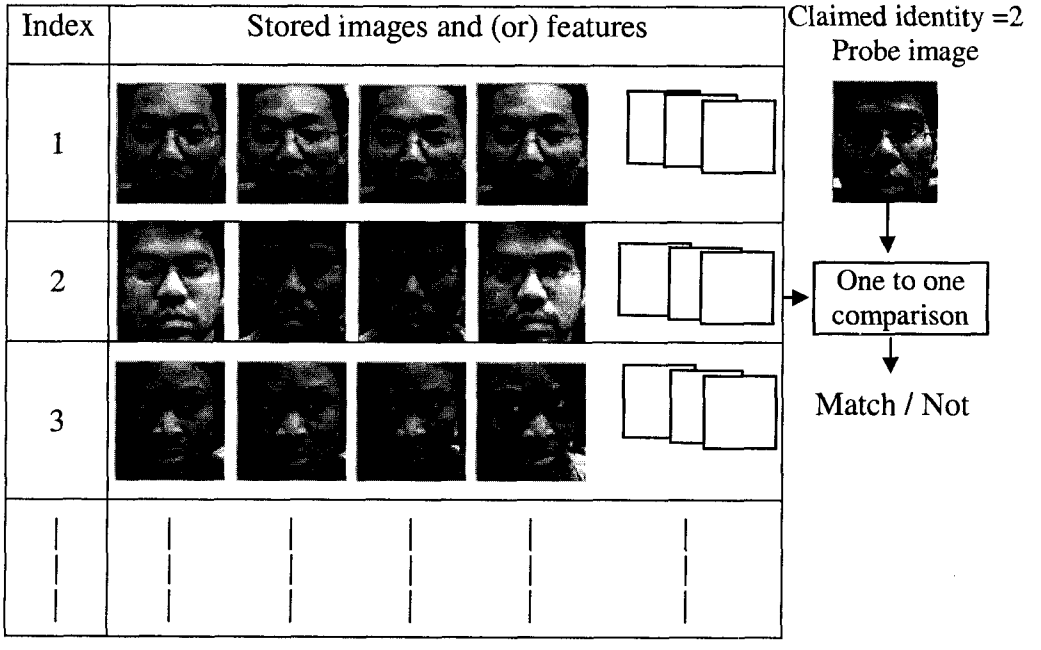


Figure 1.1. Illustration of the face verification scenario.

The block diagram in Figure 1.1 illustrates the scenario of face verification. To evaluate the verification performance, the verification rate (the rate at which legitimate users are granted access) and the false acceptance rate (the rate at which imposters are granted access) are computed. A good verification system should balance these two rates based on operational needs.

1.2 Face Identification

Face identification is a one-to-many matching process that compares a query face image against all the template images in a face database to determine the identity of the query face. Figure 1.2 illustrates the scenario of face identification.

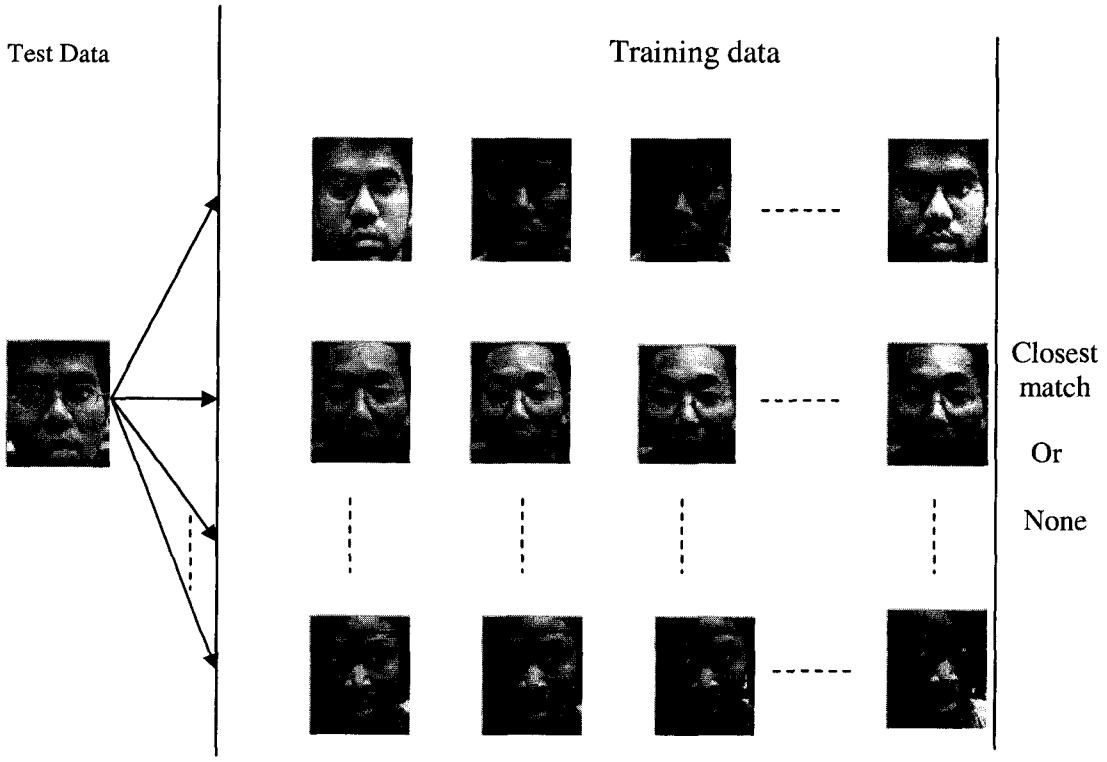


Figure 1.2. Illustration of a face recognition scenario.

The identification of the test image is done by locating the image in the database that has the highest similarity with the test image. The identification process is a closed test, which means the sensor takes an observation of an individual that is known to be in the database. The test subject's (normalized) features are compared with those in the system's database, and a similarity score is obtained for comparison. These similarity scores are then numerically ranked in descending order, and the individual with the highest similarity score is chosen if it is within a predefined threshold.

1.3 Challenges in Face Recognition

A great deal of effort has been devoted to the task of face recognition and in the process it has achieved a certain level of maturity, but it still remains a challenging problem in a general setting. Frontal face recognition under controlled environments would give near to 100% accuracy for large databases. Successful 2D face recognition systems have been deployed only under constrained situations, but there are huge challenges that need to be addressed to make face recognition technology a reality. One major factor limiting the applications of 2D face recognition systems is that human face image appearance has the potential for very large intra-subject variations due to:

1. Illumination variations
2. Facial expression variations
3. Occlusions due to facial hair, face accessories and other objects
4. Aging and
5. Pose

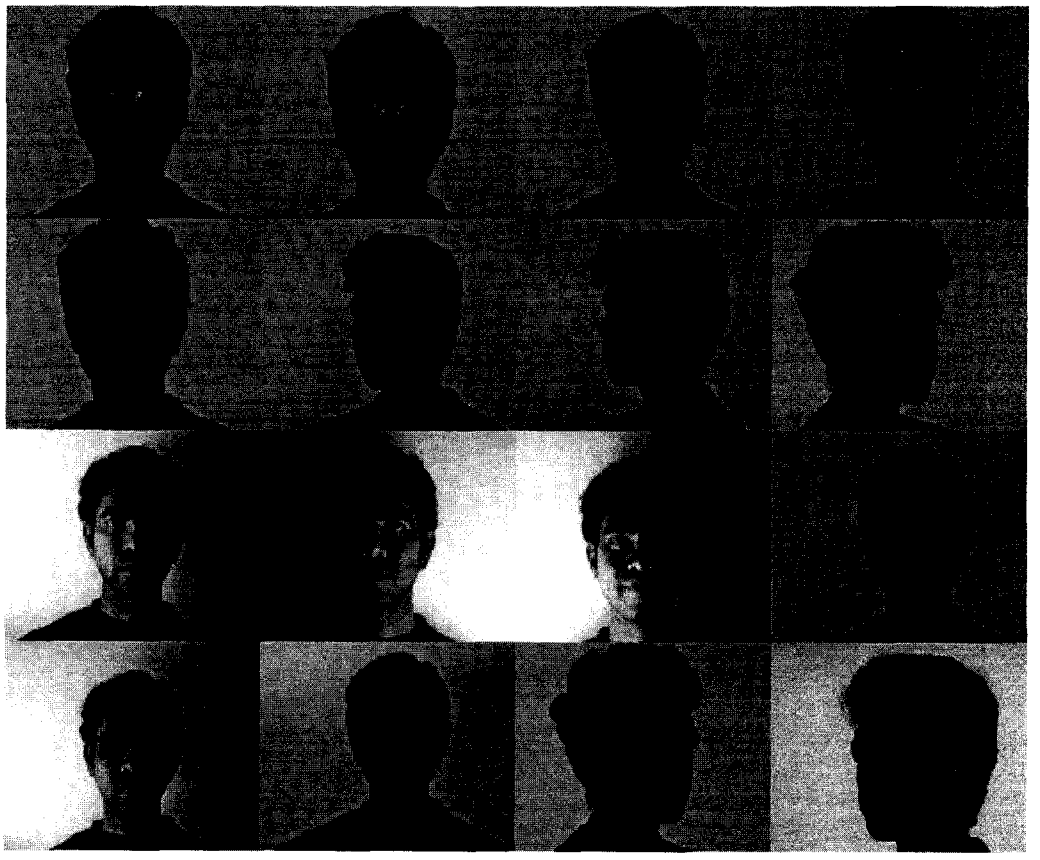


Figure 1.3. Facial appearance variations due to changes in pose illumination, expression, and facial accessories.

Figure1.3 gives examples of intra-class appearance variations. Inter-subject variations can be small due to the similarity of individual appearances. An illustration of comparison between the inter-personal and intra-personal variations is illustrated in Figure 1.4. It can be observed that, the images with larger intra-personal variations are farther from each other even though they are of the same subject, whereas images of different subjects look similar. Various techniques and methodologies have been developed over the past two decades targeting the above challenges.

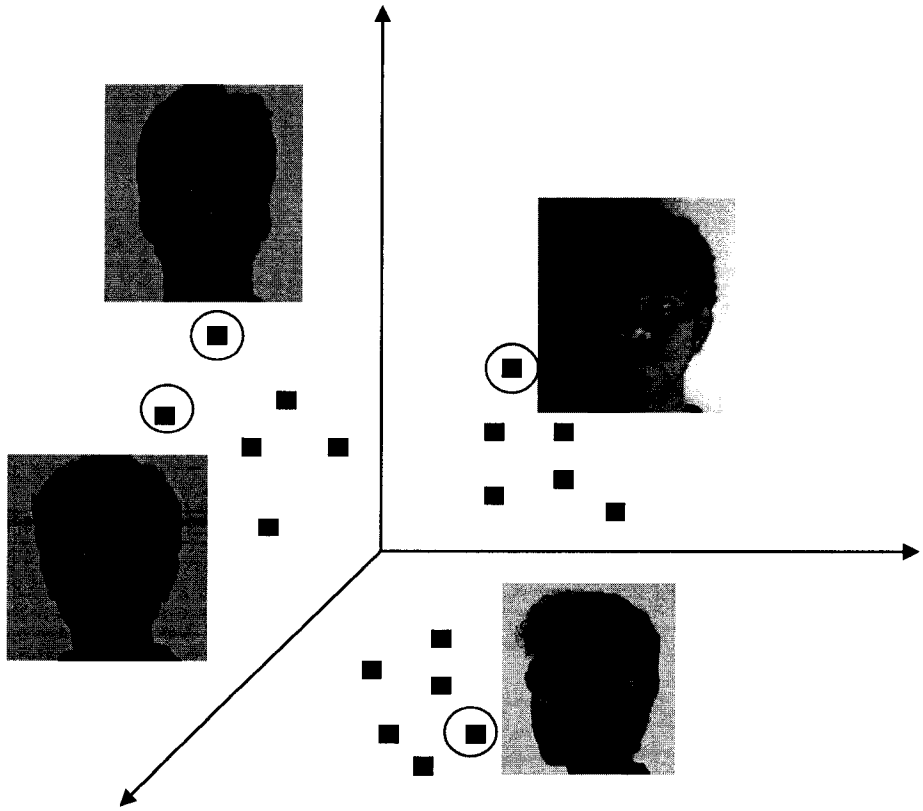


Figure 1.4. Illustration of appearance variations of different subjects.

1.4 Focus of Research and Main Contributions

The main focus of this research is to make the face recognition process robust to the first four challenges mentioned above, that is to improve the classification accuracy of face images that are affected due to illumination, and/or facial expressionss and/or partial occlusions. It is also possible that the recognition technology developed in this research for a frontal face image can be extended to other views of a face if the training sets in different views are available.

The main contributions of this dissertation are as follows:

1. Development of an adaptive weighted modular approach for improved recognition in the presence of facial variations.

2. Development of a new technique using phase congruency in feature extraction to overcome illumination variations.
3. Development of a novel modularization technique to deal with expression variations and partial occlusions.
4. Development of a feature selection strategy for improved face recognition
5. Development of a face authentication system using the above feature selection strategy in low resolution sensor images.
6. Development of a face recognition technique in thermal sensor images using the above feature extraction and selection strategies.
7. Development of a new information fusion based face recognition methodology in multi-sensor images.

1.4.1 Adaptive weighted modular approach

In this research, a methodology to determine the level of confidence of a sub-region in the overall classification of a given face image, affected due to varying expressions, illuminations and partial occlusions, is presented. The technique for obtaining the weights for each individual region of the test image is based on a measure of optical flow between that test image and a face model. Individual image regions or modules are assigned additional weights by arranging them in the order of their importance in classification. The proposed method is applicable mainly in scenarios where the samples in the training set are few in number. A K-nearest neighbor distance measure is used in classifying each module of the test image after dimensionality

reduction. A total score is calculated for each training class based on the classification result of each module and its associated weights.

1.4.2 Phase congruency in feature extraction

Applying phase congruency features for the purpose of face recognition is one of the contributions of this research. Human facial variations are identified due to illumination changes in both indoor and outdoor environments. The feature selection process is very important to overcome the effect of these variations for accurate face recognition. Many feature selection strategies have been studied and evaluated. Recent studies accompanied by psychophysical evaluations helped to implement a face recognition procedure where phase based feature selection is employed. Features perceived by the human brain are at those locations where the phase information is highly in order. Since human beings can identify faces with tremendous accuracy, the features perceived by them would provide better accuracy when applied to machine vision. Boosted by this idea, the feature selection in the face images is made according to the congruency or order in the localized phase information.

1.4.3 Novel modularization technique

In reality, facial variations are confined mostly to local regions. Modularizing the images would help to localize these variations, provided the modules created are sufficiently small. But in this process, a large amount of dependencies among various neighboring pixels might be ignored. This can be countered by making the modules larger, but this would result in improper localization of the facial variations. In order to

deal with this problem, a novel module creation strategy has been developed in this research. Considering additional pixel dependencies across various sub-regions would help in providing additional information, which in turn improves the classification accuracy. A feature selection policy based on the above technique is implemented in this dissertation. Modular spaces are created with pixels from across various local regions taking into account the locality of such regions.

1.4.4 Feature selection strategy for improved face recognition

Higher order dependencies can be obtained by projecting the data in the input space into a higher dimensional space. These dependencies could provide critical information which helps in boosting the recognition performance. The modular phase congruency features are vectorised and projected into higher dimensional spaces. The kernel trick method is used in achieving the transformation into the higher dimensional eigen spaces. Classification is carried out after this transformation. Observations indicate improvement in the recognition accuracy compared to the linear subspace based feature selection techniques.

1.4.5 Face authentication in low quality images

An application based on the new feature extraction and selection techniques is developed which can be used in a real life scenario. The objective of this research was to develop a face verification system, which could be useful for applications related with access control in a distributed environment. In this case, each subject is given a personal identification number (PIN) when enrolled. When the subject types his/her PIN, images

are acquired and stored in a database of images. The facial verification process extracts the features and compares them to the stored features in the database to verify if the PIN belongs to the actual claimant or not. The main criterion is to provide a reliable, cost effective, fast and expandable system whose inputs are images from a low cost web camera.

1.4.6 Face recognition in thermal images

Thermal images capture the shape of the object (face) under consideration; hence, feature based techniques could be better for classification tasks in this case. The phase congruency measure as explained earlier captures the local features based on the order in phase. It is independent of the contrast in the image. Hence the feature maps should provide better performance for thermal images. Also, the thermal images are opaque to glass; hence, it is common to have partially occluded regions in thermal face images. Other variations in thermal images are caused due to temperature inequalities in the face due to environmental conditions. Since these variations are confined to local regions, the proposed methodology based on modular regions could provide better performance in terms of recognition accuracy. Experimental observations indicate a significant improvement in thermal face recognition accuracy upon the use of the proposed feature selection techniques.

1.4.7 Multi-sensor image fusion for face recognition

A new multi-sensor decision level fusion based face recognition technique is developed in this research. The features obtained from the individual modalities (visible

and thermal) when fused at the decision level provide complementary information for the face recognition system. It is observed that the technique outperforms all the individual modalities as well as the DWT based data level fusion technique in terms of recognition accuracy. The effect of image registration, with respect to data and decision level image fusion techniques is also discussed.

1.5 Thesis Outline

A general survey of various technologies and methods in face recognition that are developed in the literature is presented in chapter 2. Also, a thorough survey of the existing methods corresponding to the work done in this research is presented in this chapter. The affect and analysis of local regions on face recognition is discussed in chapter 3. The technique of face recognition which is based on the adaptive modular approach is also described in this chapter. The novel feature extraction and selection strategies for improved face recognition in the presence of facial variations caused due to illumination conditions, expression, and partial occlusions are presented in chapter 4. Detailed explanation about the need and effect of phase congruency in face recognition is presented in this chapter. The details of the development and analysis of the novel modularization technique are also presented in this chapter. Discussion regarding projection of data into modular linear subspaces as well as projections into high dimensional spaces is also provided. In chapter 5 a detailed explanation of the face recognition system developed for images captured by low quality sensors is presented. Details of the algorithms for facial authentication and facial search are provided in this chapter. Experimental results and comparisons with existing methods are also included in

this chapter. Details of implementation of the developed feature selection techniques on multi-sensor image modalities for face recognition are presented in chapter 6. Analysis of the image capture modalities in various ranges of the infrared (IR) spectrum is provided. Details about the development and implementation of the data and decision level fusion techniques are also presented. Experimental results obtained from the Equinox face database, and comparisons with other methods and discussions are also presented in this chapter. Conclusions of the work performed in this dissertation and suggestions for future directions of this research are presented in chapter 7.

CHAPTER 2

LITERATURE REVIEW

Automatic human face recognition [7-10], [34] has been of interest to a growing number of research groups for the past two decades. The goal of automatic face recognition is to achieve a high level of performance in matching a given face against a database of faces. Currently, the performance of face recognition is relatively poor compared to other biometric measures such as fingerprint and iris matching. A lot of research has been done for the improvement of face recognition. Many face recognition technologies based on different methodologies have been developed and documented in the literature. A tremendous amount of work is still in progress to make the automatic face recognition technology a reality. A general survey of various technologies and methods in face recognition that are developed in the literature is presented in this chapter. Also, a thorough survey of the existing methods corresponding to the work done in this research is presented. Face recognition techniques can be broadly divided into two methods: (i) appearance based and (ii) feature based. Appearance based techniques are those in which the classification is performed by considering the holistic intensity image as the input. In the second approach, features are extracted from the given intensity images and are classified with or without an intermediate step of dimensionality reduction. In addition to this, facial identification is a pattern recognition problem, hence it can also be classified based on the type of subspace analysis and classification techniques used. The recognition technique developed in this research is mainly a 2D technique, although this can be extended to 3D [42], [43], [45] using view based analysis.

Hence, the literature review that is presented is mainly constrained to 2D face recognition.

2.1 Appearance Based Face Recognition

Some of the approaches to object recognition are based directly on the appearance of the images or the intensity of the images. These techniques depend on a representation of images that induces a vector space structure. Appearance based approaches represent an object in terms of several object views. Many view-based approaches use statistical techniques to analyze the distribution of the object image in the vector space and derive an efficient and effective representation (feature space) according to different applications. Given a test image, the similarity between the stored prototypes and the test view is then carried out in the feature space.

Image data can be represented as vectors, that is, as points in a high dimensional vector space. For example, a $p \times q$ image can be mapped to a vector $x \in R^{pq}$ by concatenating each row or column of the image. Each image, now a vector, corresponds to a point in this space. Out of the total $(p \times q)^{256}$ possible instances in this image space, human face images only reside in a very small portion. The distribution of all face images accounts for variations in facial appearance. Subspace approaches have been modeled with a view that the projection of the original image vectors into the subspaces would provide a better opportunity to perform face recognition in the reduced spaces. Three classical techniques that can be named in the category of linear subspace methods are Principal Component Analysis (PCA) [12][35], Linear Discriminant Analysis (LDA)[11][23], and Independent Component Analysis (ICA) [33].

2.1.1 Principal component analysis

PCA is used to reduce the dimensionality of a data set while retaining as much information as possible. PCA efficiently represents the data by finding orthonormal axes which maximally decorrelate the data. The directions of these basis vectors are the ones in which the data is maximally varying as shown in Figure 2.1.

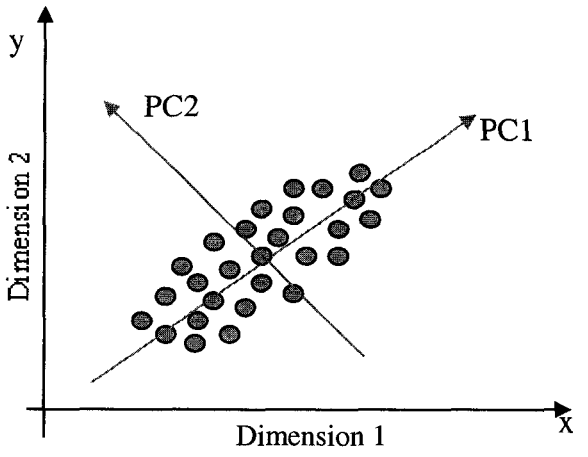


Figure 2.1. Illustration of the directions of maximum variance of the data.

The vectors onto which the data is projected are called Principal Components (PC). If $X=\{x_1,x_2,x_3....x_M\}$ are the data vectors of dimension n, $x_M \in R^n$, the eigen vectors corresponding to the non-zero eigen values of the covariance matrix S_t form the principal components.

$$S_t = \frac{1}{M}\sum_{i=1}^{M}(x_i - m_0)(x_i - m_0)^T \tag{2.1}$$

where $m_0 = \frac{1}{M}\sum_{i=1}^{M} x_i$ is the mean of the vectors X.

Hence the data can now be projected onto a space spanned by the eigen vectors of S_t as shown in Figure 2.2. The data can be represented in terms of only a few significant eigen

vectors with very low loss of information as shown in Figure 2.3. Different classification techniques such as nearest neighborhood, k-means and neural networks can be used in this reduced linear subspace. By projecting the data into this kind of linear subspace, it is possible to reduce the ill effects of the curse of dimensionality while at the same time improving the computational efficiency.

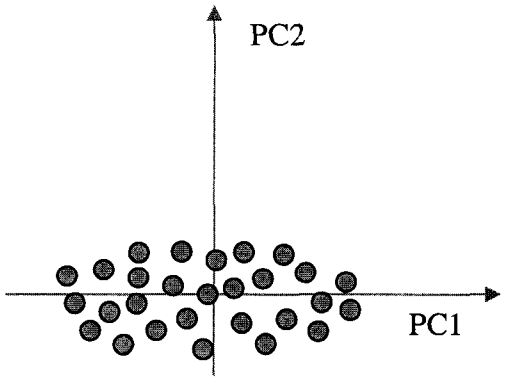


Figure 2.2. Illustration of data transformed onto the space spanned by the eigen vectors.

Figure 2.3. Data represented in terms of only the first principal component vector.

Figure 2.4. Sample face images in ORL face database.

Some of the sample face images in the ORL face database are shown in Figure 2.4, and the average of all the images is shown in Figure 2.5.

Figure 2.5. Average face image of all the faces in the ORL database.

The first seven eigen vectors (eigen faces) are shown in Figure 2.6. As explained earlier, the data (image data) is expressed in terms of linear combinations of the eigen faces.

Figure 2.6. First seven significant eigen faces.

2.1.2 Linear discriminant analysis

PCA finds directions that are useful for representing the data, which may not be good for discriminating the data, whereas LDA finds directions that are efficient for discrimination as shown in Figure 2.7. To show how to obtain the direction of the maximally discriminating vector, two classes of data are considered. $X=\{x_1,x_2,x_3....x_M\}$ are the data vectors of dimension n, $x_M \in R^n$. A subset of X belonging to class 1 is labeled as D1. A subset of X belonging to class 2 is labeled as D2, and w is the unit vector in the direction of maximum seperability.

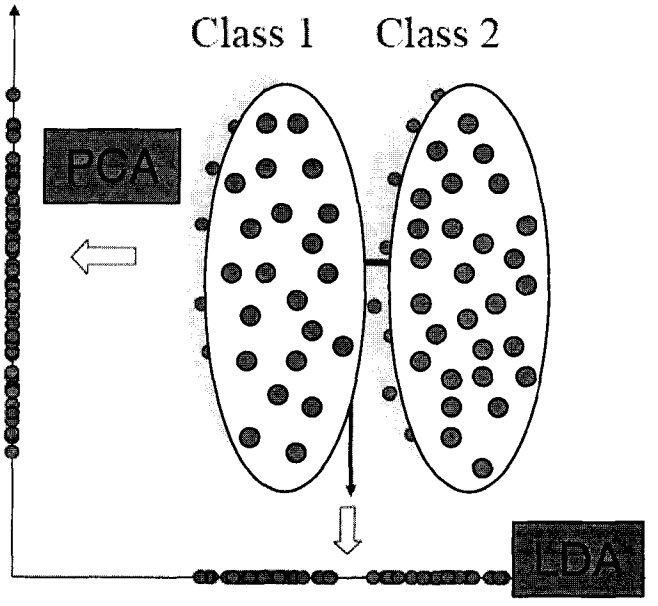


Figure 2.7. Illustration of data projection using PCA and LDA.

In this projection the examples from the same class are projected very close to each other and at the same time, the projected means are as apart as possible. Hence, a projection vector w needs to be calculated that maximizes the criterion function

$$J(w) = \frac{\left|\tilde{m}_1 - \tilde{m}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{2.2}$$

where $\tilde{m}_1$ and $\tilde{m}_2$ are the projected means. $\tilde{s}_1^2 + \tilde{s}_2^2$ is called the within class scatter of projected samples. $\left|\tilde{m}_1 - \tilde{m}_2\right| = \left|w^t (m_1 - m_2)\right|$, where m_1 and m_2 are the class means and $J(w)$ is called the Fisher criterion. The Fisher criterion can also be extended for multiple class discrimination. The criterion is given by:

$$\mathrm{J(W)} = \frac{\left|W^{\mathrm{t}} S_B W\right|}{\left|W^{\mathrm{t}} S_W W\right|} \tag{2.3}$$

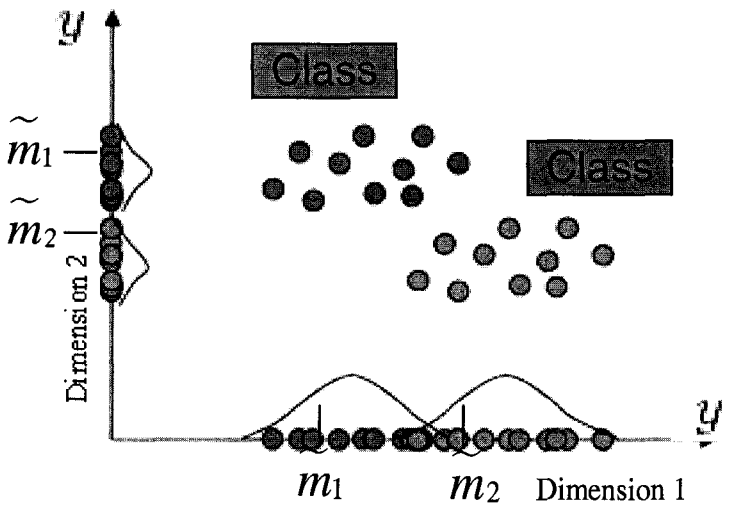


Figure 2.8. Illustration of discriminability of the data in different directions.

The task is to find out the matrix W which maximizes the above criterion function. The criterion function can be expressed as $S_B W_i = \lambda_i S_W W_i$. Eigenvectors that correspond to the non zero eigen values of $S_W^{-1} S_B$ form the discriminant vectors onto which the data is projected for maximum separability. Figure 2.8 illustrates the discrimination of two classes in two different directions.

2.1.3 Independent component analysis

Principal component analysis has transformed the data to a space where the data is maximally decorrelated, whereas linear discriminant analysis has transformed the data to a space where the data is maximally separable. On the other hand, the Independent Component Analysis [33] transforms the data to a space where the data are maximally statistically independent. It can also be interpreted as a method of separating independent sources which have been linearly mixed to produce the data.

2.1.4 Higher dimensional spaces

PCA encodes the pattern information based on second order dependencies, i.e., pixel wise covariance among the pixels and is insensitive to the dependencies of multiple (more than two) pixels in the patterns. Since the eigenvectors in PCA are the orthonormal bases, the principal components are uncorrelated. In other words, the coefficients for one of the axes cannot be linearly represented from the coefficients of the other axes. Higher order dependencies in an image include nonlinear relations among the pixel intensity values, such as the relationships among three or more pixels in an edge or a curve, which can capture important information for recognition. Explicitly mapping the vectors in input space into a higher dimensional space is computationally intensive. Kernel methods have come in handy at this juncture. Using the kernel trick one can compute the higher order statistics using only dot products of the input patterns. Kernel PCA has been applied to face recognition applications and is observed to be able to extract nonlinear features. The process of obtaining weights for the input patterns in the kernel principal component analysis transformed space is described below.

The concept of non-linear mapping is illustrated in Figure 2.9. The classes are linearly inseparable in the input space. Instead, they would have to be separated with a complex decision surface. Suppose we have the mapping $\phi : X = R^2 \rightarrow H = R^3$ $(x_1, x_2) \rightarrow (x_1, x_2, x_1^2 + x_2^2)$. Now the problem has become a linearly separable one by a simple mapping into R^3 as shown in Figure 2.10. There are also other techniques such as kernel fisher discriminant analysis (KLDA) in this category that are applied to face recognition [15], [36], [37], [38].

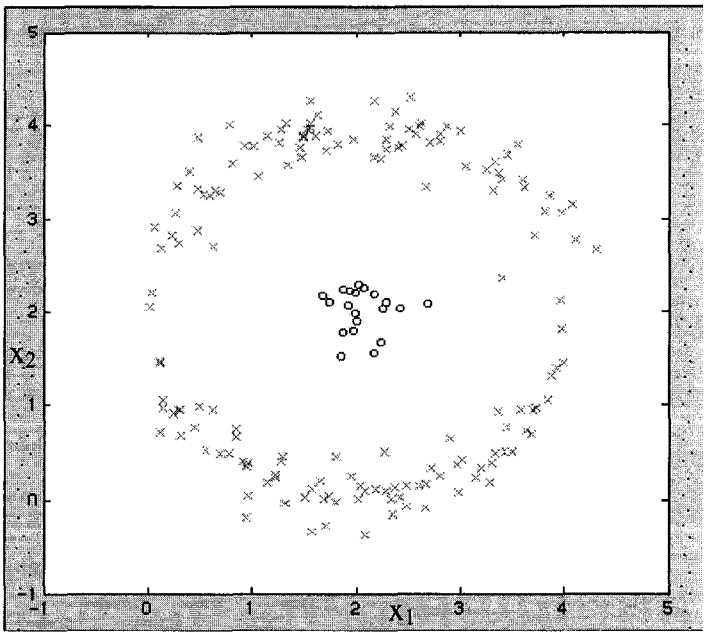


Figure 2.9. Illustration of linearly non-seperable distribution of two classes of data in the given input space.

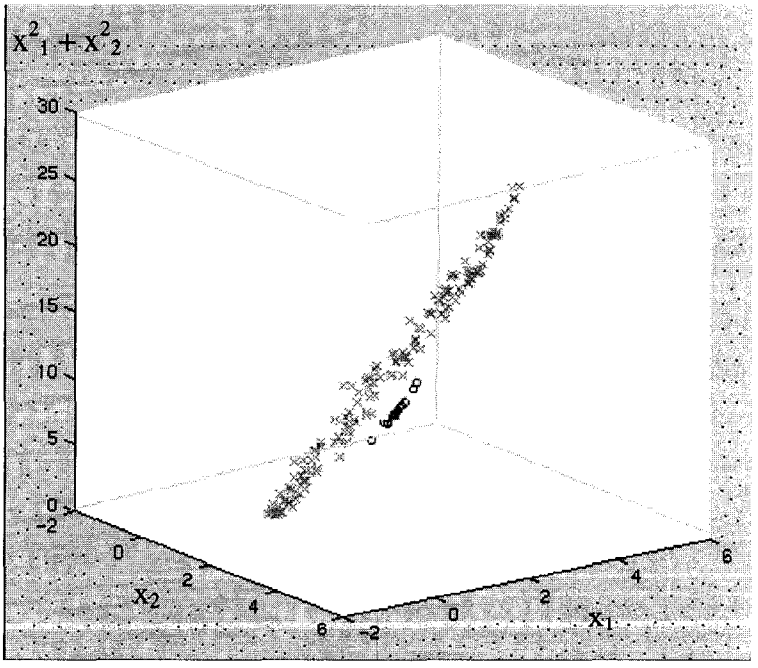


Figure 2.10. Illustration of linearly separable distribution of two classes of data in their dimensional space.

2.2 Feature Based Methods

The importance of facial features for face recognition cannot be overstated. Many face recognition systems need facial features in addition to the holistic face, as suggested by studies in psychology. Feature based techniques are targeted particularly towards achieving illumination invariance in pattern classification, although there are certain preprocessing [30],[40] and subspace techniques [31] developed exclusively for achieving lighting invariant face recognition. Many feature based face recognition techniques have been developed during recent years [19], [20], [39], [44]. Three types of feature extraction methods can be distinguished: (1) generic methods based on edges, lines, and curves; (2) feature-template-based methods; (3) structural matching methods that take into consideration geometrical constraints on the features. Early approaches focused on individual features; as with a template-based approach was described in [59] to detect and recognize the human eye in a frontal face. These methods have difficulty when the appearances of the features change significantly, for example, closed eyes, eyes with glasses, open mouth. To detect the features more reliably, recent approaches have used structural matching methods, such as, the Active Shape Model [58]. Compared to earlier methods, these recent statistical methods are much more robust in terms of handling variations in image intensity and feature shape. Typically, in these methods, local features such as the eyes, nose, and mouth are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a structural classifier. Recently there has been an interest in the research community in understanding and analyzing the concept of phase information in the image. Certain face recognition techniques have been developed based on the features extracted from the phase spectrum [41].

2.3 Local Region Based Techniques

The recently proposed modular PCA method [16] is one of the methods that try to overcome such ineffectiveness by exploring the face's local structure. In this method, a face image is first partitioned into several smaller sub-images; then a single conventional PCA is applied to each of them. However, such a local representation in MPCA ignores the mutual spatial relationship among sub-images partitioned from the original face image. There are recent publications [26][27][28][29] in this direction of expression, occlusion and lighting invariant face recognition that are based on the concept of local region analysis. In [27][28] a weighted distance measure is implemented which reduces the effect of pixels in the test image, that underwent significant movement from the corresponding positions in the training images. This technique was mainly targeted towards achieving expression invariance [47][48].

A number of other interesting approaches have been explored from different perspectives, such as local feature analysis statistical model based [60] and component-based face recognition methods [24][32]. Examples of the statistical model based scheme are Hidden Markov Model (HMM) [61] and Gaussian Mixture Model [62]. Instead of considering a face image from a global view, component–based schemes analyze each facial component separately.

2. 4 Multi-sensor Image Fusion

Even though face recognition technology has progressed from linear subspace methods such as eigen faces and fisher faces to non linear methods such as KPCA, KLDA as explained in the above sections, many of the problems are yet to be addressed completely. While the nature of face imagery in the visible domain is well-studied,

particularly with respect to illumination dependence, its thermal counterpart has received less attention. Previous studies have shown that infra-red imagery offers a promising alternative to visible imagery for handling variations in face appearance due to illumination changes. There are a few studies of the performance of face recognition algorithms in long wave infra-red spectrum, but they are not comprehensive. It is interesting to observe that face recognition on thermal images degrades more sharply than with visible images when probe and gallery are chosen from different sessions [53]. Results in this paper indicate better performances obtained with visible imagery in indoor environments under controlled lighting conditions, but in outdoor environments the thermal image based face recognition system outperformed the visible imagery based one. Also, the thermal face recognition results for both indoor and outdoor environments are comparatively less different from each other, thus reiterating that illumination has little effect on thermal imagery. The conclusion of the studies in [53][54][55] is that despite the degraded thermal recognition performance, fusion of both visible and thermal modalities yields better overall performance. Most of the studies that were conducted on thermal image face recognition relied on the conventional eigen face approach. This was particularly relevant for us because, in [53] it is observed that while multi-session thermal face recognition under controlled indoor illumination was statistically poorer than visible recognition with two standard algorithms, the significance was substantially reduced with an algorithm [53] more specifically tuned to thermal imagery. A summary of results obtained on different sensor modalities with various techniques are shown in tables 1 and 2. The results thus suggest that previous results reported on thermal face recognition may be incomplete. Hence, building of effective algorithms to fuse information from both

spectra has the potential to improve face recognition performance. It is possible to realize sensor fusion on different levels: sensor data level fusion, feature vector level fusion, and decision level fusion. In [53] fusion on the decision level is considered to have more potential applications.

Table 2.1 Recognition accuracy for visible, infrared and fusion for indoor images [53].

	Visible	LWIR	Fusion
PCA	81.54	58.89	87.87
LDA	94.98	73.92	97.36
Equinox	97.05	93.93	98.40

Table 2.2 Recognition accuracy for visible, infrared and fusion for outdoor images [53].

	Visible	LWIR	Fusion
PCA	22.18	44.29	52.56
LDA	54.91	65.30	82.53
Equinox	67.06	83.02	89.02

CHAPTER 3

ADAPTIVELY WEIGHTED MODULAR APPROACH FOR FACE RECOGNITION

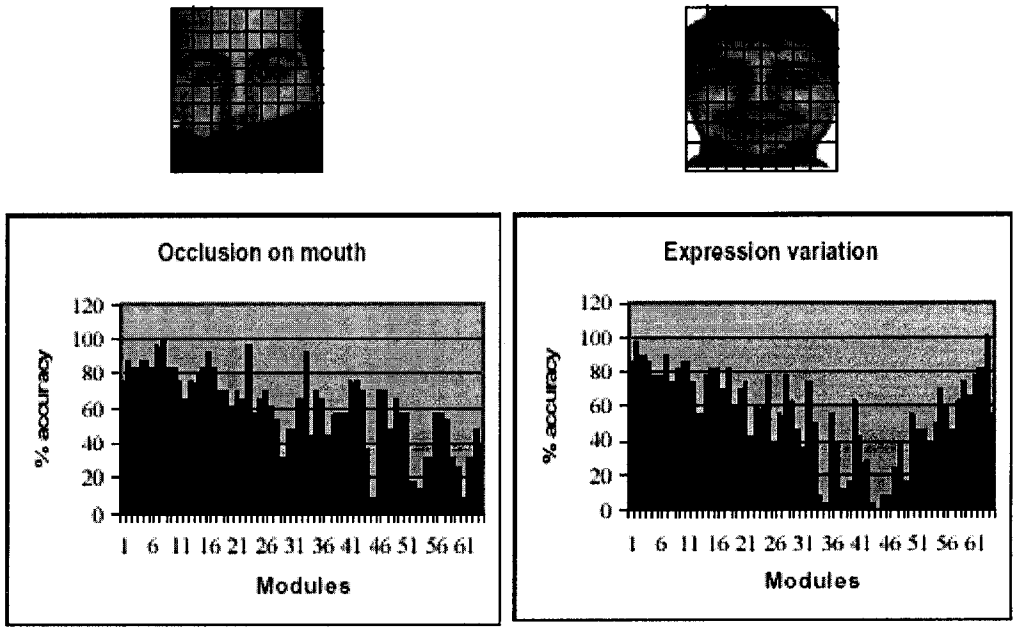
Presented in this chapter is a methodology for determining the level of confidence of a sub-region in the overall classification of a given face image affected due to varying expressions, illuminations and partial occlusions. The technique for obtaining the weights for each individual region of the test image is based on a measure of optical flow between that test image and a face model. Individual image regions or the modules are also assigned additional weights by arranging them in the order of their importance in classification. The approach presented is applicable mainly in scenarios where the number of samples in the training set is too little. A K-nearest neighbor distance measure is used in classifying each module of the test image after dimensionality reduction. A total score is calculated for each training class based on the classification result of each module and its associated weights.

A technique to identify regions that are highly affected due to facial variations is developed. The classification results of such regions are replaced with the classification result of the corresponding regions on the other half of the test image provided those regions have been determined to be less affected due to facial variations. This is implemented by assuming that the test image is a frontal face image, properly aligned with the training images. A technique based on the optical flow between the test image and a face model is implemented to determine the weights associated with the modules of the test image. The weight assigned to each module is proportional to the sum of magnitudes of the optical flow vectors within that module. An overall score is calculated

taking into consideration the classification result of each module and the weights associated with it after performing the replacement procedure. The technique presented here is computationally efficient and has achieved very high accuracy rates on standard face databases. The testing strategy is implemented such that the training set consists of face images taken under controlled conditions whereas the testing set consists of images captured in uncontrolled conditions.

3.1 Weighted Modules

Variations caused in facial images due to expression, makeup and non uniform lighting tend to move the face vector away from the neutral face of the same person both in image space and reduced linear subspace. It has been observed that the dimensionality reduction techniques on individual modules of the face images improve the accuracy of face recognition compared to applying the technique to the whole image. An experiment is conducted on the AR database [52] to show the effect of modularization of face images on recognition accuracies. Two sets of images, one with expressions mostly affecting the mouth regions and the other set with partial occlusions on the bottom half of the face as shown in Figure 3.1, are used for illustration of this concept. The two sets are tested separately using a leave one out strategy. All the training images are divided into 64 local regions. Each region of the module is projected onto a reduced eigen space. The test module is classified using a nearest neighbor algorithm. The classification accuracies of the individual modules for both the sets are shown in Figure 3.1. Sample face images with segmented regions indicated are given on top of the bar graphs.



(a) (b)

Figure 3.1. Percentage of accuracies of each module in the images affected due to (a) partial occlusion and (b) facial expression variation.

It can be observed that only the regions that are affected adversely due to facial variations have low accuracy rates. Hence, it can be inferred that in most cases it is still possible to obtain good accuracy rates if we are able to find out the amount of variation on a local region of the face image. Hence, in many of the face recognition techniques that are based on segmented human face regions [16], each module or the local region of the test image, is classified separately and the overall classification of the test image is determined by employing a voting mechanism on the classification results of all the individual modules. The classification is done in favor of the class or individual who obtains the maximum votes. Instead of a voting technique to classify the test image, a weighted module approach is implemented in this research. Two sets of weights are

assigned in this technique. The first set of weights is assigned as per the following weighting procedure:

(i) Images in the face database are divided into a predefined set of modules, m x m.

(ii) These modules are arranged in the order of importance of classification using the first set of weights.

(iii) The individual modules of all the classes are projected into the eigen subspaces.

(iv) The ratio of the within class variance to the between class variance is calculated for each module.

(v) The module with the highest ratio is given low weight and the module with the lowest ratio is given highest weight. This procedure indicates the importance of each module in classification of the overall face image.

The algorithm presented here differs from the earlier works [16][27][28] in using the image modules more effectively for the overall classification of the image. A second set of weights is assigned to the modules dynamically. A low weight, or confidence, is given to those modules of the test image that are affected due to the variations caused because of expressions, makeup or decorations, occlusions, and lighting. Determination of the weights associated with each module is achieved by the application of an optical flow algorithm between the test image and a face template.

3.2 Optical Flow

Optical flow [49] [50] between the test image and a neutral face template is calculated to determine the regions with expressions [46], partial occlusions, and extreme lighting changes. The face model that is used as a reference image for a neutral face is the mean of all the face images in the training database. Lucas and Kanade's algorithm [51]

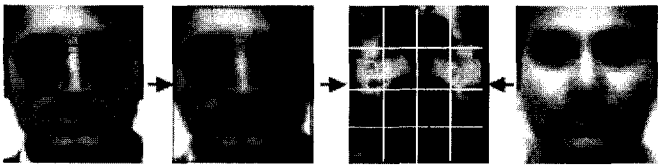
is a classical technique and is implemented to find the optical flow between the test images and the face model. A brightness constancy constraint is assumed in the calculation of optical flow as given in equation 3.1.

$$I(x,y,t) = I(x+u\delta t,\ y+v\delta t,\ t+\delta t) \tag{3.1}$$

$I(x,y,t)$ is the intensity at the image pixel (x,y) at time t. (u,v) is the horizontal and vertical velocities, and δt is a very small time interval. Tailor series expansion of equation 3.1 results in the optical flow constraint as given in equation 3.2.

$$u\frac{\partial I}{\partial x}+v\frac{\partial I}{\partial y}+\frac{\partial I}{\partial t}=0 \tag{3.2}$$

An additional smoothness constraint is assumed to solve for the velocity vectors at each pixel of the image. The test images are preprocessed before applying the optical flow algorithm.



(a) (b) (c) (d)

Figure 3.2. (a) Test image, (b)low pass filtered image, (c)optical flow magnitude image and (d) face template.

Each of the test images is passed through a low pass filter to reduce the resolution of the image. This step helps in ignoring the smaller variations between the test image and the face model. The optical flow algorithm that is applied after low pass filtering of the image, captures variations that are prominent in both images, such as expressions, decorations, and occlusions. We are only interested in the magnitudes of the optical flow

vectors and not the directions of the vectors at each pixel. It can be observed in Figure 3.2 that the regions that have undergone a lot of variations with respect to the face model have a higher sum of magnitudes of optical flow vectors. It is assumed that the face images are properly aligned prior to finding optical flow. Any change in alignment could produce false variations.

3.3 Assignment of Weights

There are two sets of weights that each individual face module is associated with. The first weight assignment scheme is based on the importance of each module for overall face recognition. The assumption is that the various facial regions have different amounts of importance, whereby the eye and mouth regions play an important role in face recognition. A weighted function defined according to the spatial position of the respective regions of the facial features is used. The second weight assignment scheme is based on the optical flow magnitudes within each module. The modules enclosing the regions with higher variations are assigned lesser weight in order to minimize the influence of such modules on the overall accuracy of classification. Equations 3.3 and 3.4 explain the process of assignment of weights.

$$w_i = \frac{\left(G_{\max} - G_i\right)}{G_{\max}} \quad for\ G_i > T \tag{3.3}$$

$$w_i = 1 \qquad for\ G_i \leq T \tag{3.2}$$

where G_i is the sum of the magnitudes of the optical flow vectors within each module. T is the magnitude threshold below which the module is given full confidence during

classification. T is specific to each module and is set in such a way that variations that exist within the neutral face images i.e., without any expressions or decorations or lighting variations are not penalized. G_i can be obtained as:

$$G_i = \sum_{p=1}^{(N/m)} \sum_{q=1}^{(N/m)} \|F_{pq}\| \qquad for\, i = 1,2,3,.....m^2 \tag{3.5}$$

Where $\|Fpq\|$ is the magnitude of the optical flow between the test image and the average face template at pixel (p,q). G_{max} represents the maximum value of G_i.

$$G_{\max} = Max(G_i)_{\forall i} \tag{3.6}$$

The weights are set in such a way that the modules that enclose maximum variations are given zero weightage or no confidence. The modules whose flow magnitude does not exceed threshold T are given a weightage of '1' and the rest of the modules that lie between the limits are assigned weights according to equation 3. Figure 3.3 shows a graphical representation of the assignment of weights.

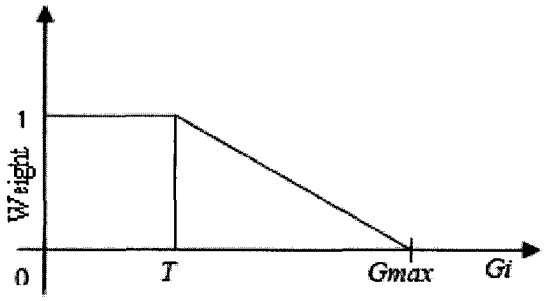


Figure 3.3. Illustration of the linear weight assignment policy.

3.4 Threshold Calculation

Threshold 'T' represents the magnitude of variations below which the weights assigned are always '1'. The variations below this threshold are not considered as the

ones that are caused by expressions, occlusions and non-uniform lighting. The mean image of the set of face images belonging to the same individual is selected to represent the neutral face of that individual. Then optical flow is calculated on the mean image of each individual and the face template. The sum of the magnitudes of flow vectors within each module is calculated. The maximum magnitude obtained for each module over all the mean images is taken as the threshold for that module. Equations 3.7 and 3.8 further explain the procedure of calculating the threshold '*T*'.

$$G_i^n = \sum_{p=1}^{(N/m)} \sum_{q=1}^{(N/m)} \left\| F_{pq} \right\| \quad for\ i = 1,2,3,.....m^2 \ \ and \ \ n = 1,2,....P \tag{3.7}$$

$$T_i = Max(G^n{}_i) \quad where \ \ i = 1,2,3,...m^2 \ and \ n = 1,2,3,..P \tag{3.8}$$

where P is the number of individuals in the database. G_i^n is the summation of the magnitudes of the optical flow vectors within each module for the mean face of the n^{th} individual with respect to the face template. T_i is the threshold for each of the module obtained by taking the maximum value of G_i^n.

3.5 Symmetrically Opposite Modules

When a very low number of training images are available, it is not possible to easily determine the intra personal subspace probability distribution for each module. In such cases it is observed that replacing the result of classification of the module, which had undergone higher variations compared to the symmetrically opposite module in the face that had undergone fewer variations, would produce better results. Figure 3.4 illustrates this principle of replacement. As shown in the figure, the modules on the left of

the magnitude image show higher variations due to non-uniform illumination. It can also be observed that the symmetrically opposite modules have undergone less or no variations. Hence if the classification results of the affected modules are replaced with those of the non affected ones then it could result in better overall accuracy. The classification result of a module is replaced by the other only when the difference between the magnitudes of variations exceeds a certain level.

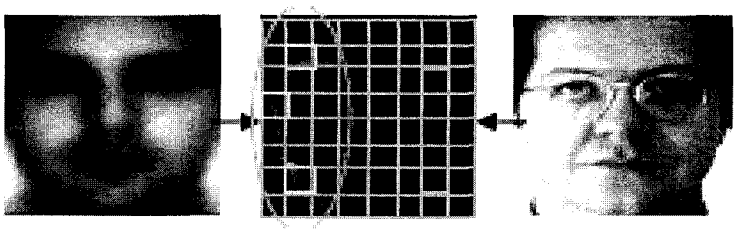


Figure 3.4. Classification result of the regions encircled are replaced with the result of the symmetrically opposite modules.

This threshold is experimentally determined in order to maximize the recognition accuracy. The weights corresponding to the modules belonging to the same class are added. Whichever class receives the highest score determines the classification result of the test image.

3.6 Experimental Results

Three different methods are implemented and compared here. One is the conventional PCA technique on holistic face images, the second is the modular PCA (MPCA) and the third is the proposed technique. In the proposed technique, the modules are first projected onto the respective eigen spaces and then classified similar to MPCA. Since weights are assigned to the modules adaptively depending upon the extent of

variations, the proposed method is called Weighted Modular Principal Component Analysis (WMPCA). Testing is carried out on AR, AT&T [64] and Yale databases. 40 individuals are chosen randomly from the AR database. 13 images of each individual are available. The AT&T database has 40 individuals with 10 images of each person. The Yale database has 150 images in total with 15 individuals. Two types of testing strategies are implemented here. The first one is the classical leave one out technique. In the second testing policy, only 4 images of each individual are chosen to form a training set. For example, in the AR database, out of the 13 images corresponding to each individual, 4 images are selected to form the training set. The remaining 9 images form the test set. Similarly, when dealing with the Yale database, out of the 11 images of each individual, 4 are assigned for training and 7 for testing. This represents the real life situation where the training set consists of images taken in controlled environments whereas the probe images were in uncontrolled environments. The images selected for training are closer to the neutral face of the individuals. The sample images for both the training and testing from the AR database are provided in Figures 3.5 and 3.6 respectively. The probe image is aligned with the face model to eliminate the possibility of false motion between the two images due to misalignment. A second step is to establish a dense correspondence between the two face images using the optical flow technique and then to calculate the magnitudes of the flow vectors.

Figure 3.5. Face images of an individual in the training set from the AR face database.

Figure 3.6. Sample test images of the individual in Figure 3.5.

The probe image is modularized and the summations of the magnitudes of the vectors within each module are calculated and the weight factor is assigned to each module. Each module of the probe image is projected onto the corresponding linear subspace created from the training set then a K- nearest neighbor distance measure is used to classify that module. A final score for each class or individual in the training set is calculated by taking into consideration the weightage associated with each module. A winner takes all strategy is followed in determining the final classification result. Figures 3.7 and 3.8 illustrate the results of the proposed technique on AR and AT&T databases respectively. The leave one out testing strategy was used for testing these sets of experiments. It can be observed that even with a comparatively large set of training images, the holistic linear subspace approach is unable to provide a high accuracy on the AR database. On the other hand, the accuracy results of both PCA and the WMPCA are close to each other in the case of the AT&T database. This is mainly due to the fact that the AT&T database has almost all the images under controlled conditions.

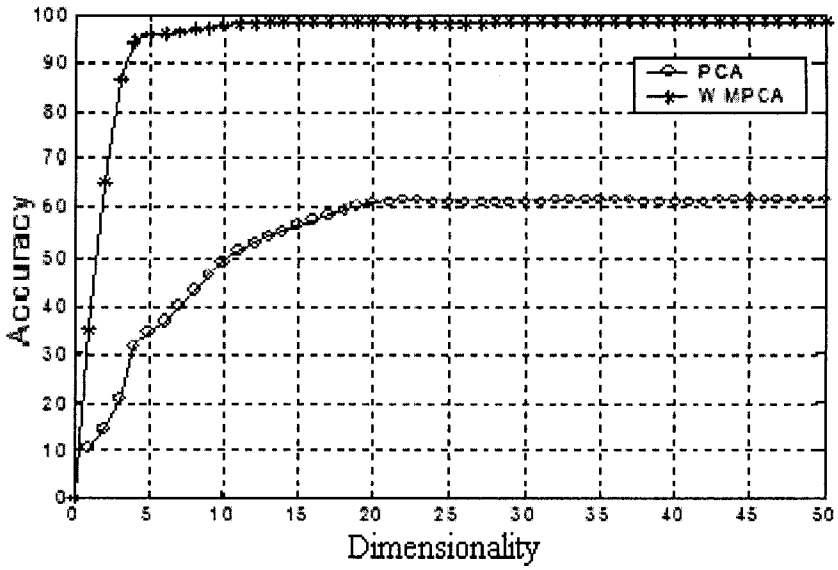


Figure 3.7. Accuracy vs. the dimensionality of the subspace corresponding to AR database using 'leave one' out testing strategy.

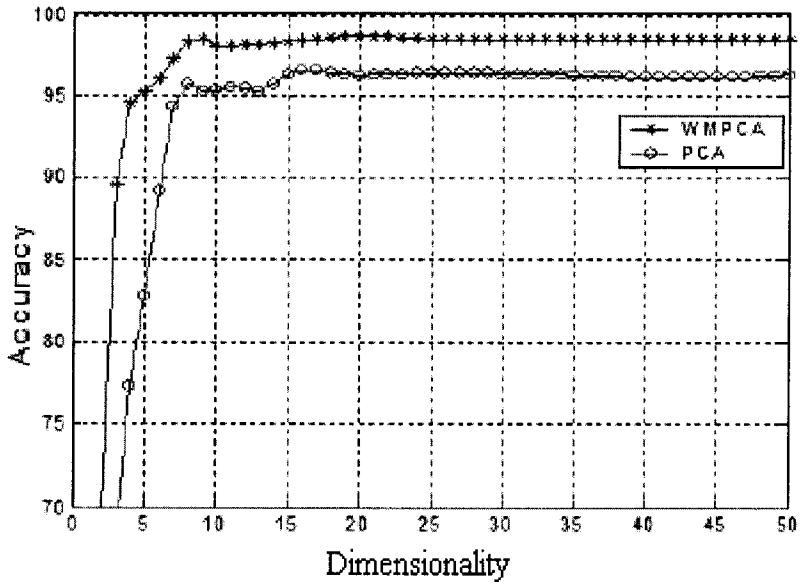


Figure 3.8. Accuracy vs. the dimensionality of the subspace corresponding to AT&T database using the leave one out testing strategy.

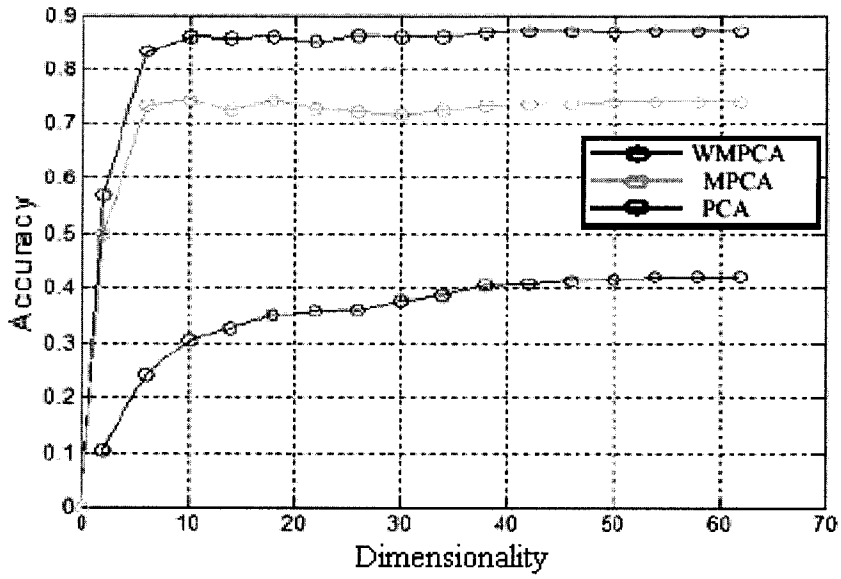


Figure 3.9. Accuracy vs the dimensionality of the subspace corresponding to AR database using the second testing strategy.

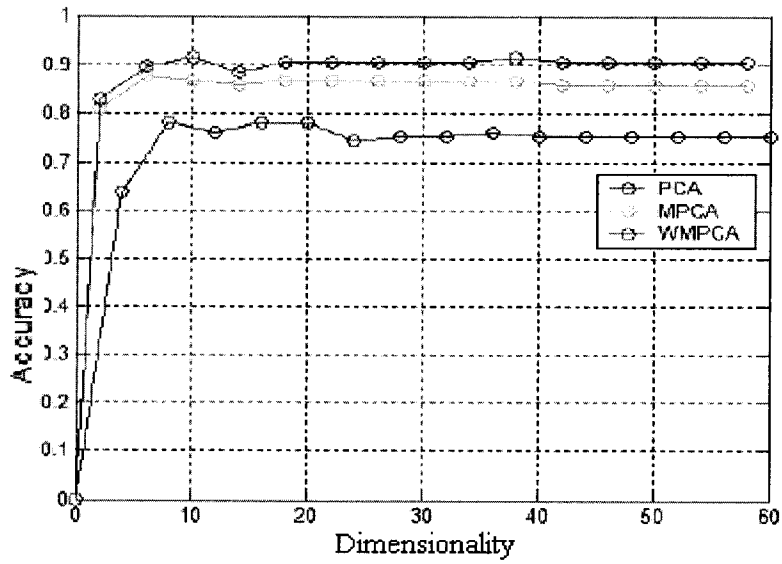


Figure 3.10. Accuracy vs the number of dimensions of the subspace corresponding to Yale database using the second testing strategy.

Figures 3.9 and 3.10 demonstrate the results of three different techniques, PCA, MPCA (modularized principal component analysis with voting), and WMPCA, using the second testing strategy on the AR and Yale databases respectively. It can be observed that the

holistic PCA has failed on the AR database. The modularized PCA approach with voting did provide better results but is still around 12% less accurate when compared to that of the proposed technique. The difference in accuracy levels between the three methods shown in figure 3.10 in the case of the Yale database is not so prominent because there are fewer variations between the training and test images.

The face recognition technique developed in this chapter performed well compared to conventional techniques, when the test images are affected due to partial occlusions and expression variations. The methodology for improved face recognition with illumination variations will be addressed in chapter 4.

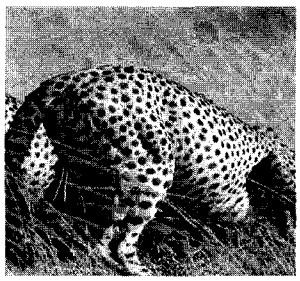
CHAPTER 4

FEATURE SELECTION STRATEGY

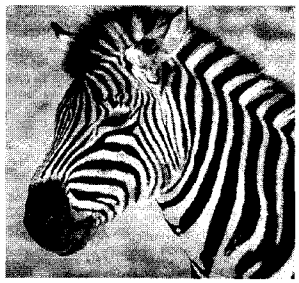
In this chapter, a face recognition technique based on novel features selected from illumination invariant feature maps and modular kernel eigen space techniques is presented. Feature based techniques are robust to illumination changes when it comes to facial recognition. A feature extraction technique based on phase is used in this research. Image features such as step edges, lines [17] give rise to points where the Fourier components of the image are maximally in phase. According to Openheim and Lim [14], the phase component is more important than the magnitude component in the reconstruction process of an image from its Fourier domain. There is also physiological evidence, indicating that the human visual system responds strongly to the points in an image where the phase information is highly ordered. Application of the PCA based recognition technique on the phase spectrum of the images resulted in better performance than on the application of intensity images and magnitude spectrum according to [41].

Figure 4.1 illustrates the concept of importance of phase. It shows images of two animals. The image shown in Figure 4.1(g) is the inverse Fourier transform of the phase spectrum from image A combined with the magnitude spectrum of image B. The image shown in Figure 4.1(h) is the inverse Fourier transform of the phase spectrum of image B combined with the magnitude spectrum of image A. It can be noticed that the processed images resemble the image to which their phase spectrums belong. For the face recognition techniques to be robust to illumination, the features extracted should be illumination invariant. A method of extracting the features based on the phase

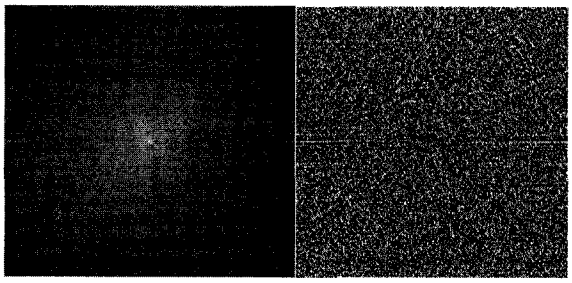
information in images for more efficient face recognition is explained in the following section.



a) Image A

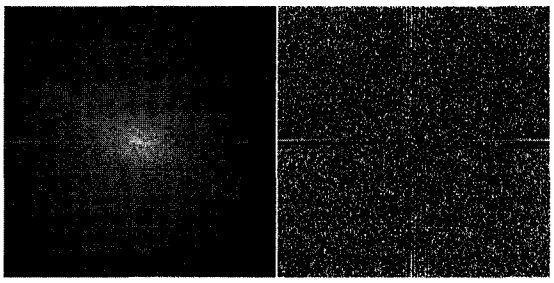


b) Image B



c) Magnitude spectrum of A

d) Phase spectrum of A



e) Magnitude spectrum of B

f) Phase spectrum of B

g) Reconstructed with d and e

h) Reconstructed with c and f

Figure 4.1. Illustration of the importance of phase information in frequency domain.

4.1 Phase Congruency Features

Edge detection schemes that use a differential approach suffer from the problem of false positives. That is, some points of maximum gradient will be marked as edges

when humans might not perceive them as such. For example, no edge is perceived in a sine wave grating at the point where the luminance crosses its mean value that is, the point at which the gradient is maximum or the second derivative is zero. Step and sinusoidal functions are perceived quite differently by the human visual system. This is illustrated in Figures 4.2 and 4.3. In Figure 4.2 we see the sine wave as a smooth gradual luminance change with no obvious features. In the second case we can see a sharp edge. The step edge is typical of the sharp changes in luminance at the boundary between two objects.

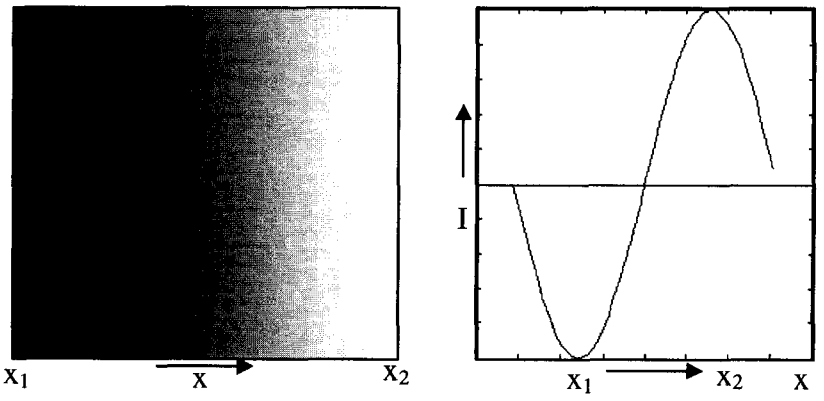


Figure 4.2. Image depicting a sinusoidal wave on the left and a sine wave grating on the right.

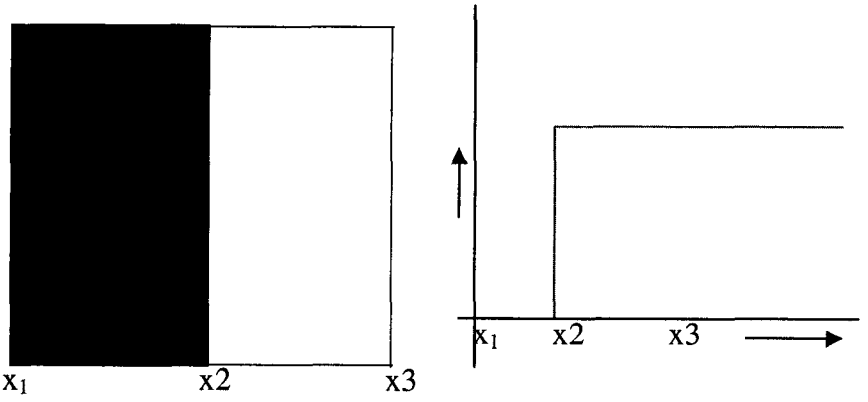


Figure 4.3. Image depicting a step edge on the left and a step wave grating on the right.

4.1.1 Phase and instantaneous phase

A general sinusoidal signal with constant amplitude is represented as :

$$S(t)=A.cos(\Phi(t)) \tag{4.1}$$

where A is the amplitude of the signal and $\Phi(t)$ is the instantaneous phase of the signal or the local phase of the signal. It can be interpreted as $\Phi(t)=\omega t+\theta$. The angular frequency of the signal is represented by 'ω' and the phase offset is given by θ.

The instantaneous or the local phase angle in this case, changes linearly with respect to the frequency. In a particular case of t=0, the instantaneous phase becomes $\Phi(t)=0$. At that instant, the sinusoidal signal has a maximum amplitude of A. Similarly when $\Phi(t)=\pi$ the signal has a minimum amplitude of '-A'. The cases explained above indicate that the knowledge of instantaneous phase gives us local information of the signal.

4.1.2 Phase congruency

It has been observed that image features such as step edges and lines give rise to points where the Fourier components of the image are maximally in phase. A periodic square wave of period '2L' is shown in Figure 4.4. The Fourier components are overlaid on the signal. It can be observed that all the Fourier components have the same phase at x=L. This explains the concept of phase congruency. That is, at the point of step, all the Fourier components have the same phase. Hence, features can be identified based on the extent to which the Fourier components are in phase.

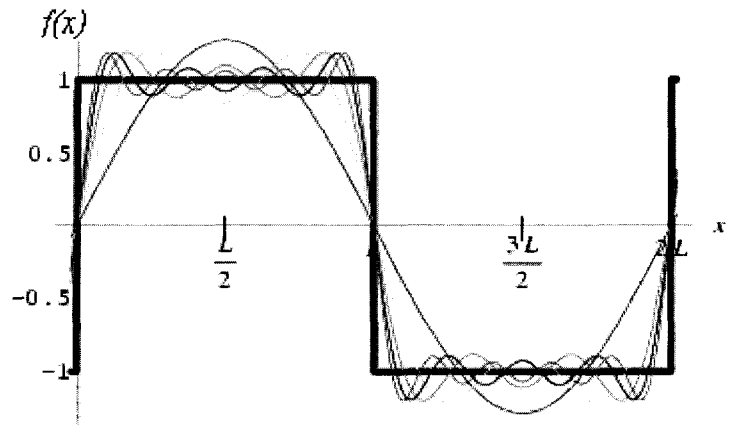


Figure 4.4. Periodic square wave with overlapped Fourier components.

4.1.3 Extraction of phase congruency features

Gradient-based operators, which look for points of maximum intensity gradient, will fail to correctly detect and localize a large proportion of features within images. Unlike the edge detectors, which identify the sharp changes in intensity, the phase congruency model detects points of order in the phase spectrum. Phase congruency provides a measure that is independent of the overall magnitude of the signal, making it invariant to variations in image illumination and/or contrast. The phase congruency technique used in this research is based on the one developed by Peter Kovesi [17]. This is in turn based on the local energy model developed in [63] to calculate the phase congruency. It has been explained that the point of strong phase congruency should relate to a maximum in the energy of the waveform. Let *I(x)* be the input periodic signal defined in the interval T= $[-\pi, \pi]$. *F(x)* is the signal with its DC component removed and *H(x)* is the Hilbert transform of *F(x),* which is a 90^0 phase shift of *F(x). F(x)* can be expressed as a Fourier series given by:

$$F(x) = \sum a_n \sin(nx + \xi_n),\ n \geq 0,\ a_n \geq 0 \qquad (4.2)$$

The Hilbert transform H of F is given by :

$$H(x) = \sum a_n \cos(nx + \xi_n),\ n > 0 \tag{4.3}$$

The local energy *E(x)* can be calculated from the signal and its Hilbert transform as :

$$E(x) = F^2(x) + H^2(x) \tag{4.4}$$

If we consider the components $a_n \sin(nx + \xi_n)$ of the Fourier transform of *F(x)* and the corresponding components $a_n \cos(nx + \xi_n)$ of the Hilbert transform *H(x)*, each of these components is a projection, in 2-dimensional space spanned by *F* and *H*, of one of a sequence of functions with amplitude a_n and argument $\phi_n(x) = nx + \xi_n (\mathrm{mod}\, 2\pi)$. These sequences of functions will sum to a function whose modulus is $\sqrt{E(x)}$.

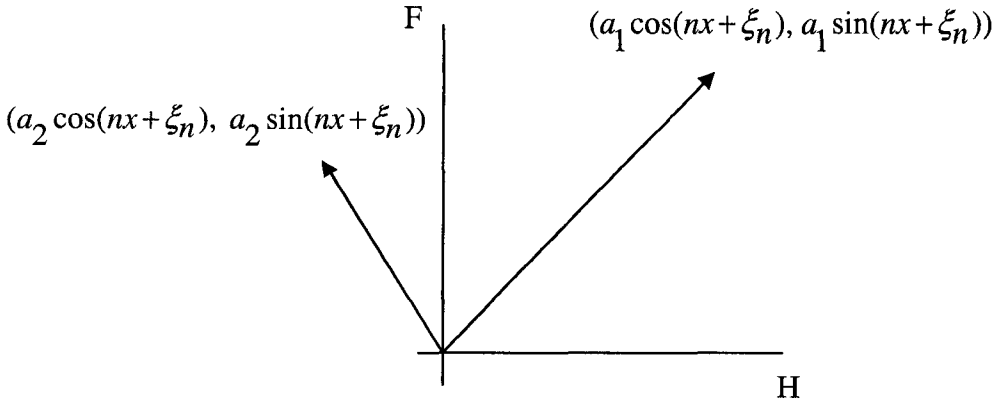


Figure 4.5. Projection of component functions onto a space spanned by F and H.

It can be observed that the amplitudes a_n are independent of *x*, and hence, for each *x* only phase relationships can affect the length of the resultant vector (modulus of the function). The resultant vector will have a higher magnitude when the phase of the vectors (component functions) has a similar phase. Hence, the phase congruency is proportional to the local energy of a signal.

It is proposed in [17] that energy is equal to phase congruency scaled by the sum of the Fourier amplitudes as shown in equation 4.5.

$$E(x) = PC(x)\sum_{n} A_n \tag{4.5}$$

This measure of phase congruency stated as the ratio of *E(x)* to the overall path length taken by the local Fourier components in reaching the end point, makes the phase congruency independent of the overall magnitude of the signal. This provides invariance to variations in image illumination and contrast.

Approximations to the components *F(x)* and *H(x)* are obtained by convolving the signal with a quadrature pair of filters. In order to calculate the local frequency and phase information in the signal, logarithmic Gabor functions are used instead of Gabor functions to obtain a non-zero DC component in the filtered signal.

If *I(x)* is the signal and M_n^e and M_n^o denote the even symmetric and odd-symmetric components of the log Gabor function at a scale *n*, the amplitude and phase in the transformed domain can be obtained as :

$$A_n = \sqrt{e_n(x)^2 + o_n(x)^2} \tag{4.6}$$

$$\phi_n = \tan^{-1}\left(o_n(x)/e_n(x)\right) \tag{4.7}$$

where $e_n(x)$ and $o_n(x)$ are the responses of each quadrature pair of filters. Equation 4.8 illustrates the response vector.

$$\left[e_n(x), o_n(x)\right] = \left[I(x) * M_n^e, I(x) * M_n^o\right] \tag{4.8}$$

F(x) and *H(x)* can be obtained from the equations 4.9 and 4.10.

$$F(x) = \sum_{n} e_n(x) \tag{4.9}$$

$$H(x) = \sum_n o_n(x) \tag{4.10}$$

If all the Fourier amplitudes at x are very small then the problem of phase congruency becomes ill conditioned. To overcome the problem a small positive constant ε is added to the denominator. The final phase congruency equation is given by equation 4.11.

$$PC(x) = \frac{E(x)}{\varepsilon + \sum_n A_n} \tag{4.11}$$

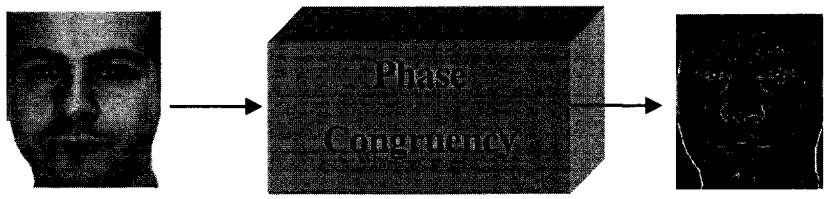


Figure 4.6. Phase congruency map obtained from the corresponding intensity image.

One-dimensional analysis is carried out over several orientations, and the results are combined to analyze a two dimensional signal (image) [17]. In this research the phase congruency maps obtained are used instead of the intensity images. Figure 4.6 shows the phase congruency image and the corresponding intensity image.

4.1.4 Steps in calculating phase congruency for images

The first step is to convolve the face image I(x,y) with a bank of 2D log Gabor filters with different orientations and scales. The log Gabor has a transfer function of the form:

$$G(w) = e^{(-\log(w/w_0)^2)/(2(\log(k/w_0)^2))} \tag{4.12}$$

where w_0 is the filter's center frequency. k/w_0 is kept constant for various w_0.

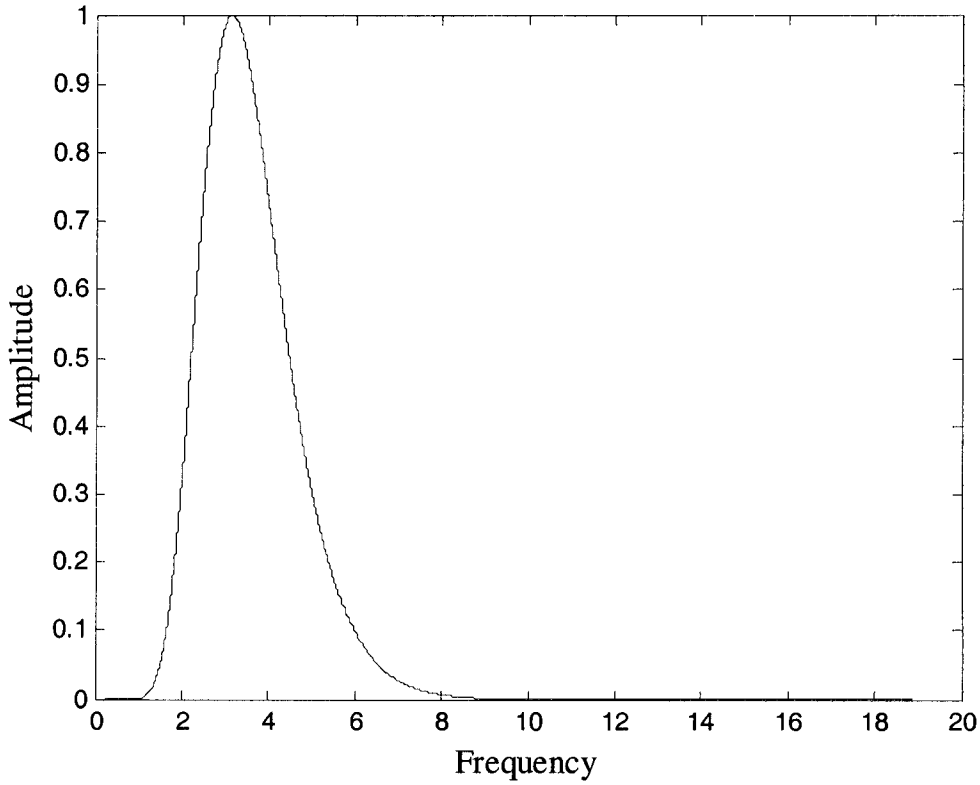


Figure 4.7. An example of a log-Gabor transfer function.

Figure 4.7 illustrates an example of a log-Gabor transfer function. The 2D log-Gabor is constructed with the cross-section of the transfer function in the angular direction of a Gaussian function and is given by :

$$G(\theta) = e^{(-(\theta-\theta_0)^2)/(2\sigma_\theta^2))} \tag{4.13}$$

where θ_0 represents the orientation of the filter and σ_θ is the standard deviation of this Gaussian function. Six orientations and 3 scales are chosen here. If M_{no}^{e} and M_{no}^{o} denote the even symmetric and odd-symmetric components at a scale n and orientation o, the response vector can be obtained by :

$$\left[e_{no}(x,y), o_{no}(x,y)\right] = \left[I(x,y) * M_{no}^{e}, I(x,y) * M_{no}^{o}\right] \tag{4.14}$$

The amplitude of the response at a given scale and orientation can be computed by:

$$A_{no} = \sqrt{e_{no}(x,y)^2 + o_{no}(x,y)^2} \tag{4.15}$$

$$PC(x,y) = \frac{\sum_o \sqrt{\left(\sum_n e_{no}(x,y)\right)^2 + \left(\sum_n o_{no}(x,y)\right)^2}}{\sum_o \sum_n A_{no}(x,y) + \epsilon} \tag{4.16}$$

Equation 4.16 gives the phase congruency of the image calculated over various scales and orientations.

4.2 Local Regions and Variations

In a face image the variations that are caused due to various reasons are in general confined to certain regions. This has already been illustrated in chapter 3. In this section the same concept is explained in a slightly different perspective. The idea of modules (non-overlapping modules) was to obtain information that is not affected due to variations. The importance of locality and modularization in face recognition is illustrated in Figure 4.8. Three different graphs of accuracy vs. dimensionality for holistic PCA, PCA on randomized pixels and Modular PCA (MPCA) are shown in this figure. The pixels in the image are indexed by row using consecutive numbers from 1 to 4096. Numbers from 1 to 4096 are randomized to form a random number set. Now the positions of the pixels are changed according to the random number set. The same randomized set of numbers is used to re-arrange the pixels in each of the training images. Figure 4.9 shows the sample training images and the corresponding randomized images.

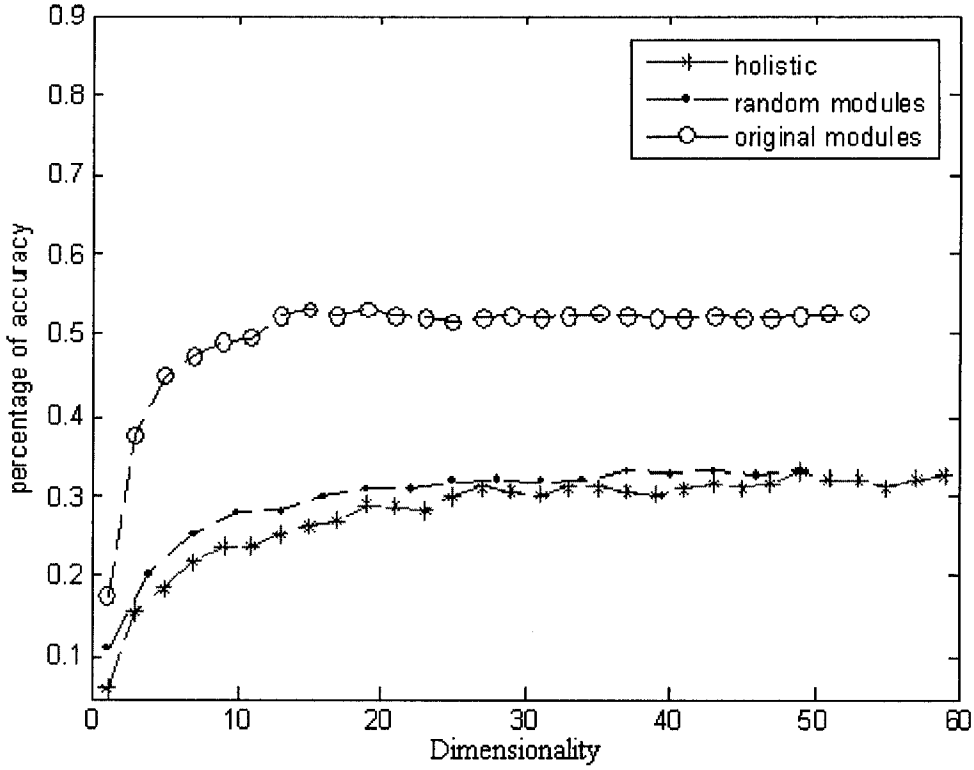


Figure 4.8. Comparison of accuracies for holistic, randomized pixel set based modular approach and conventional non-overlapping modular techniques.

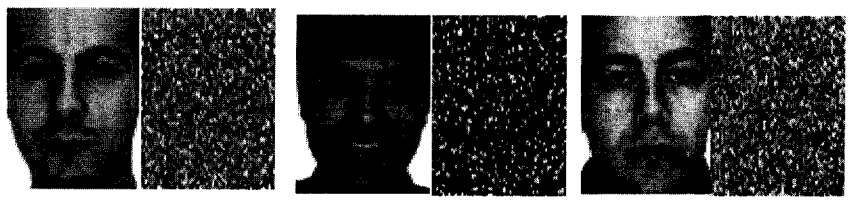


Figure 4.9. Original image and the corresponding randomized pixel position images of three different individuals.

The same randomized array is also used on the probe image and then MPCA is applied. The result is that the graph representing the accuracy against dimensionality of the subspace almost overlaps that of holistic PCA. This indicates that although the randomized pixel images contain all the pixels as the original image, due to randomization the facial variations are being spread across the image, as opposed to

being confined to a local region. Hence modularization does not have any effect. Based on the assumption that the variations in face images are local, creating more features from within the defined neighborhood, does improve the classification ability and it also suggests that, the importance of dependency between pixels decreases with increase in the distance between them.

Figure 4.10. Illustration of the concept of localization of variations and pixel dependencies.

Hence, dividing the images into sufficiently smaller modules would help in localizing these variations. Figure 4.10 illustrates the concepts of localization of variations and pixel dependencies. It can be observed that variations in this image are caused due to the face accessories (sun glasses). The localization of those variations gets better with smaller modules, but in doing so, a large amount of dependencies among various neighboring pixels is ignored. This can be countered by making the modules larger, but this will result in improper localization of the facial variations. In order to deal with this problem, a novel module creation strategy is implemented in this chapter.

4.2.1 Neighborhood defined modular spaces

Smaller sub-regions from a predefined neighborhood within the phase congruency images of the training samples are merged to obtain a large set of modules. As variations

in face images are confined to local regions, it is possible to consider additional pixel dependencies across various sub-regions and also localize the variations by merging the modules. This helps in improving the classification accuracy.

General steps for the proposed modularization technique for an image of size N×N dimensions are:

1. Each image is divided into regions of size (N/n) × (N/n) to get n×n number of neighborhoods (large modules).
2. Each neighborhood of size (N/n) × (N/n) is divided into modules of size (N/(n×i)) × (N/(n×j)), where i×j =*P* are the number of small modules within the neighborhood.
3. The *P* modules can be merged according to $\frac{P!}{k!(P-k)!}$ where *k* is the number of modules to be merged.

In this research the image size is standardized to 64 × 64 and the size of the merged module is 8 × 8. It has been proved that dividing an image of size 64 × 64 into regions of size 8 × 8 pixels is appropriate for achieving high classification accuracy [16]. The training phase of the proposed face recognition technique is illustrated in Figure 4.11. Each image under consideration is divided into small non-overlapping sub-regions of size 4 × 8 pixels. Two such 4 × 8 pixel regions from a predefined region (neighborhood) within the image can be combined to form an 8 × 8 pixel region or module. Figure 4.12 illustrates the process of obtaining such regions from each of the 16 × 16 pixels. Twenty eight different 8 × 8 modules can be created (all the combinations). A total of 448 modules are created from a 64 × 64 face image. By following the above

procedure, 448 modules of each image are produced. Eigen spaces are created for each such module. The same procedure of module creation is followed for the probe image. That is, each module is projected onto the corresponding eigen space and classified according to a minimum distance measure. A voting procedure determines the result of the overall classification by considering the individual classification results of all the 448 modules.

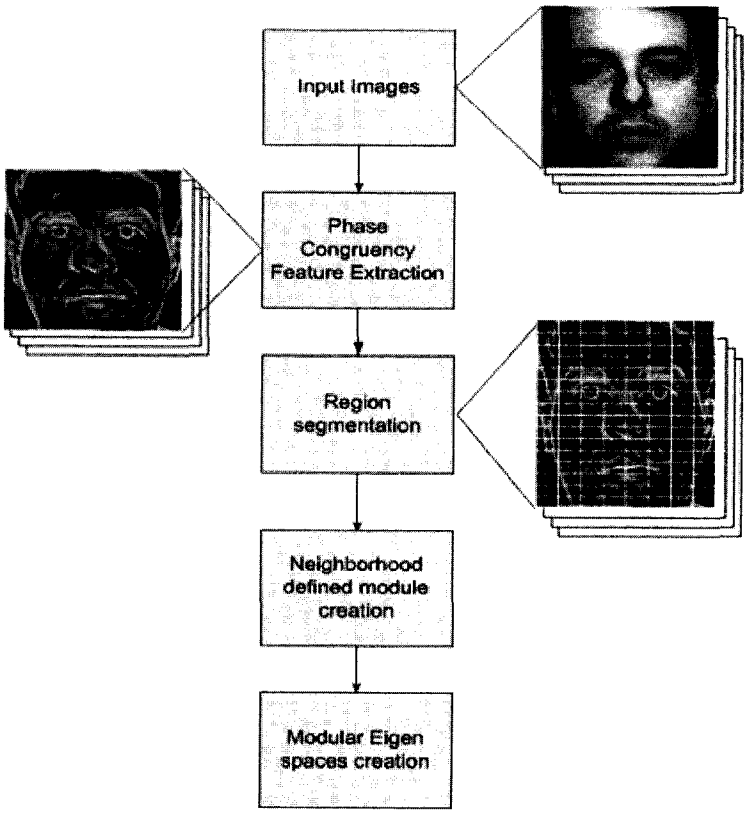


Figure 4.11. The training phase of the proposed face recognition technique.

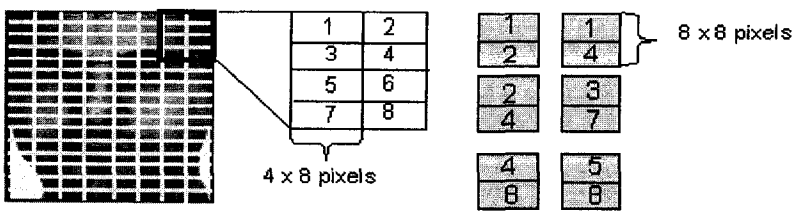


Figure 4.12. Creation of modules of size 8x8 by combining blocks of 4×8 pixels from within a neighborhood of 16×16 pixels.

4.2.2 Experimental results

The experiments are conducted on the AR database. From the AR database 40 individuals are chosen randomly to create a database. 13 images of each individual are present in the database.

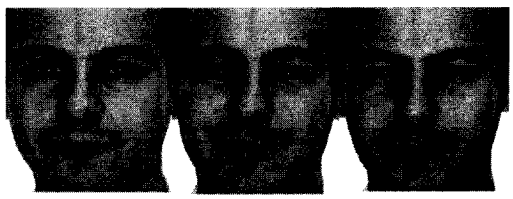


Figure 4.13. Training images of an individual in AR database.

Figure 4.14. Sample test images of the person in Figure 4.13.

Three images of each individual are used in training. Figure 4.13 shows the training images of an individual. It can be observed that all three face images are fairly neutral with little expression variations. The remaining 10 images of each individual in the database are used for testing. The sample test images are shown in Figure 4.14. The test images are affected due to lighting and/or expressions and/or partial occlusions. The graph in Figure 4.15 illustrates the relationship between percentage of accuracy and the dimensionality of the subspace for various methods namely principal component analysis

on holistic faces (PCA), modular PCA (MPCA), principal component analysis on phase congruency features (PPCA), modular subspace approach on phase congruency features (MPPCA) and the proposed method of neighborhood defined module selection on phase congruency features in the PCA domain (NMPPCA).

It can be observed that the use of phase congruency features significantly improves the face recognition accuracy. Also, modular subspaces improve the recognition for both intensity and phase congruency features. It can be observed that accuracy has risen by about 10 % in the case of NMPPCA compared to MPPCA.

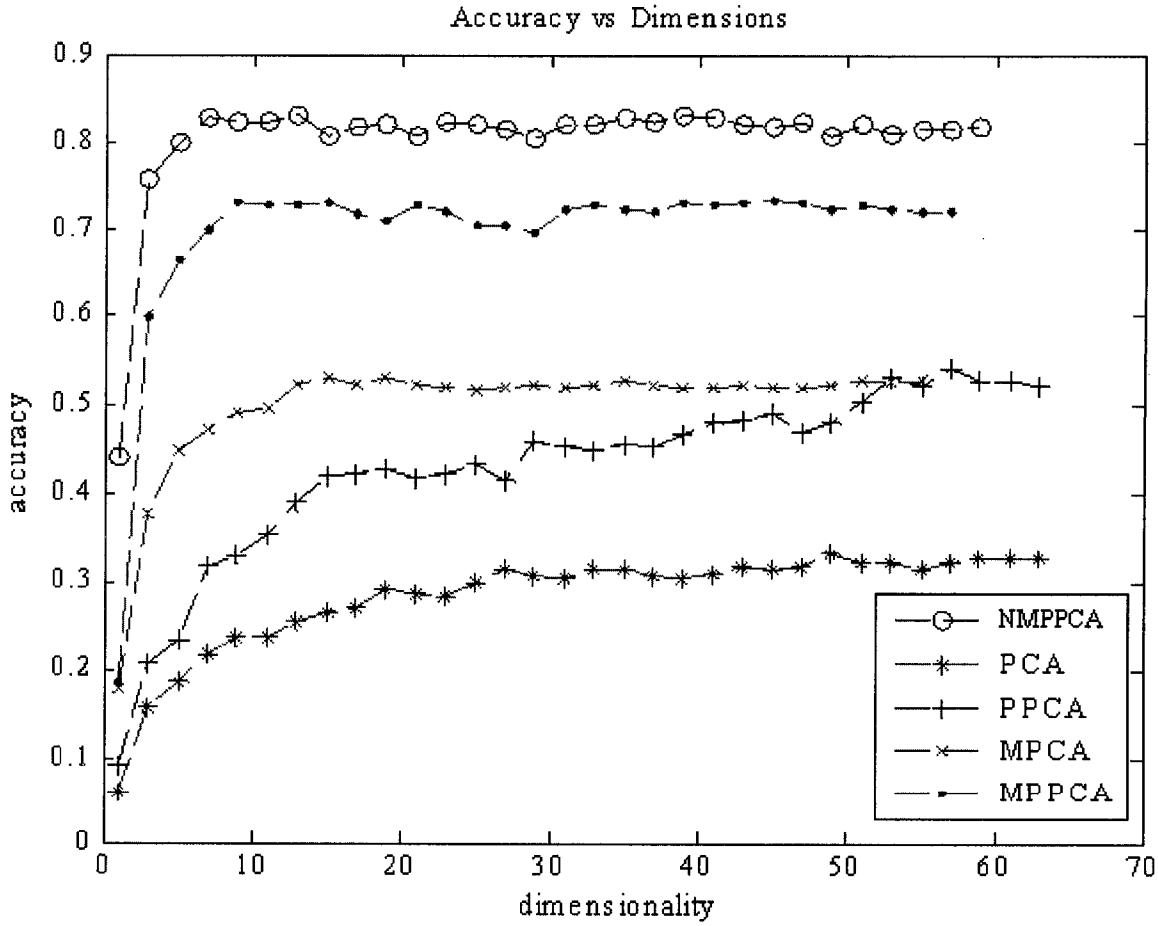


Figure 4.15. Accuracies of various methods with respect to increase in dimensionality of the subspace.

Table 4.1 A comparison of the highest accuracy achieved by several methods within the same testing environment.

	PCA	MPCA	WMPCA	PPCA	MPPCA	NMPPCA	GPCA
Highest accuracy	32%	52.5%	69.5%	52.7%	72.5%	83.0%	49.2%

Analysis of results:

It can be observed that there is a definite improvement in accuracy rate with the implementation of the phase congruency based feature extraction strategy. It can also be observed that the new modularization procedure is much better than the conventional modular PCA. It is interesting to note that the Weighted Modular PCA (WMPCA) developed in the previous chapter gives a slightly lesser accuracy rate compared to the application of modular PCA on phase congruency features (MPPCA). This highlights the importance of the lighting invariance and that the features extracted are lighting invariant. Also, the modularization procedure, when applied along with the feature extraction technique, resulted in the highest accuracy. A special comparison has been done between face recognition based on Gabor features; the phase congruency features. PCA has been applied on the Gabor features and hence, this recognition technique is termed GPCA here. It can be observed that the features generated due to phase congruency have provided better results compared to that of the Gabor features. Gabor filters localize the features of different scale and orientation.

4.3 Feature Selection Strategy in Modular Kernel Spaces

In the sections described above various features such as intensity, spectral and Gabor and the proposed techniques are compared. All these features are first projected

into linear subspaces for dimensionality reduction and then the classes are classified using a minimum distance classifier. A linear subspace approach such as PCA will not be able to capture the relationship among more than two variables. In doing so, it cannot properly depict the variations caused due to expressions, etc. In order to capture the relationships among more than two pixels, non linear spaces are utilized. In this research, a non-linear relationship among pixels is captured by projecting the data into higher dimensional spaces. The nonlinear transformation of the data into higher dimensional spaces is realized using the kernel methods to avoid computational complexities. A description of kernel principal component analysis is given below.

4.3.1 Kernel principal component analysis

PCA encodes the pattern information based on second order dependencies, i.e., pixel wise covariance among the pixels, and is insensitive to the dependencies of multiple (more than two) pixels in the patterns. Since the eigenvectors in PCA are the orthonormal bases, the principal components are uncorrelated. In other words, the coefficients for one of the axes cannot be linearly represented from the coefficients of the other axes. Higher order dependencies in an image include nonlinear relations among the pixel intensity values, such as the relationships among three or more pixels in an edge or a curve, which can capture important information for recognition. Explicitly mapping the vectors in input space into higher dimensional space is computationally intensive. The kernel trick has come in handy at this juncture. Using the kernel trick one can compute the higher order statistics using only dot products of the input patterns. Kernel PCA has been applied to face recognition applications and is observed to be able to extract nonlinear

features. The process of obtaining the weights for the input patterns in the kernel principal component analysis transformed space is described below.

Given that x_i are the vectors belonging to the training sample set X= $\{x_1,x_2,x_3,\ldots x_m\}$ where $x_i \in R^n$, there are 'c' number of classes in the training set. The data in this case is assumed to be centered. i.e., the mean of the data m_0 is given by

$$m_0 = \frac{1}{m}\sum_{k=1}^{m} x_k = 0 \tag{4.17}$$

The covariance matrix C is given by

$$C = \frac{1}{m}\sum_{j=1}^{m} x_j x_j^T \tag{4.18}$$

The eigen vectors corresponding to the non zero eigen values of the covariance matrix are calculated.

$$CV = \lambda V \tag{4.19}$$

V represents the eigen vectors and λ represents the corresponding eigen values of the matrix C. The eigen vectors arranged in descending order of the eigen values correspond to the variations in data in that order, in those directions. Let ϕ be the mapping between the input space, X, and the feature space. It is possible that the feature space H can be infinite dimensional. Hence, H can be assumed to be a Hilbert space. The samples in the feature space are $\phi(x_i) \in R^s$ where s>>n. The covariance matrix in the feature space is given by

$$C_\phi = \frac{1}{m}\sum_{j=1}^{m} \phi(x_j)\phi(x_j)^T \tag{4.20}$$

$$\phi : X \rightarrow H \tag{4.21}$$

It is assumed that the data in the feature space is also centralized, that is $\frac{1}{m}\sum_{k=1}^{m}\phi_k = 0$.

Similar to principal component analysis, V_ϕ and λ_ϕ are the eigen vectors and eigen values of the covariance matrix C_ϕ. They are obtained from the equation

$$C_\phi V_\phi = \lambda_\phi V_\phi \tag{4.22}$$

The eigen vectors of a covariance matrix that is calculated from a given set of data vectors lie within the span of those vectors. Hence, the eigen vectors can be expressed as the linear combination of the vectors in the data set. The following equation gives the relationship between the eigen vector and the sample training vectors in the feature space.

$$V_\phi = \sum_{i=1}^{m} \alpha_i \phi(x_i) \tag{4.23}$$

The projection of the sample vector $\phi(x_k)$ onto the eigen vector V_ϕ is given by

$$V_\phi \cdot \phi(x_k) = \sum_{i=1}^{m} \alpha_i \phi(x_i) \cdot \phi(x_k) \tag{4.24}$$

The above equation can be expressed as

$$V_\phi \cdot \phi(x_k) = \sum_{i=1}^{m} \alpha_i \phi(x_i)^T \phi(x_k) \tag{4.25}$$

The dot product of $\langle \phi(x_i), \phi(x_k) \rangle$ is represented with $k(x_i,x_k)$. The above equation can now be written as

$$V_\phi \cdot \phi(x_k) = \sum_{i=1}^{m} \alpha_i k(x_i, x_k) \tag{4.26}$$

$k(x_i,x_k)$ is called the Gram matrix or the kernel matrix. The elements of the kernel matrix are the dot products of the sample data vectors in the feature space. Equation 4.22 can be written as

$$\lambda_{\phi}\sum_{i=1}^{m}\alpha_i\phi(x_i)=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{m}\alpha_i(\phi(x_i)\phi(x_j))\phi(x_j{}^T) \tag{4.27}$$

Pre-multiplying the above equation with $\phi(x_k)^T$ would result in

$$\lambda_{\phi}\sum_{i=1}^{m}\alpha_i k(x_i,x_j)=\frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{m}\alpha_i k(x_i,x_j)^2 \tag{4.28}$$

After normalization equation 4.28 can be written as

$$\lambda_{\phi}K\alpha=K^2\alpha \tag{4.29}$$

This is equivalent to

$$\lambda_{\phi}\alpha=K\alpha \tag{4.30}$$

The above equation is an eigen value problem, so it can be solved to obtain the eigen vectors and eigen values of the matrix K (kernel matrix). The orthonormal vectors γ_1, γ_2, γ_3,... γ_p of K correspond to the p largest positive eigen values $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_p$. Hence

$$V_{\phi j}=\frac{1}{\sqrt{(\lambda_j)}}Q\gamma_j, j=1,2,3,4,...p \tag{4.31}$$

where $Q=\left[\phi(x_1),\phi(x_2)...\phi(x_m)\right]$, and the projection of any feature $\phi(x_i)$ vector now becomes

$$y=P^T\phi(x) \quad where\ P=V_{\phi 1},V_{\phi 2}...V_{\phi p} \tag{4.32}$$

The algorithmic steps in the training procedure are listed as follows:

(1) Phase congruency maps are obtained for each image of the training set using the method developed in section 4.1.

(2) Each training image is modularized according to the procedure developed in section 4.2.

(3) Each set of modules (corresponding modules of all the training images) is now created and processed separately.

(4) A Kernel matrix is generated for each vectorized module set after an appropriate kernel is selected.

(5) KPCA is applied for each module set and weights are obtained for all the individual modules in that set.

The steps involved in the classification of a test image are:

(1) Phase congruency features of the test image are extracted.

(2) The modular regions are created.

(3) The vectorized modules and the kernel matrix are used to obtain the weights for each individual module.

(4) Each module is classified using a minimum distance classifier on the generated weights from the training and testing phase.

In this research, a Gaussian Radial Basis Function (GRBF) kernel and a polynomial kernel function are used. It is inferred from the literature that the GRBF kernel provided better results in several classification problems [15]. Equation 4.33 represents the Gaussian RBF kernel and equation 4.34 represents the polynomial kernel function.

$$k(x, y) = \exp(-\|x - y\|^2 / \sigma) \tag{4.33}$$

$$k(x, y) = (x.y + 1)^d \tag{4.34}$$

4.3.2 Experimental results

The AR database is used again to compare and analyze the results. The same

procedure of testing is followed even in this case. That is, 3 neutral images are used for training and the remaining 10 images are used for testing. To show the effectiveness of the kernel features, ROC curves are generated for the proposed as well as other techniques. Two different kernels are considered for experimentation as explained earlier, the Gaussian RBF kernel and the polynomial kernel function. These kernels are chosen based on previous results published in the literature [15] on the facial recognition databases and character recognition databases. Although it is unclear why a particular kernel works well for a particular classification problem, based on empirical results, Gaussian RBF and polynomial kernels are the most useful. The Gaussian RBF kernel is given by equation 4.33. Each of the modules is classified separately and then the results of all those modules are combined as explained in the algorithmic steps. The kernel parameter in the case of a Gaussian RBF kernel is the standard deviation. Experimentation is carried out by varying this parameter. A false positive rate of 0.2 is set as standard, and the recognition accuracy is calculated by varying the kernel parameters.

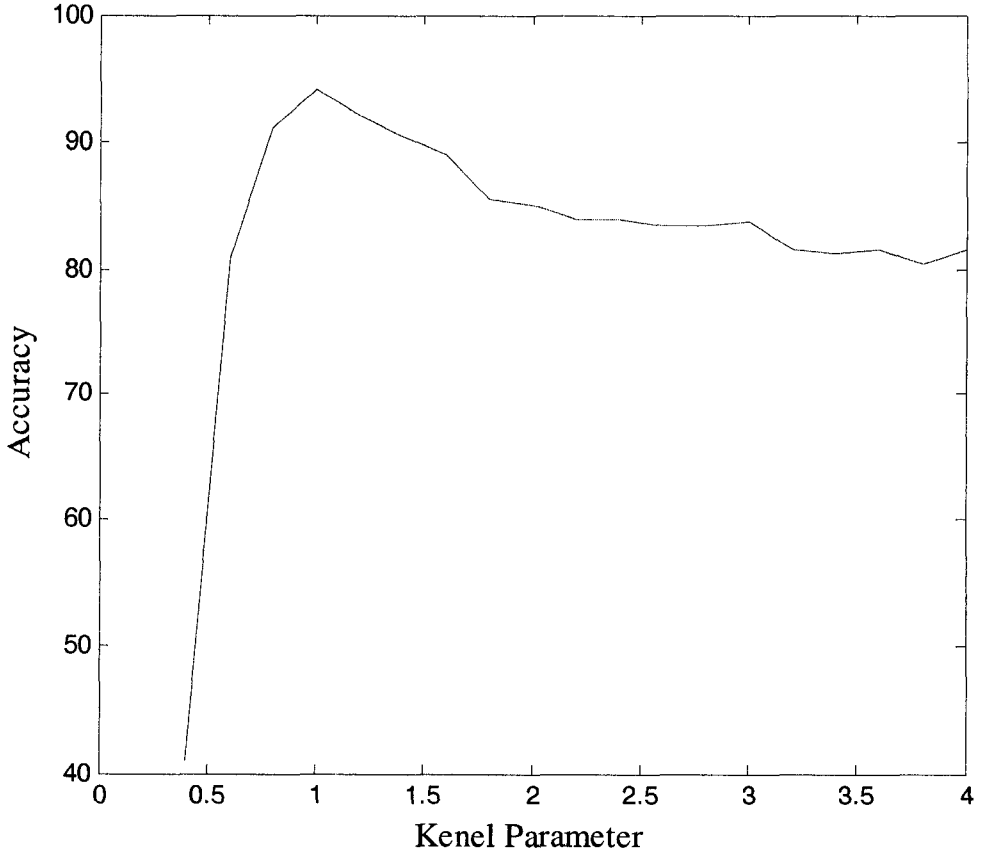


Figure 4.16. Parameter 'σ' of the RBF kernel vs. accuracy at a false accept rate of 0.2.

Figure 4.16 shows a graph of the accuracy obtained versus the kernel parameter. It should be noted that the number of dimensions is set to 20 initially. From the figure, it can be observed that for a kernel parameter of '1', there is maximum accuracy. For this kernel parameter value of 1, the optimum dimensionality is calculated by experimentation. That is, the dimensionality is varied with the kernel parameter is kept constant. The graph shown in Figure 4.17 shows the graph of accuracy versus varying dimensionality.

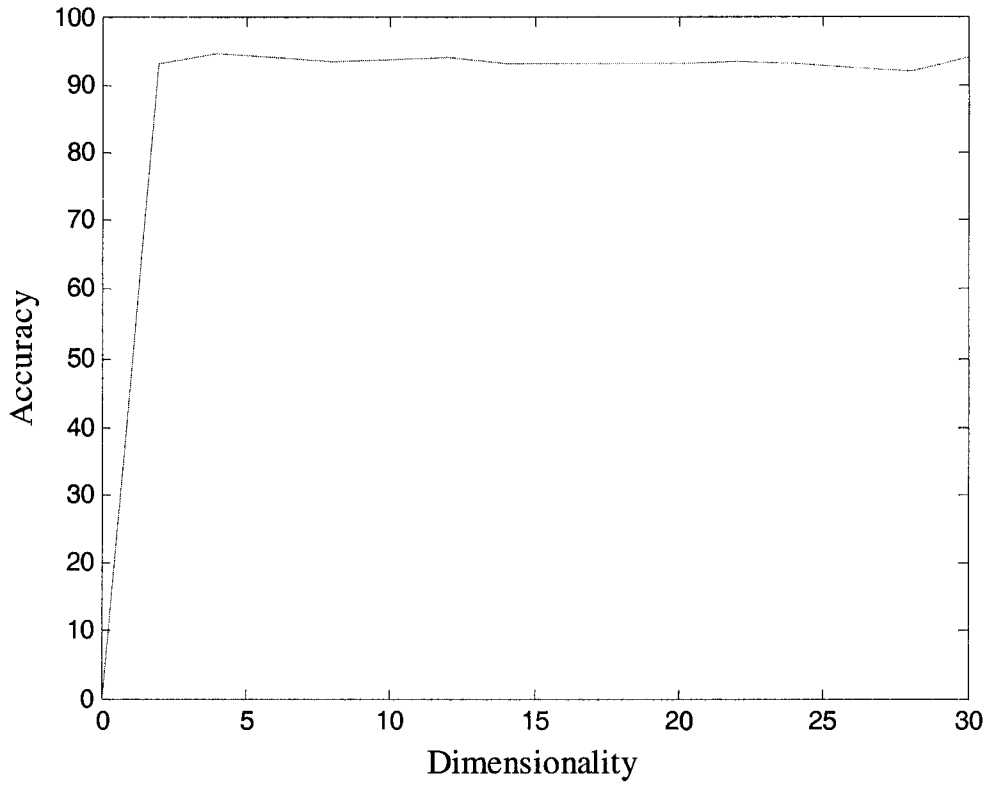


Figure 4.17. Accuracy vs dimensionality for a constant kernel parameter of '1'.

A similar procedure of optimizing the kernel parameters is carried out for the polynomial kernel. It is observed that the results obtained for the polynomial kernel, with parameter 'd' gave a maximum accuracy for a value of '1'. It is also observed that, the results obtained for the Gaussian RBF kernel outperformed those of the polynomial kernel.

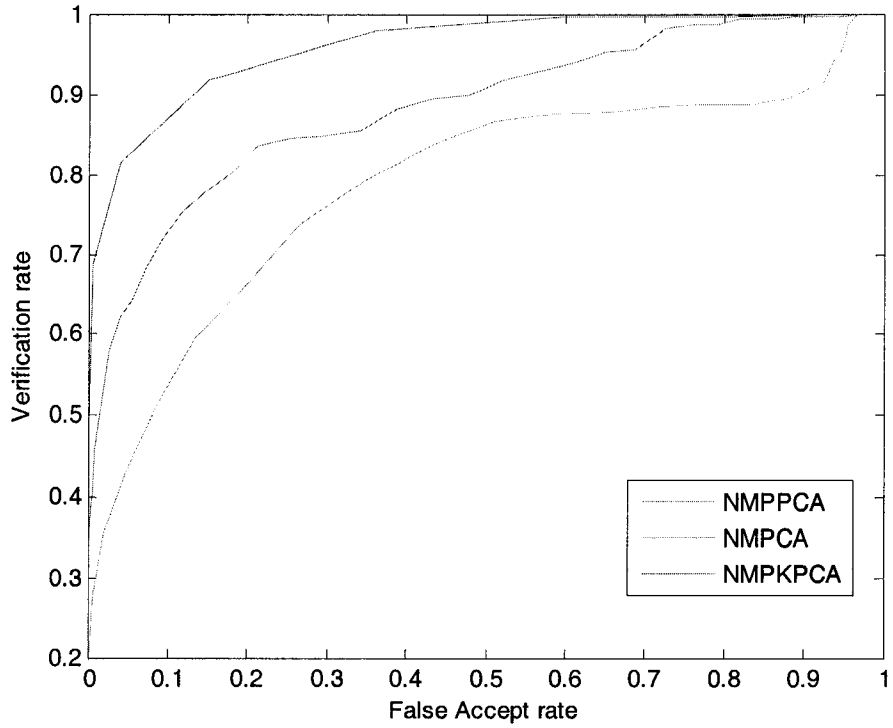


Figure 4.18. ROC's of various methods of feature extraction.

Comparison with other methods is also carried out. Three different techniques are compared here. Neighborhood defined modular PCA on Phase Congruency features (NMPPCA), Neighborhood defined Modular PCA (NMPCA) and Neighborhood defined Modular KPCA on phase congruency features (NMPKPCA). In the proposed method, the kernel parameters are initially optimized for highest accuracy before finding the ROC curve. As described earlier, the highest accuracy is obtained for a σ of 1 and a dimensionality (number of eigen vectors of the kernel matrix) of 4. Table 4.3 compares the highest accuracy achieved by those methods for a false positive rate of 0.2. Here it can be reiterated that, for a false positive rate of 0.2 the proposed method outperforms the rest of them. Figure 4.18 illustrates the ROC curves obtained for various methods.

Table 4.2 Comparison of the verification rate of different methods at a specific false accept rate of 0.2.

	False accept rate	Verification rate
NMPCA	0.2	0.67
NMPPCA	0.2	0.83
NMPKPCA	0.2	0.945

A specific set of images from the database, belonging to different individuals, is shown in Figure 4.19. The accuracies obtained for this particular image set, which is probably the most difficult to classify accurately is shown in Figure 4.20.

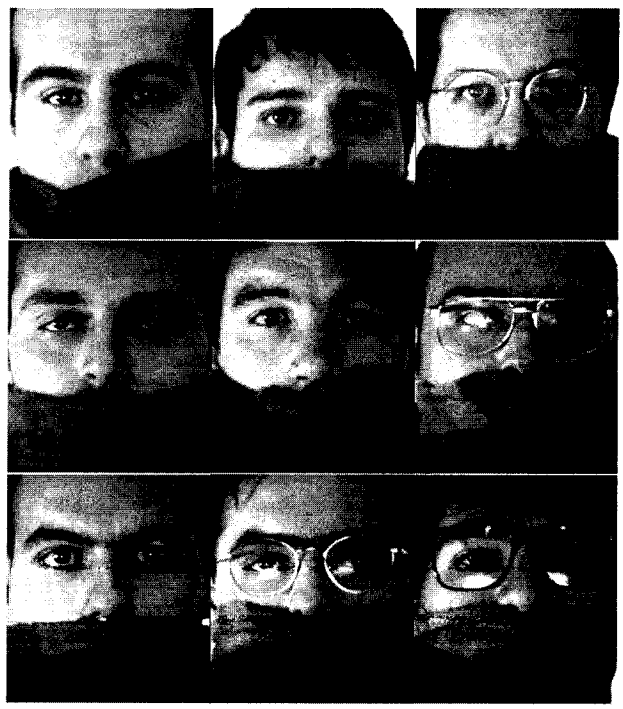


Figure 4.19. Images of individuals with the lower half of the face covered and also affected due to non uniform lighting conditions.

It can be observed that about half of the image is covered with the scarf. It can also be noticed that the face images are affected due to non-uniform lighting conditions. It can be observed that the recognition accuracy obtained for the proposed methodology of feature extraction does provide better performance when compared to the rest, although the accuracy obtained is less compared to the rest of the image sets.

This is a good result when it is noted that the face is partially occluded with glasses and that the training images do not contain any partially occluded faces.

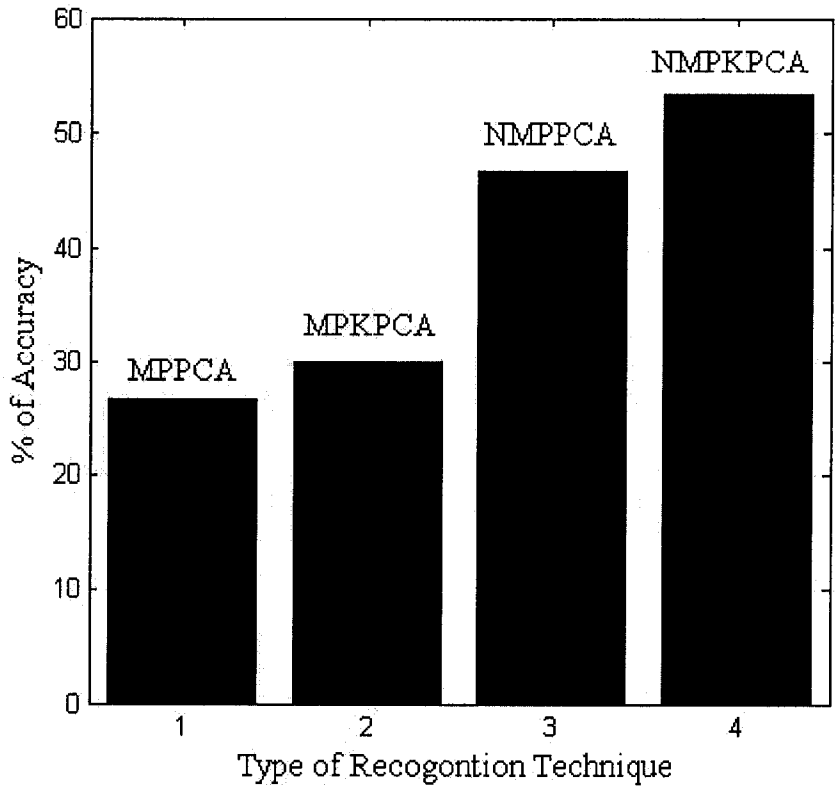


Figure 4.20. Recognition accuracy of different techniques.

A feature selection strategy based on modular kernel spaces on phase congruency images is presented in this chapter. The hybrid face recognition system developed based

on the feature selection strategy has proved to be better than conventional methods when implemented on the face images with variations.

CHAPTER 5

FACE VERIFICATION AND AUTHENTICATION USING IMAGES FROM WEB CAMERAS

Applications for facial recognition technology have been on the rise for the past few years. Surveillance and access control are two of the major applications that can benefit from an efficient recognition system. A facial verification and authentication system based on the feature selection strategies presented earlier has been developed in this chapter. This system can be installed into financial institutions or other banking firms where money transactions are crucial. A financial transaction includes opening a new account based on the information provided and depositing, withdrawing, or transferring of money based on an identity presented. So, the facial verification and authentication processes can be integrated into the system by making use of facial images of customers and the identity provided. Some of the specific problems faced by financial institutions are (i) individuals opening multiple accounts into the system and carrying out transactions with different identities, and (ii) individuals falsely portraying themselves as some one else by way of stolen identity. These are two of the major scenarios that need to be dealt with in any financial institution involving money transactions. In this chapter verification is associated with the first problem, and authentication is associated with the second. To overcome these problems an efficient facial verification and authentication system is essential. Hence, the objective of this research is to develop a face verification and authentication system that could be useful for applications related to access control in a distributed environment. The goal is to provide a reliable, cost effective, fast and expandable system whose inputs are the images from a low cost web camera. The

following sections provide the details about the facial verification and authentication system that has been developed.

5.1 Verification and Authentication System

The facial verification system developed here provides a cost effective, reliable and fast solution for the task of fraud detection in financial institutions during money transactions. The face authentication system is designed to satisfy the following requirements:

(1) Images are captured using a low cost web camera in order to keep the product cost at a minimum.

(2) Customers are not deliberately asked to pose in front of the camera.

(3) Training required to include a new individual into the system is conducted offline, during off-peak hours.

(4) The authentication process is performed online or off-line depending on the requirements.

(5) Facial verification for a new person is performed off-line.

Possible scenarios of fraud, the details of feature extraction, classification and implementation are given in the following sections.

5.1.1 Possible scenarios involving a subject in a money transaction

There are three possible situations that can be encountered at a money transaction counter. In the first case, the person is a completely new customer and he/she wants to open a new account. The claim can be either true or false. In the second situation, the

person presents an identification with which he/she accesses an account for a transaction. The identity presented may be his/her or someone else's (in which case it is a fraud). The block diagram shown in Figure 5.1 further illustrates the scenarios.

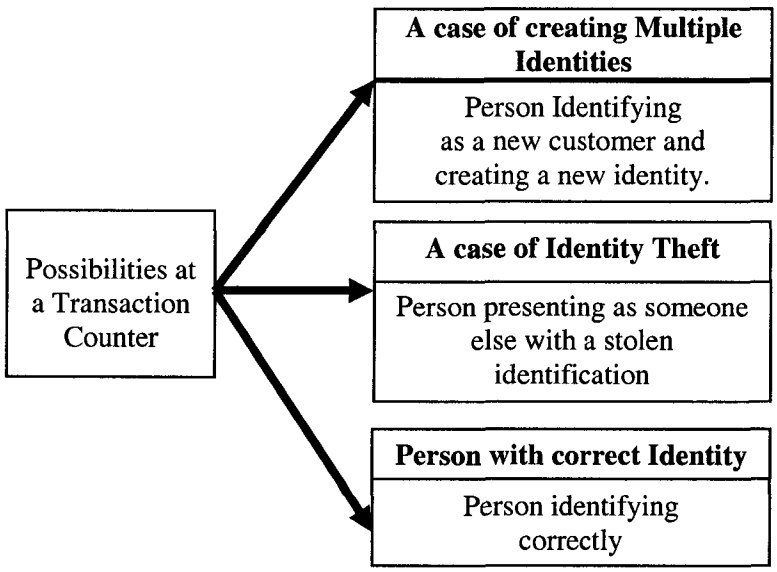


Figure 5.1. Illustration of the different possibilities at a money transaction counter.

When a person identifies himself/herself as a new customer at a transaction counter, a new identification number is provided after storing the personal information. Also, during this process face images of that person are captured from the web camera and stored against the identity that is given to the customer. The process of incorporating the individual into the system is performed offline. That is, the retraining of the system to include the new individual is performed during off peak hours. Figure 5.2 further describes the steps in dealing with a possible scenario of multiple identities. If the person identifies himself as an existing customer and presents identification, a similar procedure as above is carried out. First, the face images are captured and stored in a temporary folder.

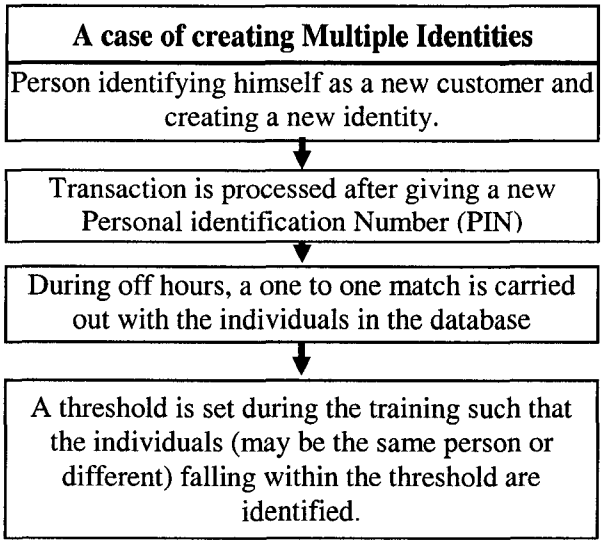


Figure 5.2. Process of identifying possible multiple identities of an individual.

Authentication is performed by comparing the stored images in the temporary folder with those that are stored in the database against the identification number presented by the customer. Figure 5.3 lists the steps in dealing with a possible scenario, of identity theft.

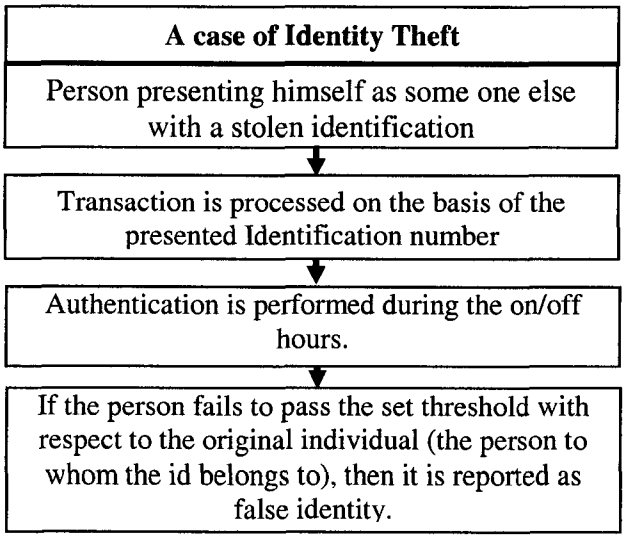


Figure 5.3. Process of identifying identity theft by an individual.

The details of the implementation of verification and authentication procedures are presented as flow charts in Figure 5.4 and Figure 5.5.

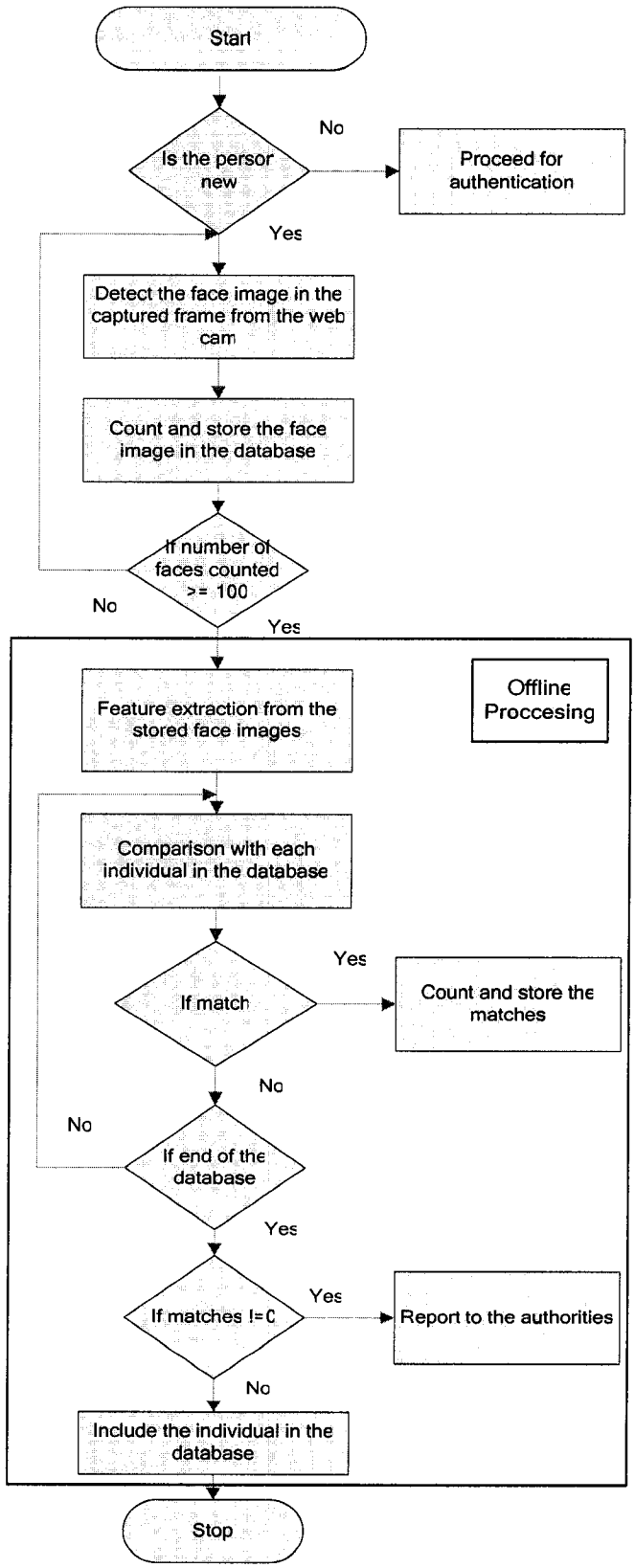


Figure 5.4. Flow chart illustrating the facial verification process.

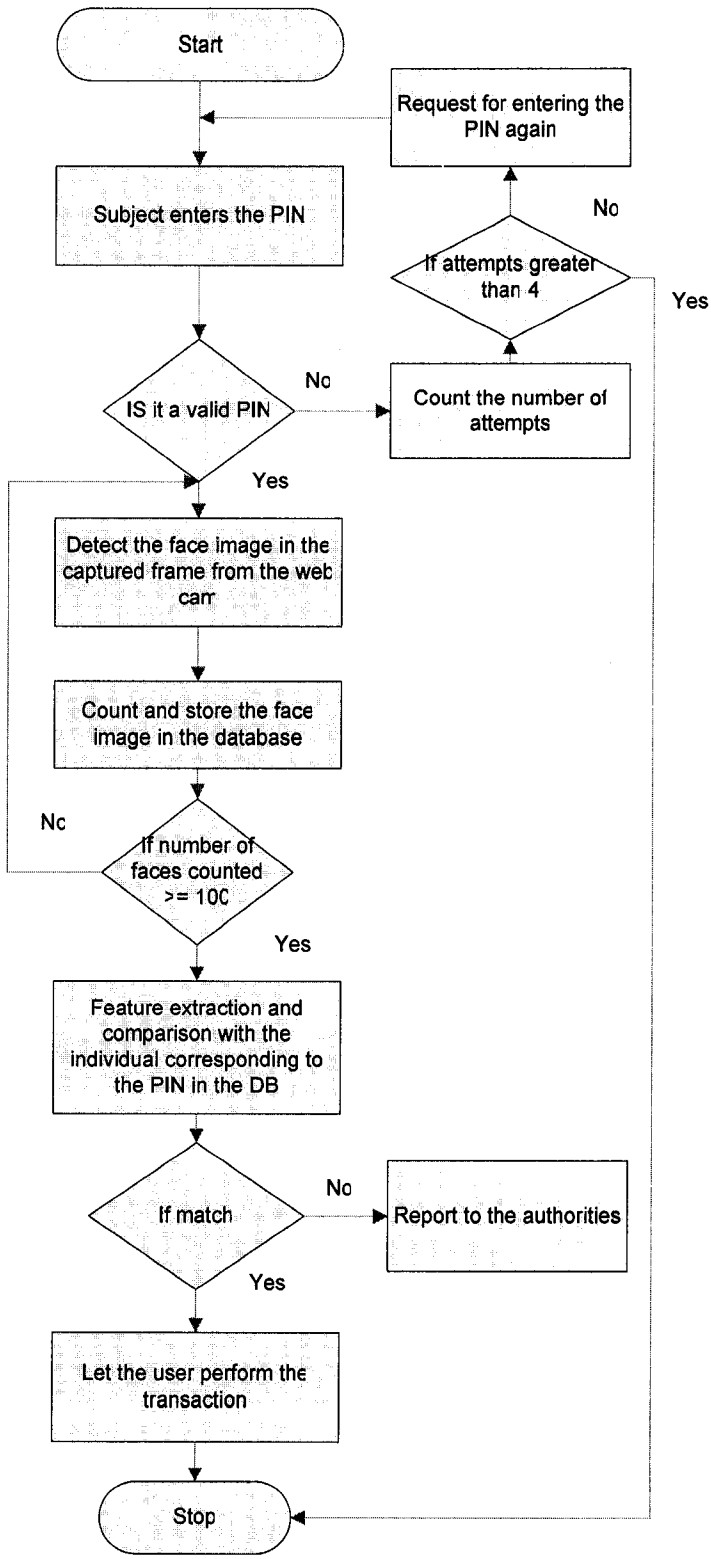


Figure 5.5. Flow chart illustrating the facial authentication process.

5.1.2 Camera and image specifications

In order to keep the cost of the verification and authentication system to a minimum, the camera is expected to be a very inexpensive one. Hence, a low cost web camera is used to capture customer images. Figure 5.6 displays the web camera that was used for the project.

Figure 5.6. Logitech web camera used for the capture of images of the customers.

The specifications of the camera are as follows:

- True 1.3 megapixel sensor with RightLight™ 2 technology,
- Live video capture: up to 640x480 pixels.
- Frame rate: up to 30 frames per second,

The specifications of the images that are captured and processed are as follows:

- Input frame rate from the camera: 30 frames / second,
- Frame of the image size captured by the camera: 640× 480 pixels,
- Minimum size of detected face within each frame: 64×64 pixels,
- Size of the face image that is processed for recognition: 64×64 pixels.

5.1.3 System specifications and processing speeds

The system specifications and the approximate times of processing various modules are as follows:

System specifications:

- Processor : Intel Xeon(TM) 1.70GHz
- Memory: 1.5 GB

Approximate times of operations:

- Time taken for detection and storage of 100 images of a customer : 10 seconds
- Time taken for training a person into the recognition system : 6 seconds
- Time taken for authentication of a person : 8.2 seconds
- Verification of a new person for the presence of multiple entries: (N×2.2 + 6) sec.

  'N' is the number of people in the database.

5.2 Feature Extraction and Classification

The feature extraction procedure developed in chapter 4 works very efficiently on face images that are affected due to partial occlusions, lighting variations and expressions, but the computational cost required for the purpose of generating those features and performing the recognition task is a very expensive process. In order to strike a balance between the computational efficiency and accuracy, the recognition technique is restricted to NMPCA (Neighborhood defined modular PCA). In this recognition technique, the intensity images are processed directly to obtain the modules, and the modules are classified individually after the projection into the respective subspaces. In order to deal with the illumination variations, a simple histogram

equalization procedure is implemented as a preprocessing step. The detailed procedure of the feature extraction and recognition is presented in chapter 4. For detection of face images, a detection algorithm developed by Viola and Jones is implemented. In ideal situations, the training and the test images should be frontal face images, properly cropped and resized. Since we are dealing with a real world scenario, this may not be possible. This is because the customer is not asked to pose before the camera. Instead, the camera is positioned in such a way that the customer's frontal face is automatically captured. This may result in the capture of non-frontal (not perfectly frontal) images, which in turn will affect the facial verification process. To overcome this difficulty multiple face images are captured over a certain interval of time. Figure 5.7 shows the detected frontal face of an individual during the authentication process.

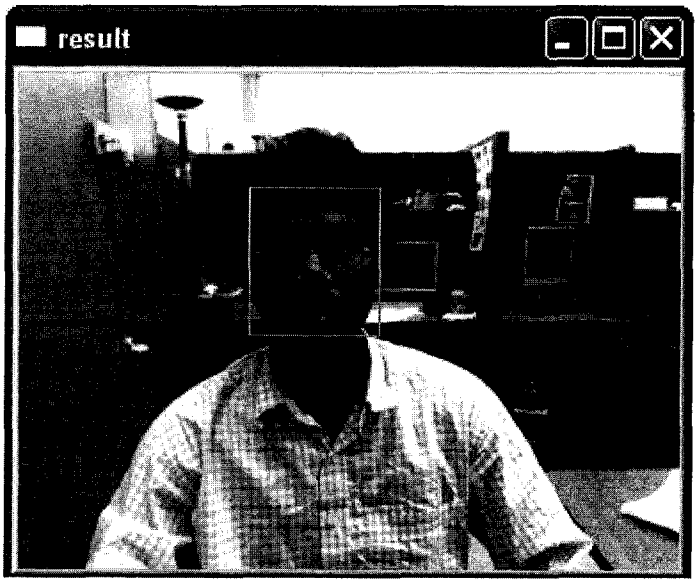


Figure 5.7. Detected face of a test person.

In this research, 100 face images of an individual are detected and stored over 10 seconds approximately. These 100 detected face images are stored in a folder. The folder is a

temporary folder if the requirement is to authenticate a person. If the purpose is to search for multiple identities, then the images are stored in a folder whose name is the identification number that is assigned to that customer. The face images captured by the web camera will vary from one image to the other even though they are of the same individual. This is especially true since images are captured over a period of 10 seconds. In addition to variations in pose and orientations of the images, the appearance of customers might change over time. Also, the sensor within the web camera is very sensitive to subtle lighting variations.

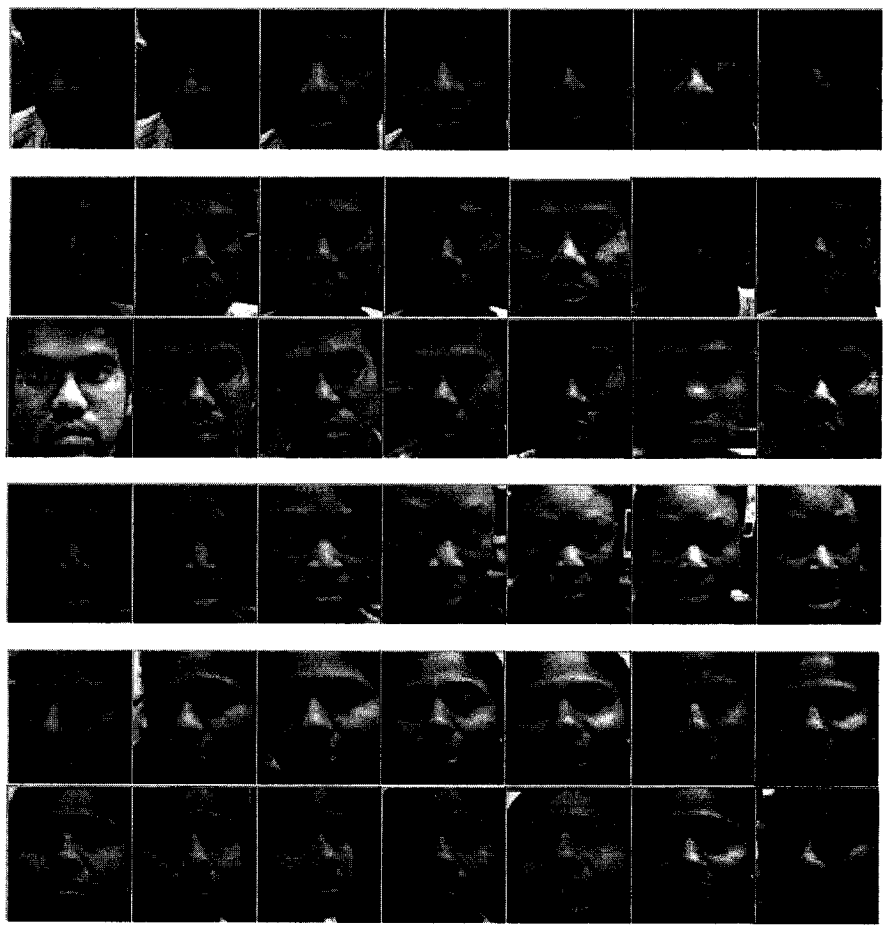


Figure 5.8. Images depicting the variations in faces due to lighting and movement of the individuals during the capture process. Also the set contains images captured during different days.

This causes the images to be captured with different shades of lighting when observed over two different instances of time. Examples of face images of the persons captured are shown in Figure 5.8. These images consist of face images captured over a small interval of 10 seconds and also over different days. Since the verification and authentication is carried out as a one-to-one matching process, the feature extraction is performed over the 100 images of an individual. The training process may not be robust because of the variations present in the captured images. One possible way to overcome this difficulty is to group similar images within the 100 images and extract the features of each group separately. To do this, a K-means clustering process is performed over the 100 captured images. The clustering process will group images into different groups. This reduces variations within each group. Details about the clustering process are presented in the following section.

5.2.1 k-means clustering

K-means clustering is an algorithm to classify or to group objects based on attributes/features into K number of groups. K is a positive integer number. The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid. Initially, the data points are the vectorised face images of a single class or an individual, and these needs to be arranged into K clusters. K data points are selected randomly and are treated as the centroids for the K groups. The distance of each point in the data set is calculated with respect to the K centroids and each is assigned to the centroid to which its distance is minimum. The centroids are re computed for each set. In the next step, every point is assigned to the cluster whose centroid is closest to that point. These steps are alternated until a stopping criterion is

met, i.e., when there is no further change in the assignment of the data points. If there are 'N' data points, the data is clustered into K predefined number of clusters such that the criterion function 'J' below is minimized. An illustration of the clustering process is shown in Figure 5.9.

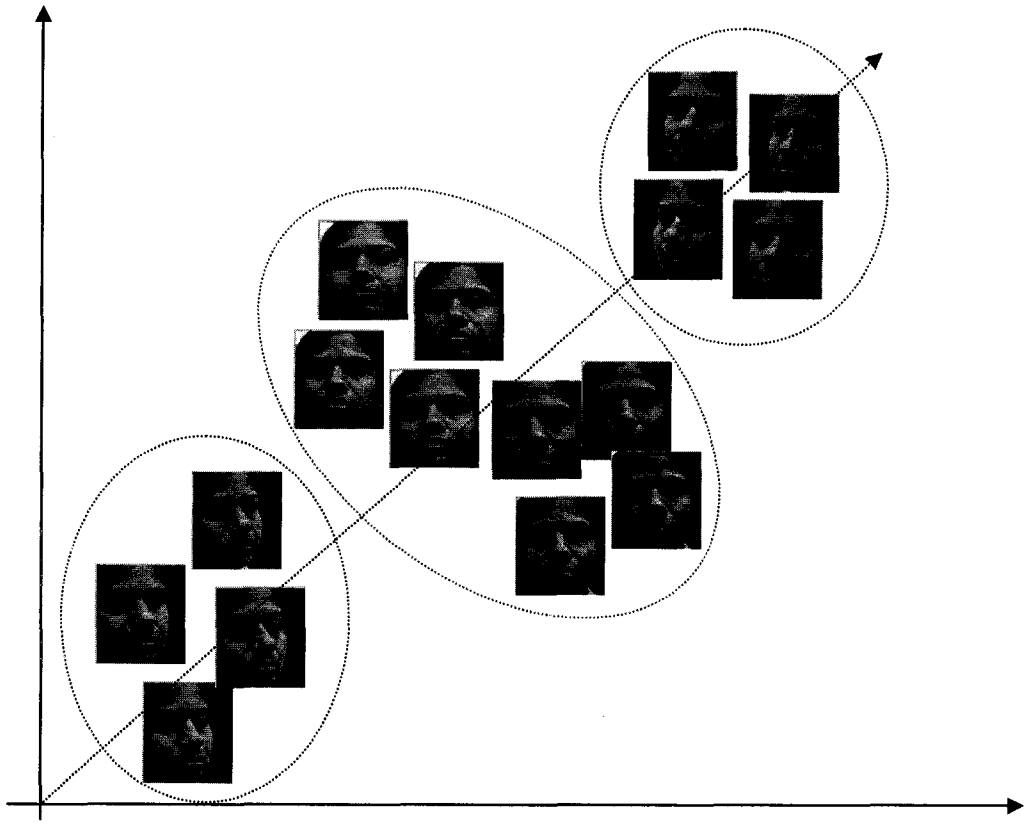


Figure 5.9. Illustration of the clustering process where similar images are grouped together.

$$J = \sum_{j=1}^{K} \sum_{n \in j} \| x_n - \mu_j \| \tag{5.1}$$

where x_n is a vector representing the n^{th} data point, and μ_j is the geometric centroid of the data points in cluster S_j. The number of a cluster is given by 'j'= 1,2 ....K. For convenience and to reduce the computational complexity, the number of clusters is

assumed to be 3. After the detected face images of the individual are captured, those images are grouped into 3 different clusters. After the clustering process, each of the groups is trained individually. That is, for each group, the features are extracted.

5.2.2 Verification and authentication process

The captured face images of a new customer are stored in a folder whose name signifies the identification number given to the individual. The folder with the images is stored on a server (a centralized location). During off peak hours, the 100 images of the individual are trained to be incorporated into the system. The training procedure has the following steps.

(1) The 100 images are clustered into three groups in the face space.

(2) Each set of images are handled separately to obtain the features.

(3) The set of features for the three groups are stored.

This procedure is followed on a daily for all new customers who register basis.

Verification process:

The new customer's features are compared against all the features of the rest of the individuals present in the database. A minimum distance classifier is used for the purpose of comparison. Since the process of verification is one-to–one, the test features (of the new individual) are compared with the features of the other individuals in the databases serially. The steps in the one-to-one comparison process are as follows.

For convenience the j^{th} image of the first individual 'A' is called I_j^A and the j^{th} image of the second individual 'B' is I_j^B (new customer) where j=*1, 2, 3,.....100.* The image set I^A are clustered into three groups I_p^{A1}, I_q^{A2}, I_r^{A3} where $p = 1, 2, 3,.. C_1^A$ $q = 1, 2, 3,.. C_2^A$, $r =$

1, 2, 3,.. C_3^A and $C_1^A + C_2^A + C_3^A = 100$. Modular features are generated for the images in each cluster separately. M_{pn}^{A1}, M_{qn}^{A2}, M_{rn}^{A3} where *n= 1, 2, 3,....448* represents the n^{th} module, for the p^{th} image in cluster '*A1*'. The mean of all features within each cluster for a module is calculated as:

$$F_n^{A1} = \frac{1}{C_1^A} \sum_{p=1}^{C_1^A} M_{pn}^{A1} \quad where\ n = 1,2,3,...448 \tag{5.2}$$

Similarly, for the rest of the clusters, the mean features are calculated and represented with F_n^{A2}, F_n^{A3}. Modular features for the images of individual B are also created by projecting the image modules onto the corresponding subspaces of 'A'. These represented as M_{jn}^{BA1}, M_{jn}^{BA2}, M_{jn}^{BA3}, are modules formed for test individuals. Since there are three clusters, three sets of features are formed. The distance between each of the test module and the corresponding modules of the three mean module features is then calculated and the minimum of those is calculated.

$$d_n(j) = \min\left\{\left\|M_{jn}^{BA1} - F_n^{A1}\right\|, \left\|M_{jn}^{BA1} - F_n^{A2}\right\|, \left\|M_{jn}^{BA1} - F_n^{A3}\right\|\right\} \tag{5.3}$$

$d_n(j)$ is the minimum distance of the n^{th} module of the j^{th} image with respect to the corresponding modules in the three clusters. A module M_{jn}^{B} (n^{th} module of the j^{th} image of B) is said to be classified in the favor of the individual A if $d_n(j) < \theta$ where θ is the threshold calculated experimentally. If V_k represents the number of modules of Image I_j^B classified in favor of A then that image is said to be the same as 'A' when $k > T$, where T represents the number of modules that need to be correctly classified. T is also

calculated experimentally. Person 'B' is said to be authenticated if at least 'P' number of the total 100 images of the person are classified in favor of 'A'. The same procedure is followed for the verification of multiple identities; the difference is that the comparison is carried out with all individuals in the database, one by one. All matches are then listed. If there is more than one match, it is treated as a case of fraud for the presence of multiple identities.

5.3 Experimental Results

Images of 35 individuals are used in the experimentation procedure. Most of the people involved belong to the ODU vision lab. The individuals selected are of different ethnicities. The experimentation process is carried out in such a way that it closely simulates the real world operational scenario. Multiple instances of each participating person are created and included in the database. This simulates the presence of multiple identities of a person. The number of instances of different individuals varies randomly. Figure 5.10 shows the face images of individuals stored in the folders. The name of the folder depicts the identification number given to the person. It can be observed that the individual with identity '16' (folder name) has two more instances or identities with numbers '28' and '29'. Similarly, the individual with identity '6' has 14 instances or identities included in the database. The minimum number of instances created for an individual is 2 and the maximum is 14. For forming different instances, face images of the same individual are captured on different days. Also, some of the face images captured contain sunglasses, facial hair and makeup to simulate the operational scenario. A total of 118 image sets are created for 35 individuals. Each image set consists of 100

face images. To test the accuracy of the algorithm developed on the data set, receiver operating curves are generated. For a particular threshold, both the verification rate and the false positive rate are calculated. As explained in the previous sections, there are three thresholds corresponding to (i) the distance threshold θ (ii) number of modules, T (iii) number of images 'P'. Thresholds (ii) and (iii) are kept constant, and only the distance threshold is varied to obtain these curves. Figure 5.11 shows the ROC for the NMPCA (proposed method). The curves show that the NMPCA technique performed very well. It achieved 100 percent recognition rate for a false accept rate of 0.12.

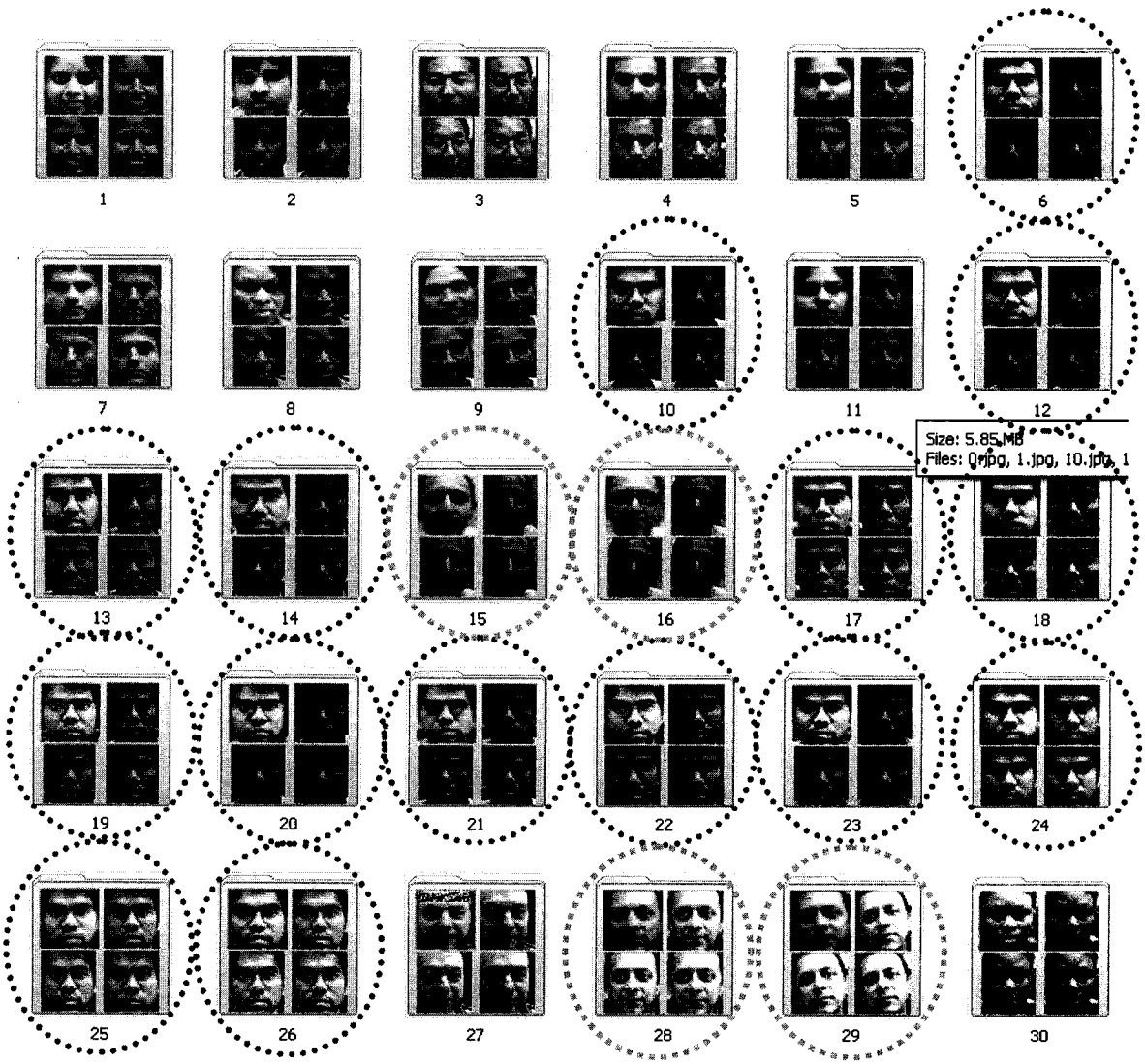


Figure 5.10. Stored face images of different individuals, similar colored circles indicate same individual with different identities.

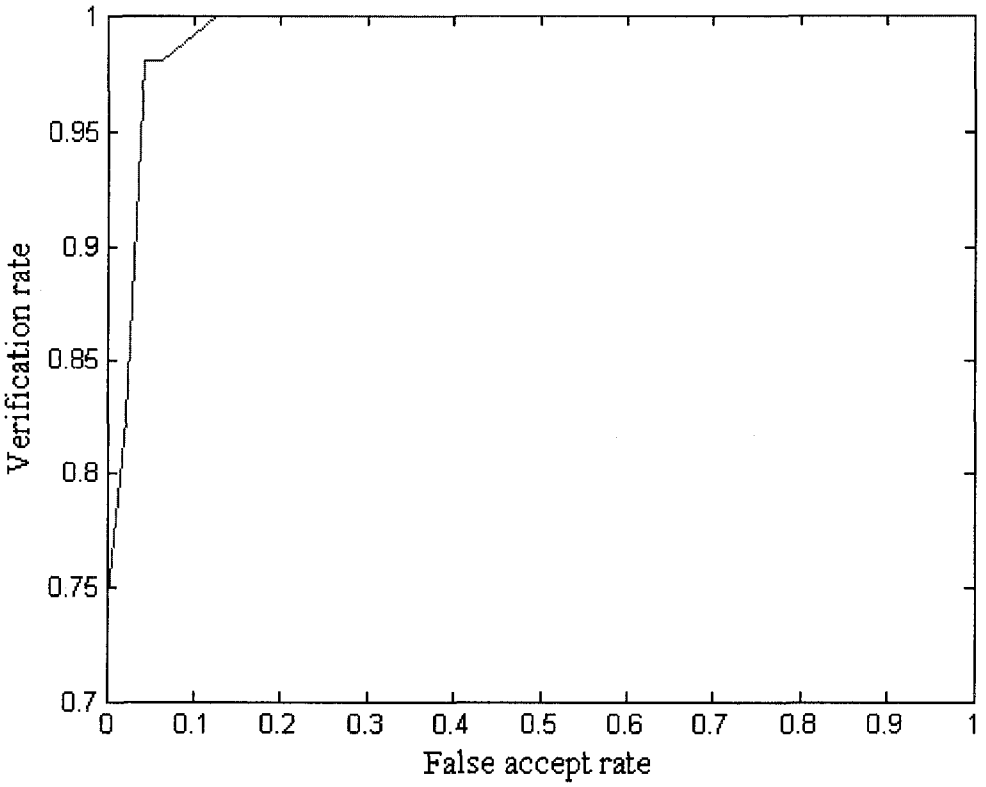


Figure 5.11. ROC obtained for the proposed face recognition technique.

CHAPTER 6

MULTI-SENSOR IMAGE FUSION FOR FACE RECOGNITION

Even though face recognition technology [7] has progressed from linear subspace methods [12] such as eigen faces and fisher faces [11][12][23] to non linear methods such as KPCA, KFD,[15][25][36-38] as explained in the previous chapters, many of the problems are yet to be addressed completely. In addition to challenges such as expression and pose variations, partial occlusions, addressed in the previous chapters, the face recognition techniques face a major bottle neck in the form of illumination variation. The feature extraction techniques presented earlier address the problems of illumination variations to some extent. This is because the lighting variations can vary from very bright to very dark conditions. Under such extreme conditions it is possible that the images captured under the visible spectrum may not be able to provide information sufficient for performing tasks such as facial recognition. A recent trend in the field of pattern recognition is to make use of information from multiple sensors for effective classification, even in extreme conditions. This chapter presents a novel method of fusion of information from multiple sensors. The fusion procedure along with the feature selection techniques, developed in earlier chapters, resulted in improved face recognition accuracy under varying illumination conditions. The recognition techniques developed in earlier chapters are applied to visual, thermal, fused image modalities and the results are evaluated. Testing is carried out on the Equinox face database, which consists of face images captured from multiple sensors. The multi-sensor images are pre-registered.

6.1 Sensors and Systems

Images from multiple sensors are utilized for the purpose of obtaining complementary information for several applications including surveillance and, remote sensing. Face recognition is one of the applications that can greatly benefit from the use of multiple sensors. Visible sensors cannot capture enough information for effective face recognition in low to very low illumination conditions. Figure 6.1 displays multi-spectral images of an individual.

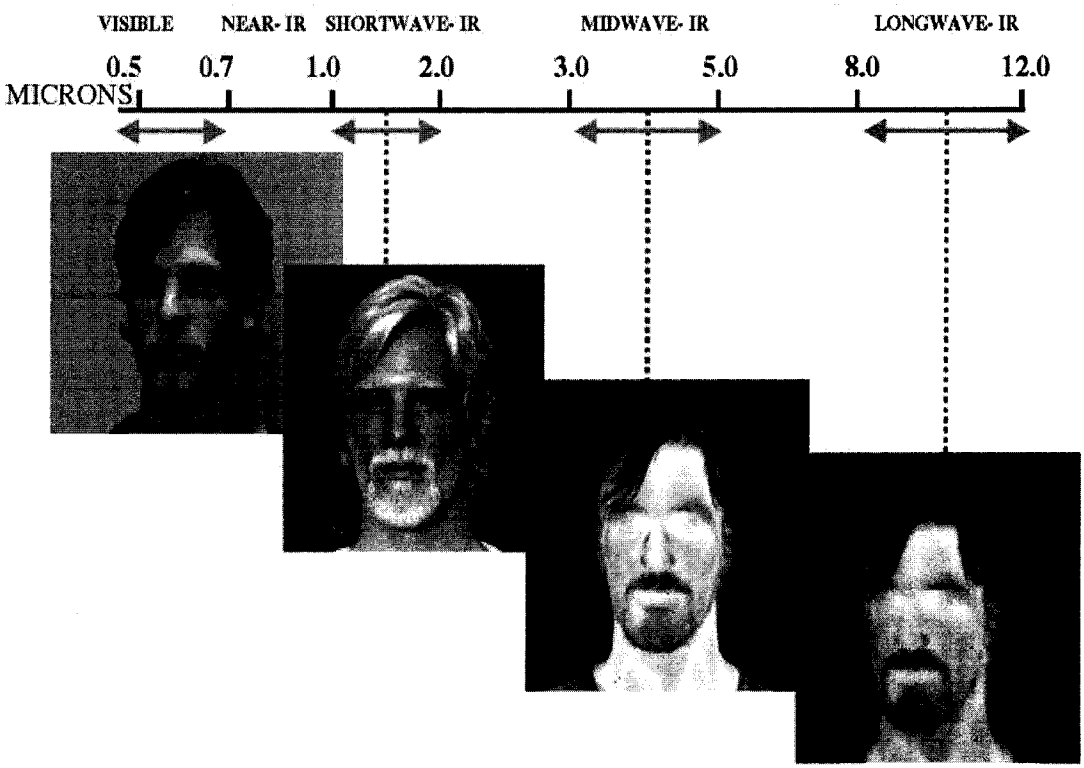


Figure 6.1. Illustration of multi-spectral face images [21].

The visible spectrum ranges from 0.5 microns to 0.7 microns. The near infrared (NIR)region ranges from 0.7 micron to 0.9 microns and the short wave infrared (SWIR) from 0.9 microns to 2.4 microns. The images that are captured in the above spectrums are

due to the reflection of light from the body in the respective spectrums. Hence, in very low illumination conditions even the NIR and SWIR images are not efficient enough for face recognition. On the other hand MWIR and LWIR sensors capture the amount of heat generated by the objects in the scene and not the light reflected from those objects. It means thermal infrared sensors are illumination independent. Hence, the information obtained from these sensors could prove complementary to the visible spectrum images and provide better face recognition in extreme illumination conditions. Figure 6.2 displays the visible image and the corresponding LWIR image of an individual.

Figure 6.2. Visible and thermal images of a person.

Despite its robustness to illumination changes, however, IR imagery has several drawbacks including that it is sensitive to temperature changes in the surrounding environment and, variations in the heat patterns of the face, and it is opaque to glass. In contrast to IR imagery, visible imagery is more robust to the above factors but as mentioned earlier, is very sensitive to illumination changes.

6.2 Related Work

While the nature of face imagery in the visible domain is well-studied, particularly with respect to illumination dependence, its thermal counterpart has received less attention. Previous studies have shown that infrared imagery offers a promising alternative to visible imagery for handling variations in face appearance due to illumination changes. There are a few studies of face recognition algorithm performance in the long wave infra red spectrum although not comprehensive. It is interesting to observe that face recognition on thermal images in [53] degrades more sharply than with visible images when probe and gallery are chosen from different sessions. Results in [53] indicate better performances obtained with visible imagery indoors under controlled lighting conditions, but outdoors the thermal image based face recognition system outperformed the visible imagery based one. Also, the thermal face recognition results for both indoor and outdoor environments are comparatively less different from each other, thus reiterating that illumination has little effect on thermal imagery. The conclusion of the studies in [53-56] is that despite the degraded thermal recognition performance, fusion of both visible and thermal modalities yields better overall performance. Most of the studies conducted on thermal image face recognition relied on the conventional eigen faces approach. That is the application of PCA on intensity images. This was particularly relevant for us because, in [53] it was noticed that while multi-session thermal face recognition under controlled indoor illumination was statistically poorer than visible recognition with two standard algorithms, the significance was substantially reduced with an algorithm [53] more specifically tuned to thermal imagery. The results suggest that previous results reported on thermal face recognition may be incomplete. Hence, the

building of effective algorithms to fuse information from both spectra has the potential to improve face recognition performance. It is possible to realize sensor fusion on different levels: sensor data level fusion, feature vector level fusion, and decision level fusion. In [53] it is considered fusion on the decision level to have more potential applications.

6.3 Image Registration

The registration of images acquired by sensors of different modalities is of special interest to remote sensing and medical imaging applications, as the information gained from such a set of images is of a complementary nature. This can also be implied of the current scenario of thermal and visual image applications.

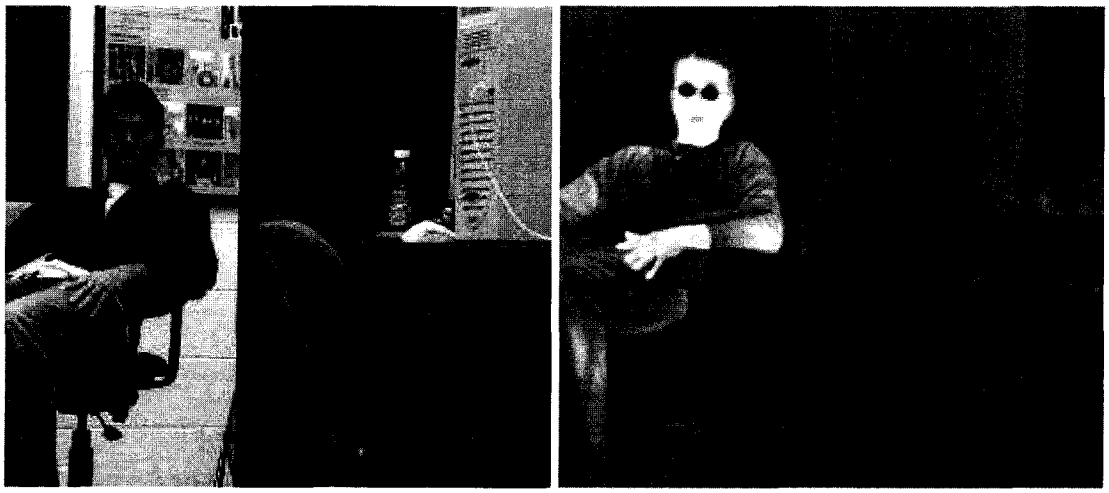


Figure 6.3. Picture on the left is captured from a visible camera and the one on the right is captured using long wave infrared camera.

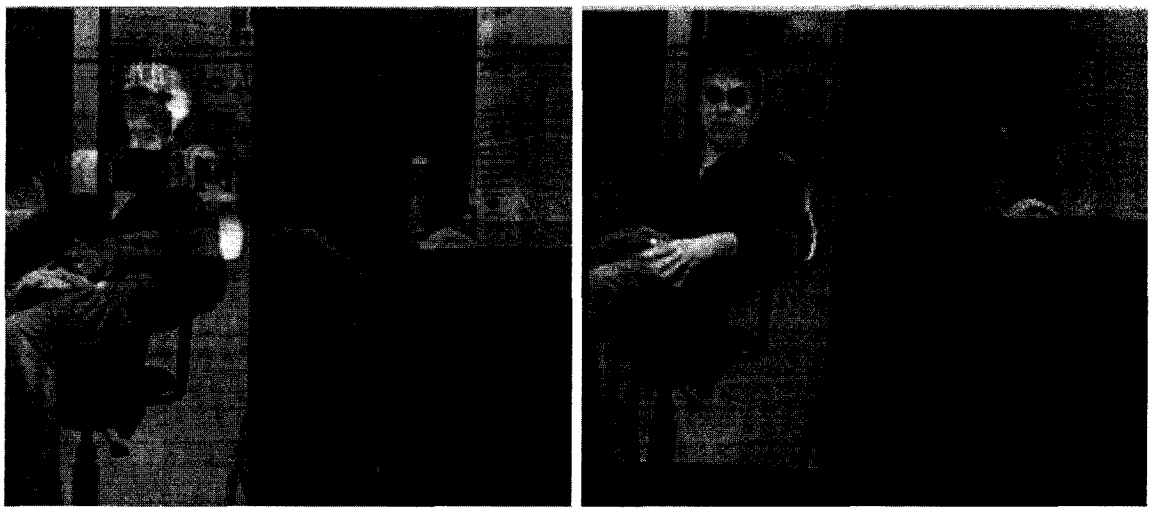


Figure 6.4. Picture on the left shows the overlap of the images from the two modalities before the registration procedure and the picture on the right after the registration.

Proper fusion of the data obtained from separate images requires accurate spatial alignment. This issue was extensively studied in the context of remote-sensing [51][57][65] and medical image registration. Due to the different physical characteristics of various sensors, the relationship between the intensities of matching pixels is often complex and unknown a-priori. Features present in one image might appear only partially in the other image or not at all. Contrast reversal may occur in some image regions while not in others; multiple intensity values in one image may map to a single intensity value in the other image and vice versa. Hence, registration of images from multiple sensors is a complex and computationally expensive process. The process of image registration may be avoided if the images obtained from the sensors are handled separately until the final decision can be made, provided there is no compromise of the recognition accuracy. Hence, in addition to the effect of fusion of multi-spectral images, the effect of image registration on recognition is also evaluated.

6.4 Image Fusion

Image fusion is the process by which two or more images obtained from multiple sensors are combined into a single image, retaining the important features from each of the original images. The easiest and the most basic image fusion technique is the average of the two images. The most popular image fusion algorithm is the one based on the discrete wavelet transform [18]. The wavelet transform of an image provides multistage pyramid decomposition for the image. This decomposition will typically have several stages. There are four frequency bands after each decomposition. These are the low-low, low-high, high-low and high-high bands. The next stage of the decomposition process operates only on the low-low part of the previous result. This produces a pyramidal hierarchy. We can think of the low-low band as the low pass filtered and sub-sampled source image. All the other bands which are called high frequency bands, contain transform coefficients that reflect the differences between neighboring pixels. The absolute values of the high frequency coefficients represent the intensity of brightness fluctuation of the image at a given scale. The larger values imply more distinct brightness changes that typically correspond to the salient features of objects.

Thus, a simple fusion rule is to select the larger absolute value of the two corresponding wavelet coefficients from each of the two source images. There are several other fusion rules [18] that can be implemented. The fusion techniques described above are implemented before any further processing is done. Hence, these are called data level or image level fusion techniques. Another type of fusion is to fuse the information that is obtained after certain processing is performed. This type of fusion is called decision level

fusion. Both data level and decision level fusion techniques are implemented and evaluated in this chapter.

6.5 Data Level Fusion

For data level fusion, we implemented a DWT based technique, with a fusion rule that selects coefficients with maximum magnitude. The obtained wavelet feature map is then used to get a fused image by using inverse wavelet transform. The block diagram shown in Figure 6.5 illustrates the fusion technique.

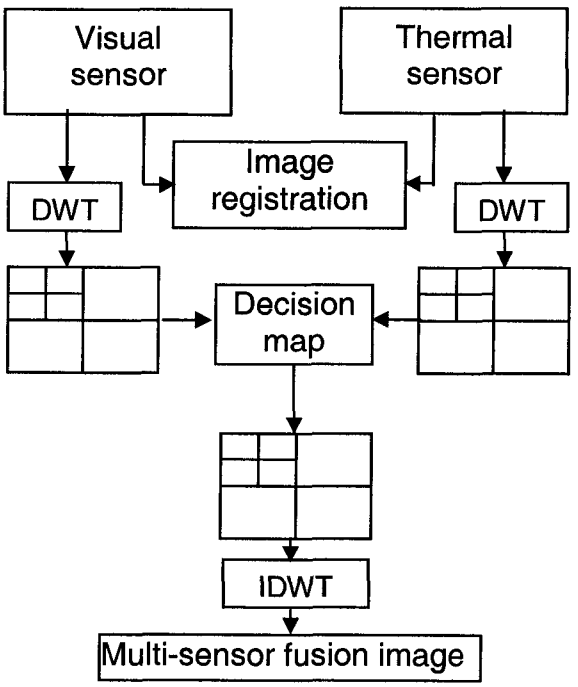


Figure 6.5. Multi- sensor image fusion using discrete wavelet transform.

Let *V(x,y)* be the visible face image and *T(x,y)* the corresponding thermal face image. Discrete wavelet transform coefficients of V(x,y) and T(x,y) are obtained as:

$$\bar{V}(u,v) = DWT\{V(x,y)\} \tag{6.1}$$

$$\bar{T}(u,v) = DWT\{T(x,y)\} \tag{6.2}$$

$\bar{V}(u,v)$ and $\bar{T}(u,v)$ are the discrete wavelet transform coefficients of the visual and thermal images respectively.

$$\bar{F}(u,v) = \max(\bar{V}(u,v),\bar{T}(u,v)) \tag{6.3}$$

$$F(x,y) = IDWT(\bar{F}(u,v)) \tag{6.4}$$

$\bar{F}(u,v)$ consists of the discrete wavelet transform coefficients selected from $\bar{V}(u,v)$ or $\bar{T}(u,v)$, based on the magnitude of the coefficients at *(u,v)*. *F(x,y)* is the fused image obtained from the inverse discrete wavelet transform of $\bar{F}(u,v)$.

6.6. Decision Level Fusion

Decision Level Fusion technique fuses the multi-sensor image after a certain amount of processing is done. Before the decision level fusion technique can be explained, it is important to take a look at the process of thermal face recognition.

6.6. 1 Proposed feature selection in thermal face recognition

The contrast of the images obtained from the thermal images is low. It should also be noted that the thermal images capture the shape of the object under consideration. Hence, a feature based technique should work better in the thermal imaging case. The phase congruency measure, as explained earlier, captures the local features based on the order I phase. It is independent of the contrast in the image. Hence, the feature maps

should provide better performance for thermal images. Since the thermal images are opaque to glass, it is very common to have partially occluded regions in a face image. Other variations in thermal images might be due to temperature inequalities of the face due to environmental conditions. Since these variations are confined to local regions, the proposed methodology of the modular regions could provide better performance in terms of recognition accuracy.

6.6.2 Decision level fusion technique

Figure 9 illustrates the decision level fusion for face recognition. Here, both visual and thermal images are subjected to phase congruency feature extraction and module creation separately, as stated in previous sections. There are a total of 896 modules that are created, i.e. 448 from the visual image and 448 from the thermal image.

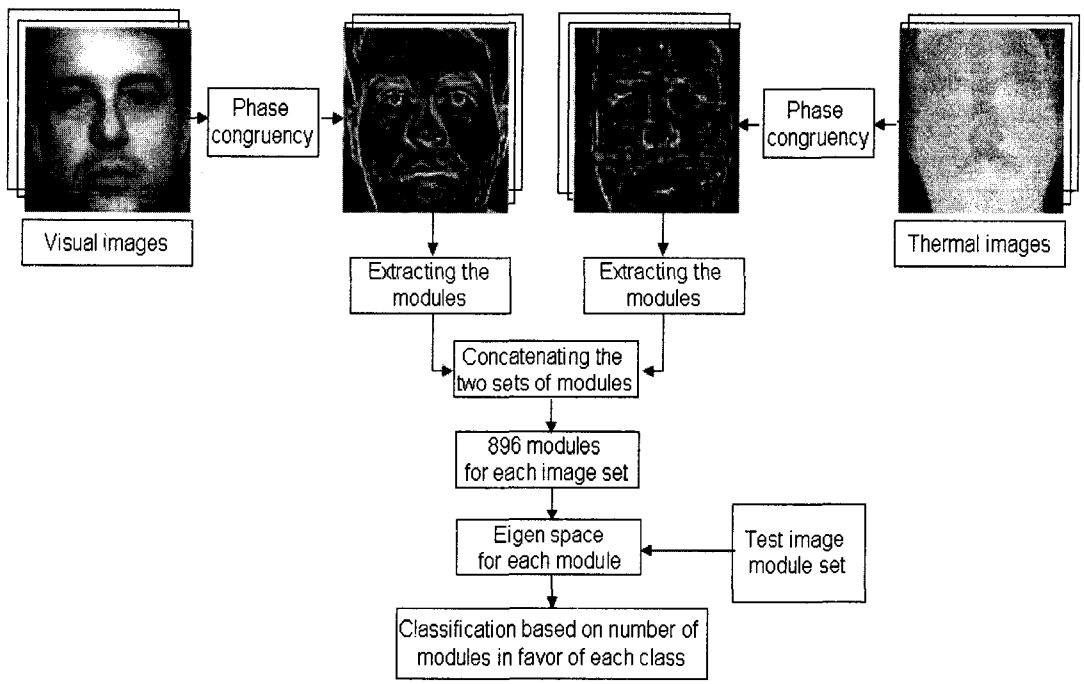


Figure 6.6. Figure illustrates the concept of decision level fusion.

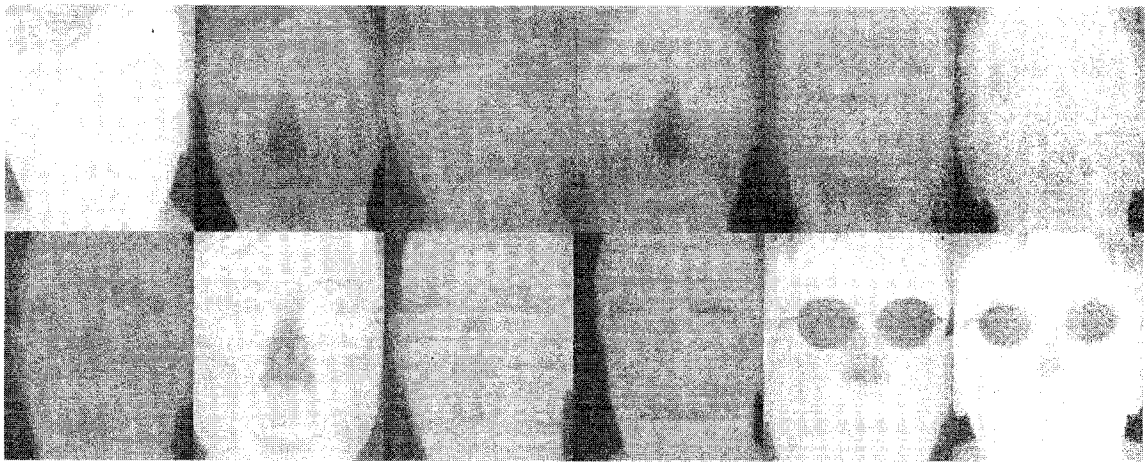
Each module is now classified individually after projecting it onto the corresponding eigen sub space. A voting procedure is now used for classification, based on the individual classification results of all the 896 modules. The overall classification is based on the number of modules obtained in favor of each class. The procedure of concatenating the modules and then classifying can be interpreted as Decision Level Fusion since the fusion is taking place only after the classification is performed on the individual modules of the respective modalities. The overall classification is based on the number of modules obtained in favor of each class.

6.7 Experimental Results

Experimentation is carried out to evaluate the performance of the proposed face recognition technique developed on fusion of multi sensor images, at data level and decision level. The equinox face database, which consists of both long wave infrared and visual spectrum face images, is used for this purpose. A subset of 34 individuals is selected for the experiments. Each individual has a total of 15 images. The thermal and visual images present in this database are pre-registered. Instead of cropping the faces manually, a face detection system developed by Viola Jones [22] is used to obtain face images to depict the real-time scenario. This face detection system is used to segment the face from the background in each visual image. The corresponding region in the thermal images is also segmented based on the coordinates of the detected faces in the visual images. Sample images from the database can be seen in Figure 6.7. Four face images of each individual are selected randomly and used for training, and the rest of the 11 images are used for testing. The experimentation process is divided into two parts. In the first

part the proposed neighborhood defined face recognition technique (NMPCA) is carried out on raw intensity images. It is observed that the recognition accuracy has improved very much on fused images when compared to either thermal or visual alone. The graphs shown in Figure 6.8 illustrate this result.

(a)



(b)

Figure 6.7. (a) Visible face images of different individuals from the equinox database (b) Corresponding thermal face (LWIR) images of those individuals.

Also, the decision level fusion process explained in the previous section, is implemented on the intensity images. It is observed that data level fusion provided better results than the decision level fusion method in this case. Table 6.3 shows the maximum accuracies obtained using the Neighborhood defined modular PCA (NMPCA) method on visual, thermal, data level fusion and decision level fusion modalities for the intensity (raw) images.

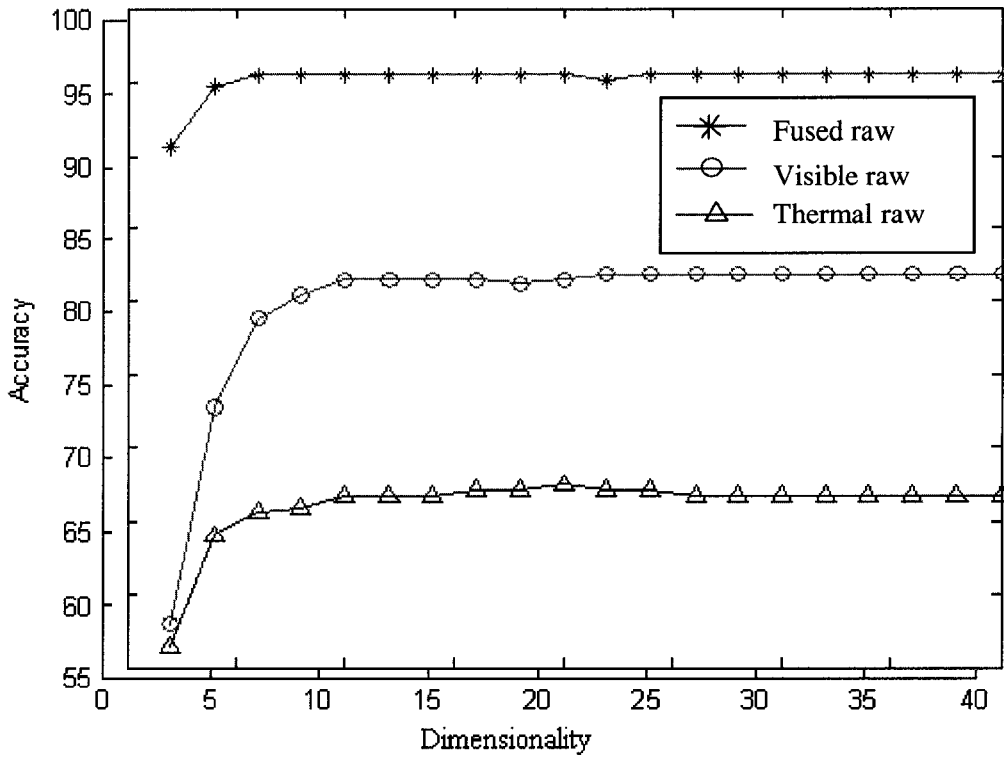
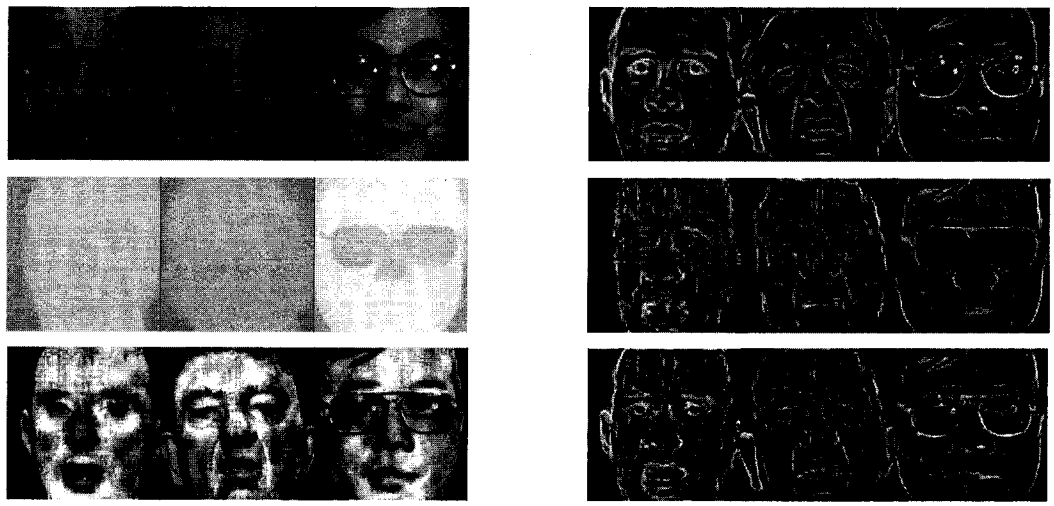


Figure 6.8. Accuracy versus dimensionality of NMPCA recognition technique for visible, thermal, fused intensity images.

Table 6.1. Accuracy for NMPCA on various sensor modalities on intensity images.

Max Accuracy using NMPCA			
Visual	Thermal	Data level fusion	Decision level fusion
82%	67.5%	95.5%	87.4%

In the second case, the recognition is performed on the phase congruency maps. The procedure is carried out on all the three modalities, that is on thermal, visual and fused (data level).The recognition accuracy has increased when applied to phase congruency images compared to intensity images. Sample visible images and the corresponding thermal and data level fused images are shown in Figure 6.9 (a). The corresponding phase congruency feature images are shown in Figure 6.9 (b).



(a) (b)

Figure 6.9. (a) Sample face images of visible, thermal and data level fused modalities of the Equinox database and (b) corresponding phase congruency features.

The graph shown in Figure 6.10 illustrates the accuracy results obtained versus the dimensionality for the visual, thermal and fused modalities. It is clear from the graph that the accuracy of the thermal modality has dramatically improved from the one that is obtained due to the intensity images shown in Figure 6.9. It is also observed that the image level fusion yielded lower accuracy than the visual modality. Recognition accuracy for the decision level fusion is calculated for the phase congruency images, and it is observed that it resulted in a higher accuracy compared to that of the data level fusion. A

comparison of the accuracies obtained for data and decision level fusion techniques for NMPPCA method is illustrated in Figure 6.11.

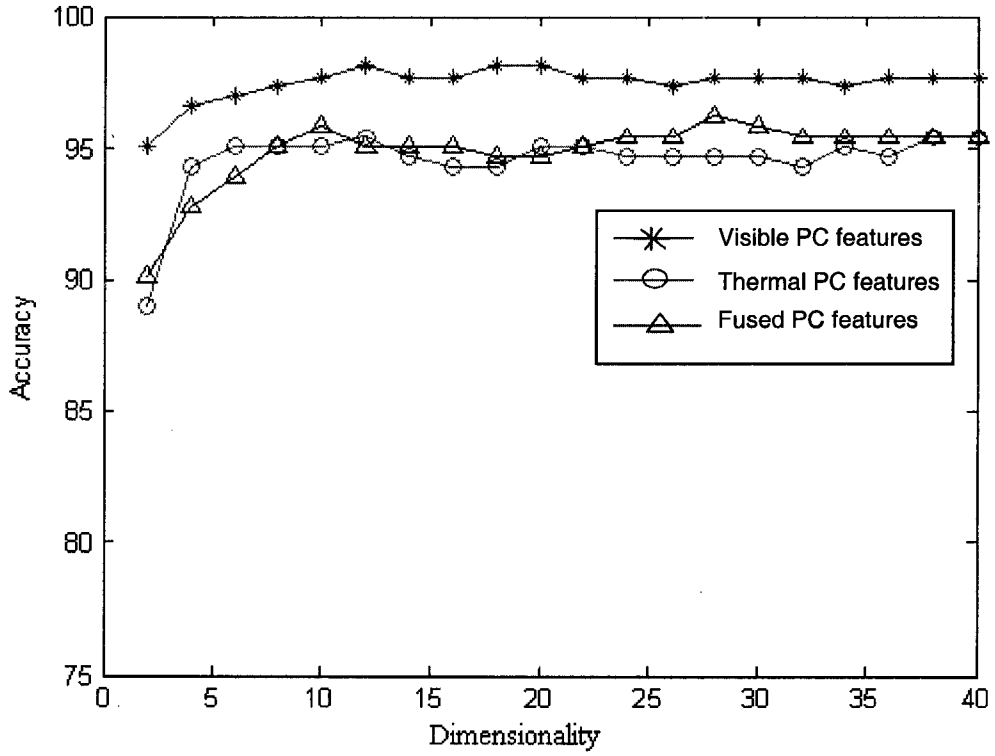


Figure 6.10. Accuracy versus dimensionality of NMPCA recognition technique for visible, thermal, fused phase congruency maps.

Maximum accuracies obtained for the visible, thermal, data level fused, decision level fused recognition procedures for the phase congruency features is tabulated in Table 6.4.

Table 6.2. Illustrates the maximum accuracy of various modalities using NMPPCA.

Max Accuracy using NMPPCA			
Visual	Thermal	Data level fusion	Decision level fusion
98.11%	95.45%	95.5%	99.24%

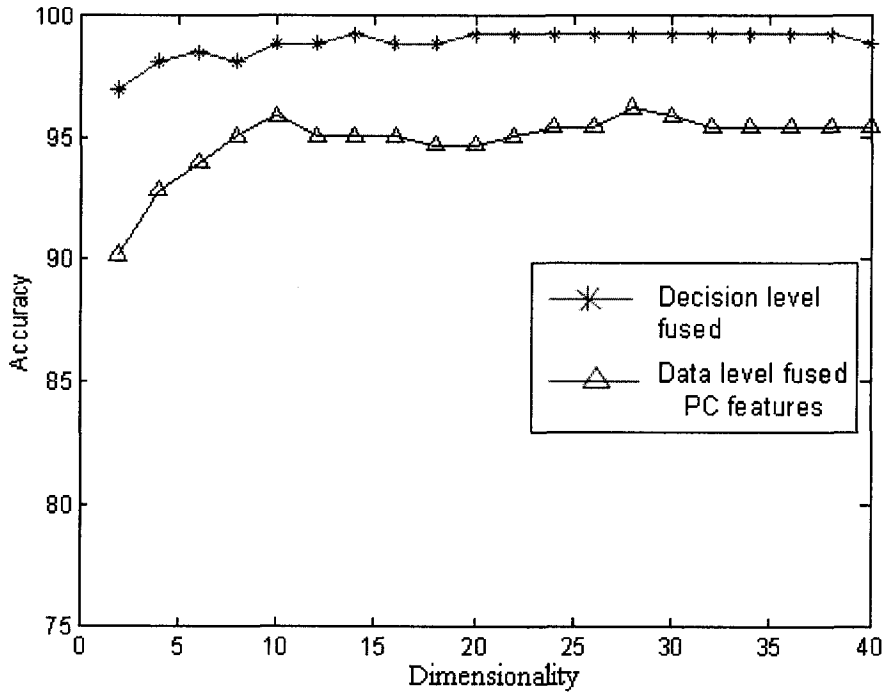


Figure 6.11. Accuracies of decision and data level fusion techniques using NMPPCA recognition technique versus dimensionality of the subspace.

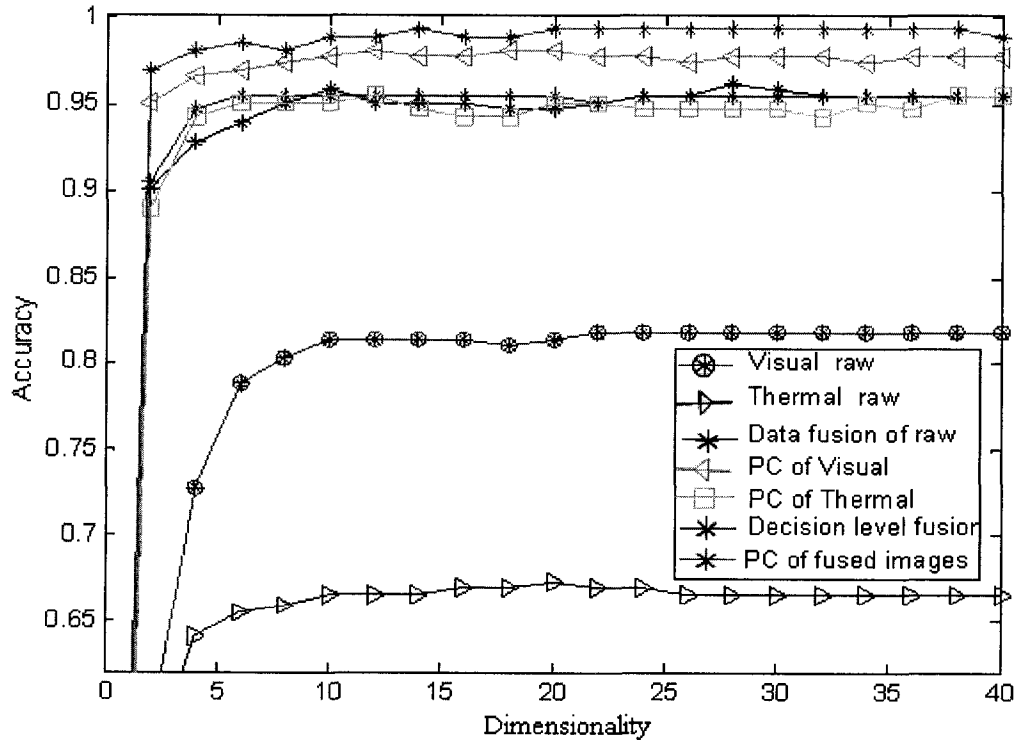


Figure 6.12. Accuracies of various methods with respect to increase in dimensionality of the subspace.

A comprehensive comparison of different modalities and recognition techniques is illustrated in Figure 6.12. It shows that the accuracy achieved by the recognition technique NMPPCA (NMPCA on phase congruency features) with decision level fusion is the highest.

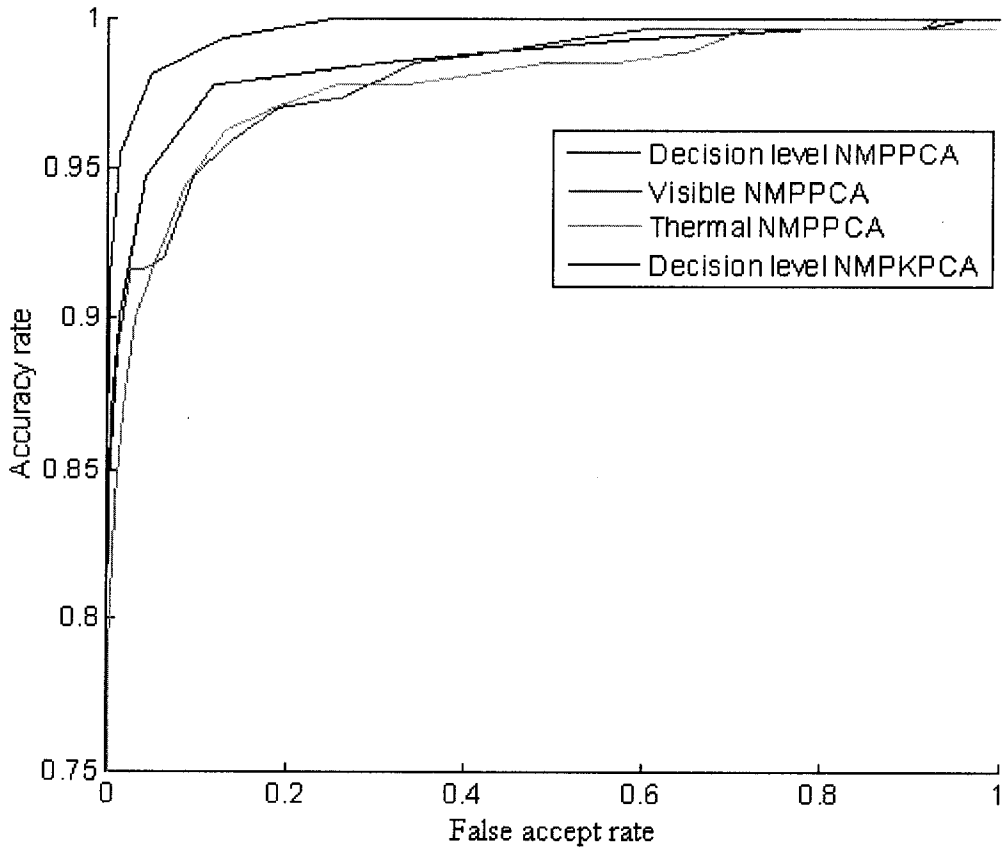


Figure 6.13. ROC's of the methods of visible, thermal and decision level fusion data modalities.

6.7.1 Modular kernel features

To study the impact of the kernel modular features, receiver operating curves are generated. Figure 6.13 shows a comparison of ROC of NMPPCA on visible, thermal and decision level fusion modalities. It also shows the ROC obtained for NMPKPCA on

decision level fusion modality. It can be observed that NMPKPCA has outperformed the rest of the methods. For a particular false positive rate of 0.2 the accuracy rate achieved is 99.5 %. This suggests that the modular kernel features further help in improving the recognition accuracy.

6.7.2 Evaluation of the effect of image registration

Image registration is essential for any application involving data level or image level fusion. This is true because the selection of features is made by comparing or combining the features from a corresponding locality of the various modalities involved in the fusion process. If the registration is not appropriate, there is a possibility that improper features may get selected from a certain region. This is particularly important for the case of face recognition involving multiple sensors. As explained earlier, the accurate registration of images captured from multiple sensors is a complex process as well as a computationally expensive one. In order to avoid or minimize the need for an accurate image registration process, it is necessary that the fusion of the modalities takes place only at a level where the point-to-point correspondence is not essential. One possible way is to combine the results of the modules of the images of the individual modalities after the decision regarding the class to which the module is associated is already taken. Hence, there is a possibility that decision level fusion could be an option to reduce the effect of point-to-point correspondence (image registration) between the modalities that are involved.

The process of complex and accurate image registration may be avoided if the images obtained from the sensors are handled separately until the final decision can be

made, provided there is no compromise on the recognition accuracy. An empirical evaluation is carried out separately on face images with and without proper registration for face recognition.

The visual and thermal face images obtained from the Equinox database are pre-registered. Hence, to simulate the effect of improper registration, a simple cropping procedure is carried out on images from one of the sensors. Here thermal images are selected. A pixel position is randomly selected and the image is cropped starting from that position and resized. The visual and thermal images do not align properly after this procedure. The random pixel point for cropping is selected between [0, 0] and [9, 9]. The procedure of image cropping, fusion and overlap is illustrated in Figure 6.14.

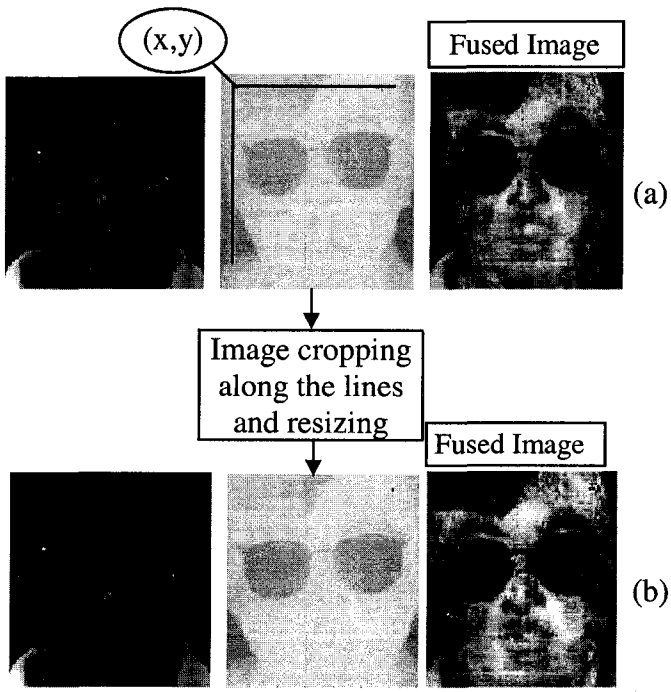


Figure 6.14. (a) Image fusion of visual and thermal images which are already registered. (b) Image fusion after the thermal image is cropped starting from point (x, y) and resized.

Face recognition accuracy for data and decision level fusion techniques is applied on properly registered images and then on images which are improperly registered (cropped and resized). Since the purpose of this experimentation is to evaluate the effect of image registration on multi-sensor images, the simpler recognition technique of NMPPCA is used. Figure 6.15 shows the face recognition accuracies obtained for the registered images, unregistered (simulated) images for both data and decision level fusion modalities. Table 6.5 gives the maximum accuracies obtained. It can be observed from the results that the difference in accuracy between registered and unregistered (simulated) images for the decision level fusion is minimal.

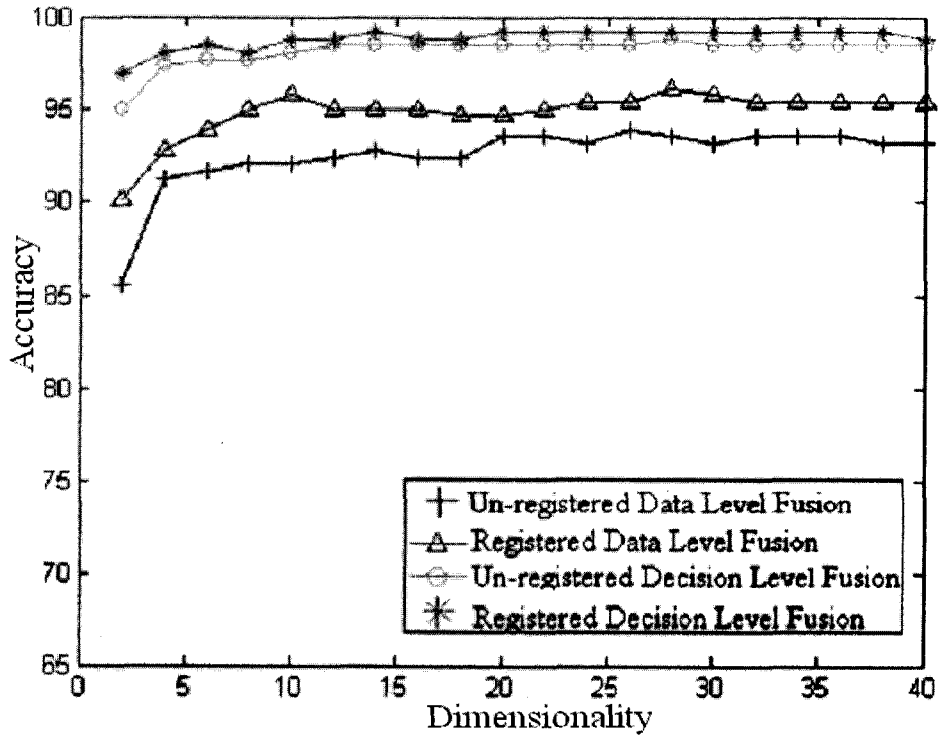


Figure 6.15. Accuracies of various methods with respect to increase in dimensionality of the subspace.

Table 6.3. Illustrates the maximum accuracy of various modalities using NMPPCA.

Max Accuracy using NMPPCA					
Data level fusion			Decision level fusion		
Unregistered	Registered	Difference	Unregistered	Registered	Difference
93.94%	95.5%	1.56%	98.86%	99.24%	0.38%

6.7.3 Analysis of results.

An analysis of the face recognition results obtained from the equinox face database with thermal and visual sensor images indicates that in the case of raw intensity images, data level fusion illustrated a better accuracy than the individual modalities. On the other hand, the result of data fusion did not yield a better result in the case of phase congruency features extracted from the fused images. It is almost similar to the accuracy obtained due to the thermal images alone, whereas the decision level fusion technique resulted in higher accuracy. The thermal facial recognition accuracy for the phase congruency features did increase by a large margin when compared to the rest of the modalities of the intensity images. The thermal and visible modalities are providing approximately the same amount of accuracy. This indicates that the information provided by each of them is equally important and discarding any of it during image level fusion would result in ignoring vital information that could help in improving the recognition accuracy. Hence, the decision level fusion provides better results as the modules of each modality are separately dealt with until the overall decision is made.

A face recognition technique based on information fusion from multiple sensors has been presented in this chapter. The methods developed in this chapter provided

evidence that multi-sensor information, when utilized properly, would help in improving facial recognition accuracies. An analysis of the face recognition results obtained from the Equinox face database of thermal and visual sensors indicated that in the case of raw intensity features, the fusion of images (data level fusion) illustrated a better accuracy than the individual modalities. It has been observed that the proposed multi sensor decision level fusion technique along with the feature selection strategy outperformed all other modalities in terms of face recognition accuracy. The empirical evaluations also indicated that the registration errors had minimal effect on the fusion technique presented in this chapter.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Feature selection has been studied and evaluated in this dissertation and has been determined to be probably the most important aspect of pattern recognition. This is in particular for face recognition, which is a complex pattern recognition problem in an operational scenario. Various aspects affect face images captured by CCD cameras. These variations have been listed and evaluated in this thesis. This thesis presented a feature selection strategy that overcomes the most common variations in the facial images as described earlier. Phase congruency maps are used for negating illumination variation effects, and a novel modular approach has been developed. Finally, the modular kernel spaces that are implemented enhance recognition accuracy. To demonstrate the effectiveness of the developed algorithms, two applications are implemented.

The multi-sensor image fusion based face recognition methodology, which has been developed in this research using the presented feature selection process, has shown extremely promising results in terms of accuracy. Experimental results show a drastic improvement in the recognition of thermal images. The decision level fusion method developed in this dissertation provided better results compared to individual as well as data level fusion modalities. It is also observed that the developed method of information fusion is robust to registration discrepancies, which is a very important factor in operational scenarios.

The modularization procedure implemented in this thesis improves the accuracy over a conventional non-overlapping modular approach, but this procedure is appropriate

only when the training and test data (face images) are of the same view. A modular manifold approach can be developed to generalize the recognition procedure over all the views and poses of the face images. It is important that a face image is good enough to provide accurate face recognition. It is possible that the regions of the face images can be assigned appropriate weightage before classification. Even though this kind of procedure is implemented in this thesis, its drawback is that it demands accurate cropping of face images (it needs proper correspondence between the face images). A learning based approach could provide a better measure of worthiness of a local region for face recognition. It is observed from the results of face recognition that kernel methods which emulate nonlinear projections into higher dimensional spaces provide better accuracy, in the case of images that are affected due to variations, than linear methods, but it is also observed that certain kernels provide better results than others. Since it is not easy to visualize the clustering of data in very high dimensional spaces, it is not easy to determine which kernel provides the desired result directly. This is still a problem to be addressed. The face recognition approach developed in this dissertation is for a frontal view system. This could be extended to include the various poses and angles.

Long wave infrared sensors are used to complement the information obtained from the visual sensors. The fusion methodologies used along with the feature selection procedures helped to provide high face recognition accuracies. Though there is a legitimate case for using LWIR sensors, because of its ability to provide images independent to illumination conditions, it has its limitations as explained in chapter 5. Research is still in progress to identify the spectrums which are most suitable (i.e invariant to various environmental changes) for the purpose of face recognition.

The classification techniques used in this research are either a minimum distance classifier or a nearest neighbor classifier. Since the focus of research is geared more towards feature selection techniques, not enough analysis was done in terms of the type of classifiers used. Although minimum distance and nearest neighbor distance measures are classical classification techniques, there are several other approaches which would work well with respect to facial recognition. Also, use of multiple classifiers combined with a boosting procedure is under investigation recently. Work is under progress in the direction using support vector machines (SVM) on the features generated. Since SVM provide better generalization, it could be expected that the recognition accuracy would improve. Research is in progress for analyzing the impact of SVM on features in terms of classification accuracy.

REFERENCES

[1] Cognitec systems < http://www.cognitec.com>

[2] Viisage, Inc <http://www.viisage.com/>

[3] A4 Vision, Inc. <http://www.a4vision.com>

[4] Gemometrix, Inc <http://www.geometrix.com/>

[5] Genex Technologies < http://www.genextech.com/>

[6] International Biometric Group http://www.biometricgroup.com/

[7] R. Chellappa, C.L Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," in *Proceedings of IEEE*, vol. 83, pp. 707-740, 1995.

[8] H. Wechsler, P.Phillips, V.Bruce, F.Soulie, and T.Huang (Eds.). "*Face Recognition : From Theory to Applications,"* Springer-Verlag, 1996.

[9] W.Zhao, R.Chellappa, and A.Rosenfeld, "Face recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, pp. 399-458, 2003.

[10] Gong S, McKenna S J, Psarrou A, "Dynamic Vision: from Images to Face Recognition," *Imperial College Press & World Scientific Publishing*, ISBN 1-86094-181-8, 364 pp. 2000.

[11] Peter N. Belhumeur, João P. Hespanhà, David J. Kriegman, "Eigenfaces vs.fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 19, pp. 711-720, 1997.

[12] M.Turk, A.Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[13] A. Pentland, B.Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 84-91, 1994.

[14] A.V. Oppenheim, and J.S. Lim, "The importance of phase in signals," in *Proceedings of IEEE*, vol. 69, pp. 529-541, 1981.

[15] Jian Huang, Pong C Yuen, Wen-Sheng Chen, J H Lai, "Kernel subspace LDA with optimized Kernel parameters on face recognition," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 327-332, 2004.

[16] R. Gottumukkal, K V Asari, "An Improved face recognition technique based on modular PCA approach," *Pattern Recoginition Letters*, vol.25, pp.429-436, 2004.

[17] P.Kovesi, "Edges are not just steps," in *Proceedings of Asian Conference on Computer Vision (ACCV)*, pp. 23-25, 2002.

[18] H. Li, B. S. Manjunath, and S. K. Mitra, "Multi Sensor image fusion using wavelet transform," in *Proceedings of IEEE International Conference on Image Processing (ICIP)*, vol. 11, pp. 51-55, 1994.

[19] Mian Zhou and Hong Wei, "Face Verification Using Gabor Wavelets and Adaboost," *18th International Conference on Pattern Recognition,* vol. 1, pp. 404-407, 2006.

[20] LinLin Shen and Li Bai, "Gabor Feature Based Face Recognition Using Kernal Methods," in *Proc. of the IEEE 6th International Conference on Automatic Face and Gesture Recognition*, 2004.

[21] http://www.equinoxsensors.com/products/HID.html

[22] Paul Viola, Michaeal J.Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.

[23] W.Zhao, A.Krishnaswamy, R.Chellappa, D.Swets, and L.Weng, "Discriminant analysis of principal components for face recognition," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.

[24] A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision an Pattern Recognition*, pp.84-91, 1994.

[25] J. Huang, P. Yuen, C. Chen, W. Sheng, J.H. Lai, "Kernel subspace LDA with optimized kernel parameters on face recognition," *IEEE International conference on Automatic Face and Gesture Recognition*, pp. 327-332, 2004.

[26] A. Martinez, "Recognition of partially occluded and/or imprecisely localized faces using a probabilistic approach," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.712–717, January 2000.

[27] Y. Zhang and A.M. Martinez, "Recognition of Expression Variant Faces Using Weighted Subspaces," *Proceedings of International Conference on Pattern Recognition*, 2004.

[28] K. Tan, S. Chen "Adaptively weighted sub-pattern PCA for face recognition," *Neurocomputing*, vol 64, pp.505-511, March 2005.

[29] A.U. Batur, "Segmented linear subspaces for illumination-robust face recognition," *International Journal of Computer Vision*, vol. 57, no. 1, pp. 49-66, April 2004.

[30] H. Wang, Z. Stan Li, Y Wang, "Face recognition under varying lighting conditions using self quotient image," *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 819-824, 2004.

[31] K.C Lee, J. Ho, D.J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684-698, May 2005.

[32] Bernd Heisele, Purdy Ho, Jane Wu and Tomaso Poggio, "Face recognition: component-based versus global approaches," *Computer Vision and Image Understanding*, vol. 91, pp. 6–21, 2003.

[33] Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski, "Face Recognition by Independent Component Analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, November 2002.

[34] W. Zhao, R. Chellappa, and A. Rosenfeld. "Face recognition: a literature survey", ACM Computing Surveys, 35:399–458, December 2003.

[35] M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp.103-108, January 1990.

[36] M.H. Yang, N. Ahuja, and D. Kriegman, "Face recognition using Kernel Eigenfaces," *Advances in NIPS*, Vol. 14, pp.215-220, 2002.

[37] M.H Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *IEEE Conference on Automatic Face and Gesture Recognition*, 2002.

[38] J. Yang, Z. Jin, J. Yang, D. Zhang, "The essence of. kernel Fisher discriminant: KPCA plus LDA," *Pattern. Recognition*, vol. 37, no.10, pp. 2097-2100, October 2004.

[39] Lai, Jian Huang Yuen, Pong C Feng, Guo Can, "Face recognition using holistic Fourier invariant features," *Pattern Recognition*, vol. 34, no. 1, pp. 95-109, Jan 2001.

[40] Z.H. Pan, G. Healey, M. Prasad, and B. Tromberg, "Illumination-invariant face recognition in hyperspectral images," *SPIE - The International Society for Optical Engineering*, vol. 5093, pp. 275-282, 2003.

[41] M. Savvides, B.V.K Vijayakumar, P. K. Khosla, "Robust, shift-invariant biometric identification from partial face images," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 5404, pp. 124-135, 2004.

[42] Won-Sook Lee, Kyung-Ah Sohn, "Database construction and recognition for multi-view face", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 17-19, May 2004.

[43] Shaohua Kevin Zhou, Rama Chellappa, "Illuminating Light Field: Image-based Face Recognition Across Illuminations and Poses," *International Conference on Automatic Face and Gesture Recognition*, pp. 229- 234, May 2004.

[44] S. Chen, B. C. Lovell, "Illumination and expression invariant face recognition with one sample image," *International Conference on Pattern Recognition*, pp. 300-303, 2004.

[45] Volker Blanz, Thomas Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063-1074, September 2003.

[46] J. Cohn, A. Zlochower, J.J Lien, and T Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 396 – 401, April 1998.

[47] P. K. Manglik, U. Misra, M. Hima Bindu, "Facial expression recognition," *IEEE International Conference on Systems, Man and Cybernetics*, pp. 2220-2224, 2004.

[48] Yu Zhangt, Edmond C. Prakashtt and Eric Sung, "Synthesis of Facial Expressions Using A 3D Anatomical Model," *Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision*, pp. 704-709, 2002.

[49] Heeger, "Optical flow using spatiotemporal filters," *International Journal of Computer Vision*, vol.1, pp. 279-302, 1988.

[50] Horn and Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, pp.185-204, 1981.

[51] Lucas and Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of DARPA Image Understanding Workshop*, pp. 121-130, 1981.

[52] A. M. Martinez and R. Benavente, "The AR Face Database", *CVC Technical Report* #24, June 1998.

[53] D. A. Socolinsky and A. Selinger, "Thermal Face Recognition in an Operational Scenario," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* Washington, D.C., June 2004.

[54] S.G. Kong, Jingu Heo, B. R. Abidi, J. Paik and M. A. Abidi, "Recent advances in visual and infrared face recognition - A review", *Computer Vision and Image Understanding*, vol. 97, pp. 103-135, 2005.

[55] Xin Chen, Patrick J.Flynn, Kevin W. Bowyer,"Ir and visible light face recognition," *Computer Vision and Image Understanding*, vol. 99, pp. 332-358, 2005.

[56] S. Singh, A. Gyaourova, G. Bebis, I. Pavlidis, "Infrared and visible image fusion for face recognition," *Proceedings of SPIE - The International Society for Optical Engineering, Biometric Technology for Human Identification*, vol. 5404, pp. 585-596, 2004.

[57] Y. Keller, A. Averbuch, "Implicit similarity: A new approach to multi-sensor image registration," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 543-548, 2003.

[58] T. F. Cootes , C. J. Taylor , D. H. Cooper , J. Graham, "Active shape models—their training and application," *Computer Vision and Image Understanding*, vol.61 no.1, pp.38-59, January 1995.

[59] P. W. Hallinan, "Recognizing Human Eyes," in *SPIE Proceedings*, Vol. 1570: *Geometric Methods in Computer Vision*, pp. 214-226, 1991.

[60] P.Penev and J.Atick. "Local feature analysis, A general statistical theory for object representation," *Network: computation in Neural Systems*, vol. 7, pp. 477-500, 1996.

[61] A. V. Nefian and M. H. Hayes, "Hidden Markov models for face recognition," IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp. 2721-2724, May 1998.

[62] F. Cardinaux, C. Sanderson, S. Marcel, "Comparison of MLP and GMM classifiers for face verification on XM2VTS," *Proceedings of the Fourth International Conference on Audio- and Video Based Biometric Person Authentication, Guildford,* pp. 911—920, 2003.

[63] M.C..Morrone, R.A.Owens, "Feature detection from local energy," *Pattern Recognition Letters*, vol. 6, pp. 303-313, 1987.

[64] F.Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," *Proceedings of 2nd IEEE workshop on Applications of Computer Vision*, 1994.

[65] B. Youcef, T. Nasreddine, K. Kidiyo, R. Joseph, "An automatic image registration for applications in remote sensing," IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 9, pp. 2127-2137, September 2005.