Graduate Theses, Dissertations, and Problem Reports

2008

# Mitigating the effect of covariates in face recognition

Richa Singh
*West Virginia University*

Follow this and additional works at: https://researchrepository.wvu.edu/etd

## Recommended Citation

# Mitigating the Effect of Covariates in Face Recognition

By

## Richa Singh

Dissertation submitted to the
College of Engineering and Mineral and Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in
Computer Science

Approved by:

Afzel Noore, PhD, Chair
Larry A. Hornak, PhD
Arun A. Ross, PhD
George E. Trapp, PhD
Keith B. Morris, PhD

Morgantown, West Virginia
2008

Keywords: Face Recognition, Covariates, Facial Disguise, Aging, Machine Learning,
Support Vector Machine

# Abstract

Mitigating the Effect of Covariates in Face Recognition

By

Richa Singh

Current face recognition systems capture faces of cooperative individuals in controlled environment as part of the face recognition process. It is therefore possible to control lighting, pose, background, and quality of images. However, in a real world application, we have to deal with both ideal and imperfect data. Performance of current face recognition systems is affected for such non-ideal and challenging cases. This research focuses on designing algorithms to mitigate the effect of covariates in face recognition.

To address the challenge of facial aging, an age transformation algorithm is proposed that registers two face images and minimizes the aging variations. Unlike the conventional method, the gallery face image is transformed with respect to the probe face image and facial features are extracted from the registered gallery and probe face images. The variations due to disguises cause change in visual perception, alter actual data, make pertinent facial information disappear, mask features to varying degrees, or introduce extraneous artifacts in the face image. To recognize face images with variations due to age progression and disguises, a granular face verification approach is designed which uses dynamic feed-forward neural architecture to extract 2D log polar Gabor phase features at different granularity levels. The granular levels provide non-disjoint spatial information which is combined using the proposed likelihood ratio based Support Vector Machine match score fusion algorithm. The face verification algorithm is validated using five face databases including the Notre Dame face database, FG-Net face database and three disguise face databases.

The information in visible spectrum images is compromised due to improper illumination whereas infrared images provide invariance to illumination and expression. A multispectral face image fusion algorithm is proposed to address the variations in illumination. The Support Vector Machine based image fusion algorithm learns the properties of the multispectral face images at different resolution and granularity levels to determine optimal information and combines them to generate a fused image. Experiments on the Equinox and Notre Dame multispectral face databases show that the proposed algorithm outperforms existing algorithms. We next propose a face mosaicing algorithm to address the challenge due to pose variations. The mosaicing algorithm generates a composite face image during enrollment using the evidence provided by frontal and semiprofile face

images of an individual. Face mosaicing obviates the need to store multiple face templates representing multiple poses of a users face image. Experiments conducted on three different databases indicate that face mosaicing offers significant benefits by accounting for the pose variations that are commonly observed in face images.

Finally, the concept of online learning is introduced to address the problem of classifier re-training and update. A learning scheme for Support Vector Machine is designed to train the classifier in online mode. This enables the classifier to update the decision hyperplane in order to account for the newly enrolled subjects. On a heterogeneous near infrared face database, the case study using Principal Component Analysis and C2 feature algorithms shows that the proposed online classifier significantly improves the verification performance both in terms of accuracy and computational time.

*Dedicated To My Family.*

# Acknowledgments

This is a great opportunity to express my thanks and respect to all those who have contributed in this journey.

First and foremost, I would like to thank my advisor, Dr. Afzel Noore, for providing a motivation, enthusiasm and critical atmosphere during many discussions we had. It was a great pleasure for me to conduct this research under his supervision. His enthusiasm and integral view on research has made a deep impression on me. Dr. Noore's instinct for "what constitutes good research" has helped me to shape my approach towards the research. I am really glad that I have learnt from his knowledge and creative experience.

I would like to thank Dr. Arun Ross for the collaborations, insightful discussions and always giving a different perspective to the problem. I would also like to thank my committee members Dr. Keith Morris, Dr. George Trapp and Dr. Larry Hornak for all the help and support I have received from them. I am greatly indebted to Dr. S.K. Singh for his collaboration and providing useful guidance. The financial support by United States National Institute of Justice is gratefully acknowledged.

I would like to acknowledge Dr. R. Chellappa, Dr. B. Abidi, Ralph Gross, CVRL University of Notre Dame, NIST, Equinox Corporation, IEEE OTCBVS WS Series Bench, Center for Biometrics and Security Research, and AuthenMetric Co. Ltd for granting us access to the face databases used in this research.

I would like to take this opportunity to thank my friends for their support. I am pleased to thank Karen, Maggie, Laura and Lucy for their administrative support.

I owe my loving thanks to my parents and sisters for so many years of patience, understanding, and unconditional love and support. Finally, I wish to thank my husband Mayank who has shared with me all the long hours of work, failures, and success. Thank you for being an inspiration to me and for always being there.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| DWT | Discrete Wavelet Transform |
| EBGM | Elastic Bunch Graph Matching |
| EER | Equal Error Rate |
| FAR | False Accept Rate |
| FFT | Fast Fourier Transform |
| FLDA | Fast Linear Discriminant Analysis |
| FRR | False Reject Rate |
| FRGC | Face Recognition Grand Challenge |
| FRVT | Face Recognition Vendor Test |
| GAR | Genuine Accept Rate |
| GSVM | Granular Support Vector Machine |
| GVF | Gradient Vector Flow |
| HTER | Half Total Error Rate |
| ICA | Independent Component Analysis |
| IPCA | Incremental Principal Component Analysis |
| LBP | Local Binary Pattern |
| LDA | Linear Discriminant Analysis |
| LFA | Local Feature Analysis |
| LWIR | Long Wave Infrared |
| MWIR | Medium Wave Infrared |
| NIR | Near Infrared |
| PCA | Principal Component Analysis |
| RBF | Radial Basis Function |
| ROC | Receiver Operating Characteristics |
| SVM | Support Vector Machine |
| SWIR | Short Wave Infrared |
| 2DLPG-NN | 2D Log Polar Gabor Neural Network |
| $2\nu$-SVM | Dual $\nu$-Support Vector Machine |
| $2\nu$-GSVM | Dual $\nu$-Granular Support Vector Machine |
| $2\nu$-GSSVM | Dual $\nu$-Granular Soft Support Vector Machine |
| $2\nu$-OGSSVM | Dual $\nu$-Online Granular Soft Support Vector Machine |

# Chapter 1

# Introduction

Humans effortlessly process information obtained from multiple sensory inputs and have the ability to recognize individuals even with limited correlation information, redundant information, or when certain features appear partially hidden, camouflaged or disguised. To recognize an individual, the visual cortex exploits spatial correlations by processing overlapping information extracted at global and local levels and effectively combines them to make a decision. The information is gathered using a set of inherent spatial filters that accurately detects any change in orientation, color, spatial frequency, texture, motion, and other pertinent features. For several years, many researchers have been motivated in developing algorithms to emulate the near perfect face recognition capability of human mind. However, human face is not a rigid object and can have different variations due to inter-personal or intra-personal transformations. Inter-personal variations can be attributed to changes in race or genetics, while intra-personal variations can be attributed to changes in expression, aging, hair, cosmetics, and facial accessories.

Automatic face recognition is a long standing problem in computer vision that requires the ability to identify an individual despite several variations in the appearance of face. As shown in Figure 1.1, current face recognition systems capture faces of cooperative individuals in a controlled environment as part of the face recognition process. It is therefore possible to control the lighting, pose, background, and quality of images. The results of the recent face recognition test reports, Face Recognition Grand Challenge 2004 [74] and Face Recognition Vendor Test 2006 [75], show that under normal changes

Figure 1.1: Face images of a cooperative user under controlled environment.

in a constrained environment, the performance of existing face recognition systems is greatly enhanced. However, in a real world application, we have to deal with both ideal and imperfect data. Figure 1.2 represents the biometric space in terms of the several challenges of face recognition. As shown in Figure 1.3, many applications require face images to be captured outdoors where the lighting conditions are unpredictable, the subjects may not be cooperative, the poses may vary, or the angles and distances from the camera may not be normal. Performance of current face recognition systems significantly deteriorates for such imperfect and challenging cases.

In the face recognition community, several covariates for face recognition have been identified such as pose, expression, illumination and aging. We expand the problem to include additional covariates of face recognition:

- **Illumination:** Images with proper illumination captured under controlled environment are ideal for face recognition. However, face images with illumination variations (Figure 1.4a) reduce the performance of recognition algorithms because illumination variations may alter the appearance and the features may be hidden.

- **Image Quality:** Quality of a face image depends on various features such as motion blur, sensor noise, environmental noise, image resolution, and gray scale/color depth. Any degradation in the quality of face images, as shown in Figure 1.4b, can lead to reduced recognition performance.

- **Expression:** Variations in expression can cause deformation in deformation in

2
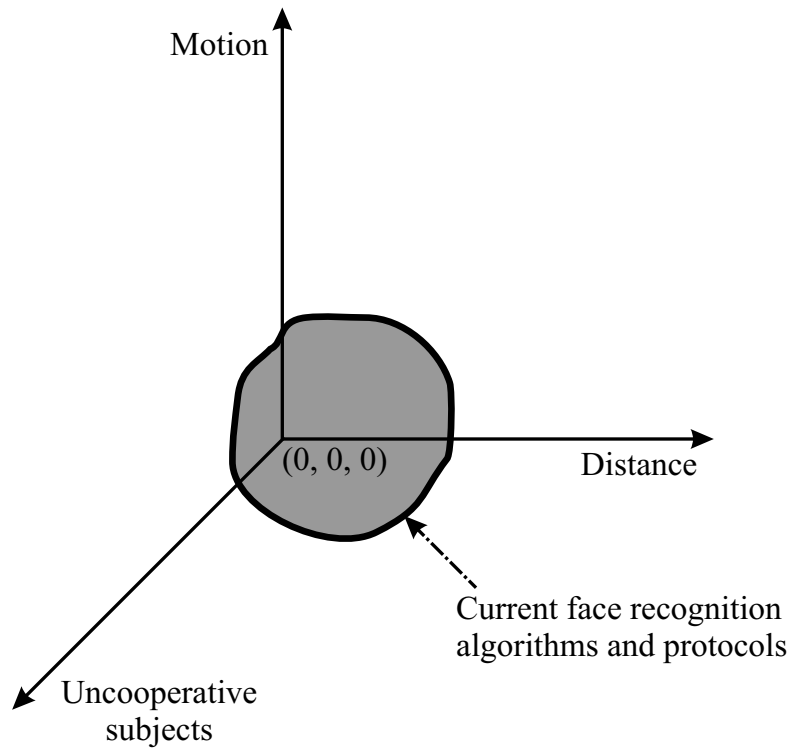
Figure 1.2: Current face recognition algorithms focus on regions near (0,0,0) but the data in real world applications stretches in the complete space.



Figure 1.3: Face images of an individual captured under varying conditions.

local facial structure and also change the facial appearance and local geometry of the face. Figure 1.4c shows example of variations in facial features which can reduce the recognition accuracy.

- **Pose:** Frontal face images contain ample information to be used for face recognition. However, in a profile or semi-profile face image, as shown in Figure 1.4d, some features are not visible and matching a frontal face image with a profile face image may produce incorrect results.

- **Aging:** Temporal variations in human face is a regular process and with age progression, these variations may cause major structural changes. As shown in Figure 1.4e, aging variations among the three face images of an individual are difficult to handle by an automated face recognition system.

- **Disguise:** The most challenging among all the covariates is the variations due to disguise in which an individual can use makeup accessories to alter the facial features and impersonate another person or hide one's identity. As shown in Figure 1.4f, simple accessories such as beard and mustache can change the appearance of an individual.

Apart from these covariates, another challenge for face recognition algorithms is the computational time required for training and updating parameters as new data is added to the database. Most state-of-the-art biometric algorithms use training for parameter estimation and learning the decision boundary. Currently, biometric systems are trained offline with available training data and domain specific knowledge. However, large scale biometric applications such as US-VISIT continuously enroll new individuals. Due to the high computational complexity required for re-training, it is not feasible to regularly update the system knowledge and decision boundary. The delay in re-training biometric systems affects the recognition performance because without re-training, the disparate characteristics of additional biometric data may lead to reduced recognition performance.

Other than these covariates, there are some challenges that are pertinent to law enforcement applications such as limited training dataset, matching a scanned face image

Figure 1.4: Covariates in face recognition: (a) illumination, (b) image quality, (c) expression, (d) pose, (e) aging, and (f) disguise.

with digitally captured face images, matching sketch images generated by forensic artists with the digital face images. Some of these challenges have a linear relation to the above mentioned covariates, e.g. matching a scanned face image with digital face images. However, challenges such as matching sketches with a database of digital images require specially designed algorithms.

## 1.1 Related Work

Over the last 40 years, researchers have proposed different algorithms to address the issues of illumination, image quality, pose and expression variations [58], [75], [120]. Several algorithms have been proposed such as Principal Component Analysis [113], Linear Discriminant Analysis [11], Local Feature Analysis [71], Elastic Bunch Graph Matching [122], wavelet features [62], local binary patterns [1] and many others [129]. Apart from these feature extraction and matching algorithms, preprocessing algorithms have also been designed such as illumination rectification algorithms and image quality enhancement for illumination correction and quality enhancement [40], [49], [88], [89] transformation algorithms for geometric deformation correction [100], active appearance [48] and morphable models [12] to handle pose variations. While illumination, image quality, pose, and expression have been well studied and documented in the literature, the challenge of face aging and disguise remains. Although few researchers have proposed age simulation algorithms [15], [39], [53], [81], these approaches are not sufficient to handle large temporal variations. Similarly, research in face recognition with disguise variations is in very early stages. Mainly, principal component analysis based algorithms have been proposed which do not provide good performance for even minor variations in disguise.

As mentioned before, existing algorithms generally require training and regular knowledge update. Traditionally, the face recognition algorithms are re-trained with the old and new database which is a computationally expensive exercise. In biometrics, very limited research has been undertaken to address this issue [79]. These algorithms address the challenge to some extent; but more research is required to circumvent the challenges

of computationally efficient knowledge update and re-training as the size of database increases.

## 1.2   Research Objectives

This research focuses on designing algorithms to mitigate the effect of covariates and improve the performance of face recognition. Specifically, the research objectives are:

- **Transformation and recognition algorithm for aging and disguise:** Design a transformation algorithm to register the gallery face image with respect to the probe face image to account for variations due to structural changes, genetics and environmental factors. Further, develop a face recognition algorithm that incorporates results from human cognition to efficiently recognize aged and disguised faces.

- **Multispectral image fusion algorithm for illumination variation:** Design an image fusion algorithm to fuse invariant information from multispectral face images in order to address the illumination and expression variations in the gallery and probe face images.

- **Face mosaicing scheme for pose invariant face recognition:** Design a face mosaicing scheme to address the variations due to differences in pose. Since the amount of information present in mosaiced and non-mosaiced face images vary significantly, develop a local feature extraction based face recognition algorithm to match non-mosaiced and mosaiced face images.

- **Computationally efficient online learning for classifier re-training and update:** Design an online classification algorithm to efficiently classify the users as genuine or impostor. The classification algorithm should be able to perform faster training and also update the decision hyperplane to account for variations

due to new enrollments in the database. Online training of the classifier should facilitate updated learning in real-time by reducing the computational cost and making the biometric system scalable.

## 1.3   Research Contributions

This section delineates an overview of the algorithms designed to accomplish the above mentioned research objectives.

**Transformation and Recognition Algorithm for Aging and Disguise**

An age transformation algorithm is developed to handle the challenge of facial aging in face recognition. The proposed algorithm registers the gallery and probe face images in polar coordinate domain and minimizes the variations in facial features caused due to aging. Further, a novel face recognition algorithm is designed that extracts non-disjoint facial features at different levels of granularity. These features are extracted using neural network architecture based 2D log polar Gabor transform. Finally, likelihood ratio based support vector machine match score fusion is developed to combine the granular information.

**Multispectral Face Image Fusion Algorithm for Illumination Variation**

Two multispectral image fusion algorithms are developed to address illumination and expression variations. The first algorithm is a simple approach in which Discrete Wavelet Transform is used for image fusion. The second algorithm uses a learning technique for image fusion. Specifically, the learning based image fusion algorithm utilizes $2\nu$-granular support vector machine to learn the properties of the multispectral face images at different resolution. It further determines optimal information in both the images and combines them to generate a fused image.

**Face Mosaicing Scheme for Pose Invariant Face Recognition**

To address the variations due to pose, a face mosaicing scheme is designed that generates a composite face image during enrollment based on the evidence provided by frontal and semi-profile face images of an individual. In the proposed scheme, the side profile images are aligned with the frontal image using a hierarchical registration algorithm that exploits neighborhood properties to determine the transformation relating the two images. Multiresolution splining is then used to blend the side profiles with the frontal image thereby generating a composite face image of the user. A texture-based face recognition technique that is a slightly modified version of the $C2$ algorithm proposed by Serre *et al.* [87] is used to compare a probe face image with the gallery face mosaic.

## Computationally Efficient Online Learning for Classifier Re-training and Update

To address the computational aspect and regular update, a new biometric classifier update algorithm is developed that incrementally re-trains the classifier using online learning and progressively establishes a decision hyperplane for improved classification. The proposed algorithm incorporates soft labels in the formulation of $2\nu$-online granular soft support vector machine to re-train the classifier using only the new data.

# Chapter 2

# Face Recognition with Aging and Disguise Variations

Recognizing individuals with altered appearances is a major challenge of face recognition. Current face recognition algorithms are not capable of detecting appearances altered due to aging or disguise; even if the images are captured in a controlled environment. The ability to recognize individuals who deliberately alter their appearance is very important in security applications such as identifying impostors and terrorists, and missing person database. The challenges due to aging and disguises cause change in data structures, alter actual data, make pertinent data disappear, mask features to varying degrees, or introduce extraneous artifacts in the face image. Many forensic and law enforcement applications have to deal with challenges due to aging and disguises. However, the performance of face recognition systems for such cases are not very encouraging. In this chapter, we specifically undertake the challenge of face verification due to age progression and variations in disguise.

Human face undergoes significant changes as a person grows older. The facial features vary for every person and are affected by several factors such as exposure to sunlight, inherent genetics, and nutrition. The performance of face recognition systems cannot contend with the dynamics of temporal metamorphosis over a period of time. Law enforcement agencies regularly require matching a probe image with individuals in the missing person database. In such applications, there may be significant differences

between facial features of probe and gallery images due to age variation. For example, if the age of an individual in the probe image is 15 years and the age of the same individual in the gallery image is five years, existing face recognition algorithms are ineffective and may not yield the desired results. One approach to handle this challenge is to regularly update the database with recent images or templates. However, this method is not feasible for applications such as border control, homeland security, and missing person verification.

Researchers have proposed several age simulation and modeling techniques to address the challenge of recognizing faces with aging. These techniques model the facial growth over a period of time to minimize the difference between probe and gallery images. Burt and Perrett [15] proposed an age simulation algorithm using shape and texture to create composite face images for different age groups. They further analyzed and measured the facial cues affected by age variations. Tiddeman *et al.* [111] proposed wavelet transform based age simulation to prototype the composite face images. Lanitis *et al.* [53]-[55] proposed statistical models for face simulation. They used training images to learn the relationship between coded face representation and actual age of subjects. This relationship is then used to estimate the age of an individual and to reconstruct the face at any age. Gandhi [39] proposed Support Vector Regression to predict the age of frontal faces. The aging function is used with the image based surface detail transfer method to simulate face image at any younger or older age. Wang *et al.* [118] obtained the texture and shape information of a face image using PCA and used this information in reconstructing the shape and texture at any particular age. Recently, Ramanathan and Chellappa [81], [82] proposed a craniofacial growth model that characterizes the shape variations in human faces during formative years. They further developed a Bayesian age difference classifier to estimate the age difference between two images and verify the identity. Park *et al.* [70] designed an aging simulation technique that learns the aging patterns of the shape and the texture based on PCA coefficients. A 3D morphable model is adapted using the 2D face image database which compensates for the aging variations between the gallery and probe faces.

The inter-personal and intra-personal characteristics can be modeled to alter the appearance of an individual to impersonate another person or hide one's identity. The ability to recognize individuals who deliberately alter their appearance is very important in security applications for identifying terrorists, controlling access to sensitive areas, or monitoring border crossing. The recognition of faces with disguise is a major challenge and has only been recently addressed by few researchers [2], [69], [80], [91]. Alexander and Smith [2] used PCA based algorithm with Mahalanobis angle as the distance metric. The results show an accuracy of 45.8% on the AR database [66]. Silva *et al.* [91] proposed using Eigen-eyes to handle several challenges of face recognition including disguise. Using the Yale database [124], the algorithm was able to achieve an accuracy of around 87.5%. The advantage of the algorithm is that alterations in facial features excluding the eye region do not affect the accuracy. Ramanathan *et al.* [80] studied the facial similarity for several variations including disguise by forming two eigenspaces from two halves of the face, one using the left half and the other using the right half. From the test image, optimally illuminated half face is chosen and is projected into the eigenspace. This algorithm has been tested on the AR face database [66] and the National Geographic database [80] which consists of variations in smile, glasses, and illumination. An accuracy of around 39% for the best two matches is reported on the AR database. Pamudurthy *et al.* [69] proposed a face recognition algorithm that uses dynamic features obtained from skin correlation and the features are matched using nearest neighbor classifier. On a database of 10 individuals, authors reported that this approach gives accurate results. The limitation of these algorithms is that the performance degrades when important regions such as the eye and the mouth are covered. Moreover, the AR and Yale databases do not contain many images with disguise and therefore are not ideal for validating algorithms under comprehensive disguise scenarios.

Generally face recognition algorithms either use facial information in a holistic way or extract features and process them in parts. On the other hand, cognitive neuroscientists have observed that humans solve problem using perception and knowledge represented at different levels of information granularity [101]. Humans recognize faces

using a combination of holistic approach together with discrete levels of information or features. They can identify specific facial features and associate a contextual relationship among them to recognize a face even with altered, occluded, and aged appearances. We have designed preprocessing and face verification algorithms to recognize faces with variations in aging and disguise. Specifically, an age transformation algorithm is proposed that registers two face images to minimize the aging variations. Unlike the conventional method, the gallery face image is transformed with respect to the probe face image and facial features are extracted from the registered gallery and probe face images. We further propose a granular approach for face verification by using dynamic feed-forward neural architecture and extracting 2D log polar Gabor phase features [97] at different granular levels. The granular levels exhibit non-disjoint spatial features which are combined using the proposed likelihood ratio based Support Vector Machine match score fusion algorithm. The proposed face verification algorithm is validated using five face databases. The Notre Dame face database [22], [37] is used for performance evaluation since it contains comprehensive variations in expression and illumination, and has been widely used for evaluating face recognition algorithms. In addition, we have used a face database which contains images with age variations and three face databases specifically aimed at validating the performance for disguised face images. The performance of the proposed face verification algorithm is also compared with existing face verification algorithms.

## 2.1 Proposed Registration based Age Transformation Algorithm

Face verification with age variations is a very challenging problem and most of the existing algorithms do not yield good recognition performance for face images with large age differences. The growth of facial features is dynamic in nature and depends on several factors such as genetics, environment, and living habits. Researchers have used anthropometric measures or learning methods to simulate the facial growth and applied

Face Image
Cartesian Coordinates

Face Image
Polar Coordinates
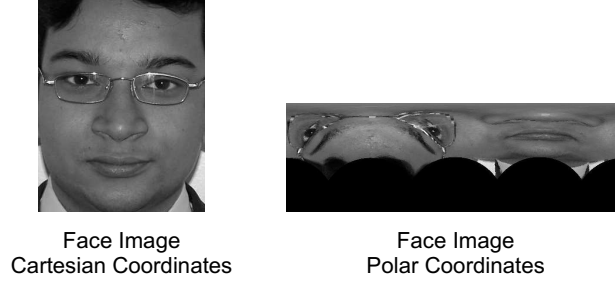
Figure 2.1: Example of cartesian to polar coordinate conversion.

face verification algorithms. However, these age simulation algorithms do not include the dynamics in the facial growth thus causing lower verification accuracy. Another way to address the issue is to minimize the aging difference between the gallery and probe face images using a registration technique. We propose the mutual information registration based age transformation algorithm which can be used to recognize face images with significant age difference between them. The details of age difference minimization using the registration of gallery and probe face images is described below:

Let $F$ be the face image and subscripts $g$ and $p$ represent the gallery and probe respectively. Let $F_g$ and $F_p$ be the detected gallery and probe face images to be matched. $F_g(x, y)$ and $F_p(x, y)$ are transformed into polar form to obtain $F_g^T(r, \theta)$ and $F_p^T(r, \theta)$ respectively. Here, $r$ and $\theta$ are defined with respect to the center coordinate $(x_c, y_c)$.

$$r = \sqrt{(x - x_c)^2 + (y - y_c)^2} \quad 0 \leq r \leq r_{max} \tag{2.1}$$

$$\theta = tan^{-1}\left(\frac{y - y_c}{x - x_c}\right) \tag{2.2}$$

The coordinates of eyes and mouth are used to form a triangle and the center of this triangle is chosen as the center point, $(x_c, y_c)$, for cartesian to polar conversion. Figure 2.1 shows an example of cartesian to polar conversion around the center point. This cartesian to polar conversion eliminates minor variations due to pose and provides robust feature mapping used in the next steps of the algorithm.

Mutual information is a concept from information theory in which statistical de-

14

pendence is measured between two random variables. Using mutual information, the proposed registration algorithm minimizes the geometric differences between the gallery and probe face images. Mutual information between two polar face images can be represented as [65], [76],

$$M(F_g^T, F_p^T) = H(F_g^T) + H(F_p^T) - H(F_g^T, F_p^T) \tag{2.3}$$

where, $H(\cdot)$ is the entropy[1] of the face image and $H(F_g^T, F_p^T)$ is the joint entropy of gallery and probe face images. Registering a gallery face image $F_g^T$ with respect to a probe face image $F_p^T$ requires maximizing the entropy between $H(F_g^T)$ and $H(F_p^T)$, and minimizing the joint entropy $H(F_g^T, F_p^T)$. Hill $et\ al.$ [44] proposed normalized mutual information which can be represented as,

$$\widetilde{M}(F_g^T, F_p^T) = \frac{H(F_g^T) + H(F_p^T)}{H(F_g^T, F_p^T)} \tag{2.4}$$

In this research, we use the normalized mutual information for minimizing aging differences in a transformation space, $S$, defined as

$$S = \begin{bmatrix} a1 & a2 & 0 \\ a3 & a4 & 0 \\ a5 & a6 & 1 \end{bmatrix} \tag{2.5}$$

This transformation space includes the parameters $a_i$ $(i = 1, \cdots, 6)$ for shear, scale, rotation, and translation. The optimal transformation parameters $S^*$ are obtained by exploring the search space $S$ using normalized mutual information. The search strategy is defined as Equation 2.6.

$$S^* = arg\ max_{\{S\}}\{\widetilde{M}(F_p^T,\ S(F_g^T))\} \tag{2.6}$$

The gallery and probe face images $F_g^T$ and $F_p^T$ are registered using the optimal transformation parameters $S^*$. This registration algorithm is linear in nature and does not

---

[1] $Entropy = \sum_i p_i log(p_i)$, where $p$ contains the histogram counts of the image.

accommodate for the non-linear dynamic variations present in face images. To address the non-linearity present in face images, we apply the multiresolution image pyramid scheme with the registration algorithm. A Gaussian pyramid of both the gallery and probe face images is constructed. Registration parameters are estimated at the coarsest level and are used to warp the gallery face image in the next level of the pyramid. The process is iteratively repeated through each level of the pyramid and a final transformed gallery face image is obtained at the finest pyramid level. In this manner, the global variations are handled at the coarsest resolution level and the local non-linear variations at the finest resolution level. Besides minimizing the age differences, this approaches also reduces the differences due to minor pose and expression variations. Thus, this algorithm can also be viewed as a preprocessing scheme to reduce the spatial variations between the gallery and probe images (Figure 2.2). The registered face images are finally transformed back to cartesian coordinates from polar coordinates. Using the FG-Net face database [35], Figures 2.2 and 2.3 show the registered gallery face images transformed with respect to the probe face images. Since the algorithm performs registration between two images, it can minimize the difference between faces of two individuals. However, as shown in Figure 2.3, the inter-personal differences cannot be exactly registered and any efficient face recognition algorithm can reject these transformed impostor images. Once the age difference between the gallery and probe face image is minimized, face verification algorithms can be applied to verify the identity of the probe image.

## 2.2   Granular Approach for Face Recognition

Sinha *et al.* established 19 results based on face recognition capabilities of human mind [101]. They suggested that humans can efficiently recognize familiar face images even with low resolution and withstand noise factors. Moreover, high and low frequency facial information are processed both holistically and locally. Shape, motion, and imaging conditions are also important factors in processing facial information by the human mind. On the other hand, researchers from psychology and neuro-cognition have simulated and

Gallery Image - 02 years     Probe Image - 07 years     Registered Gallery Image

Gallery Image - 10 years     Probe Image - 14 years     Registered Gallery Image

Gallery Image - 24 Years     Probe Image - 42 Years     Registered Gallery Image

Gallery Image - 30 years     Probe Image - 40 years     Registered Gallery Image
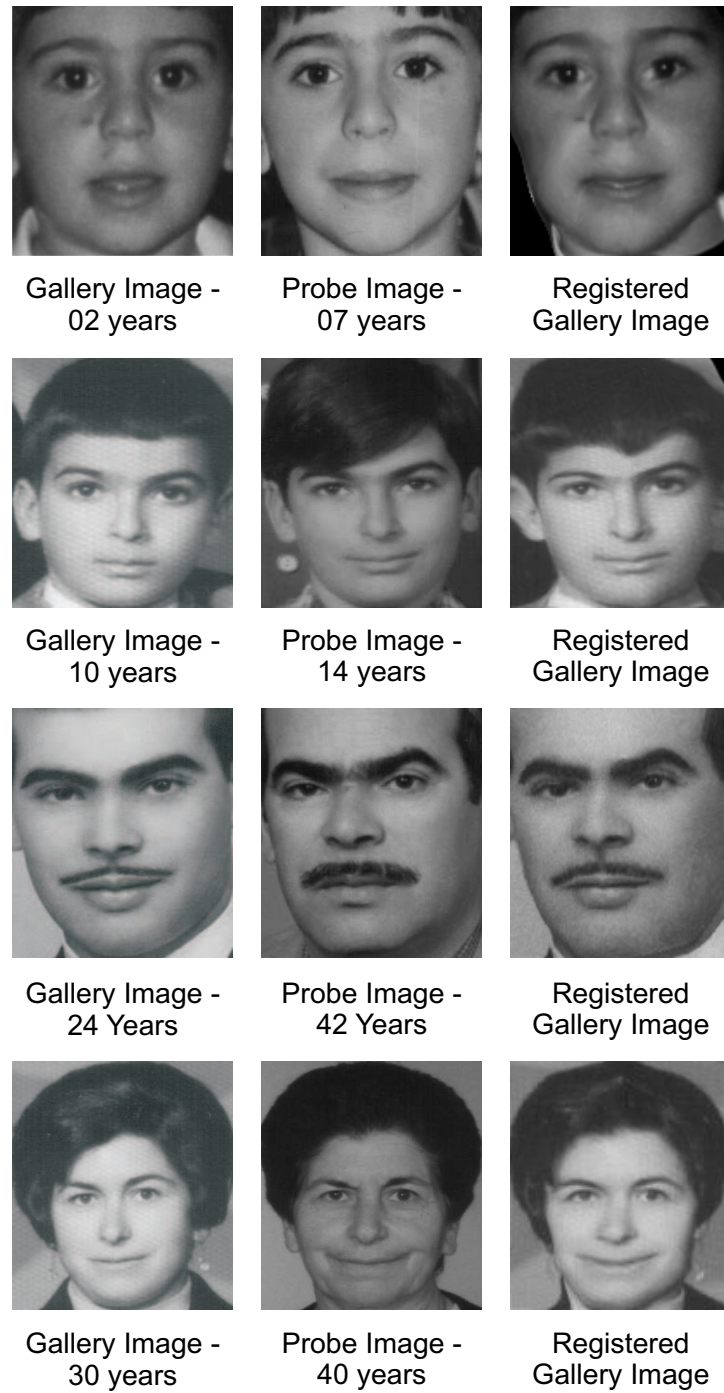
Figure 2.2: Results of the proposed age transformation algorithm on images from the FG-Net face database when the gallery and probe images belong to the same individual [35].
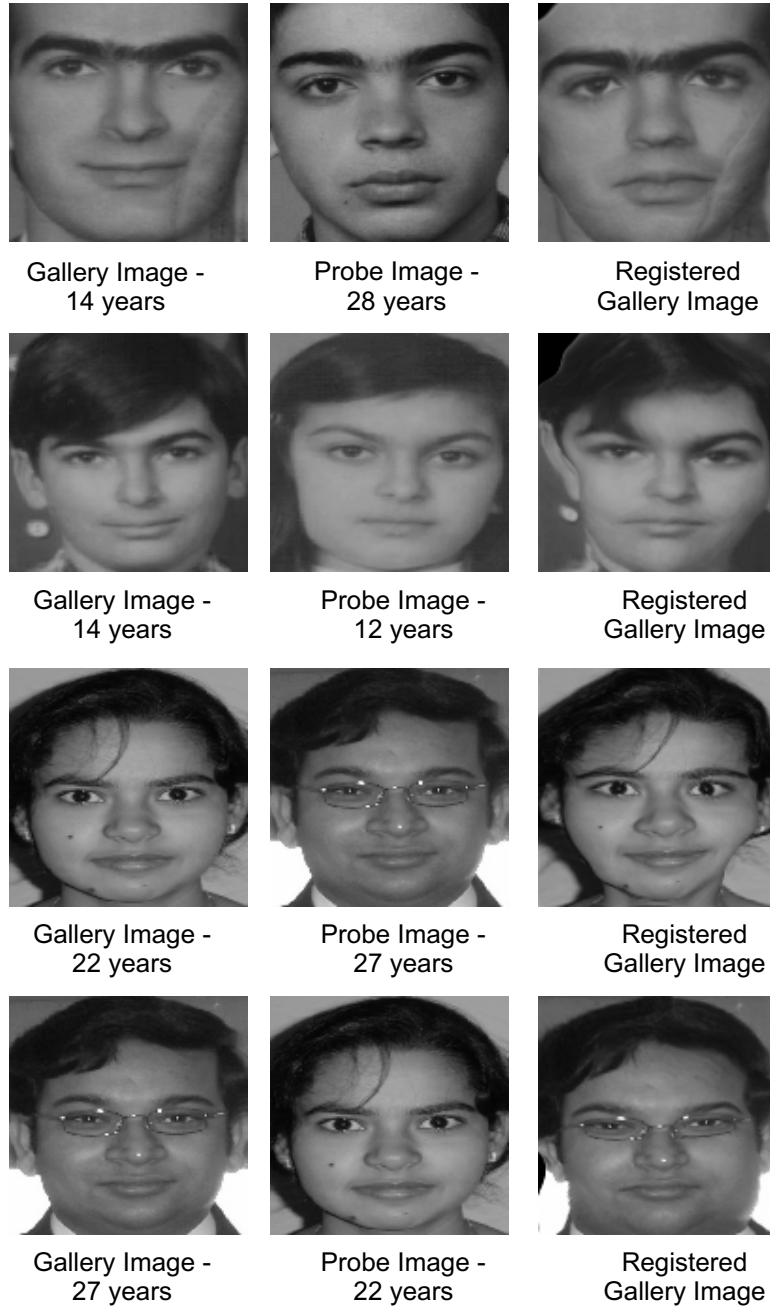
Figure 2.3: Results of the proposed age transformation algorithm on images from the FG-Net face database when the gallery and probe images belong to different individual [35]. Any efficient face recognition algorithm can identify these cases as impostors because the registration algorithm does not minimizes the inter-personal variations.

analyzed human visual cortex by conducting certain experiments. The analysis shows that Gabor transform represents the properties of human visual cortex [25], [27], [119] and exhibits log polar distribution response [28]. It is our hypothesis that if these capabilities can be encoded in an automatic face recognition algorithm, then the recognition performance of the algorithm can be comparable to the performance of human mind.

To incorporate the above mentioned research findings, we propose a granular approach for facial feature extraction and matching. The approach is based on the concept that human mind solves problems at different levels of granularity [7], [8], [21]. In this concept, known as granular computing, a unified framework is used to extract non-disjoint features at different granularity levels and are synergistically combined to obtain a composite information. Granular computing framework is more useful in solving problems when the global information is fuzzy. For example, it is difficult to produce correct verification result with a disguised face image. However, with granulated information, more flexibility is achieved in analyzing the underlying information such as nose, ears, forehead, hair, cheeks, or combination of two or more features.

There are four basic components of granular computing [7], [8], [21]: granules, granular structure, granulation, and information retrieval from granules.

- Granules are the basic component of granular computing. These are small particles which collectively provide a representation with respect to a particular level. Granules have internal, external and contextual properties which reflect interaction of elements inside a granule, interaction with other granules, and relationship among granules.

- Granular structure provides a structured description of the problem. This structure can be interpreted as a level or granulated view in the overall hierarchy.

- Granulation involves construction of granules and granular structure, and defines the representation and characterization of granules and its structure.

- Information retrieval from granules involves granular information processing such as feature extraction and matching. This component may also involve computing
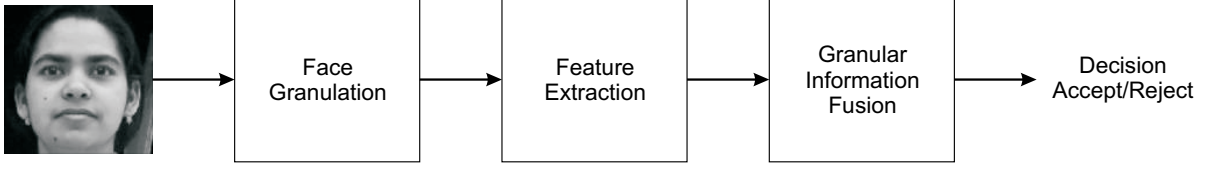
Figure 2.4: Illustrating the concept of the proposed face recognition algorithm.

the relationship among granules, processing based on priors, and preservation of invariant properties.

The proposed face recognition algorithm generates face granules by granulating face image at three different levels. The granules are generated such that non-disjoint facial features can be extracted to provide resilience to variations in disguise. Further, features are extracted from the face granules and are fused using the proposed match score fusion algorithm. As shown in Figure 2.4, the proposed face recognition algorithm thus comprises of three steps: generating granules from the face image, feature extraction from the face granules and feature matching, and fusion of granular match scores for recognition.

## 2.3   Generating Granules from Face Image

Let $F$ be the detected and preprocessed frontal face image of size $n \times n$. Face granules are generated pertaining to three different levels of granularity. The first level of granularity provides global information at multiple levels of resolution. This is analogous to a human mind processing holistic information for face recognition at varying resolutions. Campbell *et al.* have shown that inner and outer facial regions provide distinct information to recognize face images [16]. At the second level of granularity, different inner and outer facial information are extracted. Further, local facial features play an important role in face recognition by human mind. Humans efficiently process information in parts and are able to recognize occluded faces. Therefore, at the third level of granularity, we extract facial features from the local facial fragments.

20

In the first level, face granules are generated by applying the Gaussian and Laplacian operators [14]. The Gaussian operator generates a sequence of low pass filtered images by iteratively convolving each of the constituent images with a 2D Gaussian filter kernel [98]. The resolution and sample density of the image is reduced between successive iterations and therefore the Gaussian kernel operates on a reduced version of the original image in every iteration. The resultant images $I_0, I_1, \ldots, I_A$ may be viewed as a *pyramid* with $I_0$ having the highest resolution and $I_A$ having the lowest resolution. Let $\overline{w}(x, y)$ represent the Gaussian kernel of dimension $5 \times 5$ and reduction factor 4. The *Reduce* operation can be written as,

$$Reduce[F(p, q)] = \sum_{x=1}^{5} \sum_{y=1}^{5} \overline{w}(x, y) F(2p + x, 2q + y) \tag{2.7}$$

A Gaussian pyramid $I_B$ is defined as,

$$I_0 = F \tag{2.8}$$

$$I_B = Reduce[I_{B-1}], \quad 0 < B < A \tag{2.9}$$

Further, the Laplacian operator generates band-pass images and the process can be summarized as follows:

$$L_B = I_B - Expand[I_{B+1}], \quad 0 \leq B < A \tag{2.10}$$

Here, the $Expand[\cdot]$ operator interpolates a low-resolution image to the next highest resolution and can be represented as,

$$Expand[I_{B,D}(p, q)] = 4 \sum_{x=-2}^{2} \sum_{y=-2}^{2} \overline{w}(x, y) I_{B,D-1} \left( \frac{p - x}{2}, \frac{q - y}{2} \right) \tag{2.11}$$

Note that $I_{B,D}$ in Equation 2.11 denotes *expanding $I_B$ $D$ number of times*. Let the granules generated by Gaussian and Laplacian operators be represented by $F_{Gri}$, where $i$ represents the granule number. For a face image of size 128×128, Figure 2.5 represents the

$F_{Gr1}$       $F_{Gr2}$       $F_{Gr3}$       $F_{Gr4}$       $F_{Gr5}$       $F_{Gr6}$
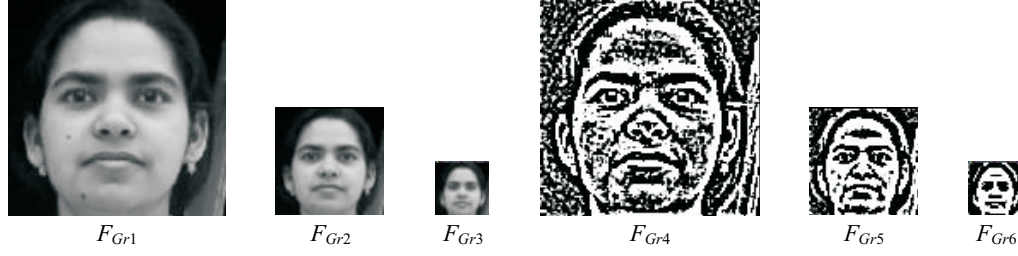
Figure 2.5: Face granules generated in the first level of granularity. $F_{Gr1}, F_{Gr2}, and F_{Gr3}$ are generated by the Gaussian operator, and $F_{Gr4}, F_{Gr5}$, and $F_{Gr6}$ are generated by the Laplacian operator.

face granules generated in the first level by applying Gaussian and Laplacian operators. $F_{Gr1}$ to $F_{Gr3}$ are the granules generated by Gaussian operator and $F_{Gr4}$ to $F_{Gr6}$ are the granules generated by Laplacian operator. The size of the smallest granule in the first level is $32 \times 32$. These operators are applied only to three levels of resolution because facial features are not useful for recognition when image size is less than $32 \times 32$. In these six granules, facial features are segregated at different resolutions to provide edge information, noise, smoothness, and blurriness present in a face image. This level of granular information thus provides resilience to disguise variations in facial features such as eyes, mouth, and nose.

Campbell *et al.* reported that different inner and outer facial regions represent distinct information which is useful for face recognition [16]. Therefore, in the second level of granularity, we generate the horizontal and vertical granules by dividing the face image $F$ into different regions as shown in Figures 2.6 and 2.7. Here, $F_{Gr7}$ to $F_{Gr15}$ denote the horizontal granules and $F_{Gr16}$ to $F_{Gr24}$ denote the vertical granules. Among the nine horizontal granules, the first three granules i.e. $F_{Gr7}$, $F_{Gr8}$, and $F_{Gr9}$ have the same size $n \times \frac{n}{3}$. The next three granules, i.e., $F_{Gr10}$, $F_{Gr11}$, and $F_{Gr12}$ are generated such that the size of $F_{Gr10}$ and $F_{Gr12}$ is $n \times (m - \epsilon)$ and the size of $F_{Gr11}$ is $n \times (m + 2\epsilon)$. Further, $F_{Gr13}$, $F_{Gr14}$, and $F_{Gr15}$ are generated such that the size of $F_{Gr13}$ and $F_{Gr15}$ is $n \times (m + \epsilon)$ and the size of $F_{Gr14}$ is $n \times (m - 2\epsilon)$. Similarly, nine vertical granules $F_{Gr16}$ to $F_{Gr24}$ are
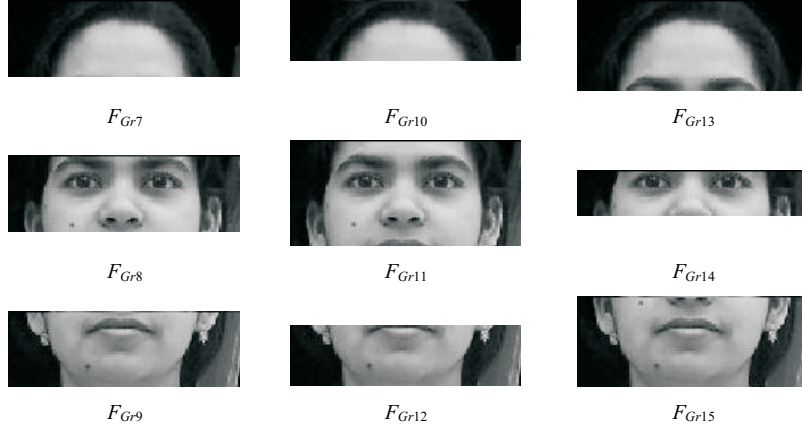
Figure 2.6: Horizontal face granules generated from a face image at the second level of granularity ($F_{Gr7}$ to $F_{Gr15}$).



Figure 2.7: Vertical face granules generated from a face image at the second level of granularity ($F_{Gr16}$ to $F_{Gr24}$).

generated. Figures 2.6 and 2.7 show horizontal and vertical granules when the size of normalized face image is $128 \times 128$ and $\epsilon = 10^2$. This level of granularity provides the relation between horizontal and vertical granules to handle minor variations in disguise such as glasses, hair style, beard, and mustache.

Researchers in cognitive science suggest that local facial fragments can provide robustness against partial occlusion and change in viewpoints [42], [86], [101]. Moreover, human mind can distinguish and classify individuals with their local facial fragments such as nose, eyes, and mouth. To incorporate this property, we extract local facial

---

[2]In our experiments, $\epsilon = 10$ yields the best verification results with face images of size $128 \times 128$.
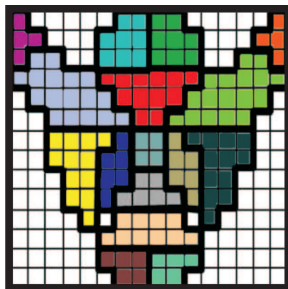
Figure 2.8: Golden ratio [3] for selecting local facial fragments.
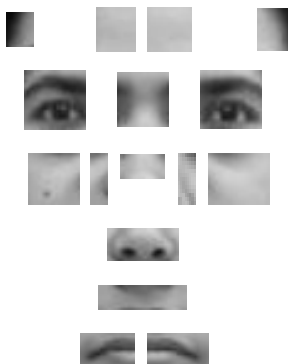


Figure 2.9: Face granules generated in the third level of granularity representing local information ($F_{Gr25}$ to $F_{Gr40}$).

fragments and use them as face granules in the third level of granularity. Given the eye coordinates, 16 local facial fragments are extracted using the golden ratio face template [3] shown in Figure 2.8. Each of these fragments is a granule representing local information that provide invariant and unique features for handling variations due to expression and disguise. Figure 2.9 shows an example of local facial fragments used as face granules in the third level of granularity.

The proposed granulation technique is used to generate 40 non-disjoint face granules from a face image of size $128 \times 128$. The technique is based on using a fixed structure (golden ratio template). However, for face images captured in a controlled setting, the images can be granulated according to local features. For face images with variations due to aging and disguise, identifying features is challenging and hence feature-based partitioning does not yield accurate results compared to fixed structure partitioning.

## 2.4 Log Polar Gabor Transform for Feature Extraction

Face recognition using Gabor features has been proposed in several research papers [6], [46], [62], [90], [122], [128]. Most of the algorithms extract amplitude and phase features at different scale and orientation. These features are either matched using a matching algorithm or the features are first subjected to techniques such as Principal Component Analysis, Independent Component Analysis, or histogram analysis and then matched. However, neural networks can also be used to combine the response of Gabor transform and obtain an optimal composite Gabor response. Further, face recognition algorithms are based on feature extraction from a cartesian perspective. The image represented in the human visual cortex is negatively correlated with retinal cell eccentricity. This property causes a formulation of spatial log polar transformation. In this transformation, the image is resampled as a logarithmic function of the distance from the center. This fundamental concept of log polar transformation has been used by Smeraldi and Bigun [103], and Tistarelli and Grosso [112]. These algorithms use a log polar transformation which simulates the retinal sampling resolution.

In this section, we present the proposed face recognition algorithm which is based on the textural information of face image extracted using 2D log polar Gabor transform. We further extend the 2D log polar Gabor transform based feature extraction algorithm to learn and combine Gabor responses using dynamic neural network.

### 2.4.1 Overview of Gabor Transform

Researchers from psychology and physics have simulated and analyzed the human visual cortex by conducting certain experiments. It has been reported that the properties of a human visual cortex can be represented by the Gabor transform [25], [27], [119]. In 1946, D. Gabor [38] established the uncertainty principle where the product of uncertainties in frequency $\Delta f$ and time $\Delta t$ must exceed a fixed constant. If the sig-

nal is transmitted through a bank of bandpass filters, the closest frequencies which can be distinguished are given by $\Delta f = 1/\Delta t$. In Fourier space, the product of time and bandwidth gives the maximum amount of information that can be transmitted. Gabor showed that the minimum area in Fourier space is achieved by the Gaussian modulated complex exponential functions of the form

$$\phi_{jk}(t) = exp\left[-\pi(t - j\Delta t)^2/\alpha^2\right] exp\left[2\pi ik\Delta f(t - j\Delta t)\right] \tag{2.12}$$

where $\Delta f\Delta t = 1$. In this equation, the first exponential term represents a Gaussian envelop centered on $j\Delta t$ and the second term represents the conjugate exponential form of the trigonometric functions of frequency $k\Delta f$ which is a periodic function, and denotes the locality of the Gaussian envelop. Equation 2.12 can also be represented as,

$$\phi_{jk}(t) = exp\left[-\pi(t - j\Delta t)^2/\alpha^2\right]\left[cos[2\pi ik\Delta f(t - j\Delta t)] + isin\left[2\pi ik\Delta f(t - j\Delta t)\right]\right]$$
$$\tag{2.13}$$
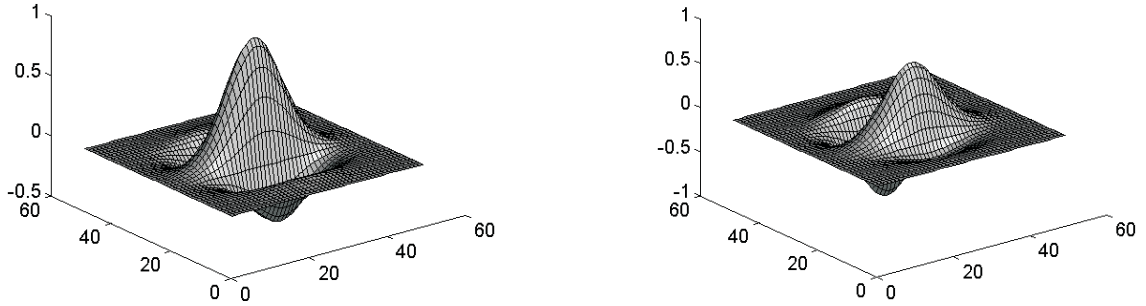


Figure 2.10: Even and odd components of the 2D Gabor filter.

As $\alpha \to \infty$, Gabor transform reduces to Fourier transform and no locality is shown at this point. Experiments performed by the researchers show that the receptive fields of simple cells in a visual cortex are similar to the Gaussian modulated sinusoids and 97% of them are statistically indistinguishable from the odd symmetric or even symmetric parts of a 2D Gabor elementary function [26]. This quadrature phase relationship can be computed from the odd and the even symmetric parts of the 2D Gabor transform

[26], as can be seen in Equation 2.13. Figure 2.10 shows the odd and the even symmetric parts of the 2D Gabor transform and Figure 2.11 shows the Gabor response at different scales and orientations. Gabor transform is a non-orthogonal transform and according to Daugman [27], non-orthogonal representations are ubiquitous in biological sensory and motor systems. Therefore, it can be used for efficiently imitating the visual cortex and coding the visual information.
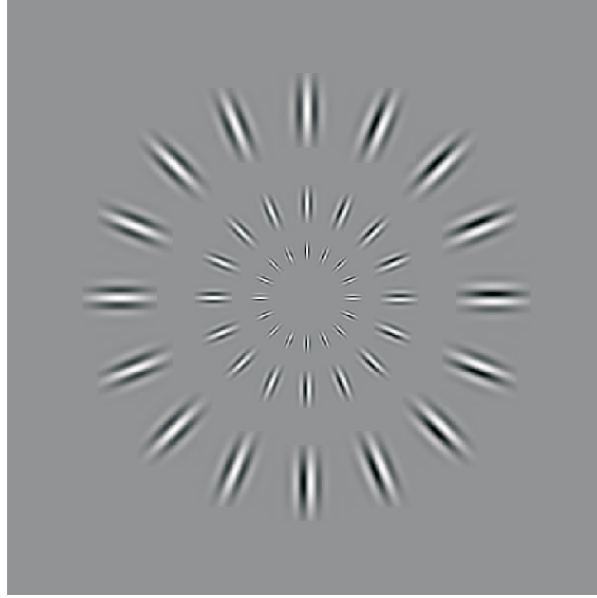


Figure 2.11: Gabor filter at different scales and orientations.

Another property of neurophysiological and psychological system is that they show a log-polar distribution of response selectivity in cells in the visual cortex, which shows the orientation half-bandwidth of $\pm 15^0$ and the frequency bandwidth of 1.5 octaves [28]. To account for this property, we have used log-polar Gabor transform for representing the face image in which radial distance represents the spatial frequency and the polar angle represents the orientation. Log polar Gabor transform is a form of Gabor transform [10], [64], [68] which is based on polar coordinates and the dependency of directional independent variance ($\sigma$) on the polar coordinate is realized by a logarithmic scale. Thus the functional form of 2D log polar Gabor filter can be represented as

$$G_{r_0,\theta_0}(r, \theta) = exp\left(-2\pi^2\sigma^2\right)\left[(ln(r) - ln(r_0))^2 + (ln(r)sin(\theta - \theta_0))\right] \qquad (2.14)$$

and the position of filter in the Fourier domain is defined by

$$r_{00} = \sqrt{2}, r_{0i} = 2^i * r_{00}, \theta_{0i} = i * \frac{2\pi}{N_\theta} \qquad (2.15)$$

where $r_{00}$ is the smallest possible frequency, $N_\theta$ is the number of filters on the unit circle, and at index $L$, $\sigma_L$ and $s_L$ are further defined by

$$\sigma_L = \frac{1}{ln(r_0)\pi sin(\pi/N_\theta)}\sqrt{\frac{ln2}{2}} \qquad (2.16)$$

$$s_L = \frac{ln(r_0)\pi sin(\pi/N_\theta)}{ln2}\sqrt{\frac{ln2}{2}} \qquad (2.17)$$

Log-Gabor transform has no DC component and has an extended tail. According to Field [34], log-Gabor functions with extended tail are able to encode the images more efficiently compared to Gabor transform, because Gabor transform would over represent the low frequency components and under represent the high frequency components. Log-polar Gabor transform also provides invariance to rotation and scaling.

## 2.4.2   Phase Feature Extraction and Matching

Let the size of preprocessed face image $F(x, y)$ be $N \times N$. The face image is transformed into its log polar form $F(r, \theta)$ [78],

$$F(r, \theta) = F\left(\left[\frac{N}{2}\right] + \left[\theta cos\left(\frac{2\pi r}{s}\right)\right], \left[\frac{N}{2}\right] - \left[\theta sin\left(\frac{2\pi r}{s}\right)\right]\right) \qquad (2.18)$$

28

where $r = 0, ..., s - 1$, $\theta = 0, ..., [N/2] - 1$, and $s$ is the factor by which the image is sampled from $0^0$ to $360^0$ to produce its equivalent polar form. This polar face image is then operated with 2D log polar Gabor transform [38] which provides both amplitude and phase information. Let $F_g(r, \theta)$ be the output of this operation. Venkatesh and Owens [115] showed that the degree of phase is independent of overall magnitude of the face image and provides invariance to the changes in illumination and contrast. Phase features are computed from $F_g(r, \theta)$ using Equation 2.19. The features are quantized using Equation 2.20 to generate a binary representation, $B$, referred to as *phase template.*

$$P(r, \theta) = tan^{-1} \left( \frac{ImF_g(r, \theta)}{ReF_g(r, \theta)} \right) \tag{2.19}$$

$$B(r, \theta) = \begin{cases} [1, 1] & if \quad 0^0 \quad < P(r, \theta) \leq 90^0 \\ [0, 1] & if \quad 90^0 \quad < P(r, \theta) \leq 180^0 \\ [0, 0] & if \quad 180^0 < P(r, \theta) \leq 270^0 \\ [1, 0] & if \quad 270^0 < P(r, \theta) \leq 360^0 \end{cases} \tag{2.20}$$

To match the two phase templates, first the templates are divided into $m$ frames each of size $p \times q$. Corresponding frames from the two phase templates are matched using hamming distance as shown in Equation 2.21.

$$D_i = min \left( \frac{\Sigma_i \left( B_i^1 \otimes B_i^2 \right)}{p * q} \right) \tag{2.21}$$

where $B_i^1$ and $B_i^2$ are the $i^{th}$ frames for the two templates, $\otimes$ represents the XOR operation, $D_i$ is the corresponding distance measure, and $min$ computes the minimum hamming distance. A circular shift of eight pixels in all directions is applied and the minimum hamming distance is computed to make the process shift invariant. The phase matching score $M^P$ is then calculated using Equations 2.22 and 2.23,

$$M_s^P = \begin{cases} M_s^P + 1, & if \quad D_i \geq \eta_1 \\ M_s^P, & if \quad D_i < \eta_1 \end{cases} \tag{2.22}$$

$$M^P = \frac{M_s^P}{m} \tag{2.23}$$

where $\eta_1$ is the frame matching threshold, $M_s^P$ is the intermediate matching score for frames, and $m$ is the number of frames. A match occurs if the matching score $M^P < \eta_2$, where $\eta_2$ is the phase matching threshold for the face template.

## 2.4.3   Learning Discriminative Log Polar Gabor Phase Features using Dynamic Neural Network Architecture

Since the log polar Gabor transform does not form an orthogonal basis, it is very difficult to obtain the expansion of the image. Daugman proposed a solution of this problem by using neural network architecture to extract the Gabor coefficients [29]. In the implementation, Daugman fixed a set of Gabor functions and expressed a pixel based image as an expression in the basis of Gabor function. A network of neuron like units with a particular learning rule is used and the architecture is designed in such a manner that the output weights after stabilization are the best least mean square approximation to the Gabor parameters. This is achieved by finding the dimension of error with respect to each of the Gabor parameters and using a gradient descent method to iteratively approximate the solution. However, classical neural networks inherently have static architecture. In such networks the number of neurons is fixed. Also adaptation is a slow process with limited connectivity between neurons, and the number of neurons needed for learning is dependent on the problem being solved.

In this research, we propose the use of four layer dynamic neural network for computing the discriminative log polar Gabor phase features. It is a fully connected, four layer feed-forward neural network with a dynamic first hidden layer architecture trained using a modified back-propagation technique. The dynamic neural network architecture [84] is defined during the feature extraction process and evolves to a quasi-optimal topology for the given task. In this network, two hidden layers are used and are central to the adaptation capabilities. The first hidden layer is dynamic in size and provides the

necessary learning support to satisfy the goals of adding knowledge capacity to the network as well as providing generalization capabilities for unknown input data. The second hidden layer is fixed in size because the need of a fixed dimension for data representation necessitates the use of a fixed second hidden layer representation. The size of input and output layers are same, and are constrained by the needs of the facial feature encoding to provide an auto-associative mapping from the input pattern to the output pattern. If the size of the dynamic hidden layer is too large, the neurons over-specialize and learn individual patterns instead of their underlying features. If the dynamic hidden layer is too small then the neurons are not able to provide the bounded distortion rates needed for facial feature extraction. We can use the traditional back-propagation techniques for training such networks but it can result in unbalanced training errors for different patterns. This is addressed in the dynamic neural network learning algorithm by the use of the descending epsilon technique [127]. Epsilon represents the margin of the tolerable error at each stage of training. It improves the correctness ratio of the network for each pattern as epsilon is reduced, and hence improves the general behavior of the trained network. The dynamic neural network algorithm also utilizes the techniques described in [45] to determine the size of the network needed for providing architectural support during learning and avoid stagnation during training. During the expansion stage, hidden layer neurons are added one at a time as required, to enable the network to escape from local minima. When a new unit is added, the dimension of the weight space increases, providing a new path for the trapped network to escape. After reaching the desired error rate, the network is pruned and tested for convergence until the optimal size is reached.

In the learning process, we train the dynamic neural network to synergistically combine the discriminating and unique log polar Gabor phase features for optimal recognition performance. Learning is performed on a labeled training database which is prepared by combining images from different existing face databases. Details of the training database are provided in Section 2.7.1.

We use a filterbank of 64 log polar Gabor filters along with 8 center surround differences of Gaussian filters and 4 low pass filters. Thus a total of 76 filters are used in

the proposed face recognition algorithm. These filters are used as the weights of dynamic hidden layer. Input to the neural network is the training face images in log polar form. Training is initialized with one filter (from the 76 filters) as the weight of the dynamic hidden layer. Suppose the log polar Gabor transform is $G_i(r, \theta)$ where $i$ = number of filters and the training polar face image is $F_t(r, \theta)$. Output of the first hidden layer is $\sum_i G_i(r, \theta) \times F_t(r, \theta)$, which is provided as input to the second hidden layer. In the second hidden layer, we synergistically combine the response of the first hidden layer with weight $w_j$ which is computed as,

$$w_j = \frac{\sum_{r,\theta} G_i(r, \theta) F_t(r, \theta)}{\sum_{r,\theta} G_i^2(r, \theta)} \tag{2.24}$$

where $j$ is the number of nodes in the second hidden layer. Finally, the output of the second hidden layer is used as input to the output layer and an auto-associative mapping is performed such that the size of output layer is the same as the input layer or the size of input image. Let the output responses be $F_{tg}(r, \theta)$, then

$$F_t(r, \theta) = \sum_i \sum_{r,\theta} F_{tg}(r, \theta) \, G_i(r, \theta) \tag{2.25}$$

Here, we represent $F_t(r, \theta)$ by finding the projection coefficients and projecting it into a set of chosen vectors $G_i(r, \theta)$ which provides the response $F_{tg}(r, \theta)$. We thus transform the face information into Gabor domain and represent face image with $F_{tg}(r, \theta)$ which is a complex valued matrix containing the amplitude and the phase information. Phase information is then computed using Equations 2.19 and 2.20.

Similarly, we compute the phase features for all the face images in the training and validation database in the first iteration of learning. Phase templates are then matched using the phase template matching algorithm described in Section 2.4.2. For the first iteration with one filter, the verification error on the validation database is 58.3%. To reduce the error rate, we repeatedly add filters to the first hidden layer until the verification error on the validation database converges to a small value. We further perform the same learning procedure for all possible combinations of filters and

the learning finally converges at the error rate of 1.9%, i.e., verification accuracy of 98.1%. Figure 2.12 shows the final quantized phase feature obtained using the proposed dynamic neural network architecture based 2D log polar Gabor transform. Once the learning is complete and discriminating log polar Gabor phase features are learned, we use the trained neural network architecture for face recognition. Figure 2.13 shows the Gabor filters selected after training.

The training process is performed separately for all 40 face granules. At the end of the training procedure, we have 40 quantized phase features, one for each granule. For testing, the trained neural network architecture is used to generate the gallery phase templates corresponding to the gallery face images and are stored in the database. During query, phase template for the probe image is generated and matched using the phase template matching algorithm. The algorithm generates match scores, $s_i$, $(i = 1, \cdots, 40)$, pertaining to 40 face granules. Each of these match scores are in the range $[0, 1]$ where 0 represents perfect reject and 1 represents perfect accept.



Figure 2.12: Face image and corresponding quantized phase feature obtained using dynamic neural network architecture based 2D log polar Gabor transform.
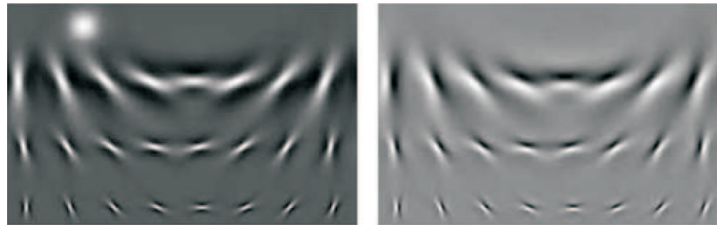


Figure 2.13: Even and odd components of Gabor filters selected after training. Size of the face image is $128 \times 128$.

## 2.5 Granular Information Fusion

The final step in the proposed granular approach for face verification is fusion of match scores obtained from different face granules. In the case of disguise and aging, face granules may provide conflicting decisions. For example, due to variations in disguise, Gaussian granules may reject the genuine subject but some of the local granules may provide a decision to accept. In such cases, a carefully designed fusion algorithm is required to efficiently combine these confounding match scores. This section presents a match score fusion algorithm to combine the match scores obtained from face granules. The novelty of this algorithm lies in integrating likelihood ratio test-statistic [67] in a SVM fusion framework. The likelihood ratio aspect of the algorithm makes it robust to uncertainties in the face granules whereas the use of SVM ensures that the algorithm is less prone to over-fitting thereby permitting it to handle conflicting match scores. The fusion algorithm consists of three steps: (1) transforming match scores into likelihood ratio, (2) integrating the verification prior of each face granule, and (3) applying support vector machine fusion. This section presents an overview of $2\nu$-support vector machine followed by the proposed fusion algorithm.

### 2.5.1 $2\nu$-Support Vector Machine

Support vector machine [116] is a pattern classifier that constructs non-linear hyperplanes in a multidimensional space. In this research, we use dual $\nu$-SVM [24] which is an attractive alternative to SVM and offers much more natural setting for parameter selection with reduced computational complexity.

Let $\{\mathbf{x}_i, y_i\}$ be the set of $N$ data vectors where $i = 1, ..., N$, $\mathbf{x}_i \in \Re^d$ and $y_i$ is the hard label such that $y_i \in (+1, -1)$. The basic principle of SVM is to find the hyperplane that separates the two classes with the widest margin, i.e., $\mathbf{w}\varphi(\mathbf{x}) + b = 0$ to minimize,

$$
\begin{aligned}
&\tfrac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \psi_i \\
&\text{subject to } y_i\left(\mathbf{w}\,\varphi(\mathbf{x}_i) + b\right) \geq (1 - \psi_i), \quad \psi_i \geq 0
\end{aligned}
\tag{2.26}
$$

where $b$ is the offset of the decision hyperplane, $\mathbf{w}$ is the normal weight vector, $\varphi(\mathbf{x})$ is the mapping function used to map the data space to the feature space and provide generalization for the decision function. $C$ is the regularization factor between the total distance of each error from the margin and the width of the margin, and $\psi_i$ is the slack variable used for classification errors [116]. The optimal SVM parameters are obtained by manually setting the parameters until an optimal error rate is achieved. This heuristic process is very time consuming. Dual $\nu$-SVM, originally proposed by [24], is a computationally efficient variant of SVM. It is more flexible in the training and overcomes the issues when the training class sizes are not same. In Equation 2.26, additional class dependent parameters ($\rho$, $\nu$ and $C_i$) are introduced such that the formulation becomes,

$$
\begin{aligned}
&min\{\tfrac{1}{2}\|\mathbf{w}\|^2 - \textstyle\sum_i C_i(\nu\rho - \psi_i)\} \\
&\text{subject to } y_i\left(\mathbf{w}\,\varphi(\mathbf{x}_i) + b\right) \geq (\rho - \psi_i), \quad \rho,\ \psi_i \geq 0
\end{aligned}
\tag{2.27}
$$

where $\rho$ is the position of the margin and $\nu$ is the error parameter that can be calculated using $\nu_+$ and $\nu_-$ which are the error parameters for training the positive and negative classes respectively.

$$
\nu = \tfrac{2\nu_+\nu_-}{\nu_+ + \nu_-}, \quad 0 < \nu_+ < 1 \ \text{ and } \ 0 < \nu_- < 1
\tag{2.28}
$$

$C_i(\nu\rho - \psi_i)$ is the cost of errors and $C_i$ is the error penalty for each class which is calculated as,

$$
C_i = \begin{cases} C_+, & if \quad y_i = +1 \\ C_-, & if \quad y_i = -1 \end{cases}
\tag{2.29}
$$

where,

$$
C_+ = \tfrac{\nu}{2n_+\nu_+},
$$

$$
C_- = \tfrac{\nu}{2n_-\nu_-}.
$$

$$
\tag{2.30}
$$

Here, $n_+$ and $n_-$ are the number of training points for the positive and negative classes respectively. Further, the $2\nu$-SVM objective function can be formulated as (Wolfe Dual

35

formulation),

$$L = \sum_i \alpha_i - \left\{ \frac{1}{2} \sum_{i,j} \alpha_i\, \alpha_j\, y_i\, y_j\, K(x_i, x_j) \right\} \tag{2.31}$$

where $i, j \in 1, ..., N$, $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function [116], $\alpha_i, \alpha_j$ are the Lagrange multipliers such that $0 \leq \alpha_i \leq C_i$, $\sum_i \alpha_i y_i = 0$, and $\sum_i \alpha_i \geq \nu$.

## 2.5.2  Incorporating Likelihood Ratio Test Statistic in a SVM Fusion Framework

Traditionally, SVM-based fusion techniques use the match scores directly in the fusion process. However, match scores may provide inaccurate or fuzzy information due to environmental dynamics or matcher limitations. By transforming the match scores into likelihood ratio, the uncertainties associated with them can be addressed better [67]. Therefore, the input to the fusion algorithm comprises of likelihood ratio induced from match scores and the verification prior of individual face granules (i.e., precision of the face granules). The procedure of transforming match scores into likelihood ratio is described as follows:

For a two-class biometrics problem, i.e. $\Theta = \{genuine, impostor\}$, the match scores corresponding to all $N = 40$ face granules are first computed, i.e. $\mathbf{s} = (s_1, s_2, \cdots, s_N)$, and then the densities of the genuine and impostor scores ($J_{gen}(\mathbf{s})$ and $J_{imp}(\mathbf{s})$, respectively) are estimated. In the proposed SVM fusion algorithm, it is assumed that the distribution of match scores is a Gaussian distribution, i.e.,

$$J_j(s_i, \overline{\mu}_{ij}, \overline{\sigma}_{ij}) = \frac{1}{\overline{\sigma}_{ij}\sqrt{2\pi}} exp\left[ -\frac{1}{2}\left\{ \frac{s_i - \overline{\mu}_{ij}}{\overline{\sigma}_{ij}} \right\}^2 \right] \tag{2.32}$$

where, $\overline{\mu}_{ij}$ and $\overline{\sigma}_{ij}$ are the mean and standard deviation of the $i^{th}$ classifier corresponding to the $j^{th}$ element of $\Theta$. While this is a very strong assumption, it does not impact the performance of the fusion system in the context of this application.

We compute the likelihood ratio $V_i = \frac{J_{gen}(s_i)}{J_{imp}(s_i)} a_i$ pertaining to each face granule where $a_i$ is the verification prior. The resultant value $V_i$ is used as input to the SVM fusion

36

algorithm. *Further, utilizing the SVM classifier for fusion addresses the limitations of the likelihood test-statistic if the input data does not conform to the Gaussian assumption* (which is usually the case).

In the training phase, likelihood values induced from the match scores and their labels are used to train the $2\nu$-SVM for fusion. Let the labeled training data be represented as $Z_i = (V_i, Y)$, where, $i$ represents the $i^{th}$ face granule. For each match score, the class label $Y \in \Theta$ (or $Y \in (+1, -1)$; where +1 represents the genuine class and -1 represents the impostor class). $N$ SVMs are trained using these labeled training data; one for each face granule. The training data is mapped to a higher dimensional feature space such that $Z \to \varphi(Z)$ and $\varphi(\cdot)$ is the mapping function. The optimal hyperplane which separates the data into two different classes in the higher dimensional feature space can be obtained using $2\nu$-SVM learning [24].

In the testing phase, the fused score of a multimodal test pattern $[V_i], i = 1, \cdots, N$ is defined as,

$$V_{fused} = \sum_{i=1}^{N} f(V_i) \tag{2.33}$$

where,

$$f(V_i) = w_i \varphi(V_i) + b_i \tag{2.34}$$

Here, $w_i$ and $b_i$ are the parameters of the $2\nu$-SVM hyperplane. Finally, to verify the identity, decision of *accept* or *reject* is made on the test pattern using a threshold $t$,

$$Decision(f(V_{fused})) = \begin{cases} Accept, & \text{if output } f(V_{fused}) > t \\ Reject, & \text{otherwise.} \end{cases} \tag{2.35}$$

## 2.6 Databases and Algorithms used for Validation

To evaluate the performance of the proposed granular approach for face recognition, five face databases are used and the performance is compared with five existing

algorithms. Details of these databases and algorithms are presented in Section 2.6.1 and Section 2.6.2 respectively.

## 2.6.1 Databases used for Validation

Currently, there is no comprehensive face disguise and aging database available in public domain. The authors have prepared three databases which contain face images with different disguise variations. These three databases are compiled using disguised faces of real people, disguised synthetic faces, and disguised faces of actors and actresses extracted from movie clips. The Notre Dame Face database [22], [37] and the FG-Net face database [35] are also used to evaluate the performance of the proposed face recognition algorithm. The characteristics of all five databases are explained below.

- **Notre Dame Face Database** [22], [37]: This database is a part of the NIST Face Recognition Grand Challenge (FRGC). We use collection B of the Notre Dame face database which contains around 35,000 high resolution frontal face images with different lighting conditions and expressions. This database also provides images of few individuals with time difference of six months to one year. It is one of the most comprehensive face databases widely used for evaluating the performance of face recognition algorithms. Sample images from the Notre Dame face database are shown in Figure 2.14.

- **Aging Database** To evaluate the performance of the proposed algorithm on aging variations, we use the face aging database which comprises of 1578 images from 130 individuals or subjects. The images are obtained partly from the FG-Net database [35] and partly collected by the authors. Figure 2.15 shows an example of images from the FG-Net face database.

- **Real Face Disguise Database**: This database is prepared by the authors. It contains face images of 25 individuals with 15-25 different variations in images of each individual. To prepare this database, we first analyze possible variations of disguises and classify them into eight categories depending on its effect on appearance

Figure 2.14: Sample images from the Notre Dame face database.



Figure 2.15: Sample images from the aging face database.

and facial features. The disguise images are collected to comprehensively include the following eight categories.

- Minimal variations

- Variations in hair style

- Variations in beard and mustache

- Variations in glasses

- Variations in cap and hat

- Variations in lips, eyebrows or nose

- Variations in aging and wrinkle

- Multiple disguise variations

Since our goal is to evaluate the performance of the proposed face recognition algorithm on disguise, the database contains frontal face images with less emphasis on variations due to illumination, expression and pose. Details of the number of images for each category of disguise variations is provided in Table 2.1. Figure 2.16 shows an example of this database.

- **Synthetic Face Disguise Database**: Creating a large real face disguise database is a challenging task. We used FACES software [33] to generate 4000 frontal face images of 100 subjects with a comprehensive set of variations for disguise. Details of the number of images in each of the eight categories are listed in Table 2.1. Figure 2.17 shows an example of 40 variations of the same face image generated using disguise accessories and makeup tools. This database is used to comprehensively evaluate the performance of the proposed algorithm for disguised images.

- **Movie Face Database**: Movies are a good source for finding face images of a large number of individuals with variations in aging and disguise. Over a period of time, an actor or actress uses makeup tools and accessories providing face images with different variations including aging and disguises. A face database is prepared
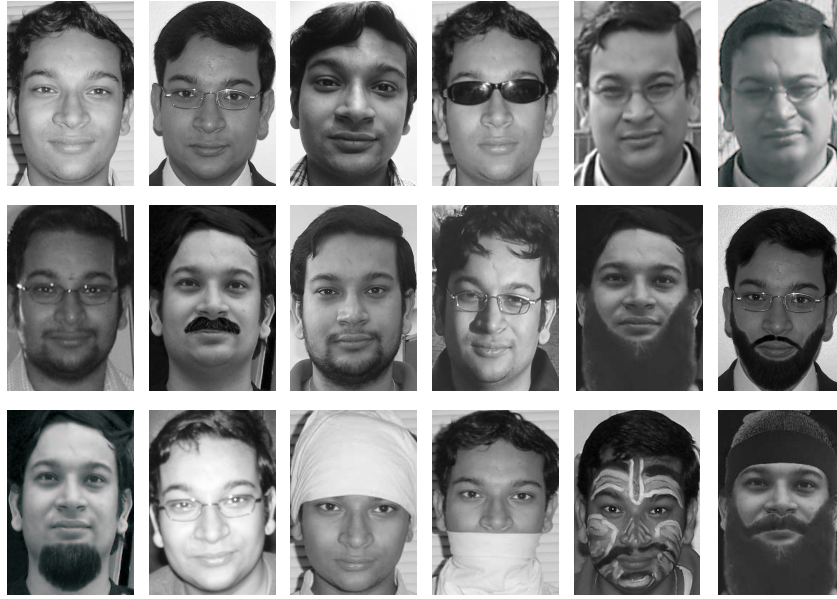
Figure 2.16: Images from the real face disguise database showing 18 variations of an individual.
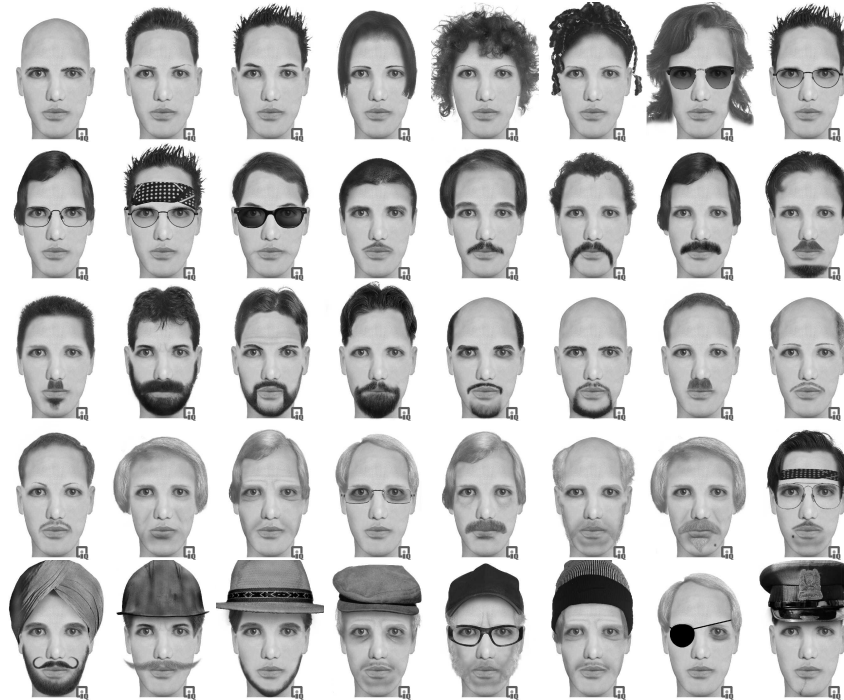


Figure 2.17: Images from the synthetic face disguise database showing 40 variations of the same face [97].

Table 2.1: Details of the face disguise databases used for evaluation.

| Disguise Categories | Disguise Database - Images per subject | |
|---|---|---|
| | **Real Faces** No. of subjects = 25 Images per subject = 15-25 Total no. of images = 475 | **Synthetic Faces** No. of subjects = 100 Images per subject = 40 Total no. of images = 4000 |
| Minimal variations | 4-5 | 4 |
| Hair style | 1-3 | 4 |
| Beard and mustache | 1-2 | 4 |
| Glasses | 1-2 | 4 |
| Cap and hat | 1-3 | 2 |
| Lips, eyebrow and nose | 1-2 | 4 |
| Aging and wrinkles | 1-3 | 4 |
| Multiple variations | 4-6 | 14 |

by collecting frontal face images of actors and actresses from movie clips. This database contains 3500 face images pertaining to 100 individuals, i.e., 35 images per individual. These images contain variations in expression, illumination, age, and disguise along with minor variations in pose. This is a diverse disguise database containing a number of real world non-ideal images.

## 2.6.2 Face Recognition Algorithms used for Comparison

To compare the performance of the proposed algorithm, we selected five existing face recognition algorithms namely Principal Component Analysis [2], Half Face [80], Eigen-eyes based algorithm [91], and Local Binary Pattern [1]. We also evaluate the performance of 2D log polar Gabor transform with neural network and when subjected to only $128 \times 128$ face image without granulation, and it is referred to as *2DLPG-NN*.

## 2.7    Experimental Results

To evaluate the performance of the proposed granular approach for face verification, the image of all five databases are first detected manually by locating the eyes and mouth coordinates. The size of detected face images is normalized to $128 \times 128$. This section is divided into four parts: (1) experimental protocol, (2) performance evaluation of face recognition algorithms, (3) effect of aging on verification performance, (4) effect of different types of disguise on verification performance.

### 2.7.1    Experimental Protocol

In most of the real world applications, a face verification system is first trained on a training database and then the trained system is used to perform recognition on the gallery-probe face database. It is very likely that there is no overlap between the subjects used in the training database and the subjects in the gallery-probe database. To evaluate the performance in such an application scenario, the databases are partitioned into two groups: training database and testing database. From each of the face databases, we select some subjects for training and the remaining images for testing. Face images pertaining to 282 subjects (40% from each database) are used to train the proposed algorithm (estimating densities, learning SVM classifier, computing granulation parameters, learning feature extraction algorithm, and computing verification priors). The remaining images pertaining to 423 subjects (60% from each database) are used as the test (gallery-probe) database for performance evaluation. This partition ensures that the verification is performed on unseen images. The train-test partitioning is repeated fifty times and the Receiver Operating Characteristics curves are generated by computing the false rejection rates over these trials at different false accept rate. The verification accuracy is computed at 0.01% FAR. The above experimental protocol is further used to compare the performance with other face recognition algorithms.

### 2.7.2 Performance Evaluation of Face Recognition Algorithms

The verification performance of the proposed granular face recognition algorithm is evaluated and a comparison is performed with existing face recognition algorithms using five face databases. The results are summarized in Table 2.2 and the ROC plots for each of the five database are shown in Figures 2.18 - 2.22.

On the Notre Dame face database [37], existing face recognition algorithms yield verification accuracy ranging from 49.1% to 93.4%. The lower performances of PCA, Half face and Eigen-eyes are due to the variations in expression and illumination present in the Notre Dame face database. Since the experiments are performed with single gallery per subject, these algorithms are not able to obtain an individualistic representation for every subject which is required to match face images with different variations. LBP and 2DLPG-NN can verify an individual if the variations are minor or moderate. However, the performance of these two algorithms suffer with multiple variations such as matching an indoor face image with an outdoor face image that have variations in facial expression.

The proposed face recognition algorithm outperforms existing algorithms by at least 5.7% and gives an accuracy of 99.1%. The proposed algorithm provides such a high level of performance because the non-disjoint texture features extracted at different levels of granularity provide local and global information at multiple resolutions. Also, the face granulation step incorporates cognitive and psychological research findings and feature extraction uses 2D log polar Gabor transform which emulates the visual cortex [25]. Finally, granular information fusion is performed using the proposed algorithm that integrates likelihood test statistics in a SVM fusion framework that makes it resilient to sensor noise and environmental dynamics.

The other four databases on the other hand contain comprehensive disguised face images and images with aging variations. Using these databases, the verification accuracy of the proposed face recognition algorithm is in the range of 81.5 - 95.7% which is at least 7.8% better than existing algorithms. Existing algorithms perform better on synthetic face disguise database compared to other databases because it contains minor variations

Table 2.2: Comparing performance of the proposed algorithm with existing face recognition algorithms.

| Database | Verification Accuracy (%) at 0.01% FAR | | | | | |
|---|---|---|---|---|---|---|
| | PCA [2] | Half Face [80] | Eigen-eyes [91] | LBP [1] | 2DLPG -NN | Proposed Algorithm |
| Notre Dame [37] | 49.1 | 61.3 | 62.8 | 81.2 | 93.4 | 99.1 |
| Aging [37] | 19.7 | 29.1 | 27.3 | 55.9 | 64.0 | 81.5 |
| Synthetic Face Disguise | 67.2 | 70.0 | 73.7 | 86.1 | 87.9 | 95.7 |
| Real Face Disguise | 19.1 | 21.6 | 17.8 | 57.5 | 74.8 | 90.0 |
| Movie Face | 30.9 | 33.2 | 43.9 | 55.7 | 72.9 | 86.4 |

in pose, expression, and illumination along with disguise and aging. The absence of all other variations in the synthetic face disguise database leads to improved verification performance. Using one gallery image per subject, the proposed face recognition algorithm shows a significantly high level of performance on the aging and disguise databases that contain images with minor variations in pose, expression, and illumination. Finally, the results of $t$-test show that the proposed granular approach is significantly different than existing face verification algorithms.

### 2.7.3   Effect of Aging on Verification Performance

The performance of the proposed algorithm is evaluated for aging variations using the face aging database. The face database is divided into three age groups: (1) 1-18 years, (2) 19-40 years, and (3) beyond 40 years. The database is divided into these three age groups because we observed that face development depends on the age of a person. For example, development in muscles and bone structure cause significant changes in the face during the age of 1-18 years. From 19-40 years, the growth rate is comparatively lower, whereas after 40 years, wrinkles and skin loosening cause major change in facial features and appearance. It is thus very difficult to accurately model an individual's face for large age variations such as from age 10 to age 60. Considering these factors, we
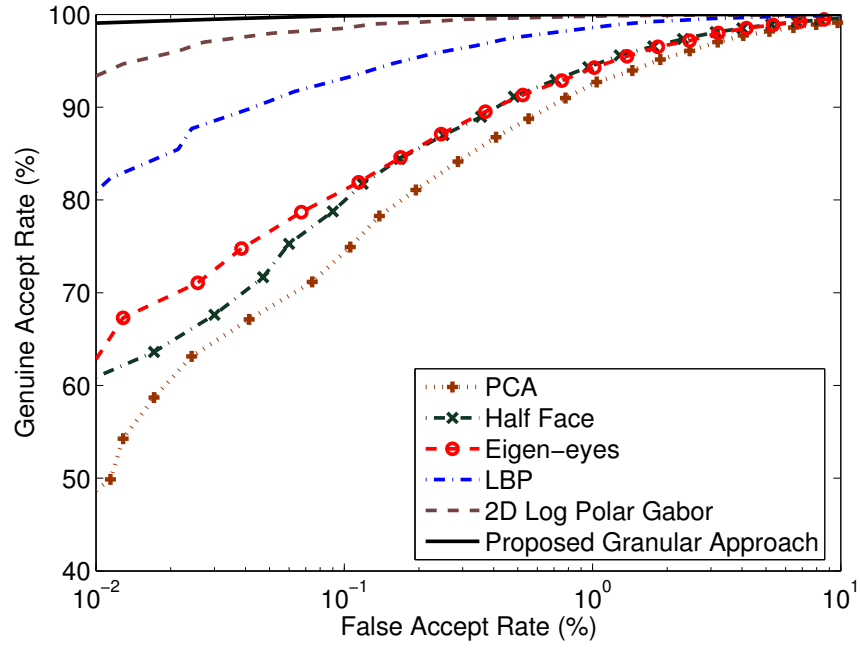
Figure 2.18: ROC plot on the Notre Dame face database [37].
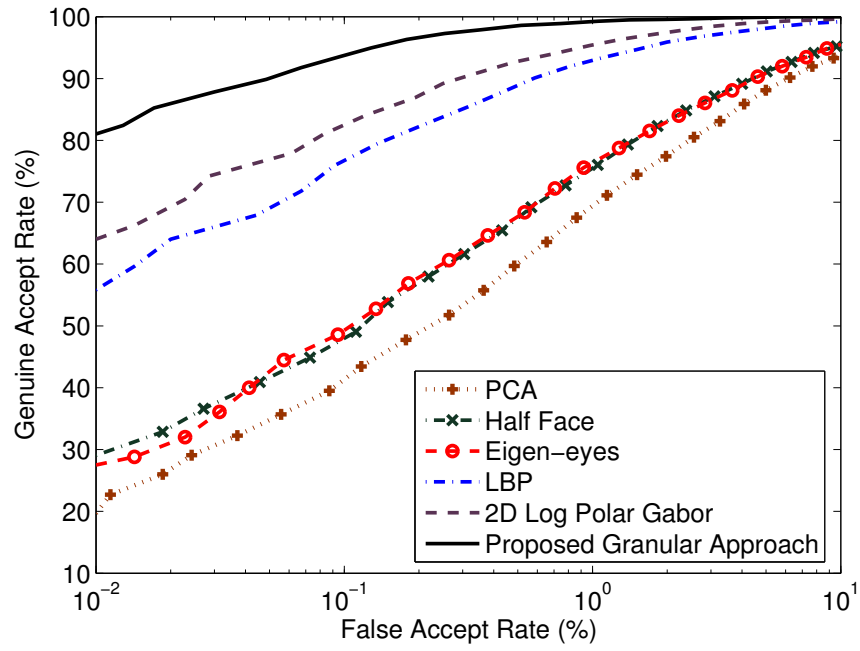


Figure 2.19: ROC plot on the aging database.
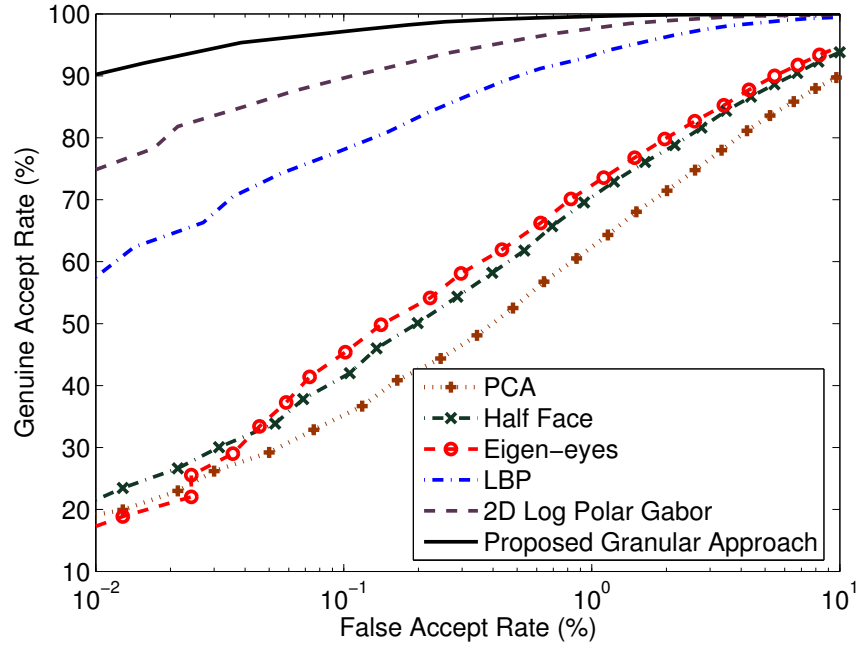
Figure 2.20: ROC plot on the real face disguise database.
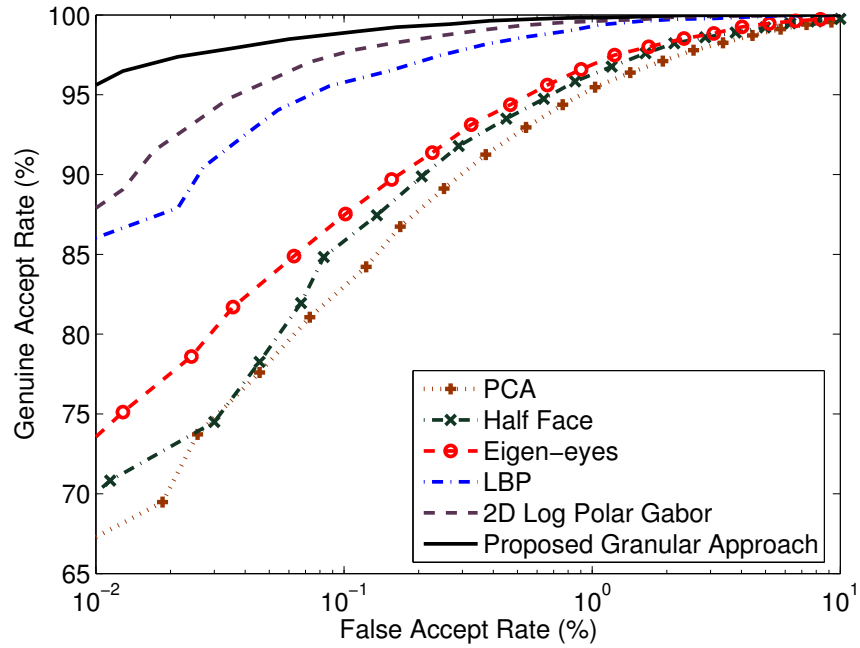


Figure 2.21: ROC plot on the synthetic face disguise database.
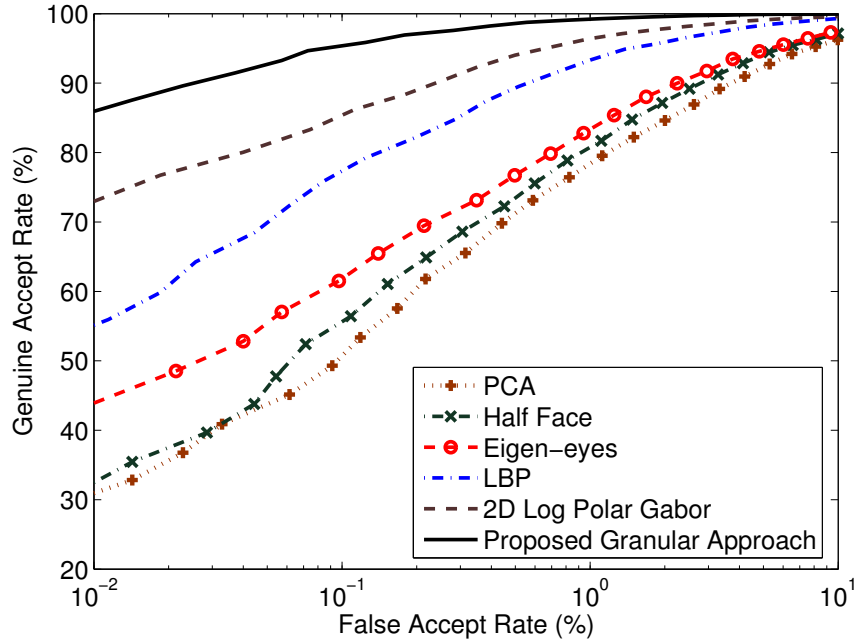
Figure 2.22: ROC plot on the movie database.

divide the database into three age groups to evaluate the performance. Details of the images in these three age groups are provided in Table 2.3.

The experiments show that for the age group of 1-18 years, the verification accuracy is 65.2% whereas with other age groups it is greater than 76% (Table 2.4). The results also suggest that the proposed algorithm yields better performance for 19-40 years, and beyond 40 years age groups compared to 1-18 years age group. This is because there is greater change in facial features during the 1-18 years age group than any other age

Table 2.3: Details of the face aging database.

| Age group | Number of subjects | Total number of images | Total number of images per subject | Average age difference in years |
|---|---|---|---|---|
| 1-18 years | 59 | 605 | 5-16 | 8 |
| 19-40 years | 68 | 735 | 8-15 | 10 |
| Beyond 40 years | 34 | 238 | 4-10 | 5 |

Table 2.4: Verification results of the proposed age transformation algorithm. Verification performance is computed at 0.01% FAR.

| Age group | Verification accuracy (%) | | |
|---|---|---|---|
| | Without registration based age transformation | With registration based age transformation | Improvement in verification accuracy |
| 1-18 years | 23.4 | 65.2 | 41.8 |
| 19-40 years | 51.7 | 76.8 | 25.1 |
| Beyond 40 years | 73.1 | 87.5 | 14.4 |

group. For example, if the age difference between the gallery and probe image is two years, the rate of growth during 1-18 years is faster than 19-40 years or beyond 40 years. Therefore, we can assert that due to feature stabilization, other age groups provide better verification performance. The results show that the age group of beyond 40 years yields the best verification accuracy of 87.5%. We also evaluate the performance of the proposed face recognition algorithm without applying the age transformation algorithm. Table 2.4 summarizes the verification performance of the proposed face recognition algorithm with and without the proposed age transformation algorithm for the three age groups. For the age group of 1-18 years, an improvement of 41.8% is observed with the use of the proposed age transformation algorithm. Since the transformation algorithm minimizes the variations by registering the gallery and probe images both locally and globally, the performance of face recognition is greatly enhanced. Similarly, for other age groups, the performance of face recognition improves by at least 14.4%.

## 2.7.4 Effect of Different Types of Disguise on Verification Performance

Currently, many security applications use human observers to recognize the face of an individual. In some applications, face recognition systems are used in conjunction with limited human intervention. For autonomous operation, it is highly desirable that the face

recognition systems be able to provide high reliability and accuracy under multifarious conditions, including disguise. In this section, we analyze the performance of the proposed and existing face recognition algorithms for each disguise category on the real, synthetic, and movie face databases.

Table 2.5 summarizes the performance of the face recognition algorithms for different categories of disguise variations. For most of the variations, appearance based algorithms [2], [80], [91] yield lower verification accuracy because these algorithms use facial appearance to determine the identity, and the make up tools and accessories significantly alter the facial appearance. On the other hand, texture based algorithms such as the proposed algorithm and LBP [1] provide significantly better verification accuracy compared to appearance based algorithms. Conversely, these algorithms do not yield good verification accuracy with moderate to large disguise variations.

The proposed face recognition algorithm emulates the problem solving approach of human mind and the results show that by fusing the granular information obtained at various levels of granularity, the proposed algorithm yields high verification accuracy. However, the performance of the proposed algorithm reduces for variations in beard, mustache, and multiple disguise variations. On a P-IV 3.2 GHz PC with 2GB RAM under MATLAB programming environment, the proposed algorithm requires around 5 seconds for face granulation, feature extraction, fusion, and decision making.

## 2.8   Summary

Generally, face recognition systems and algorithms are designed to recognize faces of cooperative individuals in controlled environment and a high level of performance has been achieved. However, it becomes a challenging problem when the faces are altered due to aging and disguise. Recognizing faces with altered appearances is a challenging task and is only now beginning to be addressed by researchers. In this chapter, we present a novel face recognition algorithm that extracts facial features at different levels of granularity. The proposed approach is based on the observation that human mind recognizes face

Table 2.5: Comparing the performance of the proposed algorithm with existing algorithms for each disguise category.

| Database | Disguise Categories | Verification Accuracy (%) | | | | | |
|---|---|---|---|---|---|---|---|
| | | PCA [2] | Half Face [80] | Eigen-eyes [91] | LBP [1] | 2DLPG -NN | Proposed Algorithm |
| Synthetic | Minimal | 89.5 | 72.7 | 93.1 | 98.5 | 98.9 | 100 |
| | Hair Style | 85.8 | 89.0 | 88.4 | 95.1 | 95.4 | 99.7 |
| | Beard + Mustache | 43.2 | 45.3 | 92.5 | 83.8 | 84.7 | 96.6 |
| | Glasses | 60.9 | 62.7 | 13.7 | 84.3 | 86.0 | 96.1 |
| | Cap/Hat | 82.7 | 85.1 | 79.2 | 93.7 | 95.1 | 99.9 |
| | Lips/Eyebrows/Nose | 86.8 | 89.3 | 79.7 | 96.9 | 97.2 | 99.4 |
| | Aging/Wrinkles | 75.6 | 77.9 | 73.1 | 94.5 | 95.6 | 99.3 |
| | Multiple Disguise | 18.7 | 21.1 | 50.3 | 63.4 | 72.9 | 89.5 |
| Real | Minimal | 31.3 | 35.5 | 26.7 | 79.1 | 96.2 | 100 |
| | Hair Style | 30.6 | 33.4 | 25.2 | 75.3 | 87.7 | 99.9 |
| | Beard + Mustache | 21.0 | 23.7 | 26.9 | 37.4 | 72.3 | 92.8 |
| | Glasses | 22.4 | 24.1 | 0.0 | 50.7 | 77.9 | 96.5 |
| | Cap/Hat | 27.1 | 28.3 | 10.6 | 72.9 | 78.2 | 99.8 |
| | Lips/Eyebrows/Nose | 28.3 | 29.2 | 19.8 | 75.6 | 81.7 | 98.3 |
| | Aging/Wrinkles | 27.2 | 28.9 | 19.5 | 66.3 | 81.1 | 95.2 |
| | Multiple Disguise | 10.3 | 12.4 | 11.9 | 29.1 | 58.3 | 81.4 |
| Movie | Minimal | 40.7 | 43.3 | 55.2 | 60.4 | 90.2 | 98.6 |
| | Hair Style | 39.1 | 41.6 | 53.8 | 76.4 | 88.7 | 98.3 |
| | Beard + Mustache | 31.3 | 34.8 | 54.9 | 37.1 | 72.5 | 87.9 |
| | Glasses | 32.0 | 35.2 | 3.7 | 48.9 | 79.4 | 93.0 |
| | Cap/Hat | 35.4 | 37.1 | 32.3 | 69.6 | 83.3 | 97.6 |
| | Lips/Eyebrows/Nose | 34.9 | 38.0 | 31.5 | 77.2 | 82.7 | 97.9 |
| | Aging/Wrinkles | 28.5 | 29.9 | 33.0 | 65.1 | 83.2 | 97.4 |
| | Multiple Disguise | 13.8 | 16.1 | 21.5 | 30.7 | 58.9 | 73.7 |

image by analyzing the relation among different non-disjoint spatial features extracted at different granular levels. These non-disjoint features are extracted using neural network architecture based 2D log polar Gabor transform. Variations in aging and disguise cause alteration in appearance and features which may lead to imprecise matching. So, to fuse the imprecise information obtained from these features, we propose the use of likelihood ratio based SVM match score fusion. The proposed granular face recognition algorithm is validated using the Notre Dame face database, FG-Net face database and three other databases specially prepared with disguised face images. The performance is also compared with existing state of the art face recognition algorithms and algorithms specifically designed to recognize disguised face images. Experimental results show that the proposed algorithm outperforms existing algorithms by at least 5.7%.

# Chapter 3

# Multispectral Face Image Fusion for Reducing the Effect of Illumination Variations

Face recognition algorithms generally use visible spectrum images for recognition because they provide clear representation of facial features and face texture to differentiate between two individuals. However, visible spectrum images also possess several other properties which affect the performance of recognition algorithms. For example, changes in lighting affect the representation of visible spectrum images and can influence feature extraction. Other variations in face images such as facial hairs, wrinkles, and expression are also evident in visible spectrum images and these variations increase the false rejection rate of face recognition algorithms. To address the challenges posed by visible spectrum images, researchers have used infrared images for face recognition [23], [51], [52], [77]. Among all the infrared spectrum images, long wave infrared images possess several properties that are complementary to visible images. Long wave infrared or thermal images are captured in the range of 8-12 $\mu m$. These images represent the heat pattern of the object and are invariant to illumination and expression. Face images captured in long wave infrared spectrum have less intra-class variation and help to reduce the false rejection rate of recognition algorithms. These properties of long wave infrared and visible images can be combined to improve the performance of face recognition algorithms.

In literature, researchers have compared the performance of visible and thermal face recognition using several face recognition algorithms. These results show that for variation in expression and illumination, thermal images provide better recognition performance compared to visible images [23], [104], [121]. Further, several fusion algorithms have been proposed to fuse the information extracted from visible and LWIR face images at image level [41], [43], [94], feature level [94], [95], match score level [95], and decision level [95]. Information fusion of multispectral images provides better performance compared to either visible or infrared spectrum images. However, research in multispectral information fusion is relatively new and intelligent techniques such as Granular computing, Support Vector Machine, and theory of evidence have not been explored. These intelligent techniques can enhance the recognition performance by providing better generalization capabilities to handle imprecise information.

In multispectral image fusion, images of two different spectrum are combined to generate a fused image followed by feature extraction and matching using the fused image. In this chapter, we have proposed two image fusion algorithms to fuse long wave infrared and visible face images. The first algorithm is a simple approach in which Discrete Wavelet Transform is used for image fusion. The second algorithm uses a learning technique for image fusion. Specifically, the learning based image fusion algorithm utilizes $2\nu$-Granular Support Vector Machine to learn the properties of the multispectral face images at different resolution. It further determines optimal information in both the images and combines them to generate a fused image. 2D log polar Gabor and local binary pattern [1] face recognition algorithms are used to extract global and local features from the fused image for matching.

## 3.1   Multispectral Face Image Fusion

Visible images provide the reflectance property and long wave infrared images provide the thermal property. In multispectral image fusion, we combine these properties to generate a fused image which possess both the properties and can be used to improve

the recognition performance. Although there are several multispectral face image fusion algorithms in literature, they have some limitations which affect the face recognition performance. Genetic algorithm based fusion proposed by Bebis *et al.* [9] suffers from making a good choice of fitness function. Fusion algorithm proposed by Kong *et al.* [52] suffers from the empirical constant weights which are assigned to the wavelet coefficients of visible and LWIR images. In this chapter, we propose two multispectral fusion algorithms. Input to both the algorithms are registered multispectral face images. Therefore, we present the mutual information based multispectral image registration algorithm followed by the proposed image fusion algorithms.

### 3.1.1   Mutual Information based Multispectral Face Image Registration

Visible and infrared images captured at different time instances can have variations due to camera angle, expression, and geometric deformations. To optimally fuse two multispectral images, we first need to minimize the linear and non-linear differences between the two images. In this research, we propose the use of mutual information based registration algorithm for registering visible and thermal face images. Similar to the age transformation algorithm described in Section 2.1, Gaussian pyramid approach registers the visible spectrum face image with respect to the infrared spectrum face image. Figure. 3.1 shows an example of multispectral image registration.

### 3.1.2   Multispectral Face Image Fusion using DWT

In this section, we propose an algorithm to fuse the visible and LWIR face image using DWT with mother wavelet *Daubechies 9/7* [4]. This mother wavelet is chosen because it is among the best filters for wavelet based operations such as image compression when operating in a distributed environment [117]. Let $V$ and $I$ be the registered visible and IR face images. First, the pixel values of $V$ and $I$ are transformed in the range of $[0, 1]$. Single level DWT is then applied on these images to obtain the detail and

Visible
Face Image

LWIR
Face Image

Registered Visible
Face Image
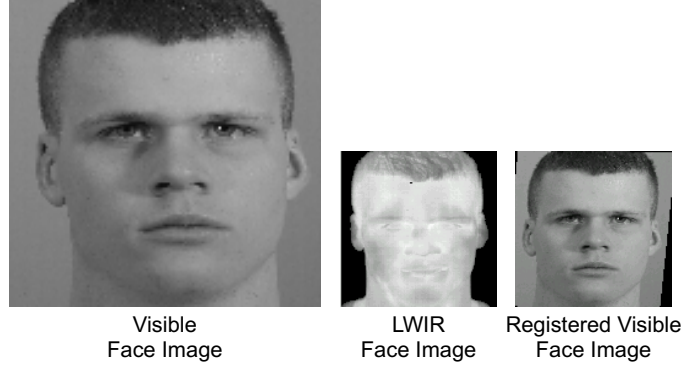
Figure 3.1: Example of visible and infrared image registration on the Notre Dame face database [37]. Visible image is registered with respect to the LWIR image. Size of the detected visible face image is $855 \times 1024$ and infrared face image is $115 \times 156$.

approximation wavelet bands for both the images. Let $V_{LL}$, $V_{LH}$, $V_{HL}$ and $V_{HH}$ be the four bands from visible face image and $I_{LL}$, $I_{LH}$, $I_{HL}$ and $I_{HH}$ be the corresponding bands from IR face image as shown in Figure 3.2. To preserve the features from both the images, coefficients from approximation band of $V$ and $I$ are averaged,

$$F_{LL} = mean(V_{LL}, I_{LL}) \qquad (3.1)$$

where $F_{LL}$ is the approximation band of the fused image. For the three detailed bands, each band is divided into windows of size $3 \times 3$ and the sum of absolute value of all the pixels in each window is calculated. Binary decision maps, $DM$, are generated for all the three detail bands using Equations 3.2, 3.3, and 3.4. A value 1 is assigned if the window from visible image has a value greater than the corresponding window from the IR image; otherwise a value 0 is assigned.

$$DM_{LH} = \begin{cases} 1 & if & max_{3\times3}V_{LH} > max_{3\times3}I_{LH} \\ 0 & if & max_{3\times3}V_{LH} < max_{3\times3}I_{LH} \end{cases} \qquad (3.2)$$

$$DM_{HL} = \begin{cases} 1 & if & max_{3\times3}V_{HL} > max_{3\times3}I_{HL} \\ 0 & if & max_{3\times3}V_{HL} < max_{3\times3}I_{HL} \end{cases} \qquad (3.3)$$

56
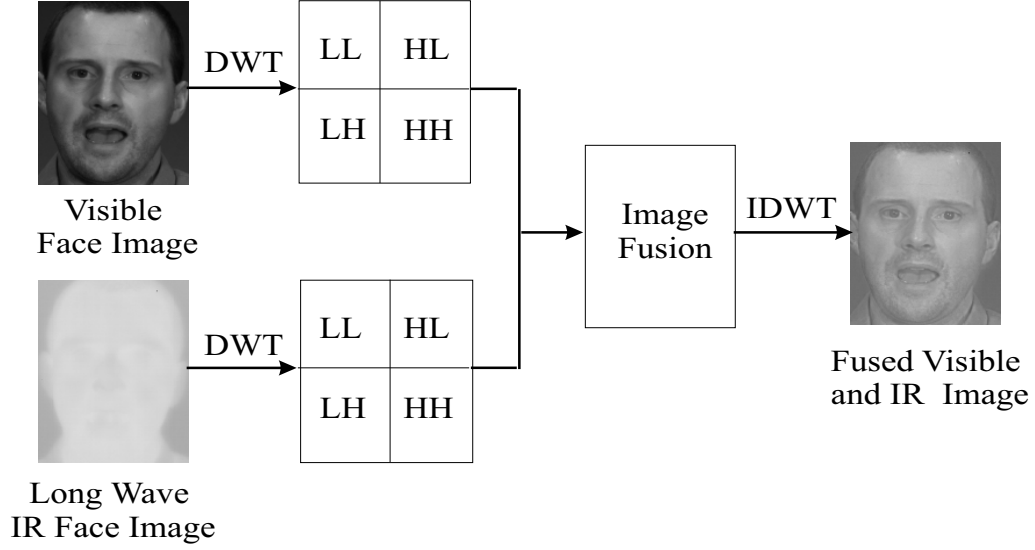
Figure 3.2: Multispectral image level fusion of visible and medium-wave infrared face images

$$DM_{HH} = \begin{cases} 1 & if & max_{3\times3}V_{HH} > max_{3\times3}I_{HH} \\ 0 & if & max_{3\times3}V_{HH} < max_{3\times3}I_{HH} \end{cases} \qquad (3.4)$$

where $DM_{LH}$, $DM_{HL}$ and $DM_{HH}$ are the binary decision maps for three detail bands, shown in Figure 3.3. Based on these decision maps, the coefficients are selected from $V$ and $I$, and the detailed bands for the fused image $F_{LH}$, $F_{HH}$, and $F_{HH}$ are generated. Inverse DWT is then applied on the four fused bands to generate the fused image.

$$F = IDWT(F_{LL}, F_{LH}, F_{HL}, F_{LH}) \qquad (3.5)$$

### 3.1.3   2$\nu$-GSVM Image Fusion Algorithm

In real world applications, weights should be dynamically and locally assigned for optimal multispectral information fusion. In this section, we propose the multispectral face image fusion algorithm which dynamically and locally computes the weights for

Figure 3.3: Decision maps for the horizontal, vertical and diagonal bands

fusion using $2\nu$-GSVM. Figure 3.4 illustrates the steps involved in the proposed image fusion algorithm.

Based on the experimental results of Chapter 2, we observed that $2\nu$-SVM provides better classification accuracy compared to classical SVM and is computationally more efficient. Recently, Tang *et al.* [105] - [108], applied the concept of granular computing [7], [8], [21], [126] to SVM and proposed Granular SVM which is more adaptive to the data distribution in comparison to SVM. Tang *et al.* have also shown that for several classification applications, GSVM outperforms SVM both in terms of classification accuracy and computational time. In this chapter, we extend the formulation to $2\nu$-Granular Support Vector Machine which embodies the properties of both GSVM and $2\nu$-SVM. Using this concept, the formulation of $2\nu$-SVM is extended to $2\nu$-GSVM.

Let the complete feature space be divided into $k$ subspaces with one $2\nu$-SVM operating on each subspace. The $i^{th}$ $2\nu$-SVM is represented by $2\nu$SVM$_i$ where $i = 1, 2, ..., k$. From each subspace, we obtain the corresponding $L_i$ using Equation 2.31. We then compute the compound margin width $W$ by using all the $L_i$ values.

$$W = \left| \sum_{i=1}^{k} \frac{t_i}{t} (2\nu SVM_i :\rightarrow L_i) - L_0 \right| \tag{3.6}$$

where $t_i$ is the number of training data in the $i^{th}$ subspace and $t = \sum_{i=1}^{k} t_i$. $2\nu$SVM$_i :\rightarrow L_i$ represents the SVM operating on the $i^{th}$ subspace. $2\nu$-SVM learning yields $L_i$ at local level, and $L_0$ is obtained by learning another $2\nu$-SVM on the complete feature space at global level. This equation provides the margin width associated to a hyperplane. How-
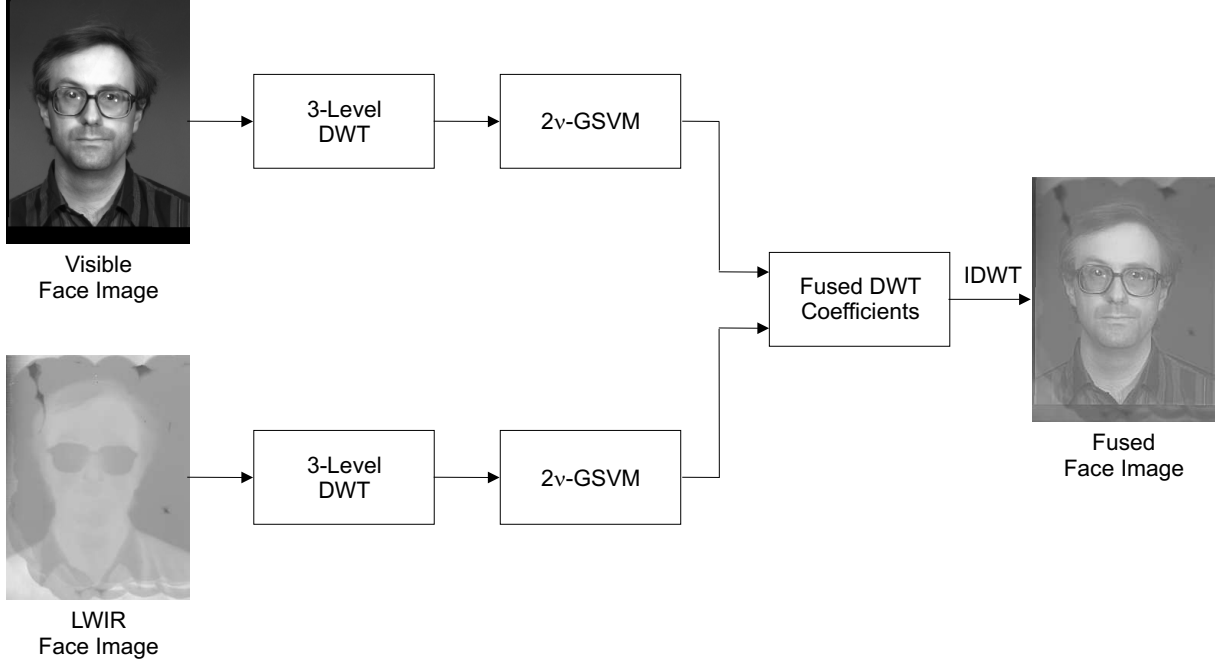
Figure 3.4: Schematic diagram of the proposed $2\nu$-GSVM image fusion algorithm.

ever, there are different ways to divide the feature space and hence different hyperplanes associated with each of the granule generation method can be obtained. We compute the classification accuracy of all the hyperplanes on the training data and then select the hyperplane that optimally classifies the training data. In contrast to a single SVM that deals with large parameter space and results in large training time, $2\nu$-Granular SVM uses multiple SVMs to learn both the local and global properties of the training data at different granularity levels. This $2\nu$-GSVM is then used for fusing multispectral face images.

In the proposed face image fusion algorithm, registered visible and infrared face images are fused using $2\nu$-GSVM and Discrete Wavelet Transform [4]. The fusion algorithm uses the activity level, $a$, of face images which is defined as,

$$a = \sqrt{\frac{1}{XY}\left[\sum_{i=0}^{X-1}\sum_{j=1}^{Y-1}\{V(i,j)-V(i,j-1)\}^2 + \sum_{j=0}^{Y-1}\sum_{i=1}^{X-1}\{V(i,j)-V(i-1,j)\}^2\right]} \quad (3.7)$$

59

where $V$ is the visible face image, and $X$ and $Y$ are the rows and columns of the face image respectively. The proposed learning based fusion algorithm is divided into two parts: (1) training and (2) classification and fusion.

**Training $2\nu$-GSVM:** We learn the $2\nu$-GSVM for image fusion by using the activity levels of labeled visible and infrared training face images. The training algorithm is described as follows:

<u>Step 1</u>: Visible and infrared training face images are decomposed using DWT to obtain 3-level approximation, horizontal, vertical, and diagonal subbands.

<u>Step 2</u>: Let $V_{LL_j}$, $V_{LH_j}$, $V_{HL_j}$, and $V_{HH_j}$ be the subbands of visible face image where $j = 1, 2, 3$ represents the decomposition levels. Similarly, let $I_{LL_j}$, $I_{LH_j}$, $I_{HL_j}$, and $I_{HH_j}$ be the subbands of infrared face image corresponding to each decomposition level, $j$. Each subband of both visible and infrared face images is divided into windows of size $8 \times 8$ and the activity level of each window is computed using Equation 3.7.

<u>Step 3</u>: The activity levels of all labeled training face images are used as input to $2\nu$-GSVM. In training, two $2\nu$-GSVMs are learned, one for visible face images and another for infrared face images.

<u>Step 4</u>: $2\nu$-GSVM trained for visible images classifies the activity levels of visible spectrum face images as *Good* or $+1$ and *Bad* or $-1$. Similarly, $2\nu$-GSVM trained for infrared face images classifies the activity levels of infrared face images into *Good* or *Bad* class.

**Classification and Fusion:** We classify the properties of visible and infrared face images using trained $2\nu$-GSVMs. This classification is used to dynamically compute the weights of visible and infrared face images in multispectral image fusion.

<u>Step 1</u>: Visible and infrared face images of an individual are provided as input. Similar to Steps 1 and 2 of the training algorithm, both the input face images are decomposed into 3-level DWT and activity levels of $8 \times 8$ windows are computed. Let $a_V$ and $a_I$ be the activity levels computed from visible and infrared face images respectively.

<u>Step 2</u>: $2\nu$-GSVM classifier is used to classify the activity levels of different subbands of visible face images as *Good* or *Bad*. A binary decision matrix, $d_V$, is generated which

contains value 1 if the activity level is *Good* and 0 if the activity level is *Bad*.

Step 3: Similar to Step 2, activity levels of infrared face image are classified and a binary decision matrix, $d_I$, is generated.

Step 4: Weight matrices $\omega_V$ and $\omega_I$ are computed using binary decision matrices ($d_V$ and $d_I$) and the following three conditions:

1. If $d_V(i) = d_I(i) = 1$, then $\omega_V(i) = \omega_I(i) = 0.5$.

2. If $d_V(i) = 1$ and $d_I(i) = 0$, then $\omega_V(i) > \omega_I(i)$ and

$$\omega_V(i) = \frac{|a_V(i) + 2a_I(i) - a_{I_{median}}|}{a_V(i) + a_I(i)} \tag{3.8}$$

$$\omega_I(i) = \frac{|a_{I_{median}} - a_I(i)|}{a_V(i) + a_I(i)} \tag{3.9}$$

3. If $d_V(i) = 0$ and $d_I(i) = 1$, then $\omega_V(i) < \omega_I(i)$ and

$$\omega_V(i) = \frac{|a_{V_{median}} - a_V(i)|}{a_V(i) + a_I(i)} \tag{3.10}$$

$$\omega_I(i) = \frac{|a_I(i) + 2a_V(i) - a_{V_{median}}|}{a_V(i) + a_I(i)} \tag{3.11}$$

where, $i$ is the window count, and $a_{V_{median}}$ and $a_{I_{median}}$ are the median values of $a_V$ and $a_I$ matrices respectively. Further, in all three cases, $\omega_V(i) + \omega_i(i) = 1$.

In condition 1, the activity levels of both visible and infrared image windows are classified as *Good* and hence equal weights are assigned. Condition 2 states that if the activity level of window corresponding to visible face image is classified as *Good* and the activity level of the window corresponding to infrared face image is classified as *Bad*, then higher weight is assigned to the visible face image window. In condition 3, higher weight is assigned to the infrared face image window because 2$\nu$-GSVM classifies the activity

Registered Visible        LWIR        Fused
Face Image       Face Image      Face Image
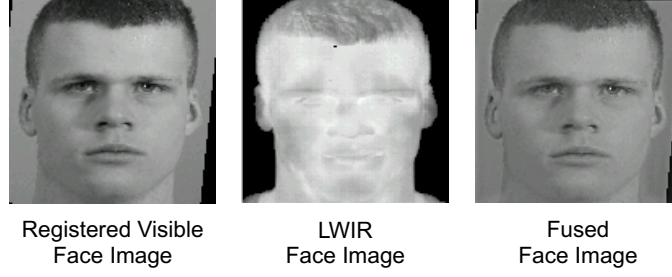
Figure 3.5: Sample result of the proposed $2\nu$-GSVM learning based image fusion algorithm.

level of visible face image window as *Bad* and the activity level of infrared face image window as *Good*.

Step 5: Visible and infrared face images are then fused using Equation 3.12.

$$F_{j_{8\times8}}(i) = \omega_V(i)V_{j_{8\times8}}(i) + \omega_I(i)I_{j_{8\times8}}(i) \tag{3.12}$$

where, $F_j$ is the fused subband, $j$ represents the approximate, vertical, horizontal, and diagonal subbands, subscript $8 \times 8$ denotes that the fusion is performed at window level of size $8 \times 8$, and $i$ represents the window count.

Step 6: Finally, inverse DWT is applied on fused subbands to generate the fused multispectral face image, $F_{SVM}$. Figure 3.5 shows an example of visible, infrared, and fused face images of an individual.

## 3.2 Databases used for Validation

To validate the proposed fusion algorithms, we used the Notre Dame [22], [37] and Equinox/NIST [32] multispectral face databases. In addition, we used 2D log polar Gabor transform and local binary pattern algorithms [1] for face verification. For comparing the performance of the proposed algorithms, we used DWT based existing image fusion algorithm to fuse visible and infrared spectrum face images [52]. In this section, we briefly describe the details of the databases used in our experiments.

- **Notre Dame Face Database:** Notre Dame face database [22], [37] contains

Table 3.1: Number of visible and infrared image pairs in the training, gallery, and probe databases.

| Face database | Number of visible and infrared image pairs in | | |
|---|---|---|---|
| | Training database | Gallery database | Probe database |
| Notre Dame | 477 | 159 | 1815 |
| Equinox | 285 | 95 | 18715 |

LWIR and visible images from 159 classes with variations in expression, lighting, and time lapse. We have chosen three visible and LWIR face images with neutral expression for the training database, one neutral visible and LWIR image for the gallery database, and the remaining images comprise the probe dataset. Table 3.1 shows the details of training, gallery, and probe images used in the experiments.

- **Equinox Face Database:** Equinox face database [32] contains long wave infrared, medium wave infrared, short wave infrared, and visible face images pertaining to 95 individuals. Long wave infrared images are captured at 8-12 $\mu m$, medium wave infrared images at 3-5 $\mu m$, and short wave infrared images at 0.9-1.7 $\mu m$. The images are captured under different illumination conditions and contain variations with expression and glasses. In our research, we have used only long wave infrared images and visible images. The number of LWIR and visible images per class vary from 43 to 516. We chose three visible and LWIR images with neutral expression and without glasses for training, one LWIR and visible image with neutral expression, uniform illumination, and no glasses for gallery, and the remaining images as probe.

## 3.3   Experimental Evaluation of Proposed Algorithms

In this section, we present the experimental results of the proposed image fusion algorithms. $2\nu$-GSVM image fusion algorithm is trained using the training images and

Table 3.2: Verification performance of the proposed 2$\nu$-GSVM and existing image fusion algorithms at 0.01% FAR.

| Face database | Recognition algorithm | Verification accuracy (%) | | | | |
|---|---|---|---|---|---|---|
| | | Visible image | LWIR image | Kong fusion [52] | DWT fusion | 2$\nu$-GSVM fusion |
| Notre Dame | 2D-LPG | 89.36 | 88.09 | 86.74 | 91.88 | 95.85 |
| | LBP | 88.20 | 87.44 | 85.87 | 91.79 | 94.80 |
| Equinox | 2D LPG | 78.91 | 82.75 | 80.83 | 90.06 | 94.98 |
| | LBP | 76.80 | 81.52 | 80.69 | 89.93 | 94.71 |

the Radial Basis Function kernel with RBF parameter as 4. The performance is evaluated in terms of verification accuracy at 0.01% False Accept Rate.

For evaluation, we separately computed the verification accuracies of visible face image and long wave infrared face image using both 2D log polar Gabor and local binary pattern face verification algorithms. The third and fourth columns of Table 3.2 summarize the verification performance of visible face image and long wave infrared face image respectively using both the verification algorithms. These results establish the baseline for evaluating and comparing the performance of fusion algorithms. We then compute the verification accuracies of the proposed DWT and 2$\nu$-GSVM multispectral face image fusion algorithms and the existing image fusion algorithm. The results summarized in Table 3.2 show that the proposed 2$\nu$-GSVM image fusion algorithm outperforms both the existing and DWT fusion algorithms by at least 3.9% for the Notre Dame database and 4.9% for the Equinox database. ROC plots in Figures 3.6 and 3.7 show the results for the Notre Dame and Equinox face databases respectively.

The proposed 2$\nu$-GSVM image fusion algorithm performs correct classification of multispectral face information at different levels of granularity which is subsequently used for computing the dynamic weights of visible and LWIR face images. This granular learning results in better generalization and fusion of high entropy visible and LWIR facial features. Further, as shown in Figure 3.8 and Table 3.3, fused face images generated from
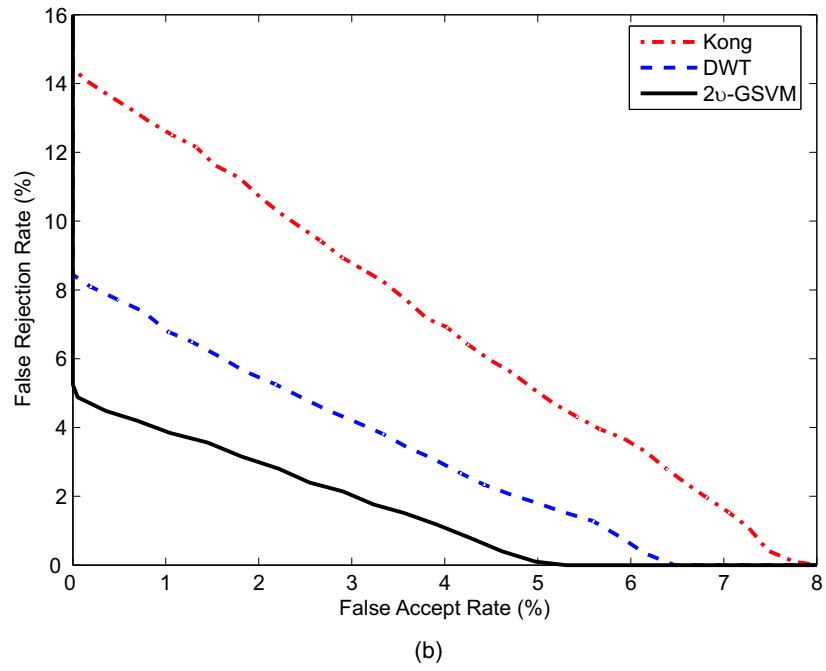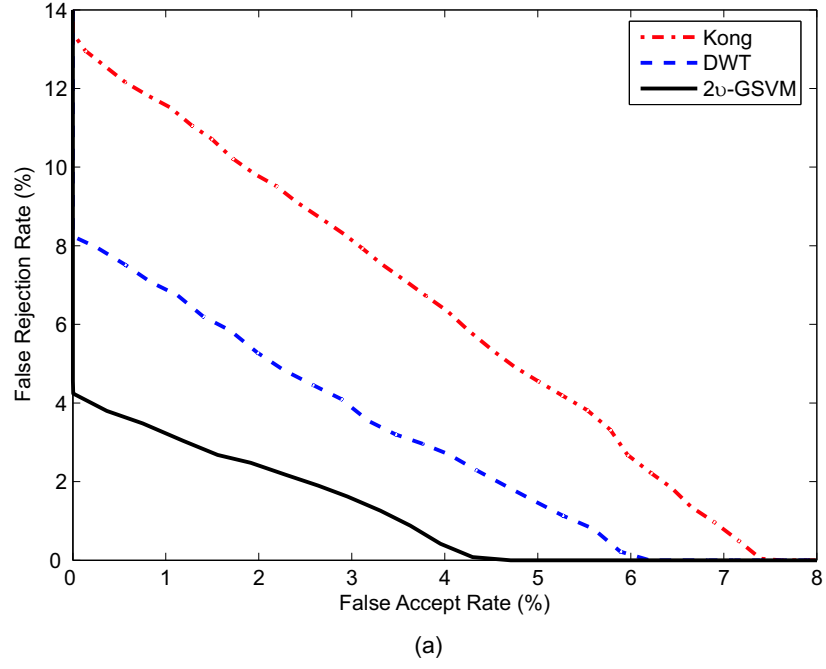
Figure 3.6: ROC plots of the proposed 2$\nu$-GSVM and existing image fusion algorithms on the Notre Dame face database. Results are computed using (a) 2D log polar Gabor (b) Local binary pattern based verification algorithms.
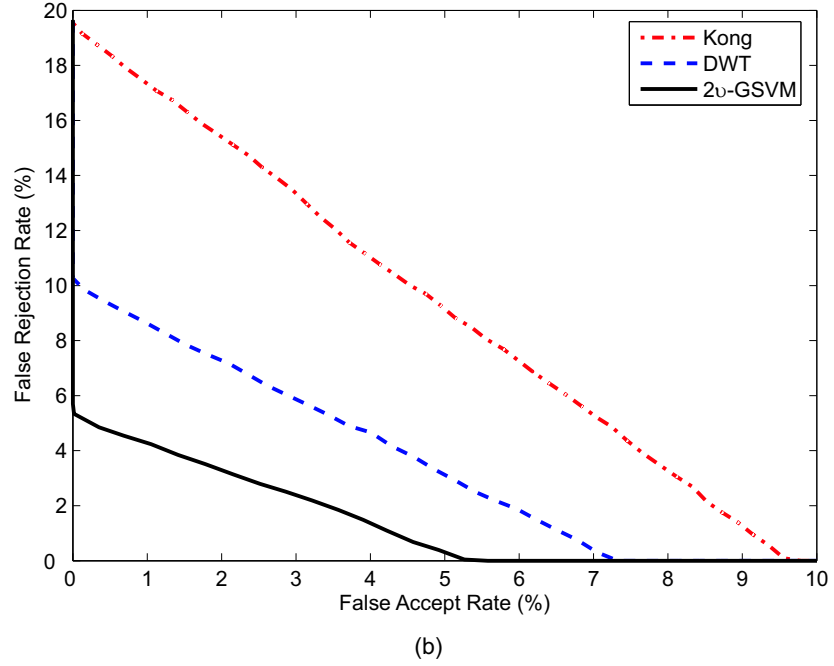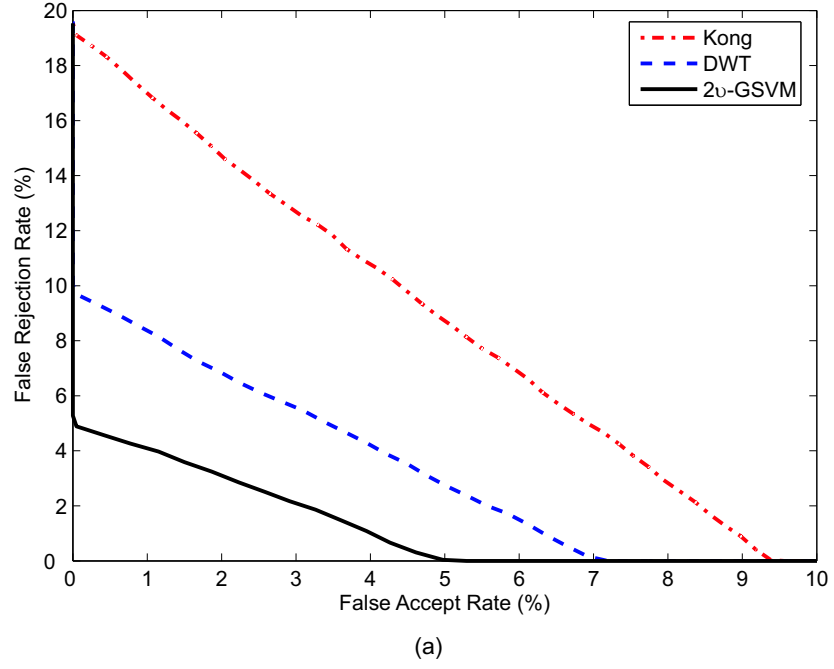
Figure 3.7: ROC plots of the proposed $2\nu$-GSVM and existing image fusion algorithms on the Equinox face database. Results are computed using (a) 2D log polar Gabor (b) Local binary pattern based verification algorithms.

Table 3.3: Performance evaluation of multispectral image fusion algorithms with expression and illumination variations at 0.01% FAR.

| Database | Classifier | Covariate | Verification accuracy (%) | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | | Visible image | LWIR image | Kong fusion [52] | DWT fusion | $2\nu$-GSVM fusion |
| Notre Dame | 2D-LPG | Expression | 88.72 | 87.96 | 85.29 | 91.03 | 95.68 |
| | | Illumination | 89.77 | 88.25 | 87.70 | 92.16 | 95.92 |
| | LBP | Expression | 87.91 | 87.62 | 84.53 | 90.98 | 94.71 |
| | | Illumination | 88.46 | 88.75 | 86.49 | 92.07 | 94.85 |
| Equinox | 2D-LPG | Expression | 78.04 | 82.68 | 78.84 | 89.33 | 94.86 |
| | | Illumination | 79.57 | 82.84 | 81.70 | 90.81 | 95.05 |
| | LBP | Expression | 76.51 | 80.79 | 80.19 | 89.23 | 94.62 |
| | | Illumination | 76.99 | 82.24 | 80.92 | 90.18 | 94.76 |

the $2\nu$-GSVM image fusion algorithm provide higher invariance to illumination compared to the visible images. The fused images also provide more distinguishing information compared to the LWIR face images. These properties of the proposed $2\nu$-GSVM image fusion algorithm leads to improved face verification performance.

## 3.4   Summary

Visible and long wave infrared images provide complementary properties which can be combined to improve the performance of face recognition and address the challenges due to illumination variations. In this chapter, we proposed two image fusion algorithms to fuse information obtained from multispectral face images. We first apply mutual information based registration algorithm to register multispectral face images and then fuse the images using the proposed DWT image fusion algorithm and $2\nu$-Granular Support Vector Machine based image fusion algorithm. The fused image contains the properties of both visible and long wave infrared images and can efficiently be used for face recog-

| Visible<br>Image | Infrared<br>Image | DWT Fusion<br>Algorithm | 2$\nu$-Granular SVM<br>Fusion Algorithm |
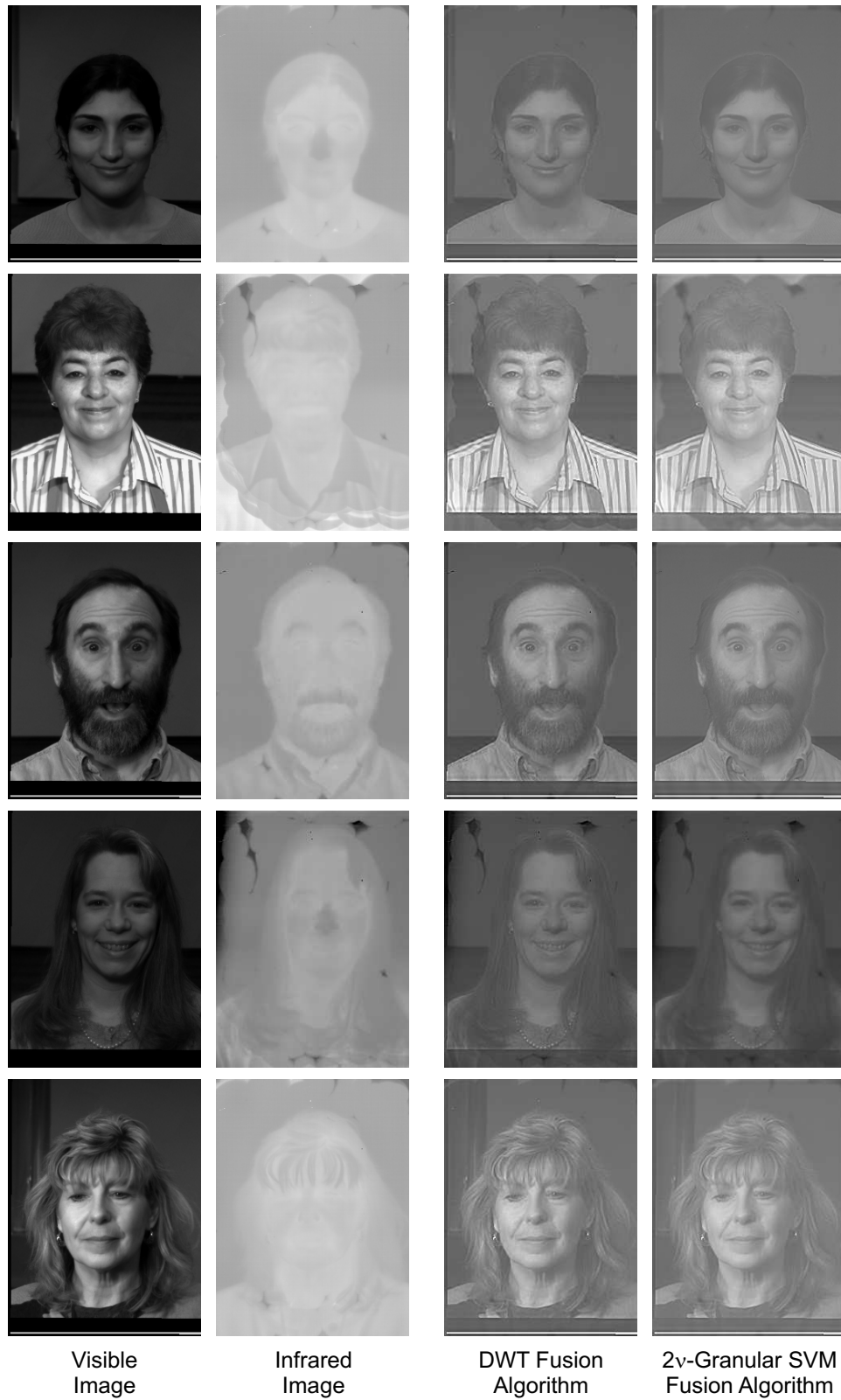
Figure 3.8: Results of the proposed image fusion algorithms on the Equinox face database [32].

nition. The performance of the proposed image fusion algorithms is validated using the Notre Dame and Equinox face databases. Experimental results show that the proposed $2\nu$-GSVM image fusion algorithm yields better performance than the proposed DWT and existing fusion algorithms even with illumination variations.

# Chapter 4

# A Mosaicing Scheme for Pose Invariant Face Recognition

The problem of 2D face recognition continues to pose challenges even after several years of research in this field [129]. State-of-the-art algorithms exhibit various degrees of sensitivity to changes in illumination, pose, facial expressions, accessories, etc. Designing pose-invariant algorithms is challenging as discussed in the Face Recognition Vendor Test 2002 report [73]. Several methods have been suggested to address the issue of pose variations including the use of active appearance models [48], morphable models [12], 3D facial imaging [19], multiple templates [114], and multiclassifier fusion [63], [85]. In this research, we propose an image fusion scheme to generate the 2D face mosaic of an individual during enrollment that can be successfully used to match various poses of a person's face during authentication. Mosaicing uses the frontal and side-profile face images (2D) of a user to generate an extended 2D image. The goal is to adequately characterize an individual's face in a 2D plane without attempting to compute the 3D structure of the face. This avoids the complexity of generating 3D structure information from multiple registered 2D images. Mosaicing also obviates the need to store multiple templates of a user during enrollment thereby optimizing storage demands and processing time.

The potential of mosaicing facial images has not received extensive attention in the literature. Table 4.1 summarizes the face mosaicing techniques proposed by researchers.

Yang *et al.* [125] propose an algorithm to create panoramic face mosaics. Their acquisition system consists of five cameras that simultaneously obtain five different views of a subject's face. In order to determine the corresponding points in multiple face views, the authors place ten colored markers on the face. Based on these control points, their algorithm uses a series of fast linear transformations on component images to generate a face mosaic. Finally, a local smoothing process is carried out to smooth the mosaiced image. Two different schemes are used to represent the panoramic image: one in the spatial domain and the other in the frequency domain. The frequency representation is observed to result in an identification accuracy of 97.46% and the spatial representation in an accuracy of 93.21% on a database of 12 individuals.

Liu and Chen [59] describe a face mosaicing technique that uses a statistical model to represent the mosaic. Given a sequence of face images captured under an orthographic camera model, each frame is unwrapped onto a certain portion of the surface of a sphere via a spherical projection. A minimization procedure using the Levenberg-Marquardt algorithm is employed to optimize the distance between an unwrapped image and the sphere. The statistical representational model comprises of a mean image and a number of eigen-images. The novelty of this technique is (a) the use of spherical projection, as opposed to cylindrical projection, which works better when there is head motion in both the horizontal and vertical directions, and (b) the computation of a representational model using both the mean image and the eigen-images rather than a single template image. Although the authors state that this method can be used for face recognition, no experimental results have been presented in the paper. In [60], the authors propose another algorithm in which the human head is approximated with a 3D ellipsoidal model. The face, at a certain pose, is viewed as a 2D projection of this 3D ellipsoid. All 2D face images of a subject are projected onto this ellipsoid via geometrical mapping to form a texture map which is represented by an array of local patches. Matching is accomplished by adopting a probabilistic model to compute the distance of patches from an input face image. The authors report an identification accuracy of 90% on the CMU PIE database [92].

Table 4.1: A comparison of three existing face mosaicing schemes.

| | Yang *et al.* [125] | Liu & Chen [59] | Singh *et al.* [96] |
|---|---|---|---|
| Subjects | 12 | 68 | 12 + 15 |
| Registration | Affine using points | Affine using triangles | Affine using regions |
| Mosaicing | Concatenation | Geometrical mapping | Multi-resolution splines |
| Representation | Spatial: PCA, Frequency: FFT amplitude | Statistical model | Local binary pattern |

Face mosaicing has also been used in non-biometric applications such as facial animation and rendering [47], and 3D face image generation [13]. However, these algorithms generate the face mosaic using complex models which do not necessarily preserve the biometric features of the face.

The concept of mosaicing may be viewed as an exercise in *information fusion*. When multiple images of a subject's face are available at the time of enrollment, a common approach is to treat these images (also known as gallery images) as independent entities; thus, when a probe (query) image is presented to the system it is compared against each gallery image *independently* and the resulting set of scores consolidated to generate a single score (e.g., via the sum rule) indicating the proximity of the probe image with the subject in the database. This is fusion at the match score level. However, in the case of mosaicing, multiple images of a subject's face are fused into a single entity in the image domain. Hence, this could be viewed as fusion at the raw data (i.e., image) level.

## 4.1 Proposed Face Mosaicing Algorithm

This section describes the face mosaicing algorithm used to consolidate the evidence presented by multiple pose images of the same face. It is assumed, therefore, that at the time of enrollment, multiple poses of an individual's face are available. The face is segmented (localized) from each image using the Gradient Vector Flow technique (see

[123] for details). A pair of face images, typically representing the frontal and profile views of an individual, are mosaiced after aligning them using a hierarchical registration algorithm. Registered images are mosaiced using the multiresolution splines algorithm based on Gaussian and Laplacian pyramids [14]. Multiresolution splines also perform blending as an integral part of mosaicing thereby offering some inherent advantages.

### 4.1.1 Hierarchical Registration Model

Before mosaicing, it is necessary to transform the images obtained during enrollment into a common image domain. The process of finding the transform that aligns one image to another is called image registration. As described earlier in Section 2.1, existing face mosaicing algorithms use some form of affine transformation for registration. However, these algorithms do not consider the non-linear deformation present in the images. In this section, we propose a hierarchical registration algorithm in which we first perform approximate registration using an affine transformation model [72]. The affine transformed images are then finely registered using a mutual information-based registration algorithm [44], [65] resulting in a more exact alignment between the images.

**Affine Transformation Model**

Let $I_1 = I(x, y, t_1)$ and $I_2 = I(x, y, t_2)$ be the two images to be mosaiced. Here, $I_1$ is the source image and $I_2$ is the target image, i.e., $I_1$ has to be suitably transformed in order to align it with $I_2$. The pixel coordinates are represented using $x$ and $y$ spanning the domain of the image. The relationship of the pixels and their intensities between the two images can be modeled as,

$$I(x, y, t_1) = \frac{I(a_1 x + a_2 y + a_3, a_4 x + a_5 y + a_6, t_2) - a_7}{a_8} \qquad (4.1)$$

where $a_1, a_2, a_4$ and $a_5$ are the affine parameters summarizing the rotation, scaling and shear; $a_3$ and $a_6$ are the translation parameters; and $a_7$ and $a_8$ are the parameters that embody changes in brightness and contrast, respectively. The following error function is minimized in order to estimate these parameters:

$$Error(\mathbf{a}) = \sum_{x,y} \left[ I(x,y,t_1) - \frac{I(a_1x + a_2y + a_3, a_4x + a_5y + a_6, t_2) - a_7)}{a_8} \right]^2, \qquad (4.2)$$

where $\mathbf{a} = (a_1, a_2, \ldots, a_8)^T$. Approximating this error function using the first order truncated Taylor series expansion gives [72],

$$Error(\mathbf{a}) = \sum_{x,y} \left[ I_t(x,y,t) - \frac{(a_1x + a_2y + a_3 - x)I_x(x,y,t)}{a_8} + \frac{(a_4x + a_5y + a_6 - y)I_y(x,y,t) - a_7}{a_8} \right]^2$$

$$(4.3)$$

where, $I_x(\cdot)$, $I_y(\cdot)$, and $I_t(\cdot)$ are the spatial and temporal derivatives of $I(\cdot)$. This error function can be minimized by differentiating with respect to $\mathbf{a}$, i.e., $\frac{dError(\mathbf{a})}{d\mathbf{a}} = 0$. The solution of Equation 4.3 is,

$$\mathbf{a} = [\sum_{x,y} A^T A]^{-1} [\sum_{x,y} Ab] \qquad (4.4)$$

where, $A = (xI_x, \ yI_x, \ xI_y, \ yI_y, \ I_x, \ I_y, \ -I, \ -1, )^T$, and $b = I_t - I + xI_x + Iy$. The output of this minimization problem gives the eight optimal parameters. Using these parameters, the source face image $I_1$ is transformed to obtain the registered face image $I_R$. In this manner, the affine model performs a coarse registration of the two images. However, local regions such as the eyes and mouth still need to be finely aligned in order to remove the degeneracies present in the transformed image.

**Mutual Information based Transformation Model**

In two coarsely registered face images, the neighborhood of the corresponding regions may not be rigorously identical. This is due to differences in the geometry and local deformations present in the constituent images. Sub-pixel shifting can also occur leading to differences in the two images. Thus, a fine alignment is necessary to account for these non-linear deformations.

Let $I_2$ be the target face image and $I_R$ be the face image which is coarsely registered with respect to $I_2$ using the affine transformation described earlier. For fine registration, we transform $I_R$ such that the mutual information between $I_2$ and $I_R$ is maximized [65].

We use the weighted normalized mutual information for face images which can be written as,

$$\hat{M}(I_2, I_R, c) = c \left( \frac{H(I_2) + H(I_R)}{H(I_2, I_R)} \right) \tag{4.5}$$

where $c$ $(0 < c \leq 1)$ is a weighting parameter that controls the amount of localization in the similarity measure. This method is a modification over the one proposed by Hill *et al.* [44]. The modified function represented by Equation 4.5 is used for the fine registration of $I_2$ and $I_R$. The transformation parameters computed from the affine transformation model, $a$, are used as the initial transformation parameters along with $c = 0.01$. Thus, a total of nine parameters are used as the initial transformation parameters for this model. A set of mutual information values between $I_2$ and $I_R$ is computed by varying the parameters with a small step size ($\approx 0.01$) in both the positive and negative directions. Parameters are varied in the range $\left[ -\frac{a}{5}, \frac{a}{5} \right]$. The mutual information is computed for all possible combinations of the parameters, and the transformation parameters corresponding to the maximum mutual information are selected for fine registration. To account for the non-linearity present in face images, the mutual information based registration is performed in blocks of size $8 \times 8$. The coarsely registered profile image, $I_R$, is transformed using these parameters to obtain the final registered image. Figure 4.1 shows a profile image ($I_1$) transformed with respect to the frontal image ($I_2$).

## 4.1.2 Mask Generation

Once a pair of images is registered, the next step involves blending the two images into a single entity. This entails the development of a spatial mask indicating the pixelwise contribution of the individual images to the final mosaiced image. Since the facial structure is different for every individual, a dynamic runtime mask generation algorithm is used. The mask is computed using the local phase correlation between the two images. The two images are first tessellated into blocks of size $8 \times 8$. Next, the phase correlation between corresponding blocks from the two face images is computed. When a correlation
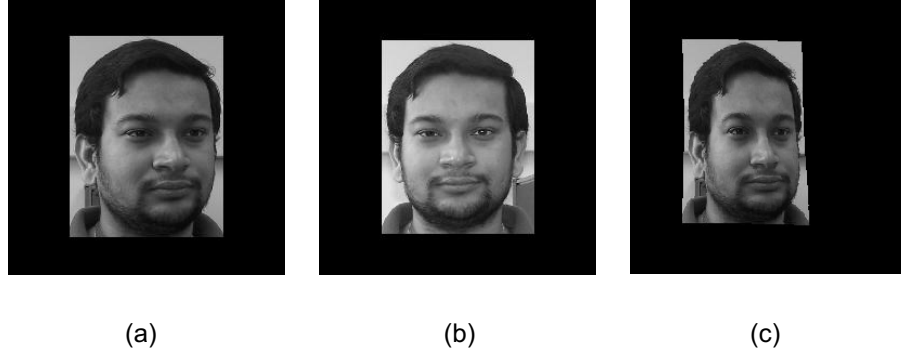
Figure 4.1: Image registration using the proposed hierarchical registration algorithm. Frontal and profile images are first placed on the center of a $256 \times 256$ image space. (a) Input profile image (b) Input frontal image (c) Profile image registered with respect to the target frontal image.

peak is observed, the block is labeled as a *match* otherwise it is labeled as a *non-match*. This results in a cluster of match/non-match blocks. The boundary of the matched *region* is selected as the boundary of the mask. The mask values on one side of the boundary is set to '0' while the other side is set to '1' (e.g., '1' may correspond to the frontal face image and '0' to the profile face image). If there are any isolated blocks with the label *match*, they are reassigned the label *non-match* and do not contribute to the mask boundary. Generally, the correlation peak is found in a thin vertical region containing the eye (left eye for left profile image and right eye for right profile image). Figure 4.2 shows the sample masks generated for a left and right profile face image with respect to an arbitrary frontal image. This mask is used during the blending process as will be seen in the next subsection.

### 4.1.3 Stitching and Blending

For blending the two images into a single mosaic we use multiresolution splines [14]. Image splining (i.e., blending) can be performed based on a simple spline-weighting
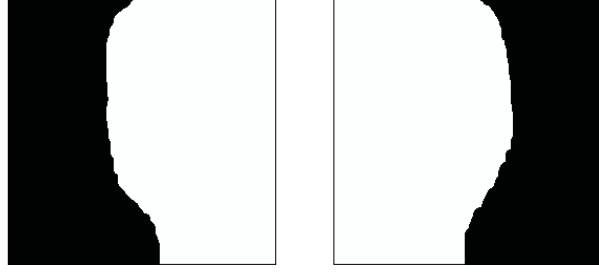
Figure 4.2: Masks generated from two profile images.

function[1] straddling the boundary of the two images, but the quality of the stitched image depends on the step size (or window) that is chosen. A large step size may lead to blurring whereas a small step size may result in discontinuities at the boundary. To overcome this problem, Burt and Adelson [14] used multiresolution splines to determine different step sizes for the various frequency components constituting the boundary. The crux of this technique involves computing a Gaussian pyramid of sub-images, followed by a Laplacian pyramid, based on the two images to be mosaiced; the pyramid structure is used to estimate the spline weighting function that relies on the frequency domain information of the image.

A sequence of low pass filtered images is obtained by iteratively convolving each of the constituent images with a 2D Gaussian filter kernel. The resolution and sample density of the image between successive iterations (levels) is reduced and, therefore, the Gaussian kernel operates on a reduced version of the original image in every iteration. The resultant images $G_0, G_1, \ldots, G_N$ may be viewed as a 'pyramid' with $G_0$ having the highest resolution (lowermost level) and $G_N$ having the lowest resolution (uppermost level) (see Figure 4.3). Let $w(m, n)$ represent the Gaussian kernel of dimension $5 \times 5$ and reduction factor 4. The *reduce* operation can be written as,

$$Reduce(I(i, j)) = \sum_{m=1}^{5} \sum_{n=1}^{5} w(m, n) I(2j + m, 2j + n) \tag{4.6}$$

A Gaussian pyramid $G_l$ is defined as,

---

[1]The term splining is used to refer to a transition function that indicates the weighting of pixels associated with the two images at the boundary (see [14] for details).
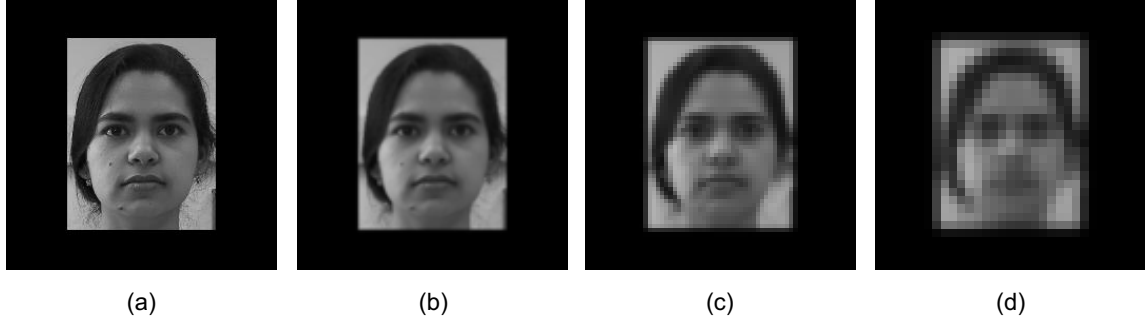
Figure 4.3: Levels in the Gaussian pyramid expanded to the original size to see the effects of the low pass filter. (a) Level 0; (b) Level 1; (c) Level 2; and (d) Level 3.

$$G_0 = I \tag{4.7}$$

$$G_l = Reduce[G_{l-1}], \ 0 \ < \ l \ < \ N \tag{4.8}$$

As shown in Figure 4.3, the effect of convolution is to blur the image, thereby reducing the filter band-limit by an octave between levels whilst reducing the sample density by the same factor. The Gaussian pyramid has the effect of a low pass filter to soften the edges of the mask.

The multiresolution spline as described in [14] requires band pass images as opposed to low pass images. Band pass images are computed by interpolating (resizing) the image at each level of the Gaussian pyramid and then subtracting it from the next lowest level. This results in a sequence of band pass images that may be viewed as a Laplacian pyramid $(L_0, L_1, \ldots, L_N)$ as shown in Figures 4.4 and 4.5. The term Laplacian is used since the Laplacian operator resembles the difference of Gaussian-like functions. These band-pass images are a result of convolving the difference of two Gaussians with the original image. The steps used to construct this pyramid can also be used to exactly recover the original image. The process described above may be summarized as follows:

$$L_l = G_l - Expand[G_{l+1}], \ \ 0 \ \leq \ l \ < \ N \tag{4.9}$$

78
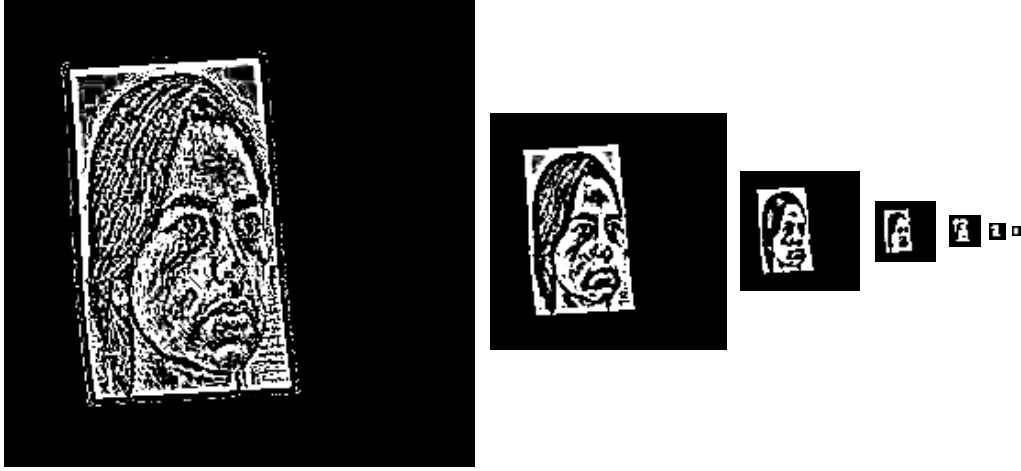
Figure 4.4: The Laplacian pyramid of a profile image from Level 0 to Level 6.

Here, the $Expand[\cdot]$ operator interpolates a low-resolution image to the next highest level and can be written as,

$$G_{l,k}(i,j) = 4 \sum_{m=-2}^{2} \sum_{n=-2}^{2} w(m,n) G_{l,k-1}\left(\frac{i-m}{2}, \frac{j-n}{2}\right). \qquad (4.10)$$

Note that $G_{l,k}$ in Equation 4.10 denotes *expanding* $G_l$ $k$ number of times. Various features of the face are segregated by scale in different levels of the pyramid. Hence, as shown in Figures 4.4, 4.5, and 4.6, the textural features of face are preserved over multiple levels of the pyramid. Let $L_1$ and $L_2$ represent the Laplacian pyramids of the two images that are being splined (i.e., blended). Let $GR$ be the pyramid associated with the Gaussian-weighted mask discussed in Section 4.1.2. The multiresolution spline, $LS$, is then computed as,

$$LS_l(i,j) = GR_l(i,j) L_{1_l}(i,j) + (1 - GR_l(i,j)) L_{2_l}(i,j) \qquad (4.11)$$

where $l$ is the level of the pyramid. The splined images at various levels are expanded and summed together to obtain the final face mosaic as shown in Figure 4.7. Gradient vector flow based active contour model is used to extract the face boundary of the mosaiced face.

Figure 4.5: The Laplacian pyramid of a frontal image from Level 0 to Level 6.
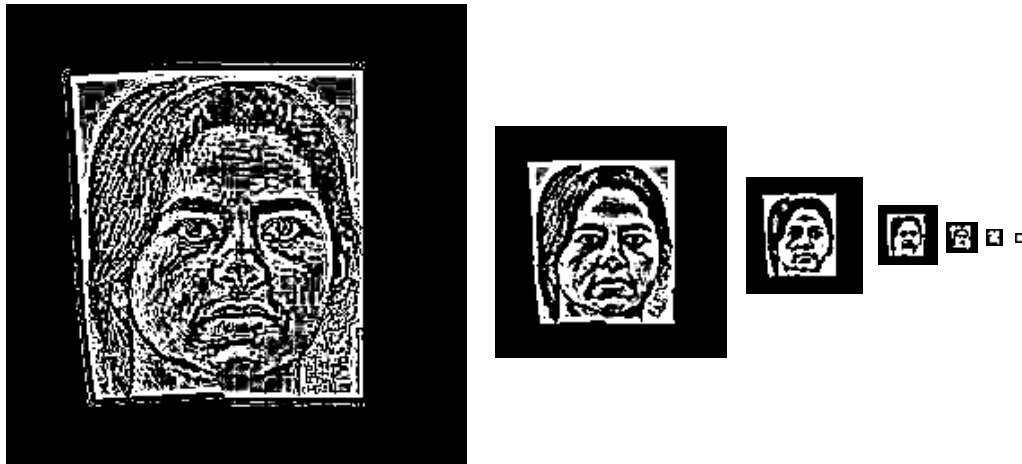


Figure 4.6: The Laplacian pyramid of the mosaiced image from Level 0 to Level 6.



| (a) | (b) | (c) | (d) | (e) |
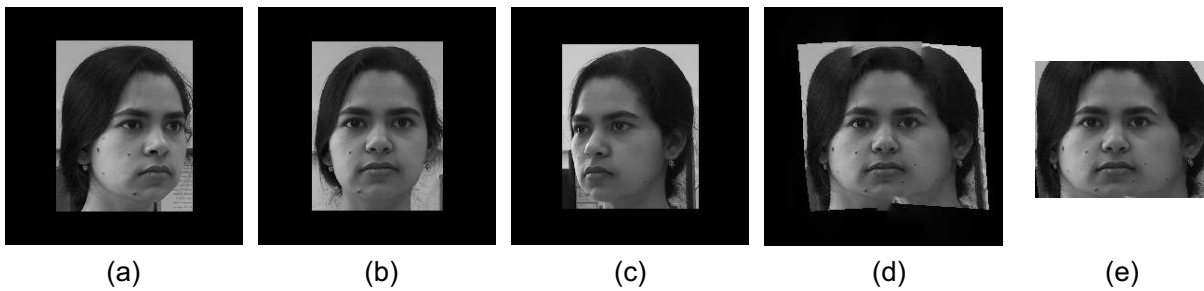
Figure 4.7: (a), (b), (c) Frontal and profile input images. (d) Mosaiced face generated using (a), (b) and (c). (e) Final mosaiced and cropped face image.

## 4.2 Face Recognition using Modified C2 Features

Several different texture-based face recognition algorithms have been proposed in the literature [30], [61]. However, most of the current texture-based face recognition algorithms fail to explicitly account for important spatial statistics between texture elements. Spatial statistics are important when analyzing facial textures that have similar texture frequency but differ in the distribution of texture elements. Face recognition algorithms should also handle small distortions and preserve the local feature geometry. In [87], a generic model-based feature extraction algorithm is proposed, which extracts visual features from an object using the fundamentals of a biological visual system. The model is a feedforward hierarchy consisting of four layers of computational units: simple $S$ units and complex $C$ units. The simple $S$ units combine their inputs with Gaussian-like tuning to increase object selectivity. The complex $C$ units pool their inputs through a maximum operation to achieve scale and shift invariance. The feature extractor is a Gabor filterbank consisting of 16 different filters and 4 different orientations resulting in 64 different maps when applied to a circular image patch. Each filter is parameterized based on its scale, width and frequency. The 64 maps are arranged in 8 bands (see [87] for details about the filterbank). The C2 algorithm operates as follows:

- The input face image is subjected to the Gabor filter bank described above resulting in 64 maps arranged in 8 bands. These maps constitute the $S1$ feature set.

- The $C1$ feature set is obtained by computing the maximum response over all scales in each band for all 4 orientations. A large pool of patches $P_{i=1,...,K}$ is extracted at random positions from all the training images. These patches are extracted for all 4 orientations and the radii of these patches varies from 4 to 16.

- For each feature set, $C1$, the value $Y$ is computed across all the bands for all image patches $(X)$ at all positions $(P)$ using the following equation:

$$Y = exp\left(-\gamma||X - P||^2\right), \tag{4.12}$$

81

where $\gamma$ is the aspect ratio. These patches are set as the prototypes or centers of $S2$ units. The $S2$ units behave as radial basis function during recognition.

- The maximum over all positions and scales for each $S2$ map gives the $C2$ features which are shift and scale invariant.

In [87], the authors have demonstrated superior recognition performance due to these features on different face and texture databases. For our purposes, we have modified this basic feature extraction algorithm to make it compatible with a face mosaicing application. In the modified algorithm, the filter bank is generated with 2D log polar Gabor transform. The bank of filter comprises of 16 filters at 4 orientations along with the 8 center surround differences of Gaussian filters and 4 low pass Gaussian filters. So, the modified filter bank consists of 76 filters which are used to compute the final $C2$ features. The size of the $C2$ feature vector varies for different face images. To efficiently match the mosaiced image with a non-mosaiced image, we use a learning based $2\nu$-SVM classifier [20], [24]. The classifier is trained with the features extracted from 148 mosaiced and 900 non-mosaiced face images of 108 individuals. This trained classifier is used to perform classification [31].

## 4.3 Database used for Evaluation

To validate the performance of the proposed face mosaicing and recognition algorithms, we used the following two databases:

- CMU PIE Face Database: The CMU PIE [92] face database contains images from 68 individuals with variations in pose, illumination, and expression. For mosaicing, we use the images with variation in pose, i.e., the images labeled as $c27$, $c37$, $c22$, $c11$, and $c34$ from both sessions. Figure 4.8 shows the two sets of images of an individual from the CMU PIE database. For each individual, that set of face images with neutral expression is used as the gallery image set. The images from
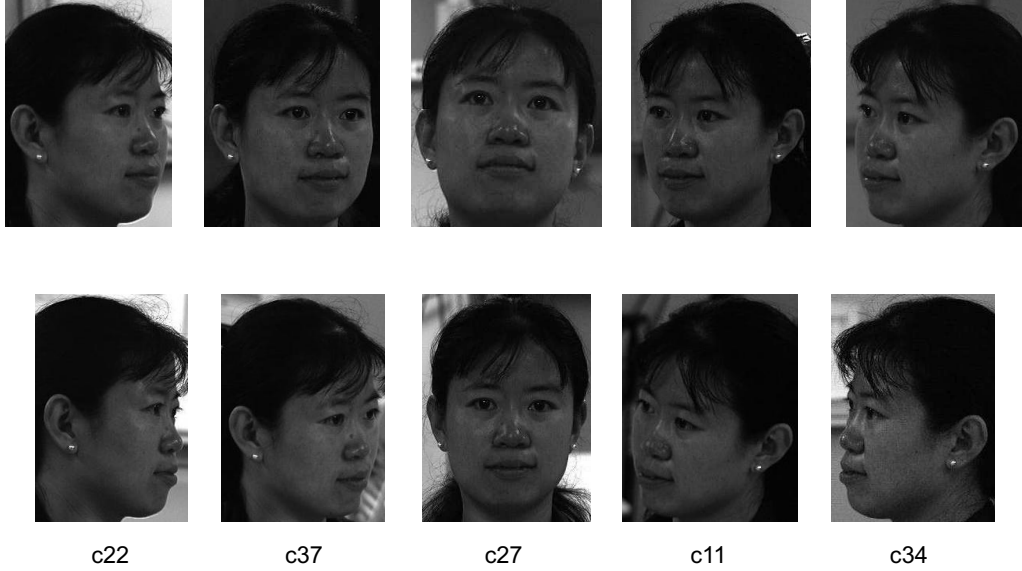
Figure 4.8: Examples of images from the CMU PIE database.

the three other sessions, which have minimum variation in lighting and expression, are chosen as probe images.

- WVU Multispectral Face Database: We assembled a multispectral face database of 40 individuals. The database contains face images in both the visible and short wave infrared light spectrum. For each spectral channel, images are captured in two different sessions, with seven images in each session. The seven images correspond to one frontal, three left profile, and three right profile images. Successive images are separated by a pose angle of approximately $20^0$. Figure 4.9 shows examples of images from the database. Various researchers have demonstrated the superior performance of face recognition on SWIR images [23], [51], [56], [77], [99].

(a)



p6       p4       p2       p1       p3       p5       p7

(b)
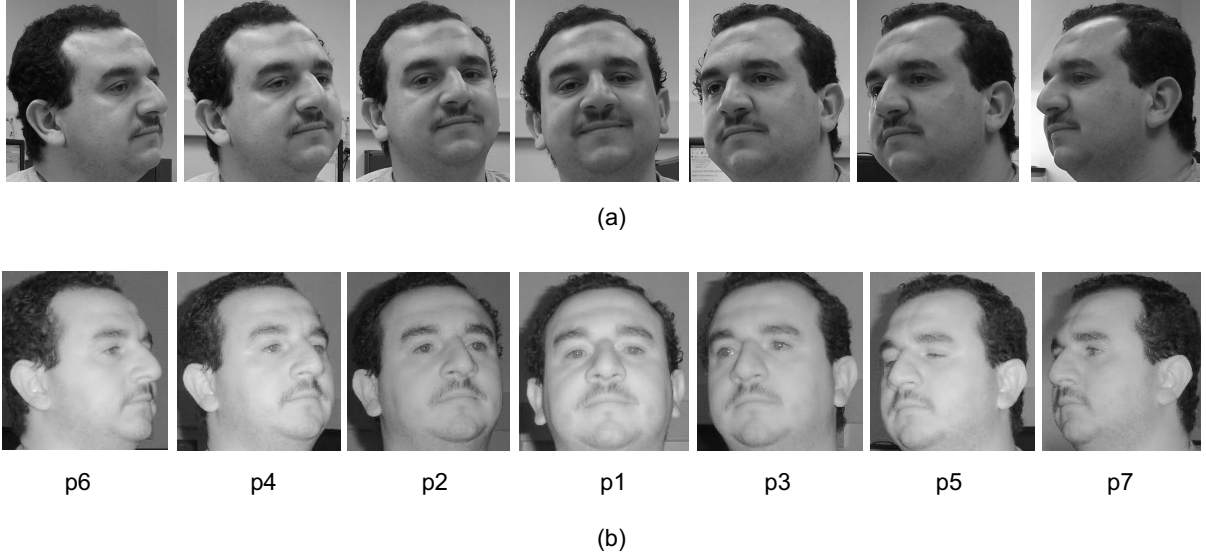
Figure 4.9: Images from the WVU multispectral face database. (a) Visible spectrum. (b) Short wave infrared spectrum.

## 4.4   Experimental Results

### 4.4.1   Performance Evaluation of Modified C2 Feature based Face Recognition Algorithm

We first validate the performance of the proposed face recognition algorithm described in Section 4.2 using the non-mosaiced face images. The $2\nu$-SVM classifier [20], [24] is trained using the C2 features extracted from the 620 labeled non-mosaiced training face images of the CMU PIE ($68 \times 5$) and the WVU visible-light ($40 \times 7$) databases. The classifier learns the genuine and impostor features from these training images prior to performing classification [87].

Recognition performance is evaluated separately for the two visible-light face databases. For evaluating the performance on the CMU PIE dataset, all the training images are used as gallery images and face images from the other sessions are used as probe images. Similarly, for the WVU visible-light face database, the training images are used as the gallery images and the remaining images are used as probe images to compute the identification

performance. The performance of the proposed recognition algorithm is also compared with five existing face recognition algorithms:

- Principal component analysis [113];

- Fisher linear discriminant analysis [11];

- Local feature analysis [71] refers to a class of algorithms that extract a set of geometrical features and distances from facial images and use these features as the basis for face representation and comparison. Hopfield network is used to extract the output which represents uncorrelated local features.

- 2D log polar Gabor: Face recognition algorithm described in Chapter 2 without granulation, only full resolution face image;

- Original C2 features [87].

A similar experiment is conducted for evaluating the performance on the SWIR database. 280 labeled non-mosaiced images from the WVU SWIR face database ($40 \times 7$) are used for training the classifier. These images are further used as gallery images and the remaining 280 SWIR images are used as probe images. Identification accuracy for all the recognition algorithms are shown in Table 4.2 and the Receiver Operating Characteristic plots are shown in Figures 4.10, 4.11 and 4.12. The results indicate that modified C2 based feature extraction algorithm outperforms the other recognition algorithms for both the visible-light and SWIR face databases. The images contain variation in pose and expression, which leads to changes in the number and position of the features present in individual images. Therefore, several face recognition algorithms that are appearance-based or feature-based do not perform well. However, the texture-based recognition algorithm results in better performance and exhibits tolerance to these variations. Table 4.2 and the ROC plots in Figures 4.10, 4.11 and 4.12 show that the modified C2 feature based face recognition algorithm provides the best performance.

A mosaiced face image is expected to contain all the features in a face while the frontal and side profile images have only a limited number of features. To match a

Table 4.2: The identification accuracy of multiple face recognition algorithms on the three databases.

| Algorithm | CMU PIE Face Database [92] | WVU Visible-light Face Database | WVU SWIR Face Database |
|---|---|---|---|
| PCA [113] | 53.81% | 49.63% | 52.94% |
| FLDA [11] | 58.24% | 54.19% | 58.38% |
| LFA [71] | 72.59% | 71.37% | 73.86% |
| 2D log Gabor | 95.32% | 96.22% | 97.19% |
| Original C2 feature [87] | 95.41% | 96.24% | 97.21% |
| **Modified C2 feature** | **96.76%** | **97.06%** | **97.92%** |
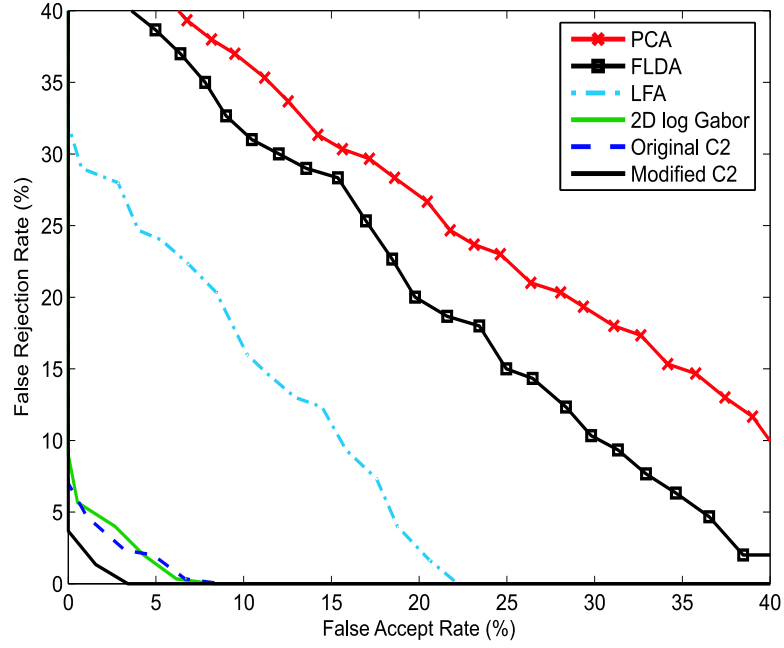


Figure 4.10: ROC indicating the performance of multiple face recognition algorithms on the CMU PIE database.
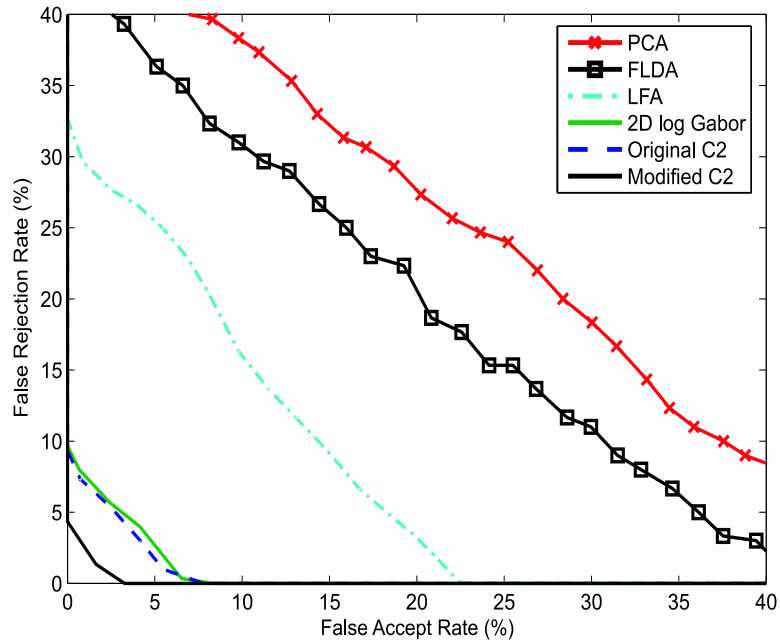
Figure 4.11: ROC indicating the performance of multiple face recognition algorithms on the WVU visible-light database.
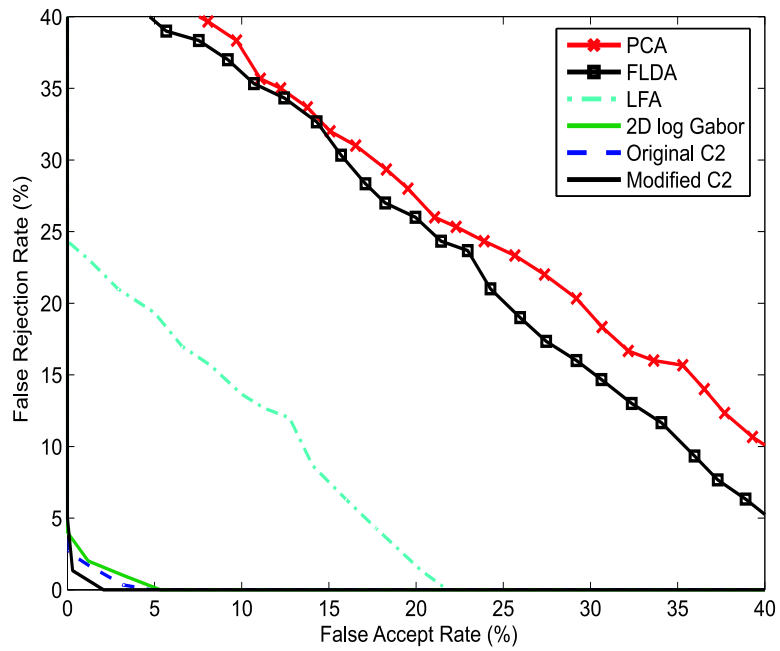


Figure 4.12: ROC indicating the performance of multiple face recognition algorithms on the WVU SWIR face database.

87

mosaiced face with a non-mosaiced face image, we need a recognition algorithm which efficiently extracts the local features from the face and compares them by assigning proper weights to the features while appropriately accounting for missing features. To facilitate this, we use a local representation based on the textural features obtained using the modified C2 feature extraction algorithm in order to evaluate the performance of the proposed face mosaicing algorithm.

## 4.4.2 Performance Evaluation of Face Mosaicing and Recognition Algorithm

The CMU PIE training set contains 68 mosaiced and 340 non-mosaiced labeled images; the WVU visible-light training set contains 40 mosaiced and 280 non-mosaiced labeled face images; and the WVU SWIR training set contains 40 mosaiced and 280 non-mosaiced labeled face images. The following experimental setup is used to compute the matching performance on all three databases separately.

- C2 features extracted from the mosaiced and non-mosaiced training images are used to train the $2\nu$-SVM based classifier.

- C2 features extracted from the mosaiced face images are used as the gallery features.

- On all three databases, the following set of experiments is then conducted:

  – The first set of experiments is performed to compute the optimal sequence and number of face images used for mosaicing (Section 4.4.2.1).

  – In the second experiment, the performance of match score fusion techniques [85] - the sum rule and the min/max rules - is compared against that of the mosaiced images (image level fusion) (Section 4.4.2.2).

### 4.4.2.1 Optimal Sequence for Face Mosaicing

The CMU PIE database contains two left and two right profile face images with successive images having a difference of approximately $22.5^0$, while the WVU multispectral database contains three left and three right profile face images with a difference of around $20^0$ between successive images. To find the optimal sequence and number of images for mosaicing, the frontal image of a subject is mosaiced with various combinations of the profile face images. In all sequences, an equal number of left and right profile images is used, and the difference in pose between successive profile images and the frontal image is approximately the same. This ensures that the extent of information on both *sides* of the mosaiced face is the same. Thus, we have four possible input sequences for the CMU PIE database [92] and eight possible input sequences for the WVU multispectral database.

Mosaiced face images are generated with the training dataset representing all possible input sequences. The verification and identification performance is evaluated by matching a mosaiced face image with all the probe non-mosaiced face images. Verification accuracy is computed at 0.001% false accept rate. Table 4.3 shows the results of this experiment. For the CMU PIE database, the input sequence 'c37 - c27 - c11' results in the best verification and identification accuracy of 96.54% and 96.88%, respectively. For both visible-light and SWIR images from the WVU multispectral database, the input sequence 'p4 - p1 - p5' results in the best performance. This suggests that increasing the number of input images for face mosaicing does not guarantee better performance. Experiments on all the databases indicate that the best result is obtained when the mosaic is generated using profile images having a difference of around $45^0$ with the frontal image. Figure 4.13 shows samples of mosaiced images generated using the optimal sequence for all three databases.

Table 4.3: Performance of face mosaicing based on different input image sequences.

| Database | Image Sequence | Verification Accuracy (0.001% FAR) | Identification Accuracy |
|---|---|---|---|
| CMU PIE Face Database [92] | c27 | 61.36% | 61.41% |
| | **c37 − c27 − c11** | **96.54**% | **96.88**% |
| | c22 - c27 - c34 | 77.54% | 75.98% |
| | c22 - c37 - c27 -c11 - c34 | 96.37% | 96.40% |
| WVU Visible Face Database | p1 | 63.05% | 64.51% |
| | p2 - p1 - p3 | 96.27% | 96.35% |
| | **p4 − p1 − p5** | **96.98**% | **97.52**% |
| | p6 - p1 - p7 | 73.30% | 71.19% |
| | p4 - p2 - p1 - p3 - p5 | 96.20% | 96.37% |
| | p6 - p2 - p1 - p3 - p7 | 93.65% | 93.44% |
| | p6 - p4 - p1 - p5 - p7 | 94.32% | 94.35% |
| | p6 - p4 - p2 - p1 - p3 - p5 - p7 | 95.82% | 95.80% |
| WVU SWIR Face Database | p1 | 64.47% | 64.67% |
| | p2 - p1 - p3 | 97.13% | 96.28% |
| | **p4 − p1 − p5** | **97.91**% | **98.16**% |
| | p6 - p1 - p7 | 74.02% | 73.76% |
| | p4 - p2 - p1 - p3 - p5 | 96.99% | 97.08% |
| | p6 - p2 - p1 - p3 - p7 | 96.26% | 96.43% |
| | p6 - p4 - p1 - p5 - p7 | 95.84% | 95.89% |
| | p6 - p4 - p2 - p1 - p3 - p5 - p7 | 94.70% | 94.81% |
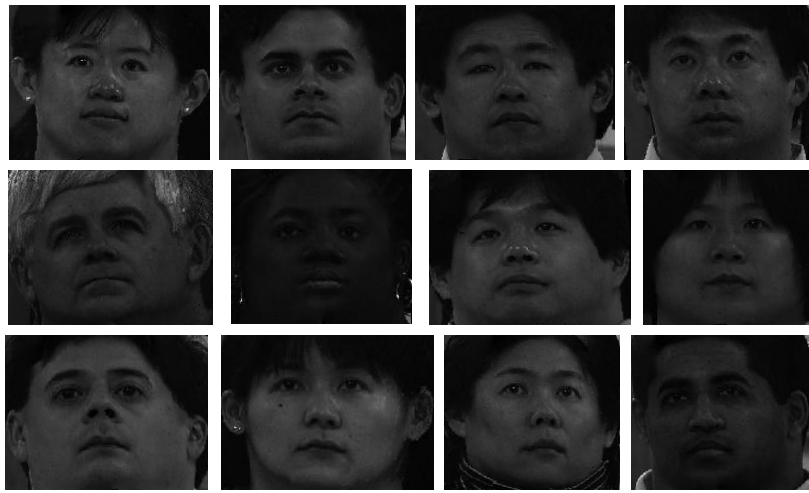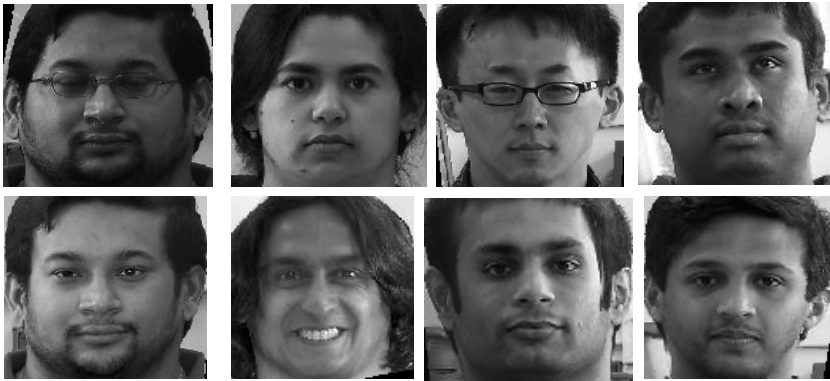
Figure 4.13: Mosaiced images generated with the optimal input sequence. (a) CMU PIE face database [92]. (b) WVU visible-light face database. (c) WVU SWIR face database.
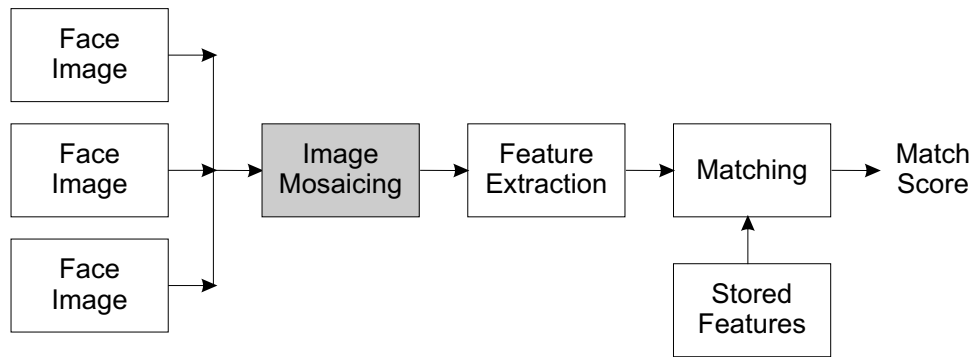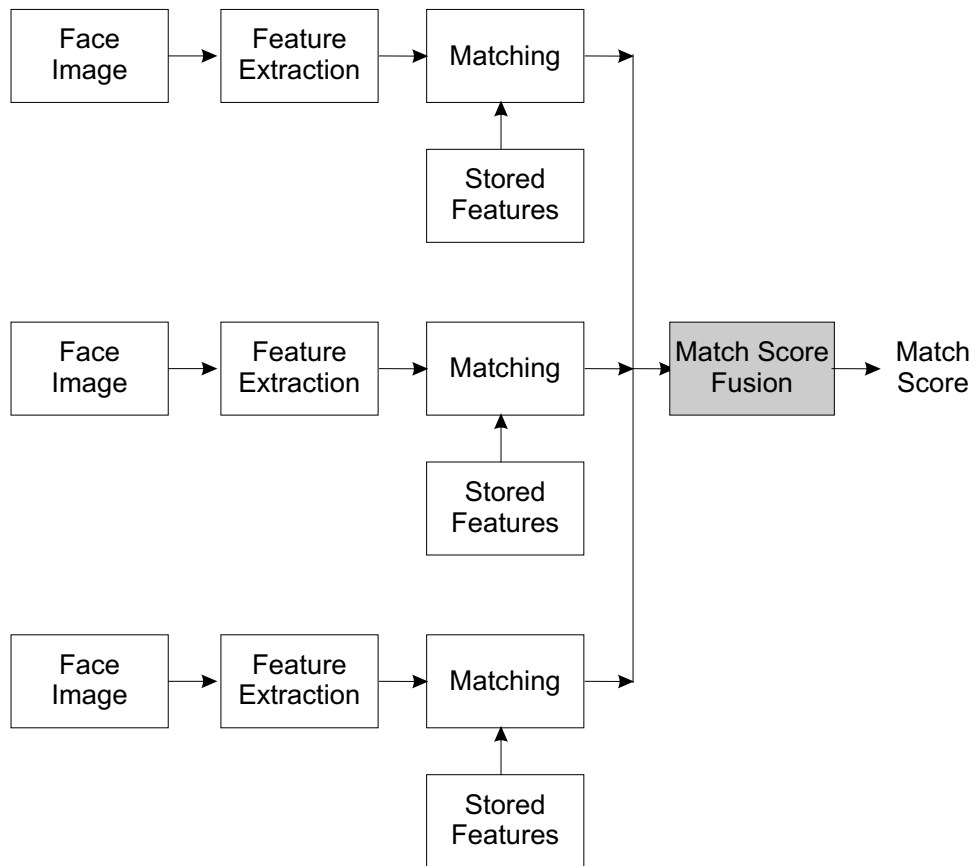
Figure 4.14: Block diagram illustrating the difference between image mosaicing (a) and match score fusion (b).

### 4.4.2.2 Comparing Face Mosaicing with Other Fusion Algorithms

We next compare the performance of fusion at the image level (mosaicing) with a few fusion operators at the match score level (sum rule and min/max rules). Figure 4.14 shows the steps involved in generating the final match score using these two methods. In the case of mosaicing, the gallery is assumed to contain only those mosaiced images corresponding to the optimal sequence obtained from the previous experiment. However, in the case of score fusion, all training images are used as gallery images. The experimental set-up is summarized below.

- Image Fusion: The gallery contains mosaiced images corresponding to the optimal sequence from one session and the probe consists of the mosaiced and non-mosaiced images from the rest of the sessions. Images are matched using the proposed face recognition algorithm. Results of this experiment for the three databases are shown in the first row of Tables 4.4, 4.5, and 4.6.

- Score Fusion: The gallery comprises of all the non-mosaiced images from one session. The probe images correspond to all the non-mosaiced as well as mosaiced images (generated using the optimal input sequence) from the other sessions. Results are shown in the second and third row of Tables 4.4, 4.5, and 4.6.

Tables 4.4, 4.5, and 4.6 show the identification accuracies for all three experiments on the CMU PIE, WVU visible-light and WVU SWIR face databases, respectively. The results from the three tables are summarized below,

- The modified C2 feature-based face recognition method gives good performance for images having variations in size and content making it particularly useful for recognizing mosaiced faces.

- A matching accuracy of 100% is obtained when both the gallery and probe images are mosaiced images.

Table 4.4: Identification accuracy on the CMU PIE database before and after fusion [92].

| Gallery Image | Test Image | | | | | |
|---|---|---|---|---|---|---|
| | Mosaiced (%) | c27 (%) | c37 (%) | c22 (%) | c11 (%) | c34 (%) |
| Mosaiced | **100** | **99.93** | **98.01** | **97.84** | **98.90** | **97.34** |
| Sum Rule [85] | 93.16 | 82.41 | 81.91 | 77.64 | 80.52 | 77.92 |
| Min/Max Rule [85] | 92.88 | 91.47 | 89.72 | 87.81 | 89.60 | 87.04 |
| c27 | 98.91 | 99.89 | 84.07 | 80.76 | 83.96 | 79.16 |
| c37 | 97.17 | 84.17 | 96.49 | 90.28 | 71.15 | 68.81 |
| c22 | 97.02 | 80.71 | 90.01 | 95.61 | 66.84 | 62.18 |
| c11 | 97.84 | 83.99 | 71.63 | 67.90 | 96.71 | 90.47 |
| c34 | 97.07 | 79.10 | 69.47 | 63.37 | 90.11 | 95.90 |

Table 4.5: Identification accuracy on the WVU visible-light face database before and after fusion.

| Gallery Image | Test Image | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mosaiced | p1 (%) | p2 (%) | p4 (%) | p6 (%) | p3 (%) | p5 (%) | p7 (%) |
| Mosaiced | **100** | **99.96** | **98.24** | **97.11** | **96.98** | **98.19** | **97.16** | **96.81** |
| Sum Rule [85] | 94.20 | 83.03 | 82.89 | 80.00 | 76.90 | 83.01 | 79.71 | 76.90 |
| Min/Max Rule [85] | 94.93 | 91.56 | 89.99 | 88.04 | 87.61 | 89.75 | 87.11 | 86.63 |
| p1 | 99.92 | 99.95 | 89.16 | 85.18 | 81.70 | 89.53 | 84.92 | 80.65 |
| p2 | 98.19 | 89.23 | 97.25 | 91.67 | 89.53 | 81.21 | 73.13 | 68.90 |
| p4 | 97.13 | 85.26 | 91.93 | 97.01 | 91.16 | 75.33 | 68.46 | 63.08 |
| p6 | 97.01 | 81.51 | 89.60 | 90.97 | 95.59 | 70.01 | 65.22 | 61.91 |
| p3 | 98.07 | 89.59 | 80.76 | 74.51 | 70.31 | 97.34 | 91.53 | 87.86 |
| p5 | 97.24 | 85.08 | 72.49 | 68.19 | 65.19 | 91.60 | 96.78 | 90.19 |
| p7 | 96.90 | 81.13 | 68.31 | 62.87 | 62.00 | 88.17 | 90.47 | 95.29 |

Table 4.6: Identification accuracy on the WVU SWIR face database before and after fusion.

| Gallery Image | Test Image | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mosaiced | p1 (%) | p2 (%) | p4 (%) | p6 (%) | p3 (%) | p5 (%) | p7 (%) |
| Mosaiced | **100** | **100** | **98.79** | **98.11** | **97.36** | **98.70** | **98.12** | **97.22** |
| Sum Rule [85] | 95.08 | 90.86 | 90.03 | 87.64 | 83.96 | 90.81 | 85.69 | 82.71 |
| Min/Max Rule [85] | 95.73 | 93.41 | 92.71 | 90.51 | 89.32 | 92.24 | 90.06 | 88.83 |
| p1 | **100** | **100** | 92.47 | 90.31 | 86.90 | 91.64 | 90.14 | 86.37 |
| p2 | 98.71 | 91.49 | 98.75 | 93.22 | 90.51 | 91.29 | 89.73 | 85.98 |
| p4 | 97.94 | 89.64 | 93.16 | 98.09 | 92.02 | 76.46 | 69.20 | 63.46 |
| p6 | 97.07 | 86.58 | 90.37 | 91.88 | 96.99 | 71.24 | 65.79 | 62.90 |
| p3 | 98.59 | 91.31 | 81.21 | 76.06 | 71.16 | **98.70** | 92.24 | 88.89 |
| p5 | 98.00 | 89.67 | 73.19 | 69.20 | 65.87 | 92.26 | 97.51 | 91.25 |
| p7 | 97.21 | 86.29 | 68.92 | 63.31 | 62.76 | 89.09 | 91.37 | 96.43 |

- Matching mosaiced images with non-mosaiced images gives higher accuracy compared to matching a non-mosaiced image with other non-mosaiced profiles.

- The proposed face mosaicing algorithm (image level fusion) gives better performance compared to the sum rule and min/max rules (score fusion algorithms).

- SWIR face recognition (using non-mosaiced as well as mosaiced images) results in better matching performance compared to visible-light face recognition. This may be attributed to the illumination invariant characteristic of SWIR images. Figure 4.15 illustrates the match scores generated using visible-light and SWIR images. The match scores are obtained using the modified C2 feature based recognition algorithm. It is evident from this example that the match scores obtained using SWIR images are more discriminating compared to that of visible-light face images.

- The performance of the proposed face mosaicing and recognition algorithm lies
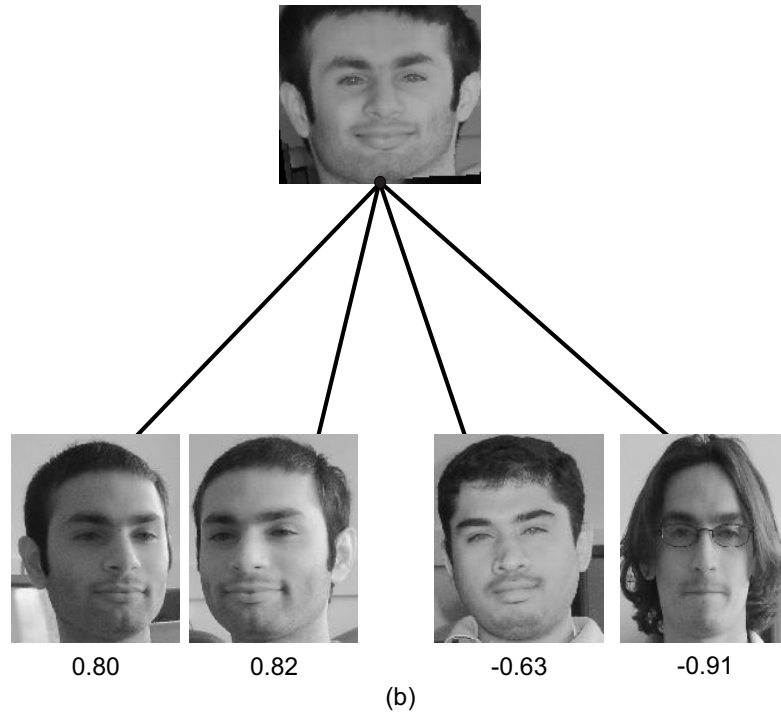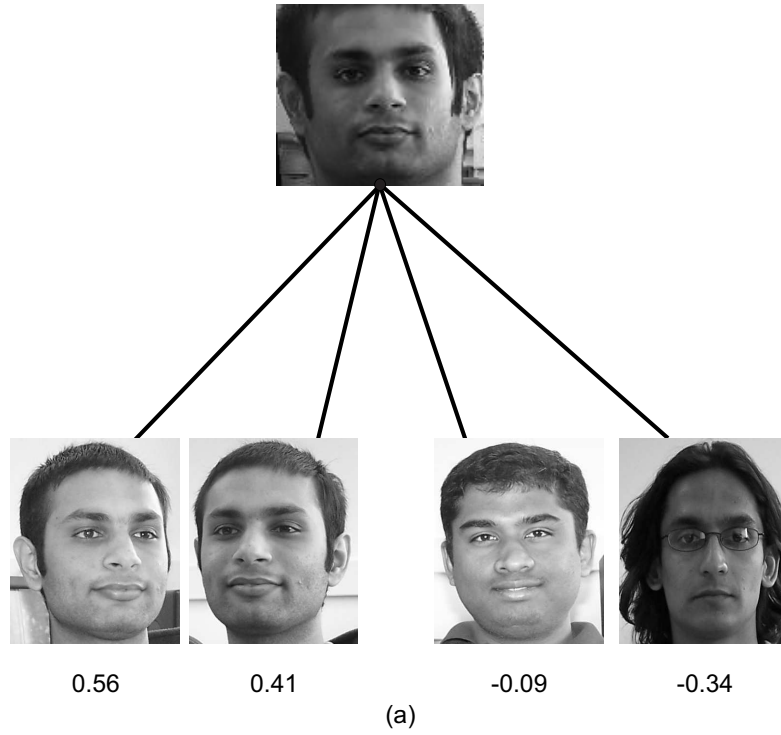
Figure 4.15: Match scores obtained using the modified C2 feature based algorithm. A value of 1 indicates a perfect match while a −1 represents a perfect reject. (a) WVU visible-light database. (b) WVU SWIR database.
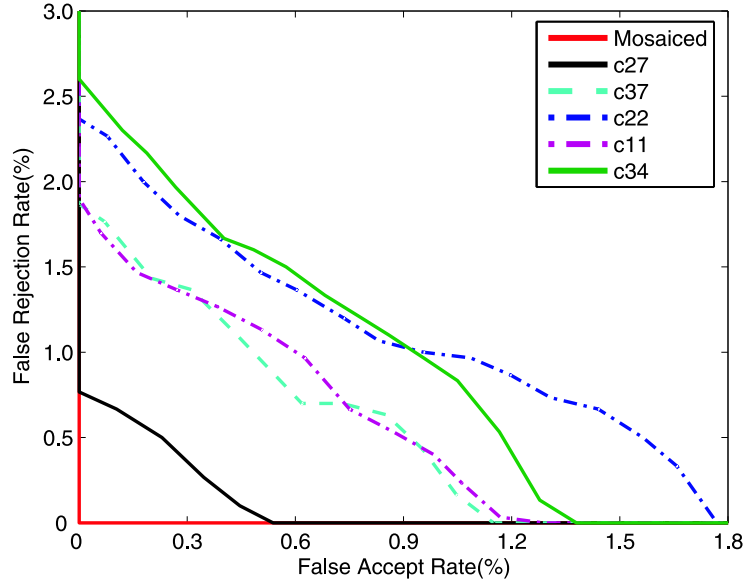
Figure 4.16: Performance of recognition algorithm with mosaiced face image as the database image on the CMU PIE database [92].

between 96.85% and 100% for all three databases.

Figures 4.16, 4.17, and 4.18 show the ROC curves as a result of comparing mosaiced face images with other mosaiced and non-mosaiced images. The verification time using a mosaiced image is $\sim 6$ seconds while that using the score fusion based approach is between 8-12 seconds (in a Matlab environment). For the score fusion scheme, the matcher has to be invoked multiple times corresponding to each subject. Further, if $M$ is the memory required to store one face image (in bytes), the memory requirement without mosaicing is between $5M$ and $7M$ whereas it is approximately $1.1M$ for the mosaiced image. These results show that image mosaicing enhances the performance of face recognition algorithms whilst reducing the memory requirement and the matching time.
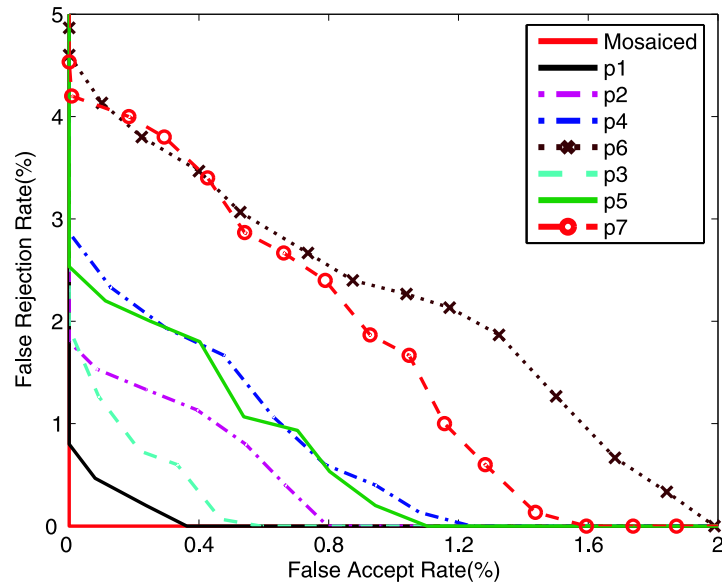
Figure 4.17: Performance of recognition algorithm with mosaiced face image as the database image on the WVU visible-light database.
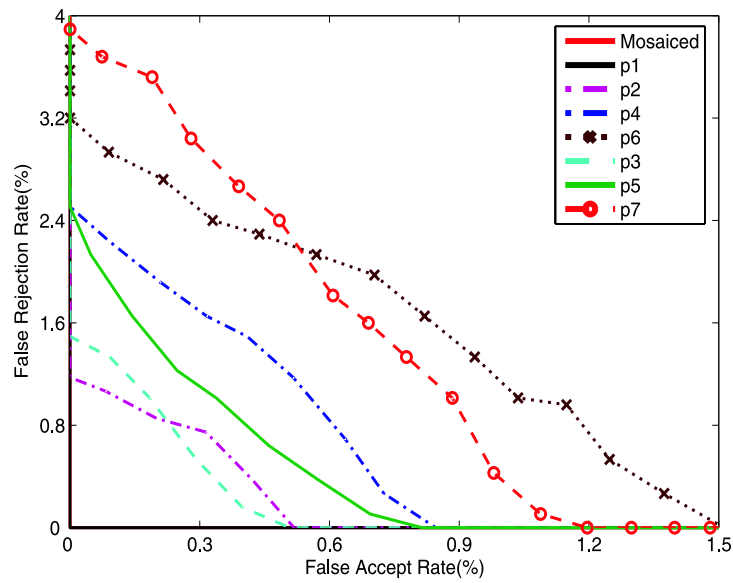


Figure 4.18: Performance of recognition algorithm with mosaiced face image as the database image on the WVU SWIR database.

## 4.5  Summary

The primary goal of this chapter was to demonstrate the role of face mosaicing in enhancing the matching performance of a face recognition system for pose variations. Given multiple images of a face during enrollment, the mosaicing algorithm blends them into a single entity by employing a multiresolution splining scheme. Experiments reported on three different databases suggest that fusing information at the image level is better than fusing information at the match score level. A modified version of the C2 algorithm, originally developed by Serre *et al.*, was used to extract features from the mosaiced and non-mosaiced face images. The modified algorithm is observed to perform very well both in the verification and identification modes of operation.

# Chapter 5

# Biometric Classifier Update using Online Learning

A carefully designed biometric system should be stable and robust to environmental dynamics and variations in the data distribution. However, in real world applications, these systems are affected by several factors including template aging, improper interaction of the user with the sensor, and noise. Furthermore, as the size of a biometric database increases, the system has to re-train itself in order to handle the variations introduced due to the newly enrolled subjects. For example, an identification system may have to re-compute its thresholds to handle the increased number of subjects, a verification system may have to re-compute its decision boundaries based on the new set of genuine and impostor scores, and a face recognition system may have to update its templates to address the aging effects of the face biometric. The re-training process can be time consuming and may not be pragmatic for large scale applications. A generic biometric system, as shown in Figure 5.1, has five modules that require regular update or re-training. Template update is required to address the issue of template aging, sensor update is necessary to accommodate advancements in sensor technology, re-training of the preprocessing and feature extraction algorithms is necessary to handle variations in data, and classifiers have to be updated to account for changes in the intra-class and inter-class dynamics of the subjects.

Currently, biometric classifiers such as support vector machine and neural net-
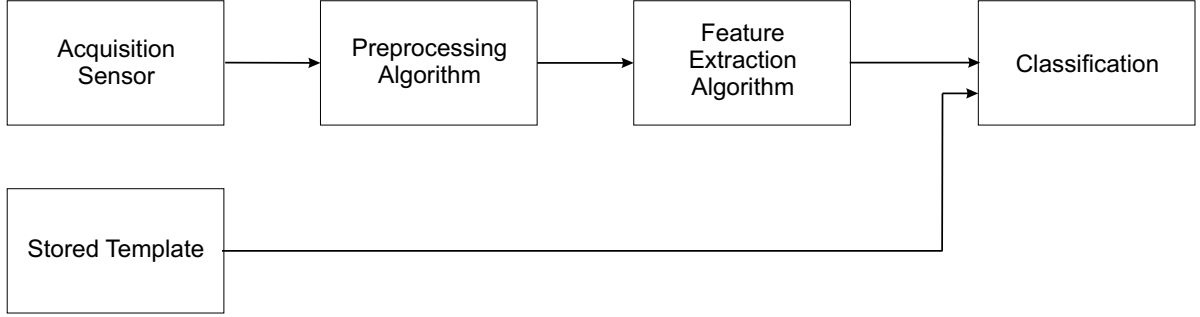
Figure 5.1: Block diagram representing the modules of a biometric system that may require regular update or re-training.

work are trained *offline* with the available training data and domain specific knowledge. However, large scale biometric applications such as US-VISIT continuously enroll new individuals. Due to the high computational complexity required for re-training in such applications, it is not feasible to regularly update the classifier knowledge and decision boundary, thereby affecting the recognition performance of the biometric system. Online learning presents an efficient alternative to offline learning and classification by updating the classifier knowledge upon the arrival of new data. While it has been extensively used in problems related to machine learning [5], [17], [50], [102], [110], *very limited research has been conducted in biometrics* [79]. Considering the high applicability of this concept in biometrics, this chapter introduces the concept of online learning for biometric classifier training and update thus reducing the computational time.

The contribution of the research lies in incorporating soft labels [109] and granular computing in the formulation of a $2\nu$-Support Vector Machine [24] to design a $2\nu$-Granular Soft Support Vector Machine and applying the online learning concept for classifier update. The introduction of soft labels and granular computing increases classification performance while online learning enables the classifier to update its knowledge as database enrollments increase. The performance of the algorithm is evaluated in the context of a near infra-red face recognition application. Facial features are computed using Principal Component Analysis [113] and modified C2 features (Section 4.2). Experiments performed on a heterogeneous NIR face database indicate that the proposed

algorithm not only improves the classification performance but also reduces the training time significantly.

## 5.1 Why Online Learning is Applicable in Biometrics?

Online learning is a concept from machine learning and inspired by human cognition. Humans are very good in adapting to the environment and learning concepts based on new information. Starting from childhood, a natural learning process constantly trains the human mind to develop new abilities and sustain old (useful) behavior. It also learns to remove behaviors or patterns that become extraneous and redundant over time. Thus, online learning has two components namely incremental and decremental learning.

Mathematically, we have the instance space $I$ consisting of patterns, label space $Y$ indicating the classes of patterns, and a learning based classifier $\Im$. The task of the classifier $\Im$ is to map $I$ to $Y$ using training examples. In general, there are two modes of learning: (1) offline/batch learning mode and (2) online learning mode. In batch mode, a set of training samples $(x_i, y_i) \in I \times Y$ are used as input ($i = 1, \cdots, N$ where $N$ is number of training samples). The learning classifier, $\Im_B$ ($B$ denotes batch mode), attempts to obtain a robust/optimal solution such that $\Im_B(x_i) = y_i$. Unlike batch learning, in the online learning mode, training samples become available in a sequential manner. In other words, at time $j$, training sample $(x_j, y_j)$ and previously processed samples $(x_1, y_1), \cdots, (x_{j-1}, y_{j-1})$ are used to train the learning classifier $\Im_O$ ($O$ denotes online mode) by predicting $y'_j = \Im_O(x_j)$ and comparing it with the true label $y_j$. The objective of online learning classifier is to accommodate new training samples and minimize the error over the whole training sequence. It is intuitive that online learning is pertinent to real world biometric applications where training samples are available sequentially (one at a time) and data representation must be updated for optimal classification performance. Another important aspect of online learning is that for a large scale application, it is

computationally efficient to solve the offline/batch training process in an online manner [50].

In large scale real world biometrics applications, it is challenging and computationally complex to train a classifier in advance. Without re-training, the disparate characteristics of additional biometric data can cause the performance to degrade. However, training the classifier dynamically online facilitates updated learning in real-time by reducing the computational cost. In a typical biometric system, online learning can be applied to any module that requires training and learning. In this chapter, we propose a $2\nu$-Online Granular Soft Support Vector Machine and apply it in the classification stage of the biometric system.

## 5.2 Formulation of $2\nu$-Online Granular Soft Support Vector Machine

As described in Section 2.5.1, $\{\mathbf{x}_i, y_i\}$ is the set of $N$ data vectors where $i = 1, ..., N$, $\mathbf{x}_i \in \Re^d$ and $y_i \in (+1, -1)$. $2\nu$-SVM objective function can be formulated as,

$$
\begin{aligned}
&min\{\tfrac{1}{2}\|\mathbf{w}\|^2 - \textstyle\sum_i C_i(\nu\rho - \psi_i)\} \\
&\text{subject to } y_i\left(\mathbf{w}\,\varphi(\mathbf{x}_i) + b\right) \geq (\rho - \psi_i), \quad \rho,\ \psi_i \geq 0
\end{aligned}
\tag{5.1}
$$

where $i, j \in 1, ..., N$, $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function [116], $\alpha_i, \alpha_j$ are the Lagrange multipliers such that $0 \leq \alpha_i \leq C_i$, $\sum_i \alpha_i y_i = 0$, and $\sum_i \alpha_i \geq \nu$.

During training, it is possible that some of the data points may be noisy or incorrectly labeled. In such cases, like any classifier, $2\nu$-SVM performs erroneous classification. To address this limitation, the formulation of $2\nu$-SVM is extended to include soft labels. Tao *et al.* [109] have shown that the use of soft labels not only reduces the classification error but also decreases the number of stored support vectors. Let $z_i$ be the soft label for the $i^{th}$ training data $x_i$. $2\nu$-SVM is transformed into $2\nu$-Soft SVM ($2\nu$-SSVM) as follows,

$$min \left\{ \tfrac{1}{2} \|\mathbf{w}\|^2 - \sum_i C_i(\nu\rho - \psi_i) \right\}$$
$$\text{subject to } z_i \left( \mathbf{w}\, \varphi(\mathbf{x}_i) + b \right) \geq z_i^2(\rho - \psi_i). \tag{5.2}$$

Even with the availability of $2\nu$-SSVM, training a large database is still time consuming. Granular computing [7] is based on a divide and conquer approach, aimed at reducing the computational time as well as increasing the adaptability to data distribution both locally and globally. In the proposed $2\nu$-Granular SSVM, the data space is divided into $c$ subspaces with one $2\nu$-SSVM operating on each subspace. Let $2\nu$-SSVM$_i$ represent the $i^{th}$ $2\nu$-SSVM, and $2\nu$SSVM$_i :\rightarrow L_i$ represents the $2\nu$-SSVM operating on the $i^{th}$ subspace $(i = 1, 2, ..., c)$. The compound margin width $W$ is computed using Equation 5.3.

$$W = \left| \sum_{i=1}^{c} \tfrac{t_i}{t}(2\nu SSVM_i :\rightarrow L_i) - L_0 \right|, \tag{5.3}$$

$$t = \sum_{i=1}^{c} t_i$$

where $t_i$ is the number of training data in the $i^{th}$ subspace. $2\nu$-SSVM learning yields $L_i$ at the local level and $L_0$ is obtained by learning another $2\nu$-SSVM on the complete feature space at the global level. Equation 5.3 provides the margin width associated with the $2\nu$-GSSVM hyperplane.

## 5.2.1   $2\nu$-Online GSSVM

Support Vector Machines, including the proposed $2\nu$-GSSVM, are trained using the training database and evaluation is performed using the testing database. Several applications including biometrics, that use SVM as a classifier, require re-training at regular intervals to accommodate the changes in data distribution. Re-training the SVM every time is computationally expensive and may not be feasible for real time applications. In this chapter, we propose an online learning scheme for $2\nu$-GSSVM and is termed as $2\nu$-OGSSVM. The main concept behind the proposed approach is to first construct the decision hyperplane using an initial training dataset and then re-train the classifier by incorporating the new training data points into the decision hyperplane. It also

removes unnecessary and irrelevant data so that the number of support vectors does not increase drastically with the increase in training samples. Thus, the proposed online learning algorithm includes both incremental and decremental learning. In this process, the Karush-Kuhn Tucker conditions [36] are maintained so that the $2\nu$-OGSSVM provides an optimal decision hyperplane. The training procedure of $2\nu$-OGSSVM is as follows:

1. $2\nu$-GSSVM is trained using an initial training database and a decision hyperplane is obtained.

2. For each new training data $\bar{x}_i$,

   (a) $\bar{x}_i$ is classified using the trained $2\nu$-GSSVM.

   (b) The classification output is compared with the associated label $\bar{z}_i$; if the classification is correct then re-training is not required.

   (c) Otherwise,

      i. The decision hyperplane is recomputed using the $m$ trained support vectors and $\{\bar{x}_i, \bar{z}_i\}$.

      ii. After recomputing the hyperplane, the number of support vectors increases. If the number of support vectors is more than $m + \lambda$, then a support vector that is farthest from the current decision hyperplane is selected.

      iii. The selected support vector is removed from the list of support vectors and stored in the list, $l$. The classifier with $m + \lambda - 1$ support vectors is used for validation and testing.

3. The support vectors in the list $l$ are used to test the new classifier. If there is any misclassification, Step 2(c) is repeated to minimize the classification error.

4. The least recently included support vectors are removed from the list, $l$, in the final classifier.

Figure 5.2 shows an example of the decision boundary generated using the proposed $2\nu$-OGSSVM approach. The two classes in the example represent the genuine and imposter match scores obtained from a face verification algorithm. Figure 5.2a shows the result with 200 training samples. In this case, the $2\nu$-SVM uses all 200 training samples as input for off-line or batch learning, while the proposed $2\nu$-OGSSVM uses the 200 training samples to perform the initial training for online learning. Figure 5.2b compares the performance when the number of training samples is increased to 400. In this case, the $2\nu$-SVM uses all 400 training samples for off-line or batch learning, while the $2\nu$-OGSSVM updates the previously trained classifier one sample at a time with the remaining 200 samples for online learning. The figure shows that the proposed $2\nu$-OGSSVM efficiently performs binary classification.

## 5.3 Application of Proposed $2\nu$-OGSSVM to Near Infrared Face Verification

Recent studies have shown the usefulness of near infrared face images for recognition [57], [99]. Previous research results from Chapter 4 also suggest that the performance of face recognition with NIR images is better than visible spectrum images. This is because the near infrared spectrum provides resilience to varying image quality, changes in illumination, and minor variations in expression. Since the proposed $2\nu$-OGSSVM can reduce the computational time required for re-training the classifiers, in this case study, NIR face verification is used for performance evaluation.

Figure 5.3 shows the steps involved in NIR face recognition. Since the proposed $2\nu$-OGSSVM is a two-class classifier, it can be used for biometric verification. Firstly, the classifier is trained for two-class classification (genuine and impostor). Next, at the probe level, facial features are extracted from the NIR face image using feature extraction algorithms (PCA/C2). Match scores are computed using the Mahalanobis distance measure by comparing the extracted features against the stored features. Finally,
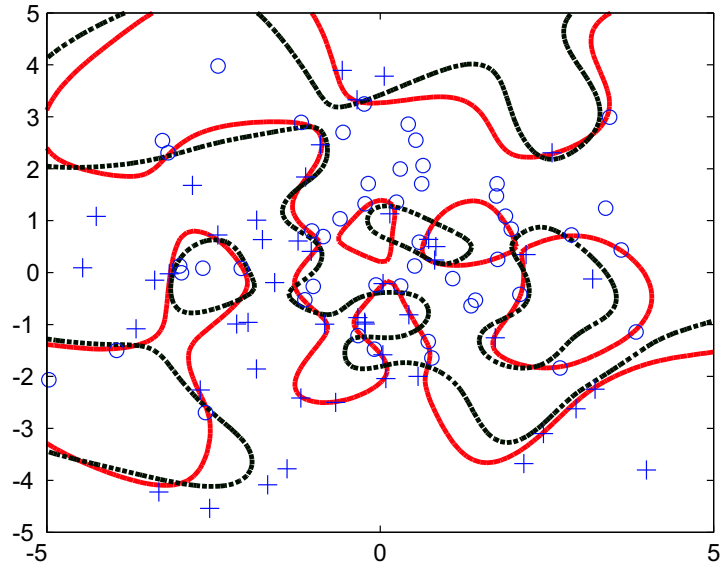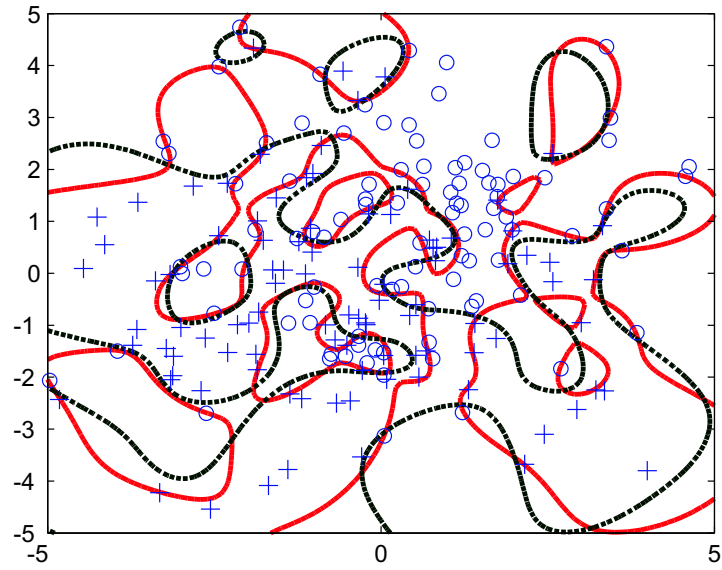
Figure 5.2: Comparison of non-linear decision boundary generated with offline learning using 2$\nu$-SVM and online learning using the proposed 2$\nu$-OGSSVM. (a) Number of training samples = 200 (b) Number of training samples = 400.

107
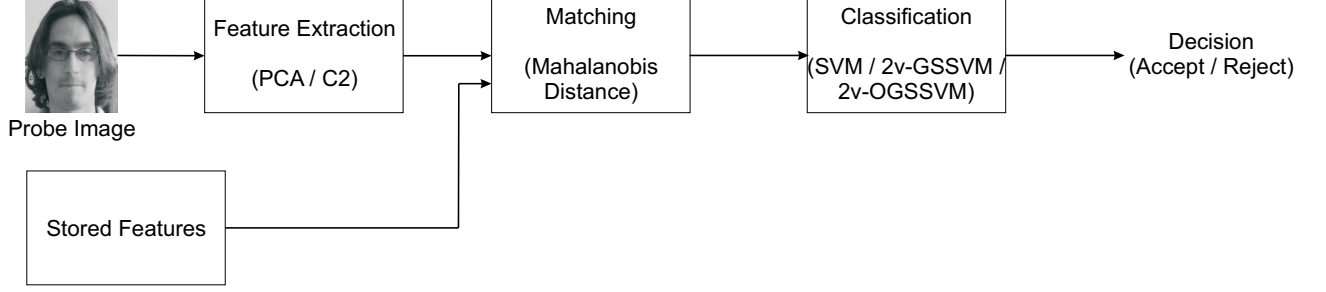
Figure 5.3: Illustrating the steps involved in the NIR face recognition case study.

classification is performed using the trained classifier.

### 5.3.1 Training 2$\nu$-OGSSVM

Let the input training data be $\{\mathbf{x}_i, z_i\}$ where $i = 1, ..., N$. $N$ is the total number of training samples and $\mathbf{x}_i$ is the $i^{th}$ match score. $z_i$ is the soft class label and labeling is performed using the density estimation approach [130]. For the $i^{th}$ training sample, likelihood ratio $P_i = \frac{g_{gen}(x_i)}{g_{imp}(x_i)}$ is computed where $g_{gen}$ and $g_{imp}$ represent the marginal densities of genuine and impostor respectively. Soft label is assigned as a $P$ value with '+' sign if $P \geq 0$ (i.e. genuine sample) otherwise it is assigned as a $P$ value with '-' sign (i.e. impostor sample). 2$\nu$-OGSSVM is trained using the radial basis function (RBF) kernel ($= exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2)$). The output of the trained 2$\nu$-OGSSVM is a non-linear decision hyperplane that can classify genuine and impostor match scores.

### 5.3.2 Probe Classification and Decision Making

At the probe level, the trained 2$\nu$-OGSSVM is used to classify the match scores. The match score obtained by matching the probe and gallery facial features, $x_p$, is provided as input to the mixture model and the likelihood ratio $P_p = \frac{g_{gen}(x_p)}{g_{imp}(x_p)}$ is computed. Finally, the trained 2$\nu$-OGSSVM is used for classifying the likelihood ratio and a decision of accept or reject is made.

Table 5.1: Composition of the heterogeneous NIR face database.

| Face Database | Number of Subjects | Number of Images |
|---|---|---|
| CBSR | 197 | 3940 |
| Equinox | 91 | 1307 |
| WVU | 40 | 560 |
| **Total** | **328** | **5807** |

### 5.3.3  Experimental Protocol

The images from three NIR face databases are combined to create a relatively large database with heterogeneous characteristics. As shown in Table 5.1, the CBSR NIR face database [18] contains 3940 images pertaining to 197 subjects, the WVU database [98] contains 560 images belonging to 40 subjects, and the Equinox database has 1307 images from 91 subjects. Combining these databases provides a wide range of interclass and intraclass variations that typically occur in real world face recognition applications. The database is divided into training database and testing database. The training database contains four images of each individual and the remaining images are used as the testing or probe dataset.

For feature extraction and evaluation, two different algorithms are used. The first facial feature extraction algorithm is the appearance based PCA algorithm [113] and the second algorithm is the modified C2 feature extraction algorithm [98] which is a local texture feature based approach. The training database is used to train these feature extraction algorithms and verification is performed using a two-class classifier such as SVM, $2\nu$-GSSVM and $2\nu$-OGSSVM. To evaluate the efficacy of the proposed approach, the experimental comparison of SVM, $2\nu$-SVM, and $2\nu$-GSSVM with online classification ($2\nu$-OGSSVM) is performed. For SVM, $2\nu$-SVM, and $2\nu$-GSSVM, the complete training database is used to train the classifier (batch learning mode). On the other hand, to evaluate the performance of the proposed $2\nu$-OGSSVM, the $2\nu$-OGSSVM classifier is initially trained with 100 subjects and then online learning is performed with training

samples from the remaining 228 subjects i.e., online training with one sample at a time. Further, the train-test partitioning is performed 20 times for cross-validation and ROC curves are generated by computing the false reject rate over these trials at different false accept rates. Finally, verification accuracies are reported at 0.01% FAR.

### 5.3.4    Experimental Results

In all the experiments, RBF kernel with $\gamma = 6$ is used for SVM, $2\nu$-SVM, $2\nu$-GSSVM and $2\nu$-OGSSVM classifiers. In the first case, PCA coefficients are extracted and matching is performed using the Mahalanobis distance measure. The match score is further classified using these classifiers separately. The verification performance of PCA with non-linear classification is also compared with the traditional PCA algorithm [113] (that uses Euclidean distance and a linear decision threshold for classification) and incremental PCA algorithm [79]. This experiment provides the baseline performance to facilitate systematic comparison with other algorithms. Furthermore, accuracies are also computed when IPCA is combined with the proposed $2\nu$-OGSSVM (i.e. IPCA + $2\nu$-OGSSVM, both feature extraction and classification algorithms are online). In the second case, modified C2 features are classified using the three classifiers separately. The ROC plots in Figures 5.4 and 5.5, and Tables 5.2 and 5.3 show the experimental results for comparison. The key results and analysis are summarized below:

- Baseline PCA (batch/offline learning mode) and IPCA (online learning mode) provide similar verification performance, but the advantage of IPCA is reduced computational cost[1] (Table 5.4). Specifically, the training time of IPCA is significantly lower compared to PCA. Experimentally, this shows that the online learning approach is a faster alternative for biometrics applications.

- PCA with SVM classifier yields the verification accuracy of 62.23% whereas with $2\nu$-GSSVM, the verification accuracy improves by 8.9%. This suggests that incor-

---

[1]Time is computed on a 2 GHz Pentium Duo Core processor with 2 GB RAM under MATLAB environment.

porating granular computing and soft labels improve the classification performance. Granular computing makes it adaptive to variations in data distribution and soft labels provide resilience to noise.

- When the feature extraction algorithm and classification algorithm both are in online learning mode, i.e. IPCA + 2$\nu$-OGSSVM, verification accuracy is about 1% greater than when both the algorithms are in batch learning mode (PCA + 2$\nu$-GSSVM). However, the training time in online learning mode, as shown in Table 5.4 is about half of the batch learning mode which shows the suitability of online learning algorithms for large scale real time applications.

- Similar improvements are observed in the case of modified C2 feature algorithm. Since the modified C2 algorithm [98] originally uses SVM for classification and decision making, the comparative study is performed with different variants of SVM. Compared to the SVM classifier, 2$\nu$-OGSSVM improves the verification accuracy by 5.77%, while the training time of the proposed 2$\nu$-OGSSVM classifier is about three times faster than the classical SVM classifier.

- From Tables 5.2 and 5.3, the covariate analysis with respect to variations in expression, illumination and pose shows that the pose variations cause a large reduction in the accuracy of appearance based PCA algorithm. On the other hand, local texture feature based modified C2 algorithm provides consistent performance for all three variations. The experiments also show that the modified C2 algorithm with 2$\nu$-OGSSVM classifier yields a verification accuracy of more than 92% which is around 21.7% better than PCA with 2$\nu$-OGSSVM classifier.

- For both the PCA and modified C2 feature algorithms, verification accuracies of the proposed 2$\nu$-OGSSVM are slightly better than 2$\nu$-GSSVM classifier. However, the advantage of the 2$\nu$-OGSSVM is significant reduction in computational time. Compared to classical SVM, 2$\nu$-GSSVM reduces the training time significantly because the dual-$\nu$ formulation requires less time for parameter estimation and

the granular computing approach reduces the time by dividing the problem into subproblems and solving it efficiently both in terms of accuracy and time. With online learning approach (i.e. $2\nu$-OGSSVM), the training time is further reduced because initial training with 100 subjects requires limited computational time (for instance, only 36.1 minutes are required in the case of PCA) and then a relatively small amount of time (22.3 minutes in the case of PCA) is required to train the remaining 228 subjects in online mode (thus the total training time for PCA + $2\nu$-OGSSVM is 58.4 minutes).

- Table 5.4 also shows that the testing time is considerably reduced when $2\nu$-OGSSVM is used as the classifier. The reason for this improved performance is same as for the reduced training time.

- As mentioned in Section 5.1, another possible advantage of online learning is to efficiently solve the offline training process in an online manner. To evaluate the appropriateness of this statement in biometric classifier training, a comparative analysis of verification accuracies obtained from $2\nu$-GSSVM and $2\nu$-OGSSVM is performed. Figure 5.6 shows that at the end of learning process with 328 subjects, verification accuracies obtained in the online learning mode converges to the accuracy obtained in the batch learning mode. Further, the time required for online learning is around half of the offline/batch training. Therefore, it is possible to apply online learning scheme in preference to offline training scheme without affecting the verification accuracy.

- Finally, t-test at 95% confidence shows that the $2\nu$-GSSVM is significantly different than the SVM classifier whereas there is no statistical difference between $2\nu$-GSSVM and $2\nu$-OGSSVM. However, as mentioned previously, the main advantage of $2\nu$-OGSSVM is reduced computational time and online classifier update.

The experiments demonstrate that the online learning approach reduces the computational cost without compromising the verification accuracy. Therefore, it is an effective alternative to traditional batch/offline learning methods.

Table 5.2: Covariate analysis of PCA and IPCA based verification algorithms with multiple classifiers.

| Feature Extraction and Classification | Covariate | | | |
|---|---|---|---|---|
| | Expression | Illumination | Pose | Overall |
| PCA [113] | 51.87 | 52.24 | 45.46 | 49.21 |
| PCA + SVM | 64.76 | 65.34 | 58.17 | 62.23 |
| PCA + $2\nu$-SVM | 64.98 | 65.83 | 58.41 | 62.95 |
| PCA + $2\nu$-GSSVM | 73.63 | 74.04 | 67.86 | 71.14 |
| PCA + $2\nu$-OGSSVM | 72.82 | 73.95 | 67.49 | 70.97 |
| IPCA [79] | 51.73 | 52.21 | 45.02 | 49.17 |
| IPCA + $2\nu$-OGSSVM | **73.67** | **74.11** | **67.92** | **71.22** |

Table 5.3: Covariate analysis of modified C2 based verification algorithm with multiple classifiers.

| Feature Extraction and Classification | Covariate | | | |
|---|---|---|---|---|
| | Expression | Illumination | Pose | Overall |
| C2 + SVM | 87.76 | 87.93 | 85.17 | 86.94 |
| C2 + $2\nu$-SVM | 87.79 | 88.02 | 85.59 | 87.21 |
| C2 + $2\nu$-GSSVM | 92.67 | 92.85 | 91.81 | 92.46 |
| C2 + $2\nu$-OGSSVM | **92.79** | **92.88** | **92.03** | **92.71** |
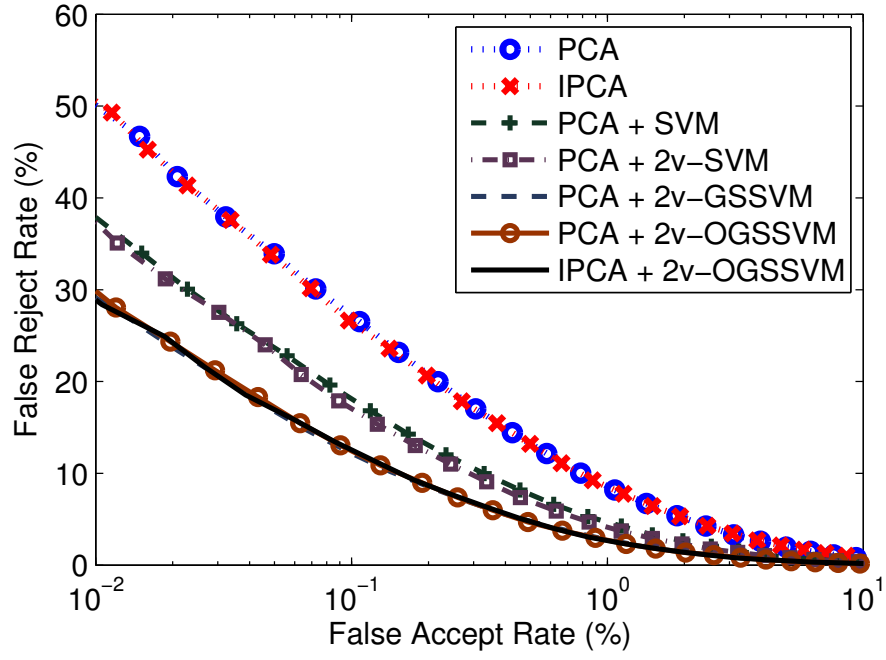
Figure 5.4: Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using appearance based PCA algorithm [113].
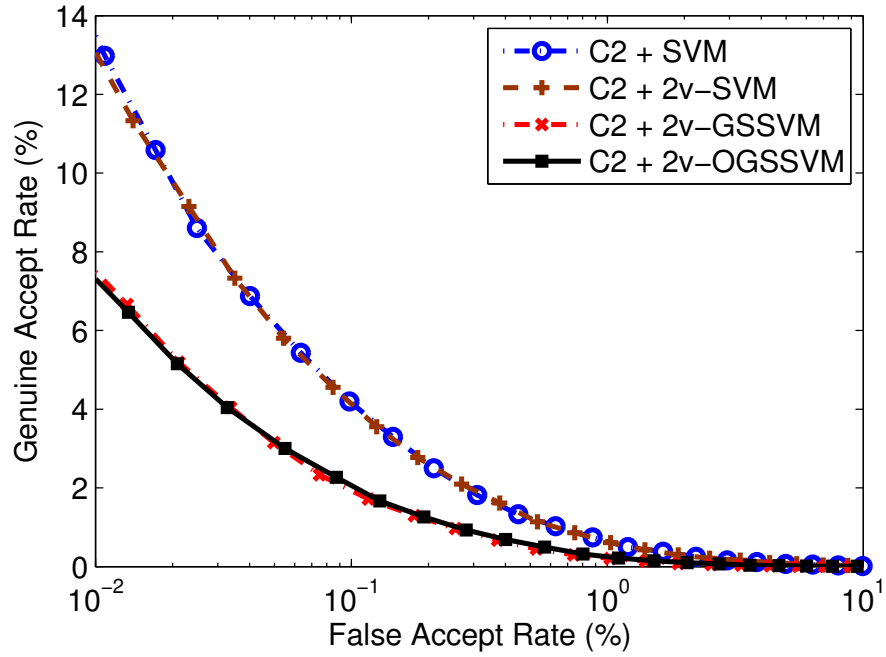


Figure 5.5: Comparing the performance of the proposed 2ν-OGSSVM (online classifier) with SVM and 2ν-GSSVM using local texture feature based modified C2 algorithm [98].
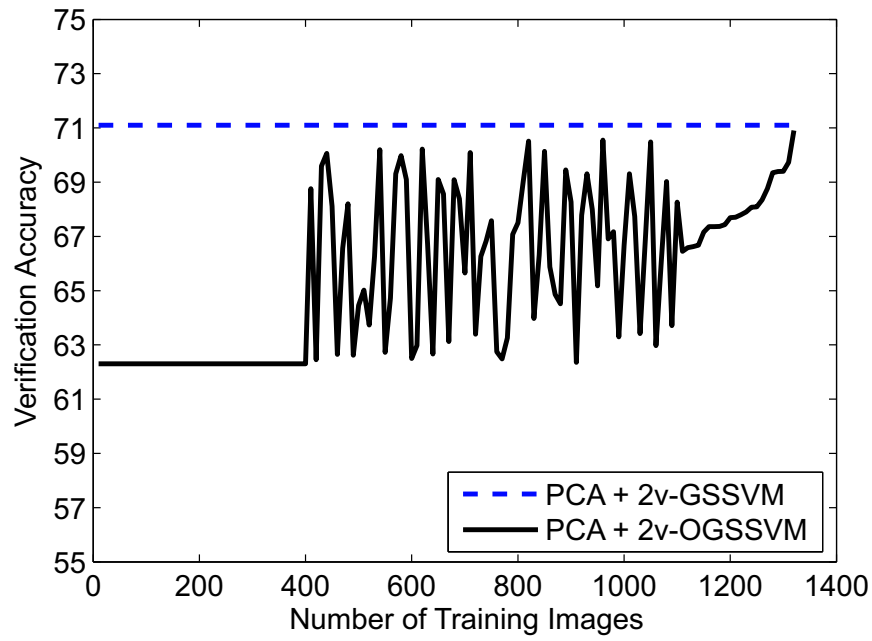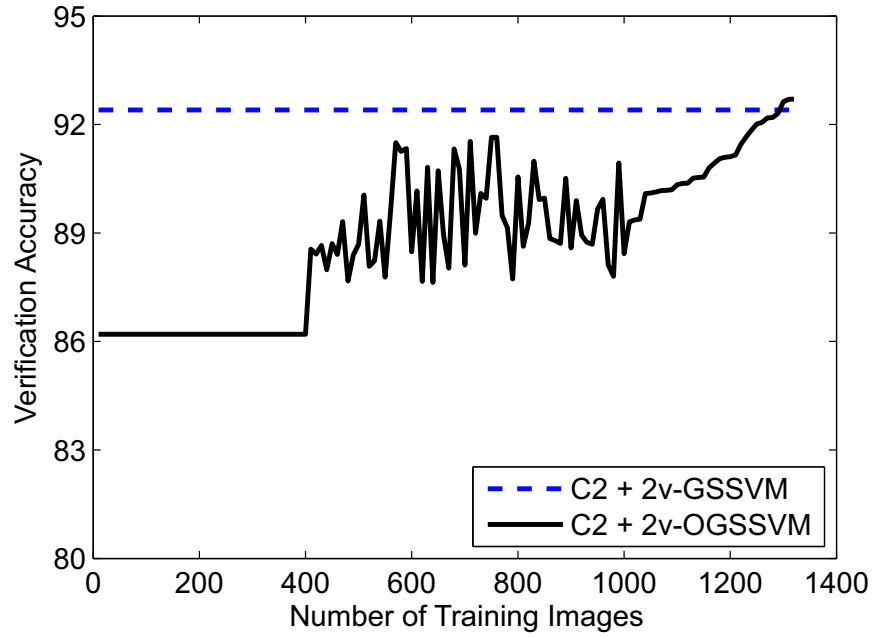
Figure 5.6: At the end of the learning process, the matching accuracy using online learning converges to the matching accuracy of the batch learning mode using (a) PCA and (b) the modified C2 algorithms.

Table 5.4: Computational time analysis for the proposed $2\nu$-OGSSVM and comparison with other classification approaches.

| Feature Extraction | Computation Time | |
|---|---|---|
| + Classification | Training Time (Minutes) | Testing Time (Seconds) |
| PCA [113] | 43.6 | 0.5 |
| PCA + SVM | 221.5 | 1.7 |
| PCA + $2\nu$-SVM | 194.8 | 1.4 |
| PCA + $2\nu$-GSSVM | 118.2 | 1.1 |
| PCA + $2\nu$-OGSSVM | 58.4 | 0.8 |
| IPCA [79] | 31.2 | 0.4 |
| IPCA + $2\nu$-OGSSVM | **52.5** | **0.7** |
| C2 + SVM | 336.8 | 2.1 |
| C2 + $2\nu$-SVM | 282.1 | 1.8 |
| C2 + $2\nu$-GSSVM | 192.4 | 1.6 |
| C2 + $2\nu$-OGSSVM | **109.7** | **1.2** |

## 5.4 Summary

Similar to template update, the parameters of the classifiers used in biometric system also require to be updated to accommodate for variations in data distribution. Current systems frequently re-train the algorithms using all enrolled subjects. This process may not be feasible for large scale systems where the number of newly enrolled subjects is significantly high. This chapter introduces the concept of online learning in biometrics to address the problem of classifier re-training and update. An online learning for $2\nu$-GSSVM is proposed to train the classifier in online mode so that it can update the decision hyperplane according to the newly enrolled subjects. This online classifier is used for feature classification and decision making in a face recognition system. On a heterogeneous NIR face database, the case study using PCA and modified C2 feature algorithms shows that the proposed online classifier significantly improves the verification performance both in terms of accuracy and computational time. Indeed, it is observed that the proposed online classifier is at least three times faster than the conventional SVM classifier.

# Chapter 6

# Conclusions and Future Work

## 6.1    Conclusions

Existing face recognition systems have demonstrated success in constrained settings with limited variability in illumination, pose, and expression. However, they are not sufficient to transcend the challenging applications such as recognizing missing persons or terrorists that require recognition of face images with aging and disguise variations. Further, large variations in pose, expression and illumination are challenges that severely affect the performance of face recognition algorithms. Along with these covariates, the computational complexity of classifier training and regular update is a challenging issue.

In this research, we first designed a face recognition algorithm that addresses two major challenges: *aging* and *disguise*. A novel age transformation algorithm is proposed to minimize the age difference between gallery and probe face images. Unlike the conventional method, we transform gallery face images with respect to probe face image using normalized mutual information based registration technique. Next, we designed a granular approach for face recognition in which the face image is decomposed at different levels of granularity to extract non-disjoint textural features from each face granule. At the first level, face granules are generated by applying Gaussian and Laplacian operators to extract features at different resolutions and image properties. The second level of granularity consists of vertical and horizontal granules of different sizes to specifically handle variations in pose and disguises. At the third level of granularity, the face image

is partitioned into small local structures to extract local features effectively. A four-layer neural network implementation of the 2D log polar Gabor transform is applied to each face granule to extract the phase features for matching. The matching scores of all face granules are effectively fused using likelihood ratio induced SVM match score fusion algorithm to arrive at a final decision. Evaluation is performed using five databases that comprises of the Notre Dame face database, FG-Net face database, disguised faces of real people, disguised synthetic face images, and disguised faces of actors and actresses from movie clips. The experimental results show that the proposed algorithm outperforms existing face recognition algorithms and is able to optimally recognize individuals despite minor to moderate level of aging and disguise variations.

We next described an image fusion algorithm that combines multispectral face images to provide invariance to illumination variations. Visible and long wave infrared images provide complementary properties that can be combined to improve the performance of face recognition. We proposed image fusion algorithms to fuse information obtained from multispectral face images. We first apply mutual information based registration algorithm to register multispectral face images and then fuse the images using the proposed algorithms. We proposed two fusion algorithms: DWT based fusion and $2\nu$-GSVM based fusion. DWT based image fusion algorithm is a simple approach whereas $2\nu$-GSVM based fusion algorithm is a learning based approach that assigns proper weights to multispectral images according to the activity level or entropy. The properties of both visible and long wave infrared images are preserved in the fused face image that can efficiently be used for face recognition. We evaluated the performance of the proposed image fusion algorithms using the Notre Dame and Equinox multispectral face databases. Experimental analysis shows that the proposed $2\nu$-GSVM based image fusion algorithm is able to sustain illumination variations that are usually encountered in visible spectral face recognition.

To account for the pose variations and to demonstrate the role of face mosaicing in enhancing the matching performance of a face recognition system, we proposed a multiresolution mosaicing scheme. Given multiple images of a face during enrollment, the

mosaicing algorithm combines them into a single entity by employing a multiresolution splining scheme. Experiments reported on three different databases suggest that fusing information at the image level is better than fusing information at the match-score level in the context of this work. A modified version of the C2 algorithm, originally developed by Serre et al. [87], is used to extract features from the mosaiced and non-mosaiced face images. The modified algorithm is observed to perform very well both in the verification and identification modes of operation.

Finally, we described a new biometric classifier update algorithm to incrementally re-train the classifier using online learning and to progressively establish a decision hyperplane for improved classification. The proposed algorithm incorporates soft labels and granular computing in the formulation of $2\nu$-Online GSSVM to retrain the classifier using only the new data. Granular computing approach makes it adaptive to local and global variations in data distribution, while soft labels provide resilience to noise. With new data, additional support vectors that are linearly independent are added and existing support vectors that do not improve the classifier performance are removed. This constrains the size of the support vectors and significantly reduces the training time without compromising the classification accuracy. The efficacy of the proposed online learning strategy is validated in a near-infrared face recognition application involving different covariates. The results obtained on a heterogeneous near-infrared face database show that in all experiments using different feature extraction and classification algorithms, the proposed $2\nu$-Online GSSVM learning approach was 2-3 times faster while achieving a high level of accuracy similar to offline training using all data.

## 6.2   Future Research Directions

We have designed reliable and fast approaches for face recognition to mitigate the effects of covariates such as aging, disguise, illumination, pose and computational time. We strongly believe that these approaches can pave the path of future research in face recognition to address non-ideal and challenging scenarios. We conclude this thesis

with possible future research directions that can be explored for further advancing and developing a highly accurate face recognition system.

- The problem becomes more complicated if we have to contend with variations in pose, expression and illumination along with aging and disguise altogether. This research can be further extended to address such type of non-ideal challenges of face recognition. However, undertaking such research would necessitate creating a comprehensive database that comprises of face images with pose, expression, illumination, aging and disguise variations.

- Image quality plays an important role in improving the recognition performance and it can be used to further extend the multispectral face image fusion algorithm. A carefully designed multispectral image quality assessment algorithm incorporated in the image fusion can further improve the recognition performance with variations in disguise, expression, and illumination.

- The role of mosaicing can be investigated in perturbing the biometric content of the human face. Also, the use of facial symmetry can be examined to create face mosaics when the frontal image is not available and only the left and right profiles are available.

- The emergence and viability of using online learning algorithms in the design of biometric classifiers addresses the real-time performance and scalability challenges, however more research is essential in order to fully understand the benefits in large-scale applications.

# Appendix

# Appendix A

# Dissemination of Research Results

1. R. Singh, M. Vatsa, and A. Noore, Face recognition with disguise and single gallery images, *Image and Vision Computing - Special Issue on Multimodal Biometrics*, Vol. 27, No. 3, pp. 245-257, 2009.

2. R. Singh, M. Vatsa, and A. Noore, Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition, *Pattern Recognition - Special Issue on Multimodal Biometrics*, Vol. 41, No. 3, pp. 880-893, 2008.

3. R. Singh, M. Vatsa, and A. Noore. Hierarchical fusion of multi-spectral face images for improved recognition performance, *Information Fusion*, Vol. 9, No. 2, pp. 200-210, 2008.

4. R. Singh, M. Vatsa, and A. Noore, Matching digital and scanned face images with age variation, *Computational Forensics, Series in Studies in Computational Intelligence*, Edited by K. Franke, S. Petrovic, and A. Abraham, Springer Verlag, 2009 (To appear).

5. R. Singh, M. Vatsa, and A. Noore, Recognizing face images with disguise variations, *Face Recognition*, Edited by M.S. Bartlett, K. Delac, and M. Grgic, ITECH Publishing, Chapter 11, pp. 149-160, 2009.

6. A. Noore, R. Singh, and M. Vatsa, Sensor level fusion, *Encyclopedia of Biometric Recognition*, Edited by S.Z. Li, Springer, 2008 (To appear).

7. M. Vatsa, R. Singh, and A. Noore, SVM based adaptive biometric image enhancement using quality assessment, *Speech, Audio, Image and Biomedical Signal Processing using Neural Networks*, Edited by B. Prasad and S.R.M. Prasanna, Springer Verlag, Chapter 16, pp. 351-372, 2008.

8. R. Singh, M. Vatsa, and A. Noore, Multiclass m$\nu$-granular soft support vector machine: a case study in dynamic classifier selection for multispectral face recognition, *Proceedings of IAPR IEEE International Conference on Pattern Recognition*, 2008 (To appear).

9. M. Vatsa, R. Singh, A. Ross, and A. Noore, Likelihood ratio in a SVM framework: fusing linear and non-linear face classifiers, *Proceedings of IEEE Computer Society Workshop on Biometrics at the Computer Vision and Pattern Recognition Conference*, pp. 1-6, 2008.

10. R. Singh, M. Vatsa, A. Ross, and A. Noore. A mosaicing scheme for pose-invariant face recognition, *IEEE Transactions on Systems, Man and Cybernetics - Part B*, Vol. 37, No. 5, pp. 1212-1225, 2007.

11. Singh, R., Vatsa, M., Noore, A.: Improving verification accuracy by synthesis of locally enhanced biometric images and deformable model, *Signal Processing*, Vol. 87, No. 11, pp. 2746-2764, 2007.

12. R. Singh, M. Vatsa, A. Noore, and S.K. Singh, Age transformation for improving face recognition performance, *Proceedings of Second International Conference on Pattern Recognition and Machine Intelligence*, LNCS Vol. 4815, pp. 576-583, 2007.

13. R. Singh, M. Vatsa, and A. Noore, Textural feature based face recognition for single training images, *IEE Electronics Letters*, Vol. 41, No. 11, pp. 23-24, 2005.

14. R. Singh, M. Vatsa, A. Ross, and A. Noore, Performance enhancement of 2D face recognition via mosaicing, *Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 63-68, 2005 **Best Paper Award**.

15. R. Singh, M. Vatsa, and A. Noore, Recognition of faces with variations in disguise, *Proceedings of Summer School on Biometric Authentication: Multimodality and System Integration*, 2005, **Best Paper Award**.

# Bibliography

[1] T. Ahonen, A. Hadid, and M. Pietikinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 12, pp. 2037-2041, 2006.

[2] J. Alexander and J. Smith, "Engineering privacy in public: Confounding face recognition, privacy enhancing technologies," *Proceedings of International Workshop on Privacy Enhancing Technologies*, pp. 88–106, 2003.

[3] K. Anderson and P. McOwan, "Robust real-time face tracker for cluttered environments," *Computer Vision and Image Understanding*, Vol. 95, pp. 184–200, 2004.

[4] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using the wavelet transform," *IEEE Transaction on Image Processing* Vol. 1, No. 2, pp. 205-220, 1992.

[5] P. Auer, N. Cesa-Bianchi, and C. Gentile, "Adaptive and self-confident on-line learning algorithms," *Journal of Computer and System Sciences*, Vol. 64, pp. 48–75, 2002.

[6] O. Ayinde and Y.H. Yang, "Face recognition approach based on rank correlation of Gabor-filtered images," *Pattern Recognition*, Vol. 35, No. 6, pp. 1275-1289, 2002.

[7] A. Bargiela and W. Pedrycz, "Granular computing: an introduction," *International Series in Engineering and Computer Science*, 2002.

[8] A. Bargiela and W. Pedrycz, "The roots of granular computing," *In Proceedings of IEEE International Conference on Granular Computing*, pp. 806-809, 2006.

[9] G. Bebis, A. Gyaourova, S. Singh, and I. Pavlidis, "Face recognition by fusing thermal infrared and visible imagery," *Image and Vision Computing*, Vol. 24, No. 7, pp. 727-742, 2006.

[10] J. Bigun and J.M. Du Buf, "N-folded symmetries by complex moments in Gabor space and their applications to unsupervised texture segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 1, pp. 80-87, 1994.

[11] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711-720, 1997.

[12] V. Blanz, S. Romdhami, and T. Vetter, "Face identification across different poses and illuminations with a 3D morphable model," *In Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 202-207, 2002.

[13] K.W. Bowyer, K. Chang, and P. Flynn, "A survey of 3D and multi-modal 3D+2D face recognition," *Notre Dame Department of Computer Science and Engineering Technical Report*, 2004.

[14] P.J. Burt and E.H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transaction on Graphics*, Vol. 2, No. 4, pp. 217-236, 1983.

[15] D.M. Burt and D.I. Perrett, "Perception of age in adult caucasian male faces: computer graphic manipulation of shape and colour information," *In Proceedings of Royal Society London - Series B*, Vol. 259, pp. 137–143, 1995.

[16] R. Campbell, M. Coleman, W. Michael, B. Jane, J. Philip, S. Wallace, J. Michelotti, and S. Baron-Cohen, "When does the inner-face advantage in familiar face recognition arise and why?" *Visual Cognition*, Vol. 6, No. 2, pp. 197–216, 1999.

[17] G. Cauwenberghs and T. Poggio, "Incremental and decremental support vector machine learning," *In Proceedings of International Conference on Neural Information Processing Systems*, pp. 409–415, 2000.

[18] CBSR NIR face dataset http://www.cse.ohio-state.edu/otcbvs-bench/.

[19] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "An evaluation of multi-modal 2D+3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, pp. 619-624, 2005.

[20] P.H. Chen, C.J. Lin, and B. Schlkopf, "A tutorial on $\nu$-Support Vector Machines," *Applied Stochastic Models in Business and Industry*, Vol. 21, pp. 111-136, 2005.

[21] Y.H. Chen and Y.Y. Yao, "Multiview intelligent data analysis based on granular computing," *In Proceedings of IEEE International Conference on Granular Computing*, pp. 281-286, 2006.

[22] X. Chen, P.J. Flynn, and K.W. Bowyer, "Visible-light and infrared face recognition," *In Proceedings of ACM Workshop on Multimodal User Authentication*, pp. 48-55, 2003.

[23] X. Chen, P. Flynn, and K. Bowyer, "IR and visible light face recognition," *Journal of Computer Vision and Image Understanding*, Vol. 99, No. 3, pp. 332-358, 2005.

[24] H. Chew, C. Lim, and R. Bogner, "An implementation of training dual-$\nu$ support vector machines," *Optimization and Control with Applications*, L. Qi, K. Teo, and X. Yang (Editors), 2004.

[25] J.G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, Vol. 20, No. 10, pp. 847-856, 1980.

[26] J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *Journal of the Optical Society of America*, Vol. 2, No. 7, pp. 1160-1169, 1985.

[27] J.G. Daugman, "Representational issues and local filter models of two-dimensional spatial visual encoding," *Models of the Visual Cortex*, D. Rose and V.G. Dobson (Editors), John Wiley and Sons, pp. 96-107, 1985.

[28] J.G. Daugman, "An information-theoretic view of analog representation in striate cortex," *Computational Neuroscience* E.L. Schwartz (Editor), MIT Press, pp. 403-423, 1988.

[29] J.G. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 7, pp. 1169-1179, 1988.

[30] B. Duc, S. Fischer, and J. Bigun, "Face authentication with Gabor information on deformable graphs," *IEEE Transactions on Image Processing*, Vol. 8, No. 4, pp. 504-516, 1999.

[31] R.O. Duda, P.E. Hart, and D.G. Stork, "Pattern classification," $2^{nd}$ *Edition, Wiley*, 2000.

[32] Equinox face database: http://www.equinoxsensors.com/products/HID.html.

[33] FACES software: http://www.iqbiometrix.com/products_faces_40.html.

[34] D.J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *Journal of the Optical Society of America*, Vol. 4, pp. 2379-2394, 1987.

[35] FG-Net Aging Database: http://www.fgnet.rsunit.com/.

[36] R. Fletcher, "Practical methods of optimization," $2^{nd}$ *Edition, Wiley*, 1987.

[37] P.J. Flynn, K.W. Bowyer, and P.J. Phillips, "Assessment of time dependency in face recognition: an initial study," *In Proceedings of International Conference on Audio and Video-Based Biometric Person Authentication*, pp. 44-51, 2003.

[38] D. Gabor, "Theory of communication," *Journal of IEE*, Vol. 93, pp. 429-457, 1946.

[39] M. Gandhi, "A method for automatic synthesis of aged human facial images," *Masters Thesis, Department of Electrical and Computer Engineering*, McGill University, 2004.

[40] R. Gross and V. Brajovic, "An image preprocessing algorithm for illumination invariant face recognition," *Proceedings of the 4th International Conference on Audio- and Video-Based Biometric Person Authentication*, Vol. 2688, pp. 10-18, 2003.

[41] A. Gyaourova, G. Bebis, and I. Pavlidis, "Fusion of infrared and visible images for face recognition," *Lecture Notes in Computer Science*, Vol. 3024, pp. 456-468, 2004.

[42] W. Hayward, G. Rhodes, and A. Schwaninger, "An own-race advantage for components as well as configurations in face recognition," *Cognition*, Vol. 106, No. 2, pp. 1017–1027, 2008.

[43] J. Heo, S. Kong, B. Abidi, and M. Abidi, "Fusion of visual and thermal signatures with eyeglass removal for robust face recognition," *In Proceedings of IEEE Workshop on Object Tracking and Classification Beyond the Visible Spectrum in conjunction with CVPR* pp. 94-99, 2004.

[44] D. Hill, C. Studholme, and D. Hawkes, "Voxel similarity measures for automated image registration," *In Proceedings of Third SPIE Conference on Visualization in Biomedical Computing*, pp. 205-216, 1994.

[45] Y. Hirose, K. Yamashita, and S. Hijiya, "Backpropagation algorithm which varies the number of hidden units," *Neural Networks*, Vol. 4, pp. 61-66, 1991.

[46] P. Kalocsai, H. Neven, and J. Steffens, "Statistical analysis of Gabor-filter representation," *In Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 360-365, 1998.

[47] S.B. Kang, "A survey of image-based rendering techniques," *Cambridge Research Laboratory Technical Report*, CRL 97/4, 1997.

[48] H. Kang, T.F. Cootes, and C.J. Taylor, "A comparison of face verification algorithms using appearance models," *In Proceedings of British Machine Vision Conference*, Vol.2, pp. 477-4862, 2002.

[49] S. Kim, S.-T. Chung, S. Jung, and S. Cho, "An improved illumination normalization based on anisotropic smoothing for face recognition," *International Journal of Computer Science and Engineering*, Vol. 2, No. 3, pp. 100-106, 2008.

[50] J. Kivinen, A. Smola, and R. Williamson, "Online learning with kernels," *IEEE Transactions on Signal Processing*, Vol. 52, No. 8, pp. 2165–2176, 2004.

[51] S. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, "Recent advances in visual and infrared face recognition - A review," *Journal of Computer Vision and Image Understanding*, Vol. 97, No. 1, pp. 103-135, 2005.

[52] S.G. Kong, J. Heo, F. Boughorbel, Y. Zheng, B.R. Abidi, A. Koschan, M.Yi, and M.A. Abidi, "Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition," *International Journal of Computer Vision*, Vol. 71, No. 2, pp. 215-233, 2007.

[53] A. Lanitis, C.J. Taylor, T.F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, pp. 442–455, 2002.

[54] A. Lanitis, "On the significance of different facial parts for automatic age estimation," *In Proceedings of International Conference on Digital Signal Processing*, Vol. 2, pp. 1027–1030, 2002.

[55] A. Lanitis, C. Draganova, C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 34, No. 1, pp. 621–628, 2004.

[56] S.Z. Li, R.F. Chu, M. Ao, L. Zhang, and R. He, "Highly accurate and fast face recognition using near infrared images," *In Proceedings of IAPR International Conference on Biometrics*, pp. 151-158, 2006.

[57] S. Li, R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 4, pp. 627–639, 2007.

[58] S.Z. Li and A.K. Jain, "Handbook of Face Recognition," *Springer*, 2005.

[59] X. Liu and T. Chen, "Geometry-assisted statistical modeling for face mosaicing," *In Proceeding of IEEE International Conference on Image Processing*, Vol. 2, pp. 883-886, 2003.

[60] X. Liu and T. Chen, "Pose-robust face recognition using geometry assisted probabilistic modeling," *In Proceedings of International Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 502-509, 2005.

[61] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model (EFM) for face recognition," *IEEE Transactions on Image Processing*, Vol. 11, No. 4, pp. 467-476, 2002.

[62] C. Liu and H. Wechsler, "Independent component analysis of Gabor features for face recognition," *IEEE Transactions on Neural Networks*, Vol. 14, No. 4, pp. 919-928, 2003.

[63] X. Lu, Y. Wang, and A.K. Jain, "Combining classifiers for face recognition," *In Proceedings of IEEE International Conference on Multimedia & Expo*, Vol. 3, pp. 13-16, 2003.

[64] B. MacLennan, "Gabor representations of spatiotemporal visual images," *University of Tennessee, Technical Report CS-91-144*, 1994.

[65] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging*, Vol. 16, No. 2, pp. 187-198, 1997.

[66] A.M. Martinez and R. Benavente, "The AR face database," *CVC Technical Report #24*, 1998.

[67] K. Nandakumar, Y. Chen, S.C. Dass, and A.K. Jain, "Likelihood ratio based biometric score fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No.2, pp. 342–347, 2008.

[68] C. Palm and T.M. Lehmann, "Classification of color textures by gabor filtering," *Machine Graphics and Vision*, Vol. 11, No. 2/3, pp. 195-219, 2002.

[69] S. Pamudurthy, E. Guan, K. Mueller, and M. Rafailovich, "Dynamic approach for face recognition using digital image skin correlation," *In Proceedings of Audio- and Video-based Biometric Person Authentication*, pp. 1010–1018, 2005.

[70] U. Park, Y. Tong, and A. K. Jain, "Face Recognition with Temporal Invariance: A 3D Aging Model", *In Proceedings of International Conference on Automatic Face and Gesture Recognition*, 2008 (To Appear).

[71] P. Penev and J. Atick, "Local feature analysis: A general statistical theory for object representation" *Network: Computation in Neural Systems*, Vol. 7, No. 3, pp. 477-500, 1996.

[72] S. Periaswamy and H. Farid , "Elastic Registration in the Presence of Intensity Variations," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 7, pp. 865-874, 2003.

[73] P.J. Phillips, P. Grother, R.J Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation Report," 2003.

[74] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, and W. Worek, "Preliminary face recognition grand challenge results," *In Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 15–24, 2006.

[75] P. Phillips, W. Scruggs, A. OToole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe, "FRVT 2006 and ICE 2006 large-scale results," *NIST Technical Report NISTIR 7408*, 2007.

[76] J.P. Pluim, J.B.A. Maintz, and M.A. Viergever, "Mutual information-based registration of medical images: a survey," *IEEE Transactions on Medical Imaging*, Vol. 22, No. 8, pp. 986-1004, 2003.

[77] F. Prokoski, "History, current status, and future of infrared identification," *In Proceedings of IEEE Workshop on Computer Vision beyond the Visible Spectrum: Methods and Applications*, pp. 5-14, 2000.

[78] C. M. Pun, "Rotation invariant texture feature for image retrieval," *Computer Vision and Image Understanding*, Vol. 89, pp. 24-43, 2003.

[79] B. Raducanu and J. Vitria, "Online nonparametric discriminant analysis for incremental subspace learning and recognition," *Pattern Analysis and Applications*, 2008 (To Appear).

[80] N. Ramanathan, A.R. Chowdhury, and R. Chellappa, "Facial similarity across age, disguise, illumination and pose," *In Proceedings of International Conference on Image Processing*, Vol. 3, pp. 1999-2002, 2004.

[81] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," *In Proceedings of IEEE Computer Vision and Pattern Recognition*, Vol. 1, pp. 387–394, 2006.

[82] N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Transactions on Image Processing*, Vol. 15, No. 11, pp. 3349–3362, 2006.

[83] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, Vol. 2, pp. 1019-1025, 1999.

[84] A. Rios and M. Kabuka, "Image compression with a dynamic autoassociative network," *In Proceedings of Artificial Neural Networks in Engineering*, pp. 503-510, 1992.

[85] A. Ross and A.K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, Vol. 24, No. 13, pp. 2115-2125, 2003.

[86] A. Schwaninger, J. Lobmaier, and S. Collishaw, "Role of featural and configural information in familiar and unfamiliar face recognition," *In Proceedings of Second International Workshop on Biologically Motivated Computer Vision*, pp. 245–258, 2002.

[87] T. Serre, L. Wolf, and T. Poggio. "Object recognition with features inspired by visual cortex," *In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 994-1000, 2005.

[88] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp.157, 2003.

[89] D. Shan and R. Ward, "Wavelet-based illumination normalization for face recognition," Vol. 2, pp. 954-957, 2005.

[90] L. Shen and L. Bai, "A review on Gabor wavelets for face recognition," *Pattern Analysis and Applications*, Vol. 9, No. 2, pp. 273-292, 2006.

[91] P.Q. Silva and A.N.C. Santa Rosa, "Face recognition based on eigeneyes," *Pattern Recognition and Image Analysis*, Vol. 13, No. 2, pp. 335-338, 2003.

[92] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 12, pp. 1615-1618, 2003.

[93] S.K. Singh, D.S. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Tamkang Journal of Science and Engineering*, Vol. 6, No. 4, pp. 227-234, 2003.

[94] S. Singh, A. Gyaourova, G. Bebis, and I. Pavlidis, "Infrared and visible image fusion for face recognition," *In Proceedings of SPIE Defense and Security Symposium (Biometric Technology for Human Identification)*, Vol. 5404, pp. 585-596, 2004.

[95] R. Singh, M. Vatsa, and A. Noore, "Intelligent biometric information fusion using support vector machine," *Soft Computing in Image Processing: Recent Advances, Springer* M. Nachtegael, D. Van der Weken, E.E. Kerre, and W. Philips (Editors), Chapter 12, pp. 327-350, 2006.

[96] R. Singh, M. Vatsa, A. Ross, and A. Noore, "Performance enhancement of 2D face recognition via mosaicing," *In Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, pp. 63-68, 2005.

[97] R. Singh, M. Vatsa, and A. Noore, "Face recognition with disguise and single gallery images," *Image and Vision Computing*, 2008 (To Appear).

[98] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, Vol. 37, No. 5, pp. 1212–1225, 2007.

[99] R. Singh, M. Vatsa, and A. Noore, "Hierarchical fusion of multi-spectral face images for improved recognition performance," *Information Fusion*, Vol. 9, No. 2, pp. 200–210, 2008.

[100] R. Singh, M. Vatsa, and A. Noore, "Improving verification accuracy by synthesis of locally enhanced biometric images and deformable model," *Signal Processing*, Vol. 87, No. 11, pp. 2746–2764, 2007.

[101] P. Sinha, B.J. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: 19 results all computer vision researchers should know about," *In Proceedings of IEEE*, Vol. 94, No. 11, pp. 1948–1962, 2006.

[102] D. Skocaj and A. Leonardis, "Incremental and robust learning of subspace representations," *Image and Vision Computing*, Vol. 26, No. 1, pp. 27–38, 2008.

[103] F. Smeraldi and J. Bigun, "Retinal vision applied to facial features detection and face authentication," *Pattern Recognition Letters*, Vol. 23, pp. 463-475, 2002.

[104] D. Socolinsky, A. Selinger, and J. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Journal of Computer Vision and Image Understanding*, Vol. 91, pp. 72-114, 2003.

[105] Y.C. Tang, B. Jin and Y.-Q. Zhang, "Granular support vector machines with association rules mining for protein homology prediction," *Special Issue on Computational Intelligence Techniques in Bioinformatics, Artificial Intelligence in Medicine*, Vol. 35, No. 1/2, pp. 121-134, 2005.

[106] Y.C. Tang and Y.-Q. Zhang, "Granular support vector machines with data cleaning for fast and accurate biomedical binary classification," *In Proceedings of International Conference on Granular Computing*, pp. 262-265, 2005.

[107] Y.C. Tang, B. Jin, Y.-Q. Zhang, H. Fang, and B. Wang, "Granular support vector machines using linear decision hyperplanes for fast medical binary classification," *In Proceedings of International Conference on Fuzzy Systems*, pp. 138-142, 2005.

[108] Y.C. Tang and Y.-Q. Zhang, "Granular SVM with repetitive undersampling for highly imbalanced protein homology prediction," *In Proceedings of International Conference on Granular Computing*, pp. 457-461, 2006.

[109] Q. Tao, G. Wu, F. Wang, and J. Wang. "Posterior probability support vector machines for unbalanced data," *IEEE Transactions on Neural Network*, Vol. 16, No. 6, pp. 1561–1573, 2005.

[110] D. Tax and P. Laskov, "Online SVM learning: from classification to data description and back," *In Proceedings of IEEE International Workshop on Neural Networks for Signal Processing*, pp. 499–508, 2003.

[111] B. Tiddeman, M. Burt, and D. Perrett, "Prototyping and transforming facial textures for perception research," *IEEE Computer Graphics and Applications*, Vol. 21, No. 5, pp. 42–50, 2001.

[112] M. Tistarelli and E. Grosso, "Active vision-based face recognition issues, applications and techniques," *Nato-ASI Advanced Study on Face Recognition, Springer*, H. Wechsler (Editors), Vol. F-163, pp. 262-286, 1998.

[113] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neurosicence*, Vol 3, No. 1, pp. 71–86, 1991.

[114] U. Uludag, A. Ross, and A.K. Jain, "Biometric template selection and update: A case study in fingerprints," *Pattern Recognition*, Vol. 37, No. 7, pp. 1533-1542, 2004.

[115] S. Venkatesh and R.A. Owens, "Energy feature detection scheme," *In Proceedings of International Conference on Image Processing*, pp. 553-557, 1989.

[116] V. Vapnik, "The nature of statistical learning theory," *Springer*, 1995.

[117] J.D. Villasenor, B. Belzer, and J. Liao, "Wavelet filter evaluation for image compression," *IEEE Transaction on Image Processing*, Vol. 4, No. 8, pp. 1053-1060, 1995.

[118] J. Wang, Y. Shang, G. Su, X. Lin, "Age simulation for face recognition," *In Proceedings of International Conference on Pattern Recognition*, pp. 913–916, 2006.

[119] M. Webster and R. de Valois, "Relationship between spatial-frequency and orientation tuning of striate-cortex cells," *Journal of the Optical Society of America*, Vol. 2, pp. 1124-1132, 1985.

[120] H. Wechsler, "Reliable Face Recognition Methods: System Design, Implementation and Evaluation," *Springer*, 2006.

[121] J. Wilder, J. Phillips, C. Jiang, and S. Wiener, "Comparison of visible and infrared imagery for face recognition," *In Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 182-187, 1986.

[122] L. Wiskott, J.M. Fellous, N. Krger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 775-779, 1997.

[123] C. Xu and J.L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Transactions on Image Processing*, Vol. 7, No. 3, pp. 359-369, 1998.

[124] Yale face database http://cvc.yale.edu/projects/yalefaces/yalefaces.html.

[125] F. Yang, M. Paindavoine, H. Abdi, and A. Monopoly, "Development of a fast panoramic face mosaicing and recognition system," *Optical Engineering*, Vol. 44, No. 8, pp. 087005/1-087005/10, 2005.

[126] Y.Y. Yao, "Perspectives of granular computing," *In Proceedings of IEEE International Conference on Granular Computing*, Vol. 1, pp. 85-90, 2005.

[127] Y. Yu and R. Simmons, "Descending epsilon in back-propagation: a technique for better generalization," *In Proceedings of International Joint Conference on Neural Networks*, Vol. 3, pp. 167-172, 1990.

[128] B. Zhang, S. Shan, X. Chen, and Wen Gao, "Histogram of gabor phase patterns (HGPP): a novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, Vol. 16, No. 1, pp. 57-68, 2007.

[129] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: a literature survey," *ACM Computing Surveys*, Vol. 12, No. 4, pp. 399-458, 2003.

[130] Z. Zivkovic and F. Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 5, pp. 651–656, 2004.