



# **Automatic Facial Age Estimation**

Thesis submitted for the degree of  
Doctor of Philosophy

Muhammad Aurangzeb Khan



School of Computing and Communications

Lancaster University

October, 2015

[This page is intentionally left blank]

Dedicated to my dear wife Dr. Aqsa and lovely son Ahmed  
for their endless love and support.

## **Acknowledgements**

I would like to express the sincerest and deepest gratitude to my supervisor Prof. Costas Xydeas for his incessant support, guidance and useful suggestions. Without his encouragement and positive attitude towards my efforts this dissertation would not have been accomplished.

Additionally, I would like to acknowledge Dr. Hassan Ahmed for being very supportive and for his very useful suggestions. Also all my friends Dr. Asmar, Dr. Amir, Dr. Asad, Haris, Izhar, Noor and Dr. Haroon who made my stay at Lancaster University so wonderful.

There are no words that can express my feelings for the love and support my family provided me during all these years.

## **Preface**

The work presented in this thesis has been carried out in the School of Computing and Communications at Lancaster University under the supervision of Prof. Costas Xydeas and Dr. Hassan Ahmed. This dissertation has not been submitted in support of an application for another degree at this or any other university. It is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated.

*Muhammad Aurangzeb Khan*

## **Abstract**

The reliability of automatically estimating human ages, by processing input facial images, has generally been found to be poor. On other hand, various real world applications, often relating to safety and security, depend on an accurate estimate of a person's age. In such situations, Face Image based Automatic Age Estimation (FI-AAE) systems which are more reliable and may ideally surpass human ability, are of importance as and represent a critical pre-requisite technology. Unfortunately, in terms of estimation accuracy and thus performance, contemporary FI-AAE systems are impeded by challenges which exist in both of the two major FI-AAE processing phases i.e. i) Age based feature extraction and representation and ii) Age group classification. Challenges in the former phase arise because facial shape and texture change independently and the magnitude of these changes vary during the different stages of a person's life. Additionally, contemporary schemes struggle to exploit age group specific characteristics of these features, which in turn has a detrimental effect on overall system performance. Furthermore misclassification errors which occur in the second processing phase and are caused by the smooth inter-class variations often observed between adjacent age groups, pose another major challenge and are responsible for low overall FI-AAE performance.

In this thesis a novel Multi-Level Age Estimation (ML-AE) framework is proposed that addresses the aforementioned challenges and improves upon state-of-the-art FI-AAE system performance. The proposed ML-AE is a hierarchical classification scheme that maximizes and then exploits inter-class variation among different age groups at each level of the

hierarchy. Furthermore, the proposed scheme exploits age based discriminating information taken from two different cues (i.e. facial shape and texture) at the decision level which improves age estimation results.

During the process of achieving our main objective of age estimation, this research work also contributes to two associated image processing/analysis areas: i) Face image modeling and synthesis; a process of representing face image data with a low dimensionality set of parameters. This is considered as precursor to every face image based age estimation system and has been studied in this thesis within the context of image face recognition ii) measuring face image data variability that can help in representing/ranking different face image datasets according to their classification difficulty level. Thus a variability measure is proposed that can also be used to predict the classification performance of a given face recognition system operating upon a particular input face dataset.

Experimental results based on well-known face image datasets revealed the superior performance of our proposed face analysis, synthesis and face image based age classification methodologies, as compared to that obtained from conventional schemes.

# Table of Contents

<b>Acknowledgements.....</b>	<b>II</b>
<b>Preface.....</b>	<b>III</b>
<b>Abstract.....</b>	<b>IV</b>
<b>1 Introduction.....</b>	<b>1</b>
1.1 Motivation of Research Work.....	1
1.2 Automatic Age Estimation (AAE).....	3
1.2.1 Feature Extraction Process.....	3
1.2.2 Age Estimation Process (ASP).....	7
1.3 Facial Aging Dataset.....	8
1.4 Research Objectives.....	10
1.5 Author's Contributions.....	11
1.5.1 Face Image Modeling and Synthesis.....	11
1.5.2 Data Variability Measure.....	12
1.5.3 Hierarchical Classification Framework for Automatic Age Estimation.....	13

Table of Contents	2
1.6 Thesis Organization.....	13
<b>2 Related Research Work and Background Material .....</b>	<b>15</b>
2.1 Related Research Work.....	15
2.1.1 Feature Extraction and Representation .....	15
2.1.2 Face Image based Age Estimation System Architectures.....	20
2.2 AAM Review .....	22
2.2.1 Active Appearance Model (AAM) .....	22
<b>3 Face Image Modeling And Synthesis .....</b>	<b>26</b>
3.1 Multi-Model AAM (MM-AAM) .....	30
3.1.1 MM-AAM Modeling Phase.....	30
3.1.2 MM-AAM Synthesis Phase .....	33
3.2 Multi-Component/Multi-Model AAM (MC/MM-AAM).....	34
3.2.1 MC/MM-AAM Modeling Phase.....	34
3.2.2 MC/MM-AAM Synthesis Phase.....	37
3.3 Experimental Results and Discussion .....	40
3.3.1 Impact of Number of Clusters.....	40
3.3.2 Synthesis Performance Comparisons.....	42

Table of Contents	3
3.4 MM-AAM as Applied to Face Recognition.....	46
3.4.1 FR-MM-AAM System Training.....	48
3.4.2 FR-MM-AAM System Testing.....	49
3.4.3 Experimental Results and Discussion.....	50
3.5 Conclusion.....	56
<b>4 Face Image Data Variability.....</b>	<b>58</b>
4.1 Variability measure (VM).....	64
4.1.1 VM-intra- and VM-inter-Subject Class Components.....	64
4.1.2 Variability Measure (VM).....	70
4.2 Experimentation & Discussion.....	71
4.2.1 Face datasets.....	71
4.2.2 Face Recognition (FR) Systems.....	74
4.2.3 Results and Discussion.....	76
4.3 Conclusion.....	86
<b>5 Face Image based Automatic Age Estimation.....</b>	<b>89</b>
5.1 Multi-level Age Estimation (ML-AE) Framework.....	93
5.1.1 System Training.....	95

Table of Contents	4
5.1.2 System Testing.....	101
5.2 Experiments.....	103
5.2.1 Input Dataset .....	103
5.2.2 Experimental Setup .....	103
5.2.3 Experimental Results .....	105
5.3 Conclusion.....	111
<b>6 Concluding Remarks and Future Work.....</b>	<b>113</b>
6.1 Conclusions .....	113
6.1.1 Face Image Modeling and Synthesis .....	114
6.1.1 Face image data variability .....	115
6.1.3 Face image based Automatic Age Estimation (FI-AAE) .....	116
6.2 Future Work .....	117
<b>Bibliography .....</b>	<b>119</b>

# List of Figures

Figure 1.1: Block diagram of a typical automatic age estimation system. ....	3
Figure 1.2: Aging effects on a human facial appearance. ....	3
Figure 1.3: Craniofacial growth (shape change) on a human face with age progression, originally given in [1] and then [3]. ....	4
Figure 1.4: Skin aging (Texture change) with age progression, originally given in [1, 3]. ....	4
Figure 1.5: Comparison of the original image and the reconstructed image from the appearance parameter of the AAM, originally given in [2]. ....	6
Figure 3.1: System Diagram of proposed MM-AAM. Here $C_1, C_2, \dots, C_P$ are $P$ clusters of face dataset, which are used to produce $P$ model matrices $R_P$ . ....	31
Figure 3.2: Some sample face images with their superimposed shape coordinates (red dots) and corresponding facial texture. ....	31
Figure 3.3: Example shapes of face images taken from three different clusters: image shown in one row are representing one cluster. Intra-cluster similarities and inter-cluster variations, in terms of both pose and expression, can be observed. ....	32
Figure 3.4: System Diagram of proposed MC/MM-AAM. Here $F_1, F_2, F_3, \dots, F_N$ are $N = 4$ components-based datasets and $C_{nk}$ 's are corresponding component clusters. These	

are used to produce component based model matrices  $Rnk$ 's. The number of clusters employed for each component can be the same i.e.  $p = q = r = 8$  or it can differ.....36

Figure 3.5: Example shapes of four components i.e. cheeks + eyebrows, nose, mouth and eyes. ....38

Figure 3.6: Example shapes of cheeks + eyebrows component taken from two different clusters: first rows shapes belong to one cluster and second row belong to another cluster. Intra-cluster similarities and inter-cluster variations can be observed here as well. ....38

Figure 3.7: Error Vs No. of Clusters. (a) Error between synthesized shape coordinates and ground truth coordinates averaged across 108 test images for each value of  $p$ , (b) error between synthesized texture intensities and ground truth texture intensities averaged across 108 test images. ....39

Figure 3.8: Sample synthesized shapes for “Seen Dataset”; first column in red is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. The difference between results obtained using MM-AAM and MC/MM-AAM is not large but they both significantly outperformed conventional AAM.....44

Figure 3.9: Sample synthesized shapes for “Unseen Dataset”; first column in red is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. Again MM-AAM and MC/MM-AAM significantly outperformed conventional AAM. ....44

Figure 3.10: Sample synthesized textures; first column is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. MM-AAM and

MC/MM-AAM outperformed conventional AAM, but the difference between the two new schemes is not significant. ....45

Figure 3.11: FR-MM-AAM Training. This involves the synthesis of face image shapes for the complete training image dataset  $F$ , using the MM-AAM procedures. MM-AAM training is a precursor to FR-MM-AAM training. ....47

Figure 3.12: Examples of synthesized shapes and their corresponding extracted textures: a) and b) are examples obtained from previously “seen” data and correspond to MM-AAM and AAM respectively, whereas c) and d) are results with MM-AAM and AAM operating on “unseen” data. ....51

Figure 3.13: Examples of manually cropped images used in the performance evaluation of conventional FR-PCA. The purpose of using cropped images is to generate input images which are therefore comparable to those used in FR-MM-AAM and FR-AAM system performance experiments, see figure 3.12. ....52

Figure 3.14: LDA Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Blue Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Green Bar).....54

Figure 3.15: Multi-Class SVM Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Blue Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Green Bar).....55

Figure 4.1: Normalized Histogram; a) represents  $C1$  that contains the NCC values corresponding to the face dataset with 10 images per subject that smoothly change pose from one image to other to cover a rotation of upto  $180^\circ$ , b) represents set  $C2$  comprised of NCC

values for the face dataset with only three images per person having approximate pose rotation of $0^\circ$ , $90^\circ$ and $180^\circ$ , respectively. ....	67
Figure 4.2: Intra-subject variability Vs face datasets Curve: Here face datasets shown on x-axis comprised of different numbers of images per person and different levels of pose variation from one image to other of same subject. ....	68
Figure 4.3: Examples of manually cropped images and their corresponding original images. ....	72
Figure 4.4: Experimental framework for evaluating the ability of VM to predict FR performance. ....	77
Figure 4.5: Recognition Rate Vs VM curves corresponding to different FR systems: A general increasing trend can be noticed. ....	77
Figure 4.6: $R_2$ and $R_2$ Vs degrees of Polynomial; a) Fisherfaces, b) PCA+SVM, c) Eigenfaces, and d) NCC. Degree $d = 2$ , in all cases, is the maximum value for which the difference between $R_2$ and $R_2$ is minimum. ....	79
Figure 4.7: Recognition Rate Vs VM a) Fisherfaces, b) PCA+SVM, c) Eigenfaces, and d) NCC. 2nd degree polynomial shown as dotted line is following the trend of original data. ....	80
Figure 4.8: Experimental framework for evaluating proposed VM using noisy data. ....	82
Figure 4.9: Recognition Rate Vs VM curves corresponding to Fisherfaces FR system: same overall increasing trend as seen in figure 4.5 can also be noticed in case of noise with all PSNR values. ....	85

Figure 4.10: Histograms of Absolute Errors for the datasets with noise at four different PSNR values: a) 55.57, b) 33.48, c) 26.86, and d) 23.98.....	86
Figure 5.1: High-level system block diagram of our proposed ML-AE framework. ....	94
Figure 5.2: Sample face images with their superimposed shape coordinates (red dots) and extracted facial texture.....	94
Figure 5.3: Flow diagram of ML-AE system training. Here L and R are the lower bounds for number of images and number of age years covered by the age-range of corresponding parent dataset. LTh and RTh are two pre-specified thresholds. Internal structure of Classifier Modelling block is explained in figure 5.4. ....	97
Figure 5.4: Internal structure of Classifier Modelling block shown in figure 5.3. ....	97
Figure 5.5: Image distribution of FG-Net face dataset according to age.....	104
Figure 5.6: Sample images of a person taken from the FG-Net aging database.....	105
Figure 5.7: Mean Absolute Errors at different age years for FG-NET dataset; blue curve is representing MAEs for ML-AE based on both shape and texture, whereas green and red curves are for ML-AE schemes based on shape or texture, respectively. ....	107
Figure 5.8: Cumulative scores comparing the proposed Multi-level Age Estimation with SVM classifier (ML-AE-SVM) framework, based on fusion of both shape and texture classifications, and ML-AE-SVMs based on shape or texture on their own. ....	107
Figure 5.9: Mean Absolute Errors at different age years for FG-NET dataset.....	108

Figure 5.10: Cumulative scores comparing the proposed Multi-level Age Estimation framework with SVM (ML-AE-SVM) and with LDA (ML-AE-LDA)..... 108

# List of Tables

Table 3.1: Average Shape and Texture Errors (Average; Standard Deviation) for both Seen and Unseen Datasets. ....	43
Table 4.1: Some of the widely used Face Datasets in Face Recognition Applications .....	61
Table 4.2: Face Datasets having Pose Variation.....	62
Table 4.3: Recognition Rate reported for different Pose Variation .....	62
Table 4.4: Parameters for 2nd-degree Polynomial Model .....	80
Table 4.5: Avg. Absolute Error (Avg. AE) at different Avg. PSNR values.....	85
Table 5.1: Look-up Table used for Decision Fusion .....	101
Table 5.2: MAE and standard deviation of error comparison of ML-AE using fusion of shape and texture classifications with ML-AE schemes using shape or texture classifications.....	106
Table 5.3: Comparison of ML-AE-SVM with ML-AE-LDA in terms of the MAE and standard deviation of error .....	109
Table 5.4: Comparison of AE methods with respect to Mean Absolute Error and Cumulative Score (MAE / CS).....	110

# List of Abbreviations

<b>AAM</b>	Active Appearance Model
<b>AAS</b>	Appearance and Age Specific
<b>AGES</b>	AGing pattErn Subspace
<b>Avg. AE</b>	Average Absolute Error
<b>BIF</b>	Bio-Inspired Features
<b>CS</b>	Cumulative Score
<b>FAE</b>	False Acceptance Error
<b>FG-NET</b>	Face and Gesture Recognition Research Network
<b>FI-AAE</b>	Face Image based Automatic Age Estimation
<b>FM/S</b>	Face Modeling / Synthesis
<b>FR</b>	Face Recognition
<b>FRE</b>	False Rejection Error
<b>GWT</b>	Gabor Wavelet Transform

<b>HCI</b>	Human Computer Interaction
<b>HHI</b>	Human-Human Interaction
<b>HMI</b>	Human-Machine Interaction
<b>IMM</b>	Informatics and Mathematical Modelling
<b>JPEG</b>	Joint Photographic Experts Group
<b>LAR</b>	Least Angle Regression
<b>LBG-VQ</b>	Linde-Buzo-Gray Vector Quantization
<b>LBP</b>	Local Binary Pattern
<b>LDA</b>	Linear discriminant Analysis
<b>MAE</b>	Mean Absolute Error
<b>MC/MM-AAM</b>	Multi-Component/Multi-Model AAM
<b>ML-AE</b>	Multi-Level Age Estimation
<b>MM-AAM</b>	Multi-Model AAM
<b>MSE</b>	Mean Square Error
<b>NCC</b>	Normalized Cross Correlation
<b>NUI</b>	Natural User Interface
<b>PCA</b>	Principal Component Analysis

<b>PSNR</b>	Peak Signal to Noise Ratio
<b>RBF</b>	Radial Basis Function
<b>SVM</b>	Support Vector Machine
<b>SVR</b>	Support Vector Regression
<b>VM</b>	Variability Measure
<b>VM-interSc</b>	Inter Subject Class Variability Measure
<b>VM-intraSC</b>	Intra Subject Class Variability Measure
<b>WAS</b>	Weighted Appearance Specific

# **Chapter 1**

## **Introduction**

### **1.1 Motivation of Research Work**

In real life, human faces have always been considered as the first source of information during Human-Human Interactions (HHI). One can gather a variety of information from another person's face, such as gender, identity, ethnicity, expression and age, that influences the course of subsequent oral communication between the two. Therefore, with the progress in the field of Human-Machine Interaction (HMI), the human face has become gradually the focus of research to develop Natural User Interfaces (NUIs) that offer natural interaction using facial characteristics. As a result, researchers have come up in last few decades with a plethora of techniques for the extraction and use of these facial characteristics in areas such as Automatic Face-based human Identification, Gender classification, Age estimation, Facial expression recognition, and Race classification. Among these, person identification based on face images is relatively a well-explored area in real-life applications such as security, surveillance, access control, image database search, identity verification, etc. Furthermore and within the context of HHI, the automatic extraction of age information from face images hasn't received enough research attention. Moreover, human based age estimation is in general less accurate than identity and gender estimation. As a result relying on humans to supply this information from face images often becomes problematic. Hence, there is a

growing need for the development of automatic facial age estimation systems that are reliable and possibly superior to human based age estimation.

Face image based automatic age estimation (FI-AAE), whose objective is to determine the specific age or age range of a person based on a given facial image, is a challenging yet attractive topic due to its roots in numerous real-life applications such as:

- **Law Enforcement:** A system equipped with good automatic age estimation module can be helpful in filtering out the potential suspects more efficiently and accurately from a database using the estimated age of the input mugshot.
- **Security Control and Surveillance:** Security control and surveillance monitoring issues are becoming more and more crucial in daily life. For example, an accurate age estimation system can prevent minors from entering bars or wine shops; stop underage smokers from purchasing cigarettes from vending machines; refuse the aged when he/she wants to try a roller coaster in an amusement park; and deny children access to adult websites or restricted movies.
- **Health Care:** Face image based automatic age estimation can also be helpful in health care systems, such as robotic nurse and intelligent intensive care unit, for customized services. For example, a personalized Avatar will be selected automatically from the custom-built Avatar database to interact with patients from different age groups with particular preferences.
- **Human-computer interaction (HCI):** The system can adjust the contents presented to a user based on his/her age. For example, a smart shopping cart can be designed to provide recommendations according to the age of the customer.
- FI-AAE system can further be used in age based indexing of face images, thereby allowing age based retrieval of face images as per need. Moreover, now different mobile

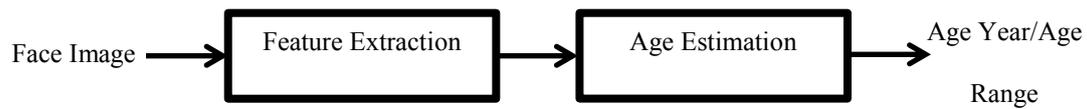


Figure 1.1: Block diagram of a typical automatic age estimation system.



Figure 1.2: Aging effects on a human facial appearance.

applications (e.g. Age Detector for iPhones) are being developed that can estimate age for entertainment purposes.

These few example application areas not only highlight the contribution of FI-AAE to real world but also inspire the need for more research work that can produce state-of-the-art systems to accurately estimate human age.

## 1.2 Face Image based Automatic Age Estimation (FI-AAE)

In general, face image based automatic age estimation (FI-AAE) systems are comprised of the following two major components (see figure 1.1).

### 1.2.1 Feature Extraction Process

Feature extraction is a process that extracts those facial features that are affected by the aging process. Aging affects the human facial appearance considerably, see in figure 1.2. Furthermore facial aging is an irreversible natural process that cannot be controlled during face image acquisition, like other facial variation due to pose and expression. Moreover, although age progression affects facial appearance of different people differently, biological

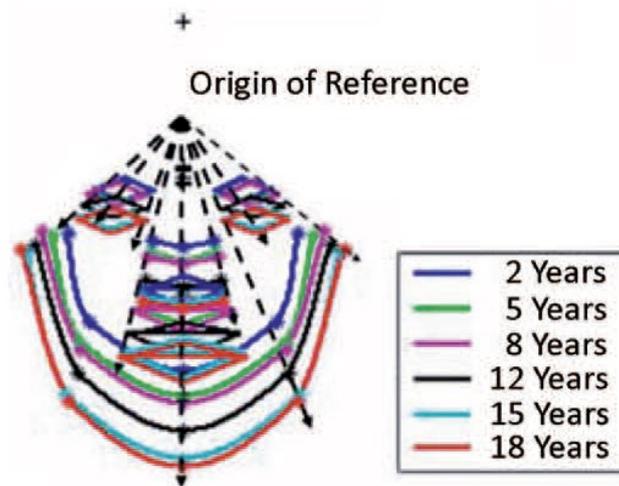


Figure 1.3: Craniofacial growth (shape change) on a human face with age progression, originally given in [1] and then [3].

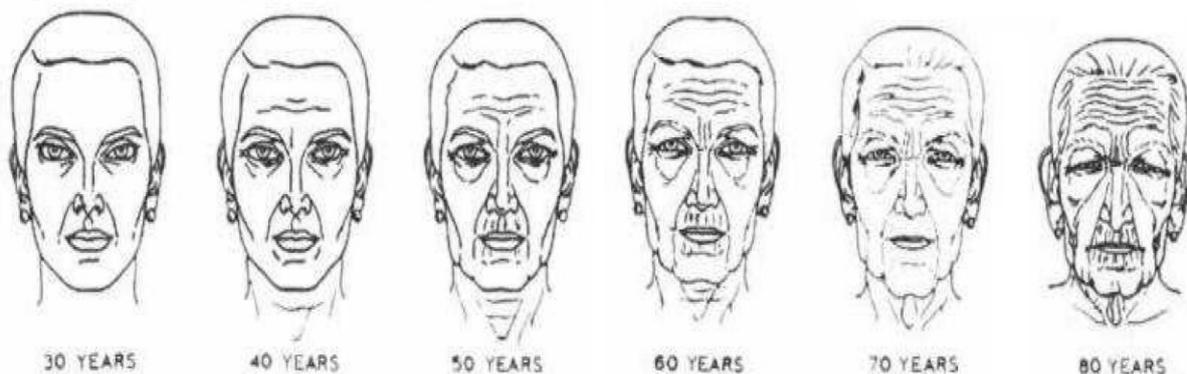


Figure 1.4: Skin aging (Texture change) with age progression, originally given in [1, 3].

or anthropometric studies suggests that on basis of some common features, facial aging can be roughly divided into two stages [3, 4]; i) *birth-to-adulthood* and ii) *adulthood-to-old*.

During *birth-to-adulthood*, usually bone growth takes place that causes major changes in the facial shape as shown by the six prototype images of figure 1.3 (originally given in [1] and then reproduced in [3]). During the *adulthood-to-old* stage the most perceptible age-related deformations are associated with texture changes (see figure 1.4).

In particular common changes in the facial shape and texture during these two stages of human life, are reported in [3] and are as follows:

1) Birth-to-Adulthood Stage:

- Forehead slopes back, shrinks and releases space on the cranium.
- Facial features such as eyes, nose, mouth and ears grow their areas.
- Cheeks expand their areas and chin becomes more bulging.
- Skin texture does not change much but facial hair become denser.

2) Adulthood-to-Old Age Stage:

- Facial skin becomes thinner, darker, less flexible, and more leathery.
- Adynamic wrinkles and blemishes due to biologic aging gradually appear.
- Dynamic wrinkles and folds due to muscle motion become more prominent.
- Cheeks start dropping, double chin and lower eyelid bags appear.
- Although the craniofacial growth is not dramatic during this aging period, the facial geometry change is still evident from 30 to 80 years, especially in the female faces. Faces change from a U-shaped or upside-down triangle shape to a trapezoid or rectangle.

Now, since aging affects both the geometry (shape) and texture of human faces, a system based on a feature set representing only one of them would not be capable of estimating human age accurately. For example, shape based features can only be used to estimate age for children. That is why researchers have come up with different techniques to merge or fuse the shape and texture information for the purpose of age estimation such as the use of Gabor Wavelet Transforms (GWT) [5, 6], Subspace Features using the image intensity [7, 8], Active Appearance Models (AAM) and image frequency [9]. Among these, the most popular technique for feature extraction is that based on Active Appearance Models (AAM) [10-14].

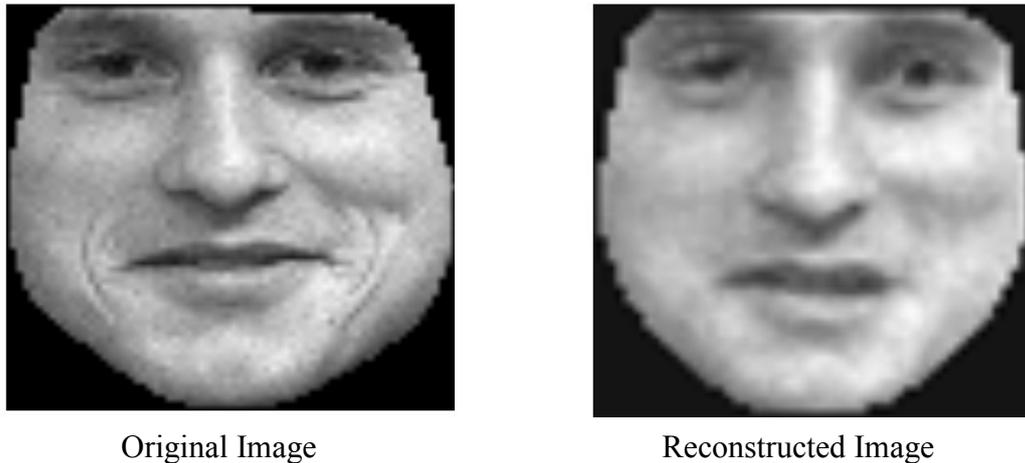


Figure 1.5: Comparison of the original image and the reconstructed image from the low dimensional set of appearance parameters of the AAM. The inability to synthesis fine aging details is obvious in reconstructed image.

The AAM is a well-known method that is used to produce a single generative parametric model that represents both face shape and texture and which is subsequently used to produce multiple instances of that face using a small number of parameters [15]. For this purpose, AAM applies Principal Component Analysis (PCA) in two stages. First, PCA is applied on shape and texture, separately, to form two sets of parameters, which are then combined to form a single set. The second PCA is applied to this combined set to generate the final set of parameters. Although this two-stage PCA reduces feature dimension substantially, while retaining general individual characteristics.

It has been reported [2], that there is loss of some critical aging details related to wrinkles and skin texture, see figure 1.5. Even though some techniques [2, 16, 17] have been proposed in an attempt to overcome this problem, by introducing separate models for the missing features, performance gains are negligible and don't justify the significant additional complexity that is often introduced in the overall system.

The above discussion suggests that despite of all previous efforts, still there exists a need for the development of appropriate methods which use effectively shape and texture aging characteristics, and thus yield improved age estimation performance. Hence the initial goal of this research work was identified as:

“The extraction of geometric (shape) and texture related information, from face images, without the loss of important aging characteristics.”

### **1.2.2 Age Estimation Process (ASP)**

The second major component of any FI-AAE system is ASP. In general, age estimation is considered as a multi-class, classification problem. According to authors in [2], the work done so far that deals with this classification problem, can be divided into systems which offer Age-group classification [5, 6, 18, 19], or those which operate using i) Single-level age estimation [7, 11, 13, 14, 16] or ii) Hierarchical age estimation [2, 11, 20-23] system architectures. An age-group classification approach is used to classify an input face into one of many age groups, for example, an input face can be classified as that of a child, adult or old person. The remaining two approaches relate to the overall classification system architecture. Furthermore, single-level age estimation methods try to find age label while using information derived from a whole input dataset and without taking into consideration group-specific characteristics.

However and as discussed earlier, shape changes the most during the child to adulthood period whereas during the adulthood to old age period all major face changes relate to texture. This implies that two separate classifiers are needed to be trained using two different face datasets, which in turn suggests a hierarchical classification system architecture. Hierarchical age estimation methods can potentially provide better age estimation

performance by using classifiers trained on sub-age group related characteristics. Moreover, as the classifiers are trained on relatively smaller groups this helps in reducing overall system computational complexity.

Note that such hierarchical classification age estimation techniques are prone to error propagation effects. As aging is a very slow process, some images at the boundaries of two adjacent groups are often quite similar and that makes it very hard for a classifier to always classify correctly these faces. Furthermore, miss-classification errors propagate within the hierarchy of groups.

This raises the need for the development of a new hierarchical face classification framework, as applied to age estimation, that operates on age partitions having large inter-group variations and mitigates miss-classification errors. Thus, the second goal of this research work can be stated as:

“The development of an advanced hierarchical age estimation / classification framework that maximizes inter-group variability.”

### **1.3 Facial Aging Dataset**

In addition to the two primary types of components discussed in previous section, the performance of FI-AAE systems is also heavily dependent on the adopted and application related face dataset generation methodology and resulting dataset characteristics. Thus face data collection is a critical element of system performance and deserves considerable attention. Note that, it is very hard to obtain in practice a large size face dataset of images taken at different ages per person and there is a scarcity of such publically available datasets. The most popular facial aging datasets used to assess FI-AAE system performance are FG-NET [24] and MORPH [25, 26]. Both datasets are publically available with FG-NET being

free to use for research purposes. FG-NET contains more than thousand images of 82 subjects of different races with high levels of non-aging variations in terms of pose, expression, and lighting image capturing conditions.

Thus the performance of any classification process involved in face image based automatic age estimation (FI-AAE) as well as face recognition (FR), is significantly affected by the intrinsic characteristics of the available face dataset which in turn is used to train/design or test FI-AAE and FR system performance.

Generally, these dataset characteristics are defined as the inter-class and intra-class variations among different dataset images. Furthermore, datasets, used for a particular application, are often captured under different capturing conditions that cause different types of data variability. For example, such conditions relate to are subject illumination, pose, expression, makeup, facial attributes (i.e. moustache, beard, glasses) and age. Also note that in addition to the type of data variation, the amount of variability allowed per type, during image capturing, is also a factor to affect the classification performance. Consider for example the severe visual changes between images of the same class that large pose variations can create, whereas, at the same time, have the potential to increase similarity between the images of different classes. This implies that a relationship exists between the amount of image variability and system classification performance and, given an appropriate measure of image data variability, this relationship could be modeled. This in turn suggests that given a classification system its performance could be possibly predicted for any given input dataset without the need to perform extensive experimentation.

Thus the ability to i) measure and then model data variability and ii) predict classification performance of a system, is an important research aim. Also note that such a capability can be used to select a classification system whose performance is suitable to a given application.

Recently in [27], authors proposed a set of different variability measures in order to represent object class properties in object classification applications. Here several variability measures are proposed which are based only on intra-class similarities. As a consequence they can only be used with binary types of classification problems and definitely not in multiclass scenarios such as those encountered in FI-AAE and FR systems. The above discussion indicative of the third and final general goal of the research work presented in this thesis, that is

“To develop i) dataset dependent multi-class data variability measure(s) and associated models per classification system and ii) given an input dataset and its data variability measure, predict system classification performance.”

## **1.4 Research Objectives**

The purpose of this investigation is to explore the possibilities and challenges in accurately estimating human age from facial images. As discussed earlier, aging affects the human face differently in different stages of human life i.e. facial geometry changes, significantly in young ages whereas in older ages most of face changes are linked to facial texture. Therefore for obtaining accurate human age estimates, face image based automatic age estimation (FI-AAE) systems are required to fully exploit these two features (i.e. face shape and texture) in their respective periods of time. Conventional systems lack this capability and accuracy during the feature extraction process which then adversely affects subsequent age classification. This raises many challenges with the selection of optimal features and subsequent classification being the most important.

In this thesis, we have focused our attention to these main age estimation research challenges and proposed novel solutions. The individual goals, discussed earlier which stem from our main age estimation research objective are listed here again:

- To extract geometric (shape) and texture information from input face images without losing important aging characteristics.
- To propose a hierarchical framework for age estimation while operating on age partitions having large inter-group variation characteristics.
- To measure and model the multi-class face data variability and to predict the classification performance of a given system.

## **1.5 Author's Contributions**

This thesis offers three major contributions which are briefly described next. Lists of publications corresponding to these contributions are also provided in this section.

### **1.5.1 Face Image Modeling and Synthesis**

Commonly used Active Appearance Model (AAM) based face image modeling techniques usually suffer in the presence of high image data variability, especially with respect to subject pose and expression. Hence, the face texture and shape parameters produced, result in large face synthesis errors and also generated texture lacks important aging features such as wrinkles. In Chapter-3, we propose novel face image modeling and Synthesis systems based on i) a Multi-Model AAM framework MM-AAM [28] and ii) a components-based, Multi-Component/Multi-Model AAM MC/MM-AAM [29] approach, with advanced performance characteristics. Thus MM-AAM performance is evaluated and compared with that of conventional AAM, when employed in face recognition and experimental results are indicative of the effectiveness of the new proposed methodology [30]. Note that in face recognition, synthesised shape has been used in both systems to extract the original face texture from the input image, so that texture synthesis errors are avoided and don't affect overall face recognition system performance.

**Related Publications**

Muhammad Aurangzeb Khan, C. S. Xydeas, and Hassan Ahmed. "Multi-Component/Multi-Model AAM framework for face image modeling." In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, pp. 2124-2128. IEEE, 2013.

Muhammad Aurangzeb Khan, C. S. Xydeas, and Hassan Ahmed. "Multi-model AAM framework for face image modeling." In Digital Signal Processing (DSP), 2013 18th International Conference on, pp. 1-5. IEEE, 2013.

Muhammad Aurangzeb Khan, C. S. Xydeas, and Hassan Ahmed. "ON THE APPLICATION OF AAM-BASED SYSTEMS IN FACE RECOGNITION." 22nd European Signal Processing Conference (EUSIPCO 2014), 2014.

**1.5.2 Face Image Data Variability Measure**

As mentioned earlier, face data variability plays an important role in the overall performance of face classification used within the context of age estimation. This in turn is indicative of the need to measure and model such variability and thus to hopefully predict classification performance. In Chapter-4 of the thesis, a novel face image data variability measure is proposed. Furthermore, due to scarcity of age related image data, the validity of this variability measure (VM) and its usage for prediction purposes is established in face recognition (FR), that is a sister application domain to age estimation. Thus the proposed variability measure is successfully used to model the recognition performance of different FR systems. FR classification system related Models are derived and tested using a total of eleven publically available face datasets.

**Related Publication**

Muhammad Aurangzeb Khan, C. S. Xydeas, and Hassan Ahmed. "Face Image data variability." To be submitted in IEEE.

**1.5.3 Hierarchical Classification Framework for Automatic Age Estimation**

This final topic of contribution offered in this thesis (see Chapter-5) takes the form of a novel hierarchical framework conceived for face image based automatic age estimation. It operates on face shape and texture separately, at each level of hierarchy, and classification estimates are fused in order to select the next branch to be followed in a classification decision tree. At each tree level and in order to avoid miss-classifications, due to smooth change in aging process, classifiers are trained on appropriately defined age-related data having large inter-class variation. Computer simulation experimental results demonstrated an advanced performance when compared to that obtained from conventional FI-AAE systems.

**Related Publication**

Muhammad Aurangzeb Khan, C. S. Xydeas, and Hassan Ahmed. "Face Image based Automatic Age Estimation." To be submitted to IEEE.

**1.6 Thesis Organization**

This thesis is organized into 6 chapters.

Chapter 1 provides an introduction to the thesis and thus presents i) the motivation and issues driving research activities and ii) basic information on face image based automatic age estimation processing.

Aspects of related existing knowledge in face image synthesis, quantifying face data variability and human age estimation techniques are presented in chapter 2.

Chapter 3 describes, in detail, the design of i) a novel Multi-Model AAM (MM-AAM) processing framework and ii) a Multi-Component/Multi-Model AAM (MC/MM-AAM) approach together with a comparative analysis of both systems with conventional Active Appearance Model (AAM).

In chapter 4, a detailed formulation of new face data variability measure (VM) along with its performance evaluation, using different face recognition (FR) systems, is presented.

Chapter 5 presents design details and performance evaluation of a proposed Multi-Level Age Estimation (ML-AE) framework, whereas chapter 6 provides overall conclusions and possible directions of future studies.

## **Chapter 2**

### **Related Research Work and Background Material**

Face image based age estimation may be considered as a two-step process: First, extraction of the age-related information from the face image and its representation using a lower dimensional feature set is performed, followed by age classification using the selected feature representation. The extraction and representation step is necessary as pixel value data of a facial image is normally of too high dimensionality and therefore complex to be used in a classification phase. This chapter presents a brief introduction of the existing work associated with both of these processes.

The chapter is organised as follows: Section 2.1 presents existing research work related to Facial Feature Extraction, Age Estimation and Classification. Note that our proposed face image modeling and representation techniques [28, 29] (see Chapter-3) are based on the Active Appearance Model (AAM) and therefore, Section 2.2 presents the design and structure of AAM as background material.

#### **2.1 Related Research Work**

##### **2.1.1 Feature Extraction and Representation**

Existing research work on feature extraction may be categorized into the following classes:

### 1) Anthropometric Models

These methods are based on face anthropometry, which is a science of measuring sizes and proportions on human faces. Some known feature extraction methods based on anthropometric models include Kwon and Lobo [31], Farkas [32], Ramanathan and Chellappa [1], Gunay and Nabiyevev [33], etc.

Kwon and Lobo [31] are considered as the pioneers in the field of human age estimation using facial images. They computed six distance ratios by dividing the distance between two features by the distance between two another features (e.g. distance between the eyes over distance between the eyes and the nose). These ratios are then used to distinguish between images of babies and adults. The features, used to compute ratios, are located by using template matching. For further classification among adults, Kwon and Lobo extracted wrinkle information from several facial regions such as forehead and around eyes using the snakelets method [34].

Hornig, Lee and Chen [18] proposed a variant of Kwon and Lobo's work for the classification of facial images into four age groups: babies, young adults, middle-aged, and old adults. They located the facial features by finding high intensity regions within an edge map of the face image. These features were then used to compute two distance ratios to distinguish between babies and adults. Secondly, they used Sobel filtered images, instead of snakelets, to measure the amount of wrinkles on a face image. The use of Sobel operator made this method simpler and faster.

Farkas [32] provided anthropometric measurements based on 57 landmarks on human face. Similarly, Ramanathan and Chellappa [1] proposed eight distance ratios to model the age progression among young faces for the purpose of face recognition across age progression.

Gunay and Nabiye [33] also proposed a variant of this approach to represent anthropometric features.

In summary, the age estimation techniques based on anthropometric models are majorly based on geometric features and can only deal with young ages, as the human facial features in terms of measurements and ratios do not change much in older ages. Furthermore and most importantly anthropometric models are restricted to only frontal images as the ratios of distances are computed from 2D face images which are sensitive to head pose.

Even the anthropometric models supported by the wrinkle information could not provide better age classification as wrinkle extraction from a facial image is a quite hard task due to lighting conditions, camera resolution, and make-up.

## 2) Active Appearance Models

The second category of age estimation methods is based on the Active Appearance Model (AAM) work originally proposed by Cootes et al. [15] for encoding facial geometry and texture. Later Lanitis et al. [11, 14] applied AAMs, for the first time, to the age estimation problem. They showed that the aging pattern could be represented by a quadratic function called aging function and proposed the Weighted Appearance Specific (WAS) method [14] and the Appearance and Age Specific (AAS) method [11]. Lanitis et al. used aging functions in the form of quadratic equations for relating the coded representation of faces to the actual age for the purpose of age estimation. According to their results the use of person specific aging functions produced improved age estimation results when compared to the use of a common aging function for all subjects.

Xin Geng et al. [10] proposed an age estimation method called AGES (AGing pattErn Subspace) to handle the highly incomplete age based face image dataset. They generated

aging patterns, i.e. a sequence of personal facial images sorted in chronological order, for each person in a dataset consisting of face images showing each subject at different ages. Although, images are represented by AAM based feature vectors in the same way as given by [11, 14]; the AGES method uses several images of the same individual, taken across different ages together to represent aging patterns. Each collection of temporal feature vectors is considered as a single sample, which can then be projected to a low dimensional space. Given a previously unseen face, the face is substituted at different positions in a pattern. The position that minimizes the reconstruction error indicates the age of the subject. Xin Geng et al. modified their earlier work AGES [10] to develop AGES<sub>LDA</sub> that additionally applies Linear Discriminant Analysis (LDA) to the AAM-based feature vectors to deal with pose, expression, and illumination variations. As AGES relies on person-specific aging patterns, it assumes that for a given input face image there exist face images of the same individual but at different ages, or at least a similar aging pattern for that face image in the training database. However, for real world application it may not be practicable to collect a large aging database having face images of same individual across many ages.

The use of AAM as a feature extraction and representation tool in age estimation field is not limited to the research work reported previously. As AAM encodes both facial shape and texture simultaneously, it is quite popular among the researchers and many image analysis schemes available in literature are based on AAM. For example, Yan et al. [13] designed a regressor based on training samples with uncertain nonnegative labels using AAM features. Karl Ricanek et al. [22, 35] applied Least Angle Regression (LAR) by Efron et al. [36] to identify the most important AAM features. Khoa Luu et al. [22] used Active Appearance Model (AAM) to extract a combined feature vector of facial images. Sethuram et al. [37] used Support Vector Regression (SVR) to learn age-based properties of AAM parameters and gradient-regression-based AAMs to represent texture information. Suo et al. [17]

designed sparse features consisting of the AAM, wrinkles, skin, hair, and the configuration of the facial components features using the hierarchical face model. In each component, four types of features were extracted: topology, geometry, photometry, and configuration features. Choi et al. [2] proposed an extraction method for wrinkle and skin features which are then combined with AAM features to estimate age in a hierarchical framework. The wrinkle features are effectively extracted by a Gabor filter set based on the direction of wrinkles on the face. The skin features used in skin aging analysis are extracted by Local Binary Pattern (LBP) [38]. Chen et al. [39] used AAM features and proposed a method of pair wise age ranking based on subspace learning for age prediction. Chao et al. [40] applied age-oriented local regression using distance metric learning and dimensionality reduction using AAM features. Notice that approaches based only on AAM-based features generally have performance limitations due to the following major factors:

- i) AAMs' lack of ability to automatically fit facial landmarks on unseen images that exhibit illumination, pose, or expression variation [41, 42],
- ii) Although AAMs are an excellent model based approach for face related problems [22, 35], they only represent holistic and not local aging information such as wrinkles [2, 3].

In this thesis, to deal with the above mentioned AAM feature issues, we have proposed two variants of AAM called Multi-Model AAM (MM-AAM) [28] and Multi-Component/Multi-Model AAM (MC/MM-AAM) [29]. These new models provide an improved synthesis capability for unseen face images (see Chapter-3).

### 3) Appearance based Features

The third category of feature extraction methods is based on a set of visual or appearance based features. For example, Günay et al. used Local Binary Patterns (LBP) [38] for appearance feature extraction in an automatic age estimation system [19]. Gao et al. [6] used

Gabor features [43] in their proposed age estimation system and reported better performance than that obtained from LBP. Gao et al. proposed an age estimation system based on Bio-Inspired Features (BIF) [44], and its different variants have been used for the purpose of age estimation [21].

### **2.1.2 Face Image based Age Estimation System Architectures**

Existing age estimation or age classification techniques generally operate in one of the following two system architectures:

#### *1) Single-Level Age classifiers*

Single-level age estimation methods aim at finding the age label of a given input face image taken from a dataset. Single-level system architectures are employed in systems such as quadratic regression [14], SVR [7], AGES [10], MLP [11], RUN [13], etc. As discussed earlier in Chapter-1, aging affects face images differently in different stages of human life, for example, shape changes the most during the child to adulthood period whereas during the adulthood to old age period all major face changes relate to texture and since single-level age estimation methods perform classification while using information derived from a whole input age based dataset, they are not able to exploit these group-specific aging characteristics of face images.

#### *2) Hierarchical or Multi-Level Age classifiers*

As discussed in Chapter-1, facial shape and texture change differently during different periods of life. Therefore, for obtaining better age estimation performance a hierarchical classification system architecture is required, in which separate classifiers are trained using different face image datasets corresponding to each period of life.

Thus Hierarchical Age Estimation methods can potentially provide better age estimation performance by using classifiers trained on sub-age group related characteristics. Moreover, as the classifiers are trained on relatively smaller groups this helps in reducing overall system computational complexity. Note however that Hierarchical age estimation methods are often prone to error propagation. Existing Hierarchical Age Estimation systems are of two types: i) systems based on hard boundaries between adjacent age groups [11, 21, 22] and ii) systems based on soft boundaries or with overlapping age groups [2, 20]. The performance of the first type of system suffers due to smooth aging variation across adjacent groups. In fact, images located at the boundaries of two adjacent groups are often so similar that it makes it very hard for a classifier to correctly classify them. Such classification errors propagate within the system hierarchy and adversely affect overall age estimation performance. To deal with this problem, researchers [2, 20] have proposed the idea of soft boundaries or the use of overlapping age groups. Choi et al. [2] designed each age group classifier to have an overlapping estimated age range, which considers the false acceptance error (FAE) and false rejection error (FRE) of each classifier. By compensating for classification errors using overlapping classes, the total age estimation performance is improved. Han et al. [20] used a similar strategy to partition the face image dataset into different age groups. These solutions reduce errors in the coarse levels of hierarchy; however, errors at finer levels are still quite large, which make overall system improvement in age estimation insignificant.

Note that in this thesis we propose a novel Multi-level Age Estimation (ML-AE) framework (see Chapter-5) that minimizes classification error significantly by overcoming the challenge of smooth variation among images of different age groups. This is achieved through a novel hierarchical method in the selection of training data which allows large inter-class variation between classifiers.

## 2.2 AAM Review

### 2.2.1 Active Appearance Model (AAM)

Active Appearance Model (AAM), originally proposed by Cootes et al. [15], is an algorithm for constructing a synthetic image, by using all the image region information related to cover a target object (a face in this case) in terms of both shape and appearance (texture), that is a close match to an input face image. Matching to an image involves finding a set of model parameters that minimizes the difference between the given image texture and the texture synthesized using the model.

This section briefly introduces the basic AAM algorithm [15], which is comprised of two operations: AAM Modeling and AAM Fitting.

#### 1) AAM Modeling

AAM models the facial shape and texture and yields a low dimensional hybrid set of parameters that can then be used to reconstruct a given facial image. The shape is represented by a vector containing the coordinates of the landmark points as given below:

$$s^i = [x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_M]^T, \quad (2.1)$$

where  $\{(x_m, y_m)\}$  are the coordinates of the  $m = 1, 2, \dots, M$  landmark points outlining different facial components of face image  $i$ . On the other hand, the texture is described by the intensity values contained within the landmark points. The modeling operation requires a training dataset of face images with corresponding labeled landmarks. As mentioned in [15], first facial shape is modeled using the labeled landmark points. The shapes corresponding to all training images are then normalized by the Procrustes Analysis [45]. Then these

normalized shape vectors are projected onto the shape subspace created by Principal Component Analysis (PCA)

$$s = s_0 + P_s \cdot b_s, \quad (2.2)$$

where  $s_0$  is the mean shape of the training dataset,  $P_s$  is the matrix that contains orthonormal base vectors derived from training set, and  $b_s$  represents a set of shape parameters in the shape subspace.

Afterwards, facial texture is extracted from the face image using corresponding landmark points. For this purpose, all the images in training dataset are warped to the mean shape to produce a shape free texture. Subsequently, a model is generated for each extracted texture. First, each texture vector  $g_i$  is linearly normalized by the parameters  $u = (\mu, \sigma^2)^T$  as given below

$$g_i = \frac{g_i - \mu}{\sigma^2}, \quad (2.3)$$

where  $\mu$  and  $\sigma^2$  are, respectively, the mean and the variance of the texture  $g_i$ . The texture is then projected onto the texture subspace obtained using PCA

$$g = g_0 + P_g \cdot b_g, \quad (2.4)$$

where  $g_0$  is the mean texture,  $P_g$  is the matrix containing orthonormal base vectors and  $b_g$  represents the set of texture parameters.

Finally, these shape and texture models are combined to generate an appearance model. It is achieved by concatenating the parameter vectors  $b_s$  and  $b_g$  to form a hybrid parameter vector  $b_{sg}$  as given below

$$b_{sg} = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix}, \quad (2.5)$$

where  $W_s$  is a weight matrix [15] to normalize  $b_s$  to same scale as of  $b_g$ . Now PCA is applied again to  $b_{sg}$  to get the final appearance model

$$b_{sg} = P_c \cdot c, \quad (2.6)$$

where  $P_c$  are the Eigen vectors and  $c$  is the set of appearance parameters. Due to the linear nature of the model, the shape and texture can be expressed as

$$\begin{aligned} s &= s_0 + P_s \cdot W_s Q_s \cdot c, \\ g &= g_0 + P_g \cdot Q_g \cdot b_g, \end{aligned} \quad (2.7)$$

where

$$P = \begin{pmatrix} Q_s \\ Q_g \end{pmatrix}, \quad (2.8)$$

## 2) Model Fitting

Once the model is created, it is important to fit the model to a given face image for obtaining the accurate model parameters. The AAM model fitting operation is an iterative procedure as given below:

**Step-1:** Extract the texture of given image using synthesized shape of that iteration and project it to the texture model space. Note: in first iteration texture is extracted on basis of mean shape of training images.

**Step-2:** Calculate the current texture error vector,  $r = g_s - g_m$ , where  $g_s$  is the normalized extracted texture sample at the currently estimated shape and  $g_m$  the normalized grey-levels of the synthesized face.

**Step-3:** Compute the current error vector using  $E = |r|^2$ , where  $|\cdot|$  represent the 2-norm.

**Step-4:** Compute the next displacement  $\delta p = -R \cdot r(p)$ , where  $R$  is the regression matrix, which is pre-computed during the AAM modeling, it describes the parameter variation that leads to convergence.  $p$  is the set of AAM parameters ( $p^T = (c^T | t^T | u^T)$  is the combination of the appearance parameters  $c$ , pose parameters  $t$ , and the texture transformation parameters  $u$ ).

**Step-5:** Update the model parameters  $p = p + k\delta p$ , where  $k = 1$  initially.

**Step-6:** Obtain the new model texture,  $g'_m$ , and extract the original texture,  $g'_s$  from the given image using new synthesized shape.

**Step-7:** Calculate the new texture error vector,  $r' = g'_s - g'_m$  and  $E' = |r'|^2$ .

**Step-8:** If the new error  $E' < E$ , then accept the new estimation, otherwise go to Step-5 and try a smaller value of step ( $k = 1.5, 0.5, 0.25, \dots$ ).

**Step-9:** Repeat until the convergence is achieved or maximum number of iterations is reached.

## **Chapter 3**

### **Face Image Modeling And Synthesis**

As discussed in Chapter-1, the first and the most important component in any face image based automatic age estimation (FI-AAE) system, to accurately estimate human age, is the extraction and representation of aging features. Therefore, researchers have done a lot of work in this area and have come up with a plethora of feature extraction and representation techniques. In general, the techniques available in literature for aging features extraction and representation are divided into three categories, i.e. i) Anthropometric Models [1, 18, 31-33], ii) Active Appearance Models (AAMs) [10, 11, 13, 14, 22, 25, 35, 37] and iii) Appearance based Features [6, 19, 21, 44]. However, as discussed in Chapter-2, among these three the most widely used are the techniques based on Active Appearance Models (AAM).

The AAM modeling involves the representation of both face geometry (i.e. shape) and texture characteristics using a small set of parameters, whereas face image synthesis is a process that recovers facial characteristics from these parameters to form face image. In addition to FI-AAE, the roots of AAM and similar face image modeling and synthesis techniques can be found in many real-life application areas such as facial expression recognition, eye tracking, visual speech understanding, video conferencing, interactive animation of cartoon characters using facial motions, etc. All these applications require facial

models that are computationally efficient and also realistic enough to synthesize the various nuances of facial structure and motion.

The AAM is considered as a general optimization process that constructs a synthetic image, in terms of both shape and texture, that is a close match to an input face image by minimizing the difference between the synthesized image and the real appearance of the input image. AAM's ability of differentiating and modeling shape and texture helps in the synthesis of more photorealistic images.

Note that there are two types of application scenarios for modeling face images [42]. One relates to applications such as gaze estimation, head pose estimation or expression recognition and involves person-specific models. The second type deals with the construction of unseen faces and involves generic face models. Modeling a face image for the purpose of age estimation comes under the later of the two scenarios discussed above. Authors in [42] have shown that person-specific AAMs are easier to build, whereas generic AAMs appear to be problematic in texture modeling, due to high data variability, especially pose and expression, among images. Hence, the face texture and shape, they produce, not only result in large synthesis errors but also texture lacks aging features such as wrinkles.

To overcome this hindrance of large data variability among images that affects the formation of generic face image models and to improve the feature extraction and representation process to achieve our main objective of improving the state-of-the-art in face-image based automatic age estimation (FI-AAE), two image modeling frameworks have been developed during this research work, i.e. i) Multi-Model AAM (MM-AAM) [28] and ii) Multi-Component/Multi-Model AAM (MC/MM-AAM) [29]. MM-AAM is a holistic approach that operates on the whole face at once, whereas MC/MM-AAM, which can be considered as an

extension to MM-AAM, models each facial component separately and then combines all the synthesised components to form a full face image. These proposed face modeling frameworks aim at the creation of generic AAM based face models, which are robust to unconstrained input conditions and can preserve discriminative information when generating “unseen” face images.

Both proposed frameworks operate in two phases: i) Modeling and ii) Synthesis phase. The modeling phase of MM-AAM involves two major steps. In the first step, face images, taken from a training dataset, are grouped into a number of clusters. This is achieved on the basis of shape similarities among images and as a result, a given face image training dataset can be represented by several subsets (or clusters) having relatively lesser data variations, which in turn facilitates the subsequent modeling process. Note that the idea of grouping face images into a number of clusters is also presented in [46, 47], but clustering is done on the basis of shape orientation (pose) only, whereas clustering here caters for both face orientation and expression. The second step involves the application of a conventional AAM to each cluster, giving rise to several AAM models. The synthesis phase of MM-AAM, allows for more than one face images to be produced as possible representations of an unseen input face image. The best synthesized face image is then selected according to the criteria explained in next section.

In contrast, in the modeling phase of MC/MM-AAM training face images are first decomposed into face related components, e.g. eyes, mouth, nose, etc., to form facial component specific datasets. Then each facial component dataset is partitioned into several clusters, which subsequently results in several AAM models for each component. This decomposition aims to exploit the local characteristics of each component and can result in better model fitting as suggested by [48, 49]. The synthesis phase of MC/MM-AAM, allows

for more than one shape for each component of the input face image to be synthesized. The best component shape is then chosen for each component on basis of the criteria explained later. Finally, the selected shapes of all components are combined to form a whole face shape, which then presented into a whole face conventional AAM [15] that delivers the synthesized texture of the final reconstructed face image.

Computer simulation experiments, performed on two different sets of face images, show that the proposed MM-AAM and MC/MM-AAM approaches produce more accurate representations of unseen face images, in terms of both shape and texture, as compared to conventional AAM. Although, out of the two proposed methodologies, MC/MM-AAM has performed slightly better, the difference may not be significant when considering the extra system complexity introduced by MC/MM-AAM.

Furthermore, as the goal of the work in this chapter is to develop a novel and effective method of aging feature extraction and representation that can be helpful in achieving our main research objective of developing a state-of-the-art face image based automatic age estimation (FI-AAE), therefore to further examine and ascertain the effectiveness of the proposed MM-AAM system in generating face image models, it is employed under face recognition (FR) application. Thus a FR system framework i.e. FR-MM-AAM [30] is developed and studied. Here, we have used only MM-AAM synthesized shape to extract the original texture of the face image that is subsequently used for the recognition purpose. The performance of the proposed FR-MM-AAM is compared against the recognition performance of the FR system based on the conventional AAM.

In this chapter, the design and structure of the above mentioned two improved AAM-type of systems are first presented, in sections 3.1, 3.2 respectively. Section 3.3 presents a comparative analysis, with respect to face image synthesis, between the proposed systems

and conventional AAM. Section 3.4 provides a detailed formulation of FR-MM-AAM together with a system performance comparison with FR-AAM. Finally, concluding remarks are presented in section 3.5.

### 3.1 Multi-Model AAM (MM-AAM)

The proposed MM-AAM framework, which operates on the face image as a whole, comprises of a Modeling and a Synthesis phase.

#### 3.1.1 MM-AAM Modeling Phase

The MM-AAM Modeling phase involves two major steps, as it is explained below and shown in figure 3.1.

STEP-1: Involves clustering, using only the shape information taken from a training dataset of face images. This dataset contains both shape information  $\mathbf{S}$ , in the form of landmark points, and texture  $\mathbf{G}$  in the form of intensity values (see figure 3.2).

Consider that i) shape in a face image  $i = 1, 2, \dots, L$  is represented by a vector  $f^i$  where:

$$f^i = [x_1, x_2, \dots, x_M, y_1, y_2, \dots, y_M]^T \quad (3.1)$$

and  $\{(x_m, y_m)\}$  are the coordinates of the  $m = 1, 2, \dots, M$  landmark points outlining different facial components, and ii) texture information is represented by a vector  $g^i$ . Then the sets of shape information  $\mathbf{S}$  and texture information  $\mathbf{G}$  from all  $L$  training images can be represented as

$$\begin{aligned} \mathbf{S} &= [f^1, f^2, f^3, \dots, f^L], \\ \mathbf{G} &= [g^1, g^2, g^3, \dots, g^L]. \end{aligned} \quad (3.2)$$

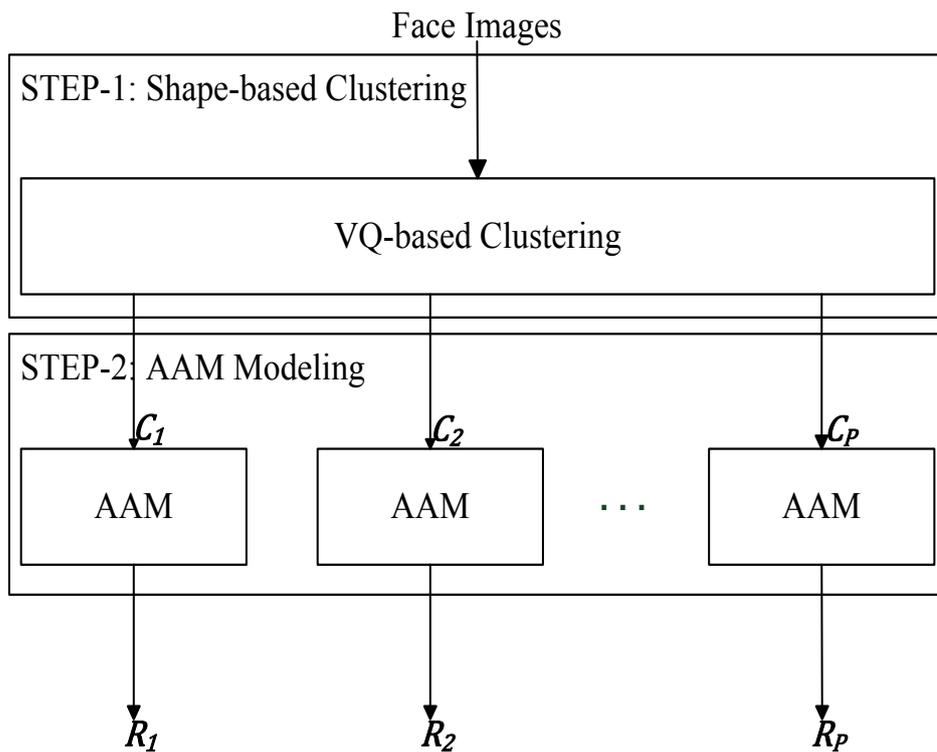


Figure 3.1: System Diagram of proposed MM-AAM. Here  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_P$  are  $P$  clusters of face dataset, which are used to produce  $P$  model matrices  $R_P$ .



Figure 3.2: Some sample face images with their superimposed shape coordinates (red dots) and corresponding facial texture.

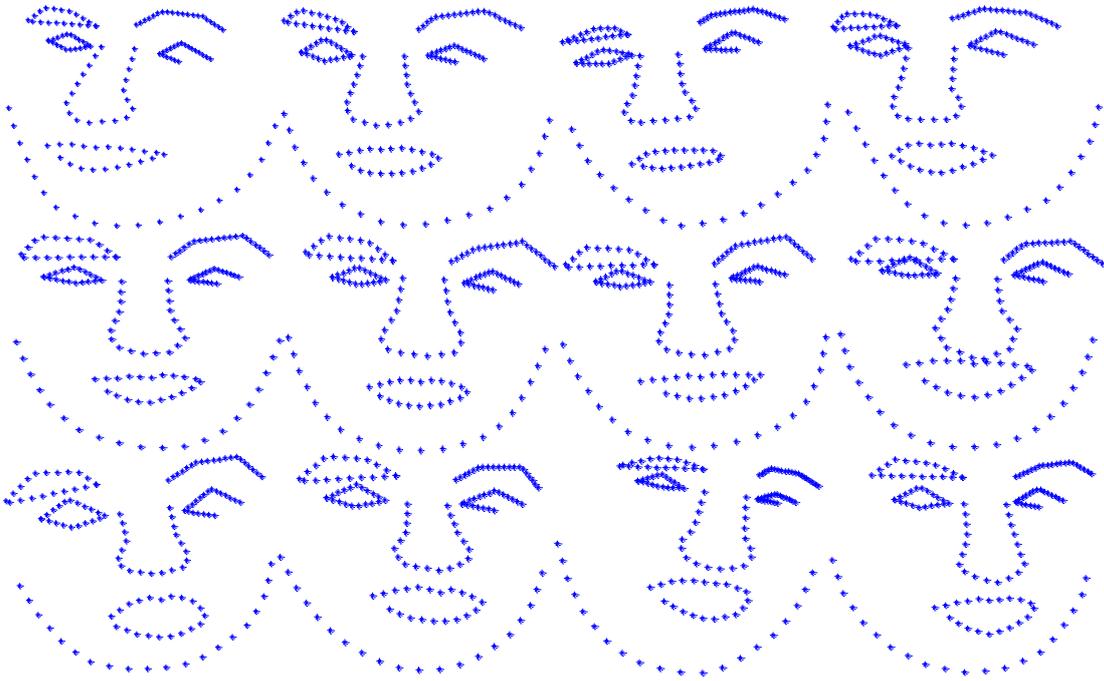


Figure 3.3: Example shapes of face images taken from three different clusters: images shown in one row are representing one cluster. Intra-cluster similarities and inter-cluster variations, in terms of both pose and expression, can be observed.

Face images in the training dataset  $\mathbf{F}$  are then divided into  $P$  clusters via LBG-Vector Quantization (VQ) [50] operating on shape information  $\mathbf{S}$ , i.e.

$$\mathbf{C}_p = VQ\{\mathbf{S}\}, \quad (3.3)$$

where  $\mathbf{C}_p$  for  $p = 1, 2, 3, \dots, P$  is  $p$ th cluster training dataset and is obtained by employing VQ on shape information set  $\mathbf{S}$ .

LBG-VQ is an iterative algorithm that starts with taking the average of whole training set to be the initial code vector. This is subsequently split into two code vectors which are then optimized and divide the initial set into two clusters. These two clusters are split into four and the LBG-VQ process continues until the desired number of clusters is obtained.

Note that following clustering, corresponding texture information is attached to cluster members, i.e. clusters contain both shape and texture information and thus the union of  $P$   $C_p$  clusters gives  $F$ . Figure 3.3 shows sample shapes of face images taken from three different clusters having low intra-cluster and high inter-cluster variation.

STEP-2: Here a corresponding parametric model and a model matrix  $R_p$  is constructed for each cluster using conventional AAM optimization [15], see figure 3.1. Thus a training process produces  $P$  model matrices  $R_p$  which are stored and can be subsequently used in the synthesis of an unseen input face image.

The above steps increase modeling accuracy by exploiting similarities in the shape characteristics of different person's face images. This in turn can be viewed as an attempt to bridge the existing gap between the observed relatively low modeling accuracy of generic AAMs with the much higher accuracy of specific AAMs.

### 3.1.2 MM-AAM Synthesis Phase

Following the previously generated appearance model matrices  $R_p$ , one model for an unseen input face image is selected to obtain a single set of parameters that can give the best representation of the face image. The proposed model fitting or synthesis process can be explained as follows:

- 1) For a given input face image  $t$ , apply  $P$  conventional iterative AAM fitting algorithms based on model matrices  $R_p$  ( $p = 1, 2, \dots, P$ ) and obtain same number of model parameters' vectors  $c_1^t, c_2^t, \dots, c_p^t$ .
- 2) Select the best model parameters' vector  $c$  on the basis of a minimum average Mean Square Error (MSE). MSEs are formed between the original input face texture and the

textures associated with the  $P$  models and averaged across all iterations of the fitting algorithm [15].

- 3) Finally synthesize face texture and shape using the best model parameters' vector and their corresponding Eigenspace obtained in previous step.

### 3.2 Multi-Component/Multi-Model AAM (MC/MM-AAM)

The proposed Multi-Component/Multi-Model AAM (MC/MM-AAM) framework is an extension to the MM-AAM that models different facial components (i.e. eyes, mouth, nose, etc.) separately and produces the best synthesised shapes and textures for all components, which are then combined to form face images. The rationale behind this is to study and analyse the synthesis performance of the framework on basis of different local features (i.e. components) and compare it with the performance obtained using the global feature (i.e. face). The proposed MC/MM-AAM, like MM-AAM, also comprises of a Modeling and a Synthesis phase that are explained below.

#### 3.2.1 MC/MM-AAM Modeling Phase

In contrast to MM-AAM, the Modeling phase of MC/MM-AAM involves one extra step, as it is explained below and shown in figure 3.4.

STEP-1: Involves the component-based decomposition of images into facial components. Face images taken from a training dataset are decomposed on the basis of  $N$  facial components ( $N = 4$  in our case i.e. cheeks + eyebrows, eyes, mouth and nose, see figure 3.5). This component-based decomposition is being used to account for the local shape and texture variability that characterizes different facial components. As given in Eq. 3.1, the shape in a face image  $i = 1, 2, \dots, L$  is represented by a vector  $f^i$  that contains the  $M$  landmark points

outlining the different facial components. In this step, the shape vector  $f^i$  for each  $i$ th face image is decomposed into  $N$  sub vectors of different lengths, such that

$$\begin{aligned}
 f_1^i &= [x_{11}, x_{12}, \dots, x_{1a}, y_{11}, y_{12}, \dots, y_{1a}]^T, \\
 f_2^i &= [x_{21}, x_{22}, \dots, x_{2b}, y_{21}, y_{22}, \dots, y_{2b}]^T, \\
 &\vdots \\
 f_N^i &= [x_{N1}, x_{N2}, \dots, x_{Nc}, y_{N1}, y_{N2}, \dots, y_{Nc}]^T,
 \end{aligned} \tag{3.4}$$

where  $f_n^i$  is the shape vector of  $n$ th facial component of the  $i$ th image. After decomposition of all training face images, shape and texture vectors belonging to the same component, are grouped into separate sets to form  $N$  component-based datasets as given by:

$$\begin{aligned}
 \mathbf{S}_n &= [f_n^1, f_n^2, f_n^3, \dots, f_n^L], \\
 \mathbf{G}_n &= [g_n^1, g_n^2, g_n^3, \dots, g_n^L], \\
 \mathbf{F}_n &= \{\mathbf{S}_n | \mathbf{G}_n\},
 \end{aligned} \tag{3.5}$$

where  $\mathbf{F}_n$  is the  $n$ th component-based dataset containing shape vectors  $\mathbf{S}_n$  and texture vectors  $\mathbf{G}_n$  from all  $L$  training images.

STEP-2: Each component-based dataset  $\mathbf{F}_n$  is then partitioned, on basis of shape vectors  $\mathbf{S}_n$ , into a number of clusters  $\mathbf{C}_{nk}$  (for  $k = 1, 2, 3, \dots, K$ ) by using same LBG-VQ process as explained earlier. In general, the number of clusters for each component i.e.  $p, q$ , etc. can be different as shown in figure 3.4. Every cluster contains both shape and texture information so  $\mathbf{C}_{nk} = \{\mathbf{S}_{nk} | \mathbf{G}_{nk}\}$  and the union of  $p$   $\mathbf{C}_{nk}$  clusters gives  $\mathbf{F}_n$ . Figure 3.6 shows sample shapes of cheeks + eyebrows taken from two different clusters.

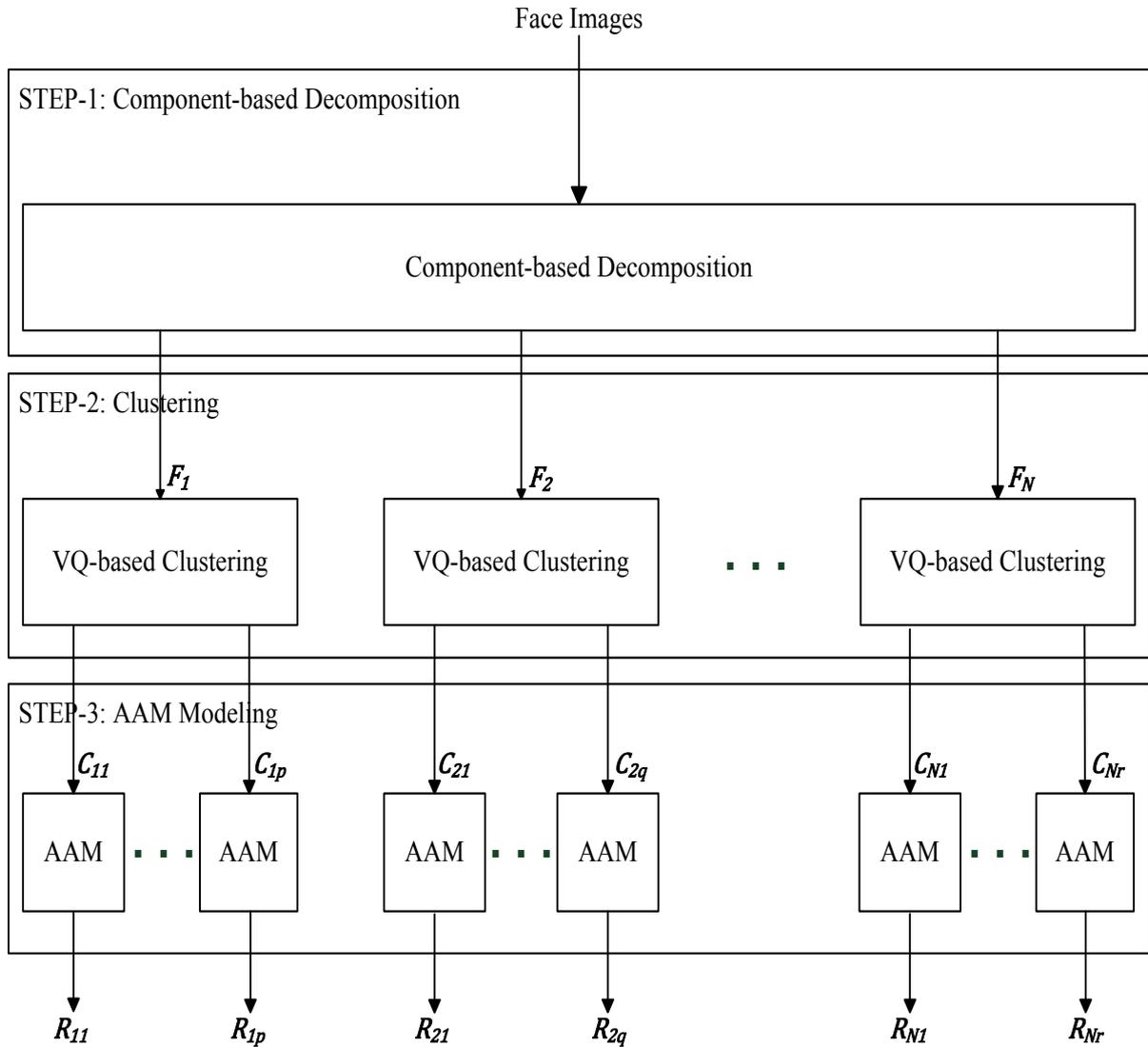


Figure 3.4: System Diagram of proposed MC/MM-AAM. Here  $F_1, F_2, F_3, \dots, F_N$  are ( $N = 4$ ) components-based datasets and  $C_{nk}$ 's are corresponding component clusters. These are used to produce component based model matrices  $R_{nk}$ 's. The number of clusters employed for each component can be the same i.e.  $p = q = r = 8$  or it can differ.

STEP-3: A corresponding component model and a model matrix  $R_{nk}$  is constructed for each cluster using conventional AAM optimization, see figure 3.4. Thus a training process produces  $(p + q + \dots + r)$  model matrices  $R_{nk}$  which are stored and can be subsequently used in the synthesis of an unseen input face image.

### 3.2.2 MC/MM-AAM Synthesis Phase

Following the previously generated component models, the best model for each component is selected and these are subsequently fused to form a single set of parameters that represents the complete face. The proposed synthesis process can be explained as follows:

- 1) Given face image  $t$  is decomposed into components to form  $f_1^t, f_2^t, \dots, f_N^t$  shape vectors and their corresponding texture vectors  $g_1^t, g_2^t, \dots, g_N^t$ .
- 2) Apply  $p$  conventional iterative AAM fitting algorithms based on model matrices  $R_{nk}$  ( $k = 1, 2, \dots, p$ ) for each  $n = 1, 2, \dots, N$  component. Obtain same number of model parameters' vectors  $c_{n1}^t, c_{n2}^t, \dots, c_{np}^t$  for each component.
- 3) For each component, select the best model parameters' vector on the basis of a minimum average Mean Square Error (MSE). MSEs are formed between the original texture and the textures associated with the  $p$  models of each component and calculated across all iterations of the fitting algorithm as given in case of conventional AAM.
- 4) Synthesize only the shape vectors  $\hat{f}_1^t, \hat{f}_2^t, \dots, \hat{f}_N^t$  of all  $N$  components with the best model parameters selected above. Combine all component-based shape vectors to form one single vector that represents the whole face shape i.e.

$$\hat{f}^t = [\hat{f}_1^t, \hat{f}_2^t, \hat{f}_3^t, \dots, \hat{f}_N^t]. \quad (3.6)$$

- 5) Finally in the last step, the whole face texture is synthesized for the shape vector obtained in previous step. For this purpose, the corresponding texture of the best shape vector is projected in the Eigenspace obtained from a whole face conventional AAM. The resulting model parameters are then used to synthesize the whole face texture.

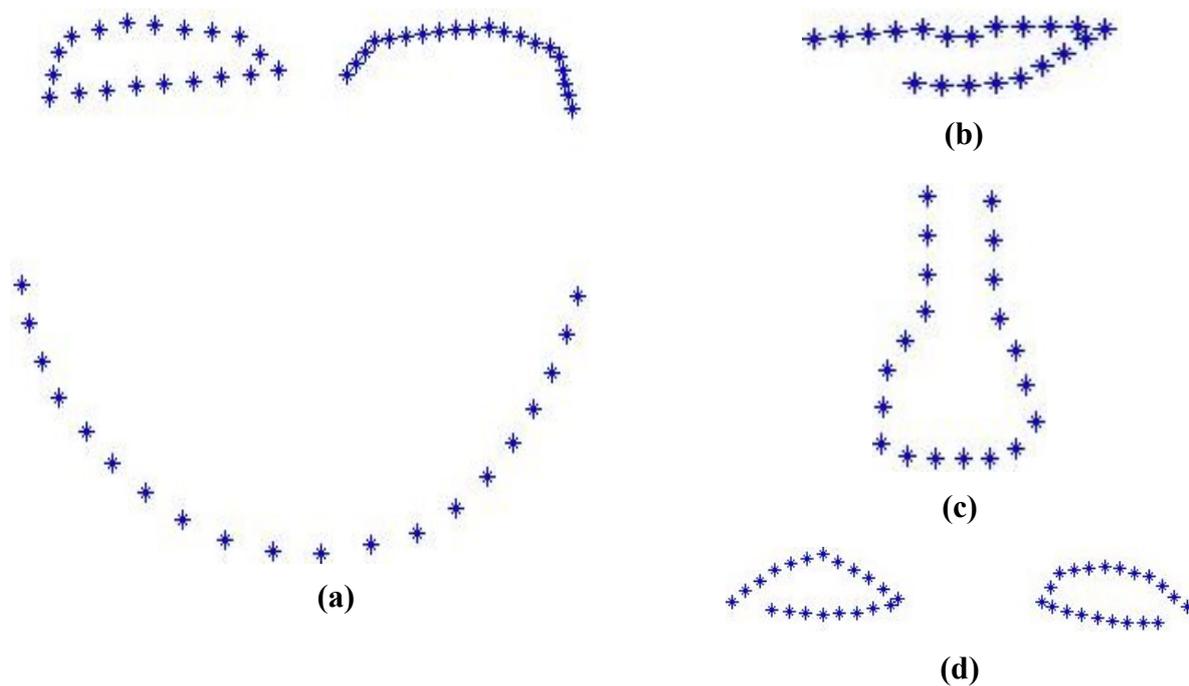


Figure 3.5: Example shapes of four components i.e. cheeks + eyebrows, nose, mouth and eyes.

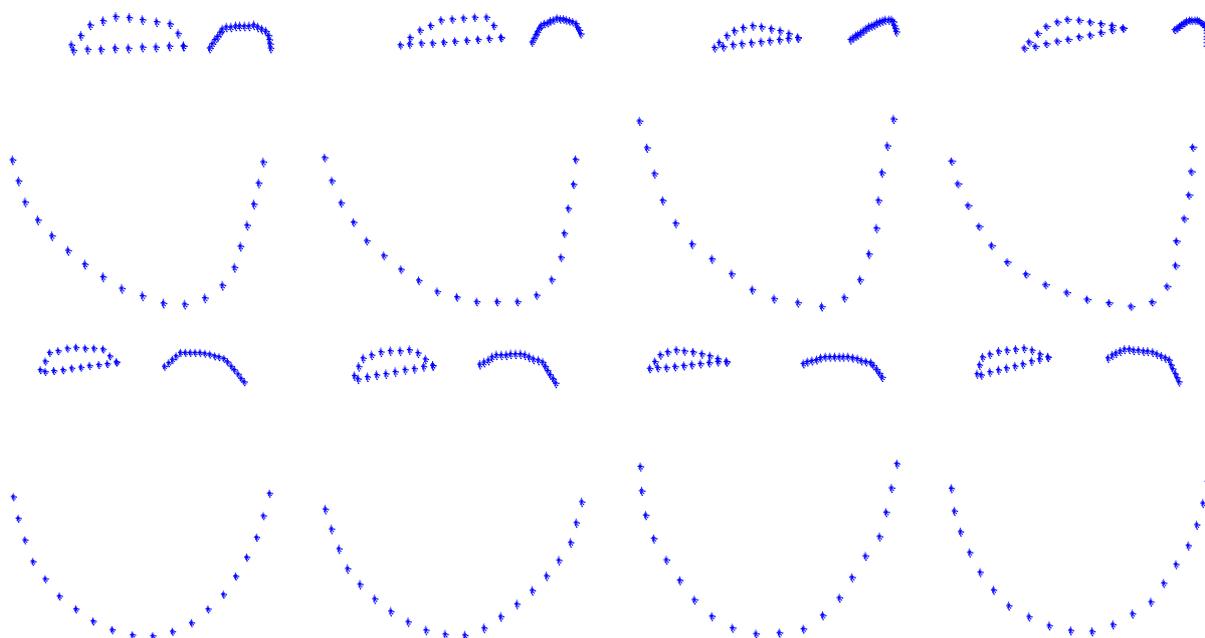


Figure 3.6: Example shapes of cheeks + eyebrows component taken from two different clusters: first row shapes belong to one cluster and second row belong to another cluster. Intra-cluster similarities and inter-cluster variations can be observed here as well.

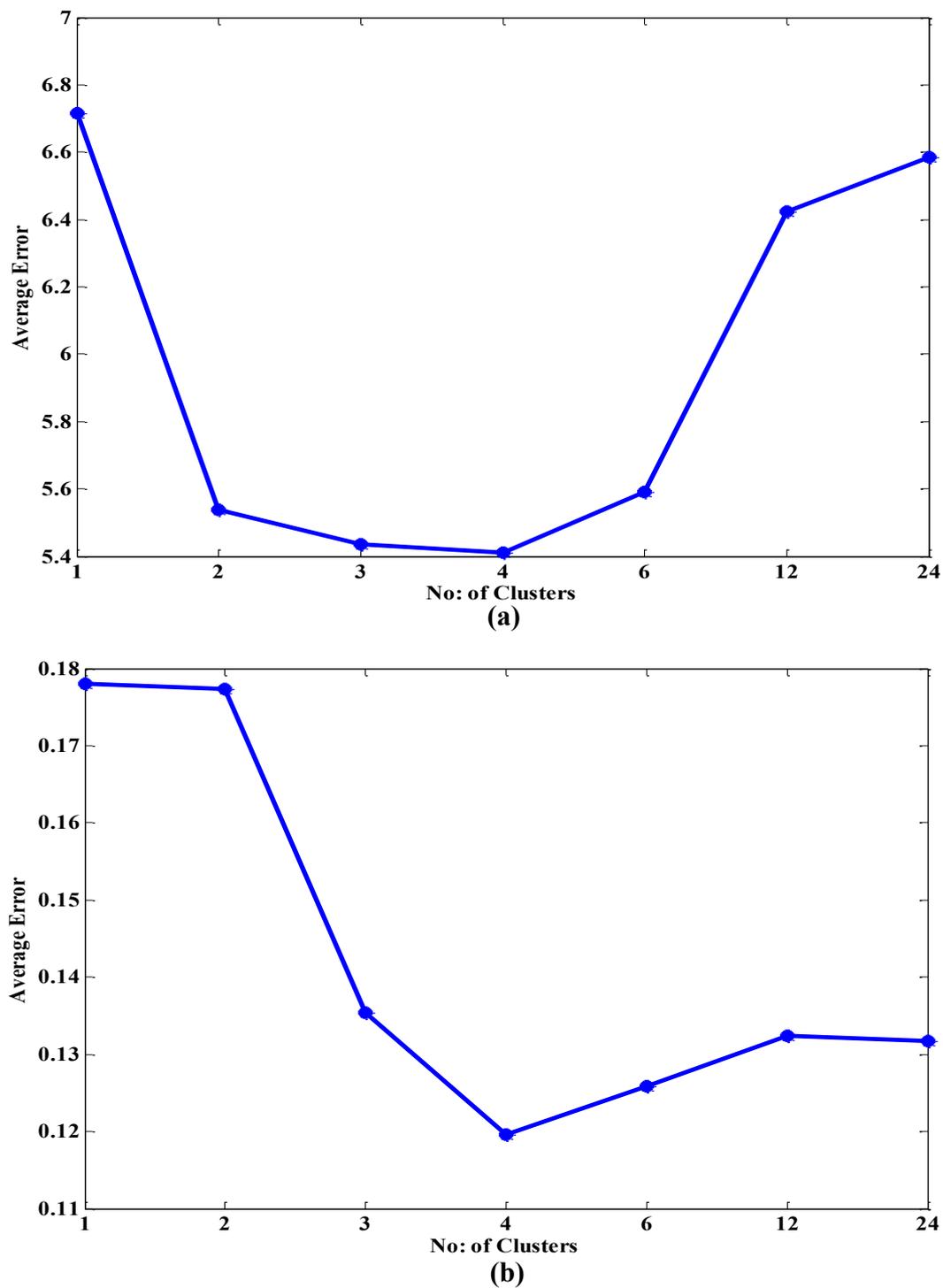


Figure 3.7: Error Vs No. of Clusters. (a) Error between synthesized shape coordinates and ground truth coordinates averaged across 108 test images for each value of  $p$ , (b) error between synthesized texture intensities and ground truth texture intensities averaged across 108 test images.

### 3.3 Experimental Results and Discussion

Performance assessment and thus comparisons between MM-AAM, MC/MM-AAM and conventional AAM were performed in two experimental setups: i) a first experiment, which involved only MM-AAM due to its lesser complexity, explains the impact of number of clusters on overall synthesis performance and ii) a second experiment which compares the performance of all three frameworks i.e. conventional AAM, MM-AAM and MC/MM-AAM.

#### 3.3.1 Impact of Number of Clusters

This experiment investigates how the proposed MM-AAM system performance changes with respect to number of clusters. For this purpose, a subset of publically available facial dataset IMM [51] was used for modeling and synthesis purpose. IMM consists of 240 annotated images (6 images per person). Each image is  $640 \times 480$  pixels in size and comes with  $M = 58$  hand labeled shape points which outline face contours. From these images, in order to constrain experimental complexity, only 72 images (12 persons with 6 images per person) have been used to train the system for different number of clusters i.e.  $p = 1,2,3,4,6,12,24$ . Note that the number of training images per cluster varies, for a given value of  $p$ . Furthermore from the remaining 168 images 108 images of 18 persons with 6 images per person were used for each value of  $p$  to test the system performance.

Note that model accuracy or fitting was evaluated with respect to both shape and texture. In the case of shape and for each input test image  $t$ , an average point-to-point error  $E(\hat{f}^t, f^{gt})$  between modeled shape and ground truth shape coordinates was calculated as suggested in [52], i.e.:

$$E(\hat{f}^t, f^{gt}) = \frac{1}{M} \sum_{m=1}^M \sqrt{(\hat{x}_m^t - x_m^{gt})^2 + (\hat{y}_m^t - y_m^{gt})^2}, \quad (3.7)$$

where  $M$  is total number of landmark points. Similarly, a normalized Mean Square Error (MSE)  $E(\hat{g}^t, g^{gt})$  was calculated between modeled texture  $\hat{g}^t$  and the ground truth texture  $g^{gt}$  which is effectively contained by ground truth shape points  $f^{gt}$ . In particular,

$$E(\hat{g}^t, g^{gt}) = \frac{1}{N} \sum_{n=1}^N (\hat{g}_n^t - g_n^{gt})^2, \quad (3.8)$$

where  $N$  is the length of the texture vectors  $\hat{g}^t$  and  $g^{gt}$ . Then  $E(\hat{f}^t, f^{gt})$  and  $E(\hat{g}^t, g^{gt})$  are averaged over the 108 test images available for each value of  $p$ , see figure 3.7.

In general, figure 3.7 suggests that MSE decreases as number of clusters are increased and thus overall modeling system performance increases by applying the proposed MM-AAM approach. It also suggests that there is an ‘‘optimum’’ value of number of clusters  $p_{op}$ , in this experiment, for which MSE is minimum.

Note that MSE values start to increase for  $p > p_{op}$ , due to the population of images per cluster becoming relatively small and as a consequence, model fitting accuracy per cluster deteriorates. Furthermore system training using a considerably larger number of input images, that will allow large number of clusters with enough number of images per cluster for better AAM modeling, is expected to reveal a type of convergence towards a minimum MSE floor value in system performance behaviour.

### 3.3.2 Synthesis Performance Comparisons

The second experiment employs a different and much larger input image dataset and thus provides a more accurate performance comparison between MM-AAM, MC/MM-AAM with  $p = 8$  and conventional AAM systems. Thus all three systems have been trained using face images taken from the IMM dataset [51] and also from another publically available FG-NET facial dataset [24]. FG-NET is a database of subject faces at different ages and contains 1002 color or gray-scale face images of 82 persons (12 images per person on average) with an age range from infant to 69 years. Note: the images from both datasets are converted to gray scale and extracted textures are made equal in size using cubic spline interpolation.

Experimentation during training involved 175 images of 35 persons (5 images per person) from the IMM dataset and 480 images of 60 persons (8 images per person) from FG-NET dataset. Note that the number  $M$  of hand labeled face shape points is kept to  $M = 58$  in both datasets. Experimentation during testing involved two different test datasets. One dataset, named as “seen dataset”, contains 35 images from IMM dataset and 240 images from FG-NET dataset (4 images per person on average) of persons having example images included in the training set, whereas the second test dataset contains 30 images of 5 completely unseen persons (6 images per person) from IMM dataset and 160 images of 20 completely unseen persons (8 images per person on average) from FG-NET dataset, and hence it is named as “unseen dataset”. Again, point-to-point MSE error  $E(\hat{f}^t, f^{gt})$  and  $E(\hat{g}^t, g^{gt})$  values (see Eq. 3.7 and 3.8) were calculated for the images of both types of test datasets. Table 3.1 shows shape and texture errors, averaged across all the test images, for both “seen dataset” and “unseen dataset” respectively, note that both the proposed frameworks, i.e. MM-AAM and MC/MM-AAM, outperform the conventional AAM system. Moreover, although the component based system has performed better than the system based on whole face image,

the difference in performance may be considered too small, as shown in table 3.1, when taking into consideration the added complexity of MC/MM-AAM. System complexity with respect to separate model initialization and optimization for each individual component restricted us to use MC/MM-AAM for face recognition or automatic age estimation. However, in future if somehow this complexity is reduced, MC/MM-AAM may be used effectively in real-life applications. A further illustration of this fact is shown in figure 3.8 and figure 3.9, here sample synthesized shapes from all three systems are compared with the corresponding ground truth shapes, and the better model fitting performance of MM-AAM and MC/MM-AAM over AAM becomes apparent.

**Table 3.1: Average Shape and Texture Errors (Average; Standard Deviation) for both Seen and Unseen Datasets.**

	Unseen Dataset		Seen Dataset	
	Shape Error (Avg; Std)	Texture Error (Avg; Std)	Shape Error (Avg; Std)	Texture Error (Avg; Std)
<b>AAM [14]</b>	(12.1174; 6.7521)	(0.4966; 0.2107)	(10.8399; 4.3520)	(0.4957; 0.2416)
<b>MM-AAM [33]</b>	(7.5937; 8.5774)	(0.1647; 0.0938)	(6.1006; 2.1624)	(0.1306; 0.0504)
<b>MC/MM-AAM [32]</b>	(7.1256; 6.2458)	(0.0583; 0.0919)	(6.0292; 2.3074)	(0.0332; 0.0154)

Finally, figure 3.10 illustrates visually and possibly more effectively, the MM-AAM and MC/MM-AAM advantage over AAM, by offering a comparison between “Target”, MM-AAM modeled textures, MC/MM-AAM modeled textures and AAM modeled textures of some of the test images from both types of dataset. Again a small improvement in performance, offered by MC/MM-AAM, over MM-AAM is evident.

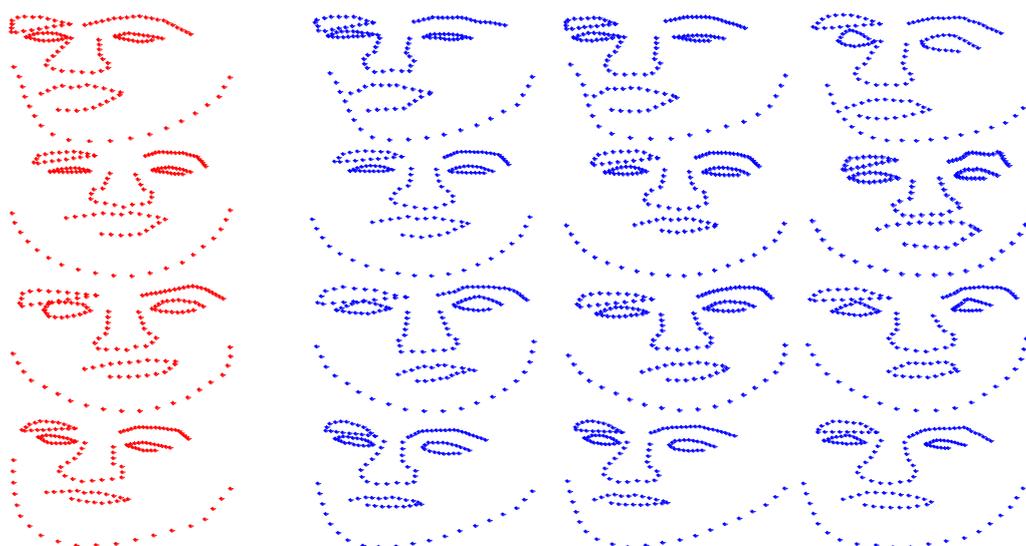


Figure 3.8: Sample synthesized shapes for “Seen Dataset”; first column in red is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. The difference between results obtained using MM-AAM and MC/MM-AAM is not large but they both significantly outperformed conventional AAM.

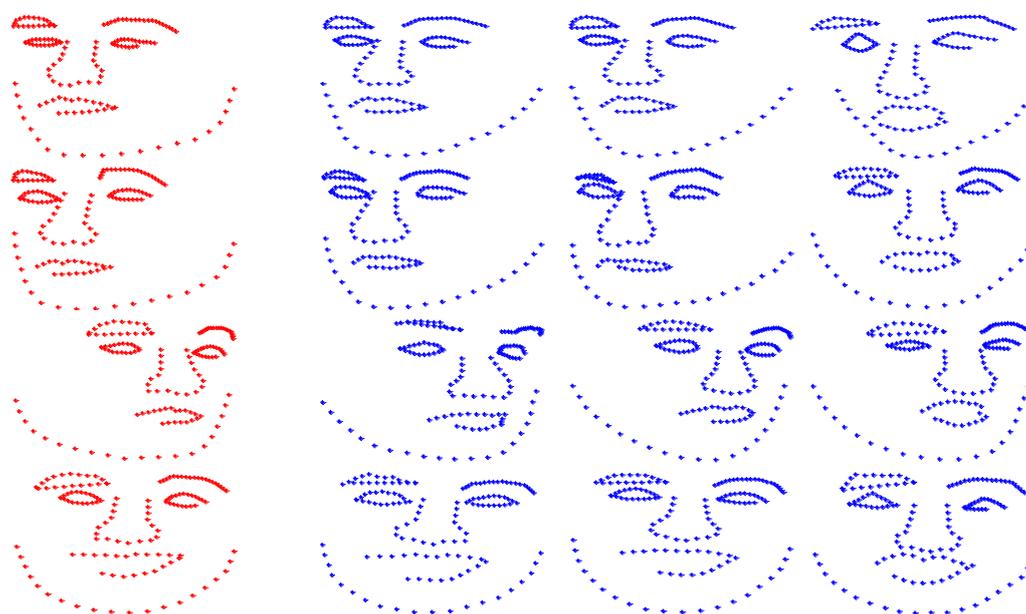


Figure 3.9: Sample synthesized shapes for “Unseen Dataset”; first column in red is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. Again MM-AAM and MC/MM-AAM significantly outperformed conventional AAM.



Figure 3.10: Sample synthesized textures; first column is Ground Truth, second column is MM-AAM, third column is MC/MM-AAM and the last column is AAM. MM-AAM and MC/MM-AAM outperformed conventional AAM, but the difference between the two new schemes is not significant.

### 3.4 MM-AAM as Applied to Face Recognition

This section examines the applicability of AAM and MM-AAM in face recognition and proposes an FR-MM-AAM [30] system that is far less affected by input data variability and thus outperforms FR-AAM. Here, input data variability relates to differences in face illumination, pose and expression [53]. Furthermore, input data variability is generally been accepted as having an adverse effect on the overall performance of both face modeling/synthesis (FM/S) procedures and on face recognition (FR) systems. Note that due to the excessive complexity characteristics of MC/MM-AAM only MM-AAM has been used for the face recognition experiments.

This proposed FR-MM-AAM framework has been developed and evaluated in two phases, that is: i) a design phase based on system training and using a wide range of input face images and ii) a system performance testing phase, using both “seen” and “unseen”, by the face modeling and synthesis processes, input image data.

System training involves three major steps: Firstly, face image models, created according to MM-AAM procedures given in previous sections, are used to model the shape information of all faces included in the training input dataset. Shape information is then employed to extract, from corresponding training images, the actual facial texture, i.e. those face pixels contained within corresponding shapes. It is this actual face texture information that forms the basis for information discrimination and face recognition.

Secondly, Principal Component Analysis (PCA) [54] is employed on this texture information in order to obtain a lower dimensionality feature/texture space than that created in the previous step.

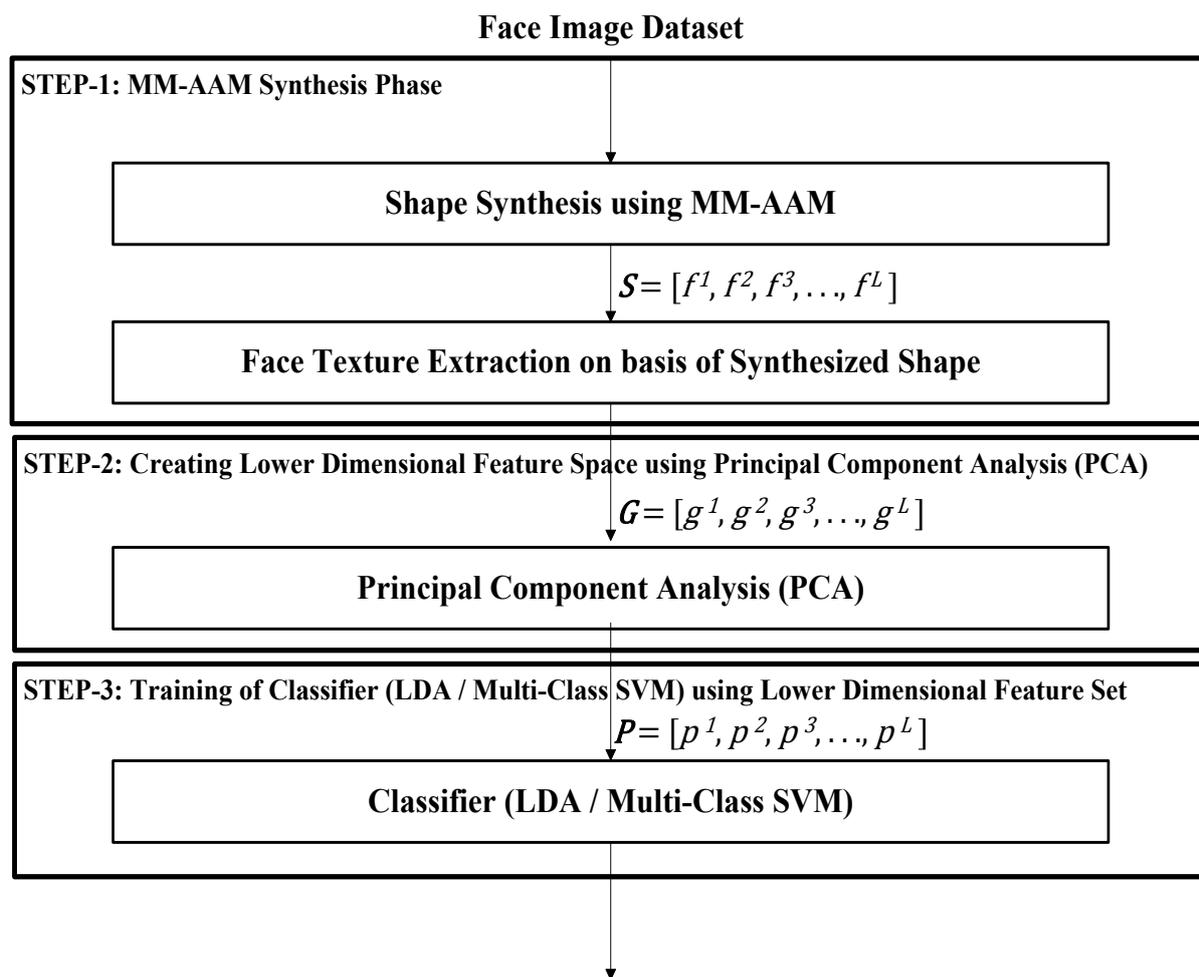


Figure 3.11: FR-MM-AAM Training. This involves the synthesis of face image shapes for the complete training image dataset  $F$ , using the MM-AAM procedures. MM-AAM training is a precursor to FR-MM-AAM training.

Finally, a given classification system is designed via training to operate on PCA derived information. Note that two classification methods have been employed in this work, i.e. Linear Discriminant Analysis (LDA) [55] and Multi-Class, Radial basis function (RBF) kernel Support Vector Machines (SVM) [56].

Testing system performance involves the synthesis of face shape information for a given input image and the subsequent extraction of corresponding texture information. This face

texture is projected on the PCA Eigen-Space that is created during the training phase, to yield a lower dimensional feature vector, which then is classified to one of a number of possible subjects.

Face recognition system performance has been evaluated using i) “seen” and ii) “unseen” input face images. In the first case, face shape models were defined during training using a set of 655 input images (taken from 95 subjects), whereas system recognition performance was evaluated based on a set of images of 35 out of the 95 subjects. In the second case the previous images of 35 subjects were used only during recognition and they were therefore “unseen” by the shape modeling process.

### 3.4.1 FR-MM-AAM System Training

Training encompasses three major steps. These are shown in figure 3.11 and discussed below. Note that MM-AAM operation is a precursor to FR-MM-AAM training.

1) Consider that the MM-AAM synthesized shape information of the  $i$ th input face image is represented as

$$\hat{f}^i = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_M, \hat{y}_1, \hat{y}_2, \dots, \hat{y}_M]^T \quad (3.9)$$

where  $\{(\hat{x}_m, \hat{y}_m)\}$  are the synthesized coordinates of the  $m = 1, 2, \dots, M$  landmark points outlining different facial components. Furthermore, the set  $\hat{\mathbf{S}}$  of shape information obtained from all the  $L$  training images of dataset  $F$  is given by

$$\hat{\mathbf{S}} = [\hat{f}^1, \hat{f}^2, \hat{f}^3, \dots, \hat{f}^L], \quad (3.10)$$

Now, for each face image  $i$ , face texture information  $g^i$  is extracted (i.e. pixels contained within shape defined face outlines) using corresponding shape coordinates  $\hat{f}^i$ . Thus, the set  $\mathbf{G}$  of face texture information obtained from all  $L$  training images is represented by

$$\mathbf{G} = [g^1, g^2, g^3, \dots, g^L]. \quad (3.11)$$

- 2) Principal Component Analysis (PCA) is applied on dataset  $\mathbf{G}$ , to obtain a lower  $N$ -dimensional representation, texture only related,  $p^i$  for each texture vector  $g^i$ . This is achieved by PCA selecting a relatively small number of  $N$  Eigen-Vectors on the basis of percentage energy captured by corresponding Eigen-Values [54].
- 3) Finally, a classifier (LDA or Multi-Class SVM) is trained using the set  $\mathbf{P}$  of feature vectors  $p^i$   $i = 1, 2, \dots, L$ .

### 3.4.2 FR-MM-AAM System Testing

System operation involves the following three procedures:

- 1) Given an input face image  $t$ , MM-AAM synthesis is applied that derives shape coordinate information  $\hat{f}^t$ . This information allows the corresponding face texture  $g^t$  to be obtained from the input image.
- 2) Texture  $g^t$  is projected onto the  $N$ -dimensional Eigenspace produced during training, in order to obtain the  $N$ -dimensional feature set  $p^t$ .
- 3) Finally,  $p^t$  is given as input to the classifier, that has been designed during training, the output of which returns a label class i.e. person assigned to image.

### 3.4.3 Experimental Results and Discussion

Experimentation was performed using computer simulation of FR systems and involved two different facial datasets i) FG-NET and ii) the IMM face dataset that are employed in two different experiments to assess the FR performance of systems.

In the first experiment, MM-AAM and AAM training, involved i) 480 images of 60 persons (8 images per person) from FG-NET dataset and ii) 175 images of 35 persons (5 images per person) from the IMM dataset. The number  $M$  of hand labeled face shape points was kept to  $M = 58$  in both datasets. Recall that MM-AAM and AAM training, as specified previously, is a precursor to the FR-MM-AAM three steps training process described in the previous section. In this experiment, face recognition system performance was evaluated using a “seen” dataset composed of the above 175 IMM images (used in shape modeling) plus another 35 image version of the same IMM subjects.

In the second experiment, MM-AAM and AAM training involved only the 480 images taken from FG-NET, whereas during face recognition the “unseen” dataset of 210 IMM images was employed as input data. Examples of MM-AAM and AAM synthesized shapes with their corresponding extracted textures, for both “seen” and “unseen” input datasets are shown in figure 3.12. These images are indicative of the improved shape modeling accuracy of MM-AAM, as compared to AAM, particularly in the case of “unseen” images. Furthermore, and in both experiments, FR system training and testing has been performed using the Leave One Image Out (LOIO) fold approach, i.e. in each fold one image taken from each person is left out for testing whereas all remaining images are used for system training.

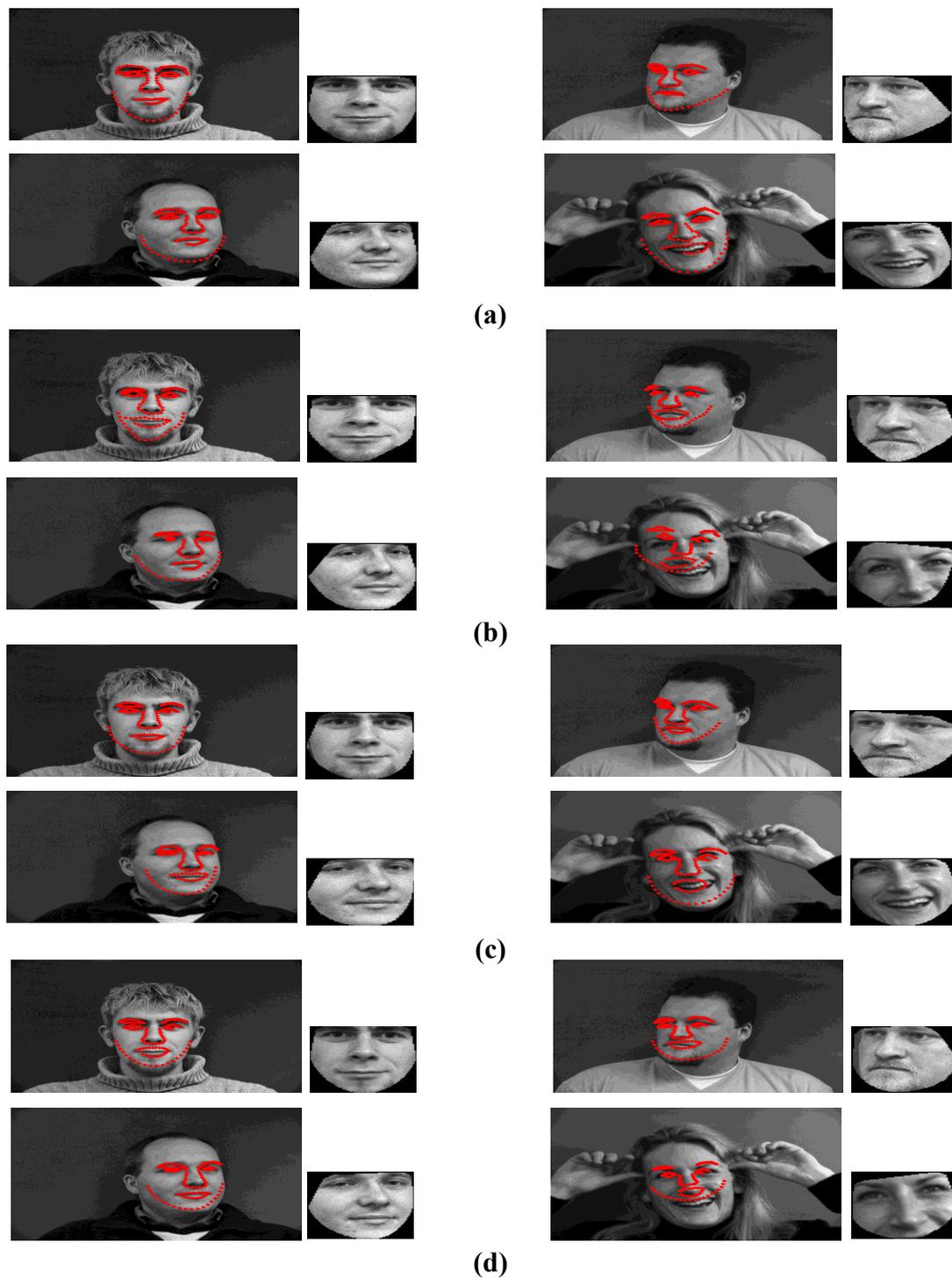


Figure 3.12: Examples of synthesized shapes and their corresponding extracted textures: a) and b) are examples obtained from previously “seen” data with images that are known to modeling process and correspond to MM-AAM and AAM respectively, whereas c) and d) are results with MM-AAM and AAM operating on “unseen” data with images that are unknown to modeling process.



Figure 3.13: Examples of manually cropped images used in the performance evaluation of conventional FR-PCA. The purpose of using cropped images is to generate input images which are therefore comparable to those used in FR-MM-AAM and FR-AAM system performance experiments, see figure 3.12.

Notice that in addition to the above techniques, a conventional FR-PCA system has been simulated and tested. Here cropped image face information i.e. pixels, see figure 3.13, are used directly in dimensionality reduction (PCA) and resulting coefficients are classified. The purpose for using manually cropped image data is to compare the FR-PCA system performance with that obtained from the FR-AAM and FR-MM-AAM systems, where texture is obtained via shape modeling.

FR system performance bars are plotted in figure 3.14 and figure 3.15 that show recognition rates for systems using two different classification techniques i.e. LDA and Multi-Class SVM. Note that throughout this work, retained Eigen-Vectors correspond to the largest Eigen-Values that capture 90% of cumulative signal energy. In both figures, results indicate clearly that FR performance based on texture extracted using MM-AAM modeled shapes and with the system operating on “seen” input data is significantly better than that obtained from both FR-AAM and FR-PCA.

In the case of LDA classification, see figure 3.14, FR-MM-AAM offers a recognition rate of 94.29% and 69.05% for “seen” and “unseen” datasets, respectively. Furthermore FR-MM-

AAM outperforms FR-AAM by 42.39% and 25.50%, for “seen” and “unseen” datasets, respectively. This is indicative of AAM’s inability to cope with input data variability, even in the case of previously “seen” inputs. Furthermore, and in the case of “seen” input images, FR-MM-AAM outperforms FR-PCA by 21.91%. The above noted general trends in FR performance remain valid in figure 3.15. Note that the implementation of Multi-Class SVM, used in this work, is from the statistical pattern recognition (STPR) toolbox for Matlab [57] and furthermore the RBF kernel is optimized for each FR system, separately, for both types of datasets.

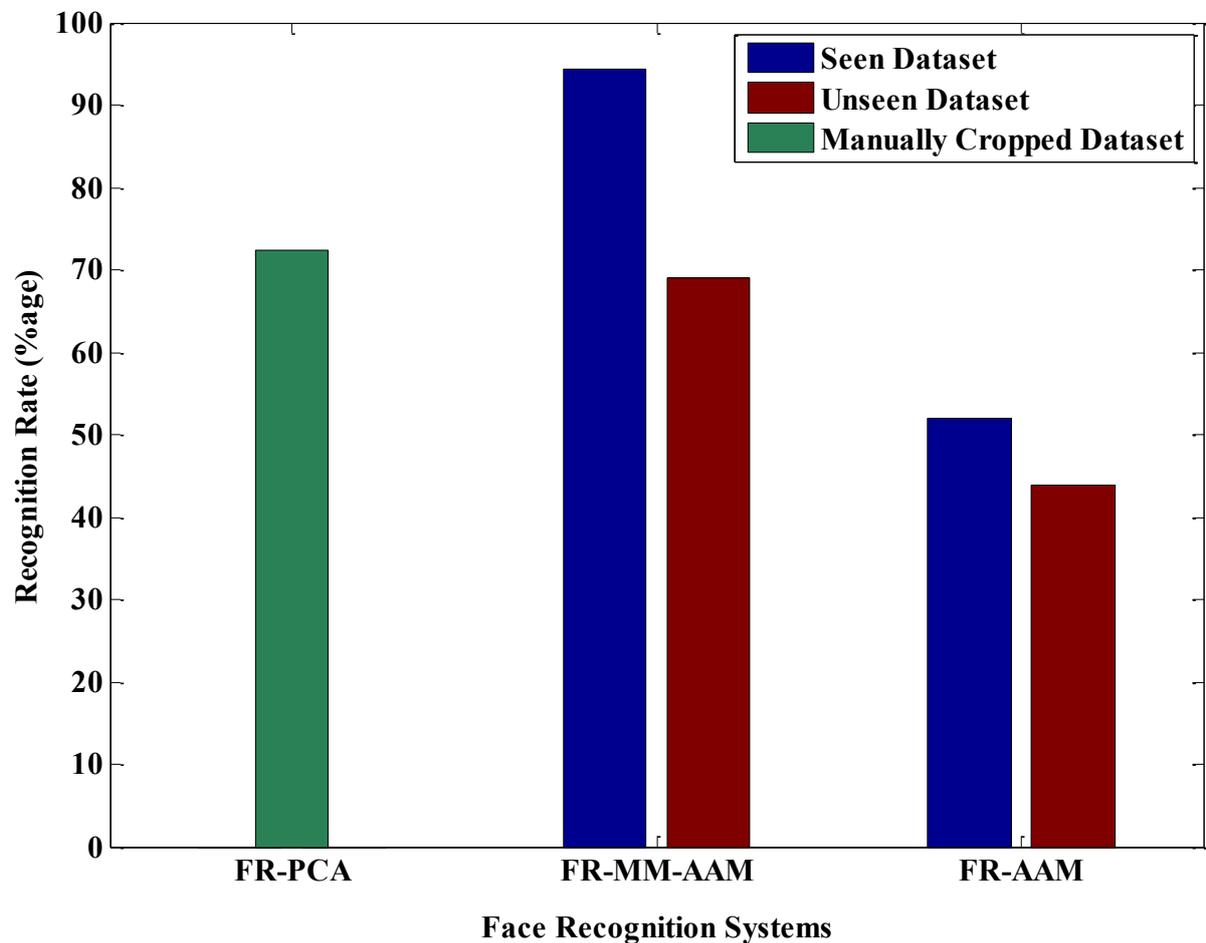


Figure 3.14: LDA Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Blue Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Green Bar).

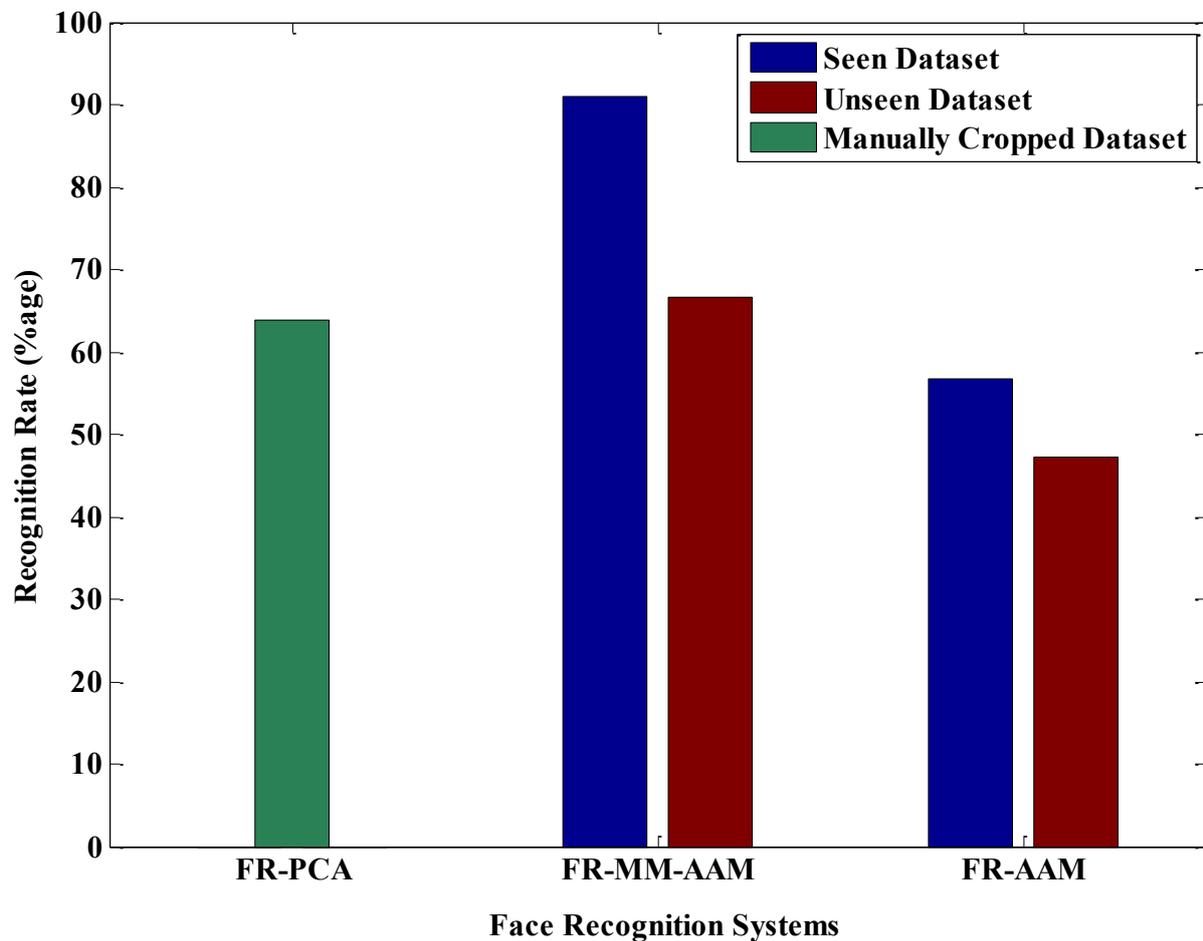


Figure 3.15: Multi-Class SVM Classifier: Bars show recognition rates achieved by the three FR methods under consideration for all three types of datasets i.e. Seen Dataset (Blue Bar), Unseen Dataset (Red Bar) and Manually Cropped Dataset (Green Bar).

### 3.5 Conclusion

The proposed face image modeling/synthesis methodologies, MM-AAM and MC/MM-AAM, improve image synthesis performance as compared to existing and commonly used active appearance model (AAM) technique and provide a novel solution to the problem of feature extraction and representation with respect to FI-AAE application. Furthermore, both proposed systems are generic in nature and can model/synthesise even face images that are unseen for the system. MM-AAM operates on the whole face image whereas MC/MM-AAM exploits the local information by using each face component separately. They cluster face image data into small groups on basis of shape similarities and yield multiple models, out of which the best one is selected for the synthesis of unknown test image. The novelty of the proposed frameworks stem from the notion that face data variability (i.e. due to pose, expression, illumination) can be reduced by splitting data into smaller groups of similar characteristics, an action that eventually facilitates the subsequent modeling and synthesis process. Although MC/MM-AAM performs better than MM-AAM in terms of synthesis performance, computational complexity with respect to separate initialization and optimization of AAM for each component is prohibitively large for use in applications such as face recognition or automatic age estimation. However, if somehow this complexity is reduced in future, MC/MM-AAM may be a better option to use in real-life applications due to its better synthesis performance.

In conclusion computer simulation based experimental results, obtained using two different types of input databases, show that MM-AAM and MC/MM-AAM can deliver improved face image modeling/synthesis performance of the order of 37%, when compared to conventional active appearance model (AAM). Furthermore before employing MM-AAM as a tool for feature extraction in the proposed FI-AAE system, discussed in Chapter-5, its capability of

synthesize unseen facial shapes is ascertained by employing it in FR application and obtained simulation results shown that FR-MM-AAM recognition rates outperform those obtained from FR-AAM by as much as 42.39% and 25.50%, for “seen” and “unseen” datasets, respectively.

## Chapter 4

### Face Image Data Variability

In addition to the age related feature extraction and representation discussed in Chapter-3 under face image modeling and synthesis, the performance of face image based automatic age estimation (FI-AAE) systems is also heavily dependent on the application related face dataset generation methodology and resulting input dataset characteristics. Thus face data collection is a critical element of system performance and deserves considerable attention.

In general, datasets associated with any type of classification system of a particular application domain come with some undesired variation among their samples that adversely affects performance of that system. For example, an age-related face image dataset is desired to contain only aging variation among face images of different age groups. However it is very common for these datasets, such as FG-NET dataset [24], to have high levels of non-aging variation. Such variability is due to capturing conditions such as:

- Illumination
- Pose
- Expression
- Makeup
- Facial attributes i.e. moustache, beard, glasses,

The data variation among different samples of a dataset is, generally, divided into two categories, i) inter-class variation or between class variation and ii) intra-class variation or

within class variation. It is believed that to distinguish between two different classes of a dataset, intra-class variation must be smaller than the inter-class variation [58]. However, the variation caused due to the factors mentioned earlier may potentially increase the intra-class variation and thus create overlap between different class distributions which in turn may make class separability an even more difficult task. In the case of FI-AAE and since aging is a slow process, face image datasets, used to train/test the age classifiers, usually suffer with small inter-class/inter-group age variation and large intra-class variation. This is due to the non-aging variation factors given above, which make it quite hard for classifiers to distinguish between images of two different age groups and cause a decrease in overall system performance. This implies that a relationship exists between the face image data variability and classification performance of a given system. Therefore, the ability to somehow measure dataset variability can lead to prediction of classification performance and thus to the selection of an appropriate system suitable to a given application. Furthermore an appropriate face image variability measure can be also used to rank face image datasets in terms of their classification difficulty level.

Now several researchers have proposed a variety of techniques to cope with face image data variability and thus improve age classification performance, see Chapter-5. However, to the best of our knowledge no one has come up with a measure of face data variability that can be used to i) predict system classification performance and ii) select appropriate face image datasets to train/test a given age estimation application. The conceptually nearest publication [27] proposes a set of different variability measures in order to represent object class properties in object classification. Here several variability measures are proposed which are based on intra-class similarities. As a consequence they can only be used with binary types of

classification problems and definitely not in multiclass scenarios such as those encountered in FI-AAE.

In this chapter, we propose a method to i) quantify the overall variation present in a face image dataset and ii) model the relationship between proposed variability measure and system classification performance. However, modeling of this relationship requires a large number of face image datasets to be used for model training and testing. Furthermore in the case of FI-AAE where the face images correspond to different ages, there is a scarcity of publically available age based face image datasets. Therefore in this chapter, instead of FI-AAE type of data we have studied variability using face image datasets employed in face recognition (FR). Thus, the capturing conditions mentioned earlier, which affect FR related face datasets, are expected to have similar detrimental effects on FI-AAE systems, while increasing intra-class variability to significant levels as compared to inter-class variations. It is for this reason that the proposed in this chapter formulation of FR “related” variability measure should in general be applicable to FI-AAE. Of course this logical assumption, which should be verified in future work, allowed us to work with easily available FR datasets and associated systems. Table 4.1 presents several well-known face image datasets, each created with its own image capture specification.

In addition to the above types of variation, the amount of variability allowed per type, during image capturing, is also of importance and emphasizes the need of a single variability measure to be defined. Consider for example the type of variability “pose” (see table 4.2) in FR related face datasets which can vary from 0 to  $\pm 90$  degrees. Large variations in pose can create severe visual changes between images taken of the same person, whereas, at the same time, have the potential to increase similarity between the images of different subjects. Of course in both cases recognition/classification becomes a more challenging task with adverse

implications in system performance. This general dependency of system performance upon specific input image sets and their associated types and levels of variability is also discussed in [59], see table 4.3.

**Table 4.1: Some of the widely used Face Datasets in Face Recognition Applications**

Database	RGB/Gray	Image Size	No: of Subjects	No: of Images / Subject	Variation
AT&T Face Dataset [60]	Gray	112x92	40	10	Pose, Illumination, expression
IMM Face Dataset [51]	RGB/Gray	640x480	40	6	Pose, Illumination, expression
The Extended Yale Face Dataset [61]	Gray	168x192	28	~576	Illumination, pose
Georgia Tech. Face Dataset [62]	RGB	640x480	50	15	Pose, Illumination, expression
Stirling Face Dataset [63]	Gray	269x369	35	9	Pose, expression
Indian Face Dataset [64]	RGB	640x480	61	11	Pose, Illumination, expression
FEI Face Dataset [65]	RGB	640x480	200	14	Pose, Illumination, expression
XM2VTSDB [66]	RGB	576x720	295	8	Pose
UMIST Face Dataset [67]	Gray	220x220	20	19-36	Pose

**Table 4.2: Face Datasets having Pose Variation**

Database	No: of Subjects	Pose Variation
AT&T Face Dataset [60]	40	10 random poses within $\pm 20$ in Yaw and Tilt
Bern Uni Face Dataset [68]	30	5 poses: $0^\circ$ , $\pm 20$ in Yaw and Tilt
XM2VTSDB [66]	125	5 poses: $0^\circ$ , $\pm 30$ in Yaw and Tilt
WVU [69]	40	7 poses: $0^\circ$ , $\pm 20$ , $\pm 40$ , $\pm 60$ in Yaw
MIT Face Dataset [70]	62	10 random poses within $\pm 40$ in Yaw and Tilt
Asian Face Dataset [71]	46	5 poses: $0^\circ$ , $\pm 20$ , $\pm 25$ in Yaw

**Table 4.3: Recognition Rate reported for different Pose Variation**

Database	No: of Subjects	Pose Difference among Images	Reported Recognition Rate
FERET [72]	100	$22.5^\circ / 67.5^\circ / 90^\circ$	100 / 99 / 92 [72]
CMU PIE [73]	68	$16^\circ / 45^\circ$	99.85 / 89.7 [74]
CMU PIE [73]	34	$45^\circ / 67.5^\circ / 90^\circ$	100 / 80 / 40 [75]

Here the authors employed their proposed Tied Factor Analysis based FR algorithm and results show i) that an FR system trained and optimized using a specific type of image variability performs differently when operating over datasets having different variability

characteristics and ii) that a relationship exists between the amount of image variability and system recognition performance.

All of the above discussion suggest that the ability to i) measure and then model image face data variability and ii) predict system performance, is an important research aim. Therefore, in this chapter, based on the above background and for a given dataset of face images, inter- and intra-subject dataset measures are first defined for face image datasets. These are subsequently combined to form a single variability measure (VM) which can be used to quantify the overall level of image variability in the dataset. Furthermore the relationship between VM values and face recognition rates is modeled using  $n$ th-degree polynomials. Thus VM/FR performance models are derived for four different face recognition (FR) systems and eleven publically available face image datasets.

Experimental results show that the modeling of FR performance in terms of VM allows relatively good performance prediction estimates. That is to say, given an input face dataset and its VM value as well as an FR versus VM model, FR system performance can be predicted reasonably well. Furthermore, the prediction capability of our proposed VM/FR models is evaluated using the face image datasets that were coded using JPEG at four different PSNR values as the test samples. Results show ability for VM/FR performance models to operate well even under noisy input conditions. .

The chapter is organized as follows: Section 4.1 explains in detail VM formulation, whereas, Section 4.2 describes the experimental set up used to produce computer simulation results. These results are then presented and discussed in the second part of Section 4.2. Concluding remarks are given in Section 4.3.

## 4.1 Variability measure (VM)

The proposed overall variability measure VM of an image dataset is composed of two components i.e. an inter- and an intra-Subject Class, denoted as VM-interSC and VM-intraSC respectively.

### 4.1.1 VM-intra- and VM-inter-Subject Class Components

Here the Normalized Cross Correlation (NCC-AB) [76] is used as a generic similarity measure between two face images A and B. In VM-intraSC, NCC is calculated among all the available images of each subject, whereas in VM-interSC, NCC is calculated among all images of one subject with respect to all images of all other subjects.

#### 1) VM-intraSC

First step to calculate VM-intraSC is to create a matrix  $\hat{\mathbf{C}}$  of order  $P \times Q$  which contains NCC values for all the subjects as given below:

$$\hat{\mathbf{C}} = \begin{bmatrix} \vartheta_{12}^1 & \vartheta_{12}^2 & \cdots & \vartheta_{12}^M \\ \vartheta_{13}^1 & \vartheta_{13}^2 & \cdots & \vartheta_{13}^M \\ \vdots & \vdots & \vdots & \vdots \\ \vartheta_{1N}^1 & \vartheta_{1N}^2 & \cdots & \vartheta_{1N}^M \\ \vartheta_{23}^1 & \vartheta_{23}^2 & \cdots & \vartheta_{23}^M \\ \vartheta_{24}^1 & \vartheta_{24}^2 & \cdots & \vartheta_{24}^M \\ \vdots & \vdots & \vdots & \vdots \\ \vartheta_{2N}^1 & \vartheta_{2N}^2 & \cdots & \vartheta_{2N}^M \\ \vdots & \vdots & \vdots & \vdots \\ \vartheta_{(N-2)(N-1)}^1 & \vartheta_{(N-2)(N-1)}^2 & \cdots & \vartheta_{(N-2)(N-1)}^M \\ \vartheta_{(N-2)(N-1)}^1 & \vartheta_{(N-2)(N-1)}^2 & \cdots & \vartheta_{(N-2)(N-1)}^M \\ \vartheta_{(N-1)N}^1 & \vartheta_{(N-1)N}^2 & \cdots & \vartheta_{(N-1)N}^M \end{bmatrix}, \quad (4.1)$$

here number of columns  $P$  are equal to the number of subjects  $M$  and number of rows  $Q$  are equal to  $\frac{N(N-1)}{2}$ ; where  $N$  is the number of images per subject in a particular face dataset.

Each element  $\vartheta_{nk}^m$  of matrix  $\hat{\mathbf{C}}$  is representing the maximum NCC value between two images  $n$  and  $k$  of  $m$ th subject which is calculated as

$$\vartheta_{nk}^m = \text{maximum element}\{\boldsymbol{\gamma}_{nk}^m(u, v)\}, \quad (4.2)$$

here, the matrix  $\boldsymbol{\gamma}_{nk}^m$  contains all the normalized cross correlation NCC values between the two images  $\mathbf{I}_n^m$  and  $\mathbf{I}_k^m$  of same subject. Furthermore  $\boldsymbol{\gamma}_n^m$ , originally given in [76], can be written as

$$\boldsymbol{\gamma}_{nk}^m(u, v) = \frac{\sum_{x,y} [\mathbf{I}_n^m(x, y) - \overline{\mathbf{I}_n^m}] [\mathbf{I}_k^m(x - u, y - v) - \overline{\mathbf{I}_k^m}]}{\sqrt{\sum_{x,y} [\mathbf{I}_n^m(x, y) - \overline{\mathbf{I}_n^m}]^2 \sum_{x,y} [\mathbf{I}_k^m(x - u, y - v) - \overline{\mathbf{I}_k^m}]^2}}, \quad (4.3)$$

where  $x$  and  $y$  are the pixel coordinates while  $u$  and  $v$  refer to the shift at which the NCC value is calculated. Moreover,  $\overline{\mathbf{I}_n^m}$  and  $\overline{\mathbf{I}_k^m}$  are the means of the overlapped regions of the two images. Once,  $\hat{\mathbf{C}}$  is populated with NCC values of a whole dataset, it is used to calculate VM-intraSc  $\hat{\varphi}$  in the way given below:

$$\hat{\varphi} = \hat{\mu} \times \hat{\sigma}^2,$$

where

$$\hat{\mu} = \frac{1}{(P \times Q)} \sum_{p=1}^P \sum_{q=1}^Q \hat{\mathbf{C}}(p, q), \quad (4.4)$$

and

$$\hat{\sigma}^2 = \frac{1}{(P \times Q) - 1} \sum_{p=1}^P \sum_{q=1}^Q (\hat{\mathbf{C}}(p, q) - \hat{\mu})^2.$$

In Eq. 4.4, a large value of mean  $\hat{\mu}$  represents a larger similarity or lesser variation among images of each subject, whereas a large value of variance  $\hat{\sigma}^2$  corresponds to the whole range of variation covered by all the images of a subject. Therefore, a face dataset having large number of images per subject, changing smoothly from one image to other to cover a larger variation, will produce a larger value of VM-intraSC  $\hat{\phi}$ .

To validate the above mentioned assumptions, we performed two different experiments. Both of these experiments involve Part-2 of the FEI face dataset [65]. FEI is a publically available face dataset that comes in four different parts. Each part contains 50 subjects with 14 color images per subject. 10 out of these 14 images smoothly cover a profile rotation of upto  $180^\circ$ , whereas remaining 4 images contain illumination and expression variation. In the first experiment, two  $\hat{C}_1$  and  $\hat{C}_2$  are formed from two different subsets of FEI face dataset (Part-2). The first subset is comprised of all those 10 images per subject that contain smooth rotational variation, whereas the other subset contains only 3 images per subject having approximate rotation of  $0^\circ$ ,  $90^\circ$  and  $180^\circ$  respectively. In figure 4.1, two normalized histograms corresponding to both  $\hat{C}_1$  and  $\hat{C}_2$  are shown, respectively. The histogram corresponding to dataset having a smooth pose variation from  $0^\circ$  to  $180^\circ$  (figure 4.1(a)) covers a larger range of NCC and hence yields a larger intra-subject measure value  $\hat{\phi}_1 = 0.0119$  as compared to  $\hat{\phi}_2 = 0.0061$  that corresponds to the dataset with larger variation among images of a subject (figure 4.1 (b)).

In the second experiment, VM-intraSC values are calculated for four different datasets named as DS1, DS2, DS3 and DS4 to produce a curve shown in figure 4.2. All four datasets, used here, are different from each other with respect to both number of images per subject and pose variation form one image to other image.

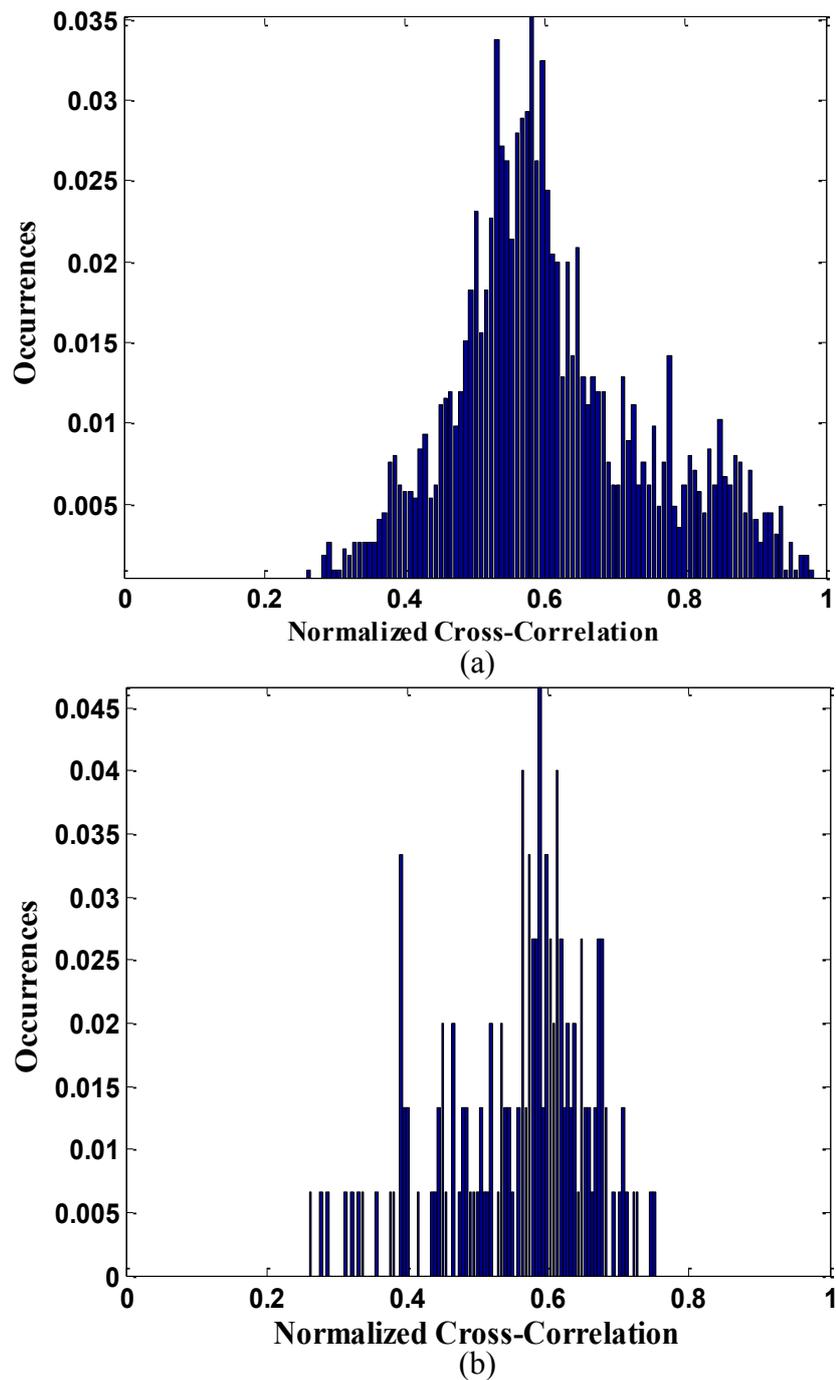


Figure 4.1: Normalized Histogram; a) represents  $\hat{\mathbf{C}}_1$  that contains the NCC values corresponding to the face dataset with 10 images per subject that smoothly change pose from one image to other to cover a rotation of upto  $180^\circ$ , b) represents set  $\hat{\mathbf{C}}_2$  comprised of NCC values for the face dataset with only three images per person having approximate pose rotation of  $0^\circ$ ,  $90^\circ$  and  $180^\circ$ , respectively.

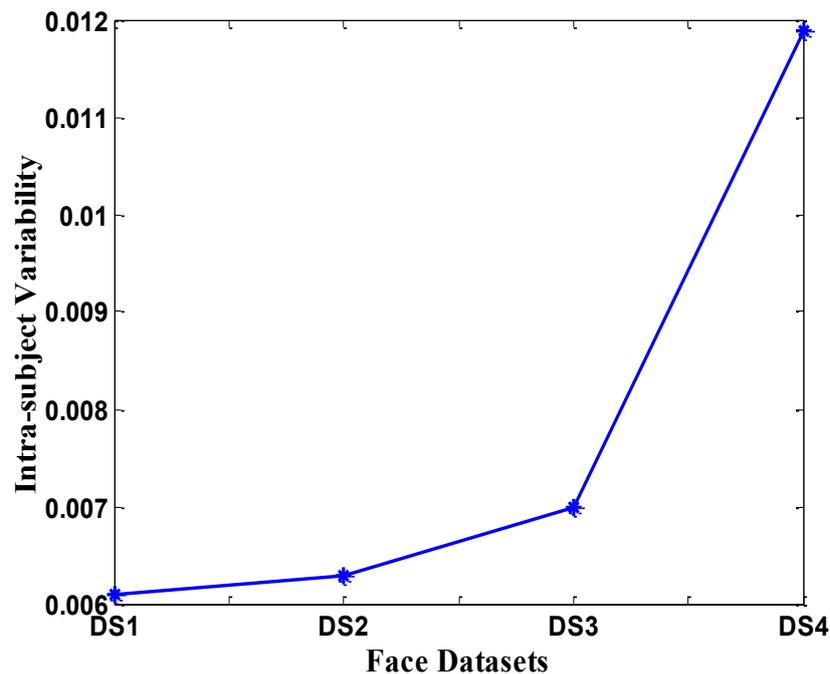


Figure 4.2: Intra-subject variability Vs face datasets Curve: Here face datasets shown on x-axis comprised of different numbers of images per person and different levels of pose variation from one image to other of same subject.

DS1 dataset contains three images per person with approximate rotation of  $0^\circ$ ,  $90^\circ$  and  $180^\circ$  respectively, DS2 comprises of four images per person with approximate rotation of  $0^\circ$ ,  $60^\circ$ ,  $120^\circ$  and  $180^\circ$ , respectively, DS3 contains five images per subject with approximate rotation of  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$  and  $180^\circ$ , respectively and DS4 contains all the ten images per subject. It is obvious from the figure 4.2 that an increase in number of images per subject, that smoothly cover a large range of variation, increases VM-intraSc. Hence it can be concluded from both experiments that: i) VM-intraSC is an effective representation of intra-subject variation among images of a face dataset and ii) A face image dataset with a large number of images per person, changing smoothly from one image to other, has a better capability to cope with environment conditions.

## 2) VM-interSC

As discussed in the previous section, quantifying intra-subject variation alone is not sufficient since inter-subject dataset properties are also equally important and should also be quantified. To easily/successfully distinguish one subject from others, there must be a large variation among the images of different subjects. Thus to quantify such inter-subject variability, another matrix  $\check{C}$  is created that contains Normalized Cross-Correlation values among images of one subject and all other images.  $\check{C}$  is populated in the way shown below:

$$\check{C} = \begin{bmatrix} C^{12} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ C^{13} & C^{23} & \mathbf{0} & \dots & \mathbf{0} \\ C^{14} & C^{24} & C^{34} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C^{1M} & C^{2M} & C^{3M} & \dots & C^{(M-1)M} \end{bmatrix},$$

and

$$C^{ml} = \begin{bmatrix} \vartheta_{11}^{ml} \\ \vartheta_{12}^{ml} \\ \vdots \\ \vartheta_{1N}^{ml} \\ \vartheta_{21}^{ml} \\ \vartheta_{22}^{ml} \\ \vdots \\ \vartheta_{2N}^{ml} \\ \vdots \\ \vartheta_{N1}^{ml} \\ \vdots \\ \vartheta_{NN}^{ml} \end{bmatrix}, \quad \text{with } m < l. \quad (4.5)$$

In Eq. 4.5  $\vartheta_{nk}^{ml}$  is the maximum NCC between  $n$ th image of subject  $m$  i.e.  $I_n^m$ , and  $k$ th image of subject  $l$  i.e.  $I_k^l$ . The value of  $\vartheta_{nk}^{ml}$  is calculated in the same way as given in Eq. 4.2 and Eq. 4.3. The order of matrix  $\check{C}$  is  $G \times H$  where the number of rows  $G$  is equal to  $(M - 1) \times N^2$  and number of columns are equal to  $M - 1$ . Once the matrix  $\check{C}$  is obtained, it is used to

form VM-interSC ( $\check{\phi}$ ) as given in Eq. 4.4. The corresponding  $\check{\mu}$  and  $\check{\sigma}^2$  are calculated using only elements present in the lower triangle of the matrix  $\check{\mathbf{C}}$ .

In case of inter-subject variability, a face dataset with large variation among the images, of different subjects, yields smaller NCC values which in turn result in smaller mean and variance values and hence in a small  $\check{\phi}$  value.

Since it is believed that a face dataset with large inter-class variation (i.e. small value of inter-SC  $\check{\phi}$ ) and small intra-class variation (i.e. large value of intra-SC  $\hat{\phi}$ ) always yields better classification performance, thus, it is expected that a face dataset which fulfils the following condition:

$$\check{\phi} \ll \hat{\phi}, \quad (4.6)$$

will produce relatively high classification results. Consider for example, the proposed VM-interSCs for two subsets used in first of the two experiments given above are  $\check{\phi}_1 = 0.0068$  and  $\check{\phi}_2 = 0.0071$ , respectively and hence the first subset with  $\check{\phi}_1 < \hat{\phi}_1$  can yield better recognition performance for any FR system as compared to second subset where  $\check{\phi}_2 > \hat{\phi}_2$ .

#### 4.1.2 Variability Measure (VM)

Both VM-intraSC and VM-interSC, defined in the previous section, are combined to form a single image dataset variability measure (VM). That is:

$$\phi = \hat{\phi} \times \sqrt{\hat{\phi}^2 - \check{\phi}^2}, \quad \text{for } \hat{\phi} > \check{\phi} \quad (4.7)$$

$\check{\phi}$  and  $\hat{\phi}$  are the previously defined inter- and intra-subject measures, respectively. The purpose of this scaling factor is to distinguish between the two datasets for which difference

term i.e.  $\sqrt{\widehat{\varnothing}^2 - \check{\varnothing}^2}$  yields same output for different values of VM-intraSC. Consequently, the dataset with higher value of VM-intraSC yields a larger value of VM  $\varnothing$  as compared to the dataset with a lower value of VM-intraSC.

Moreover, the above equation will produce the VM value  $\varnothing$  to rank only those datasets for which  $\widehat{\varnothing} > \check{\varnothing}$ , and all the datasets with  $\widehat{\varnothing} < \check{\varnothing}$  will be rejected straight away to be used in any classification system, as they violate the primary condition i.e. inter-class variation must be greater than intra-class variation.

## 4.2 Experimentation & Discussion

In order to investigate the effectiveness and validity of our proposed variability measure VM in representing facial data variation, we have performed a number of experiments. These are based on eleven different publically available face datasets and four different face recognition (FR) systems. In this section, firstly these datasets and FR systems are briefly introduced and then a discussion on experimental setup and results is given.

### 4.2.1 Face datasets

Face datasets used in experiments are as follow:

1) AT&T Face dataset [60]:

Dataset contains a total of 400 grayscale images; ten images of each of 40 different subjects. Images of each subject differ from each other with respect to the lighting, facial expressions and facial details. Size of each image is  $112 \times 92$  pixels.



Figure 4.3: Examples of manually cropped images and their corresponding original images.

#### 2) IMM Face Dataset [51]:

IMM consists of 240 annotated images (6 images per person). Each image is  $640 \times 480$  pixels in size and comes with 58 hand labeled shape points which outline face contours. Images of each subject vary in lighting, pose and facial expression. Out of all 40 subjects, 37 consist of RGB images whereas remaining three subjects have grayscale images.

#### 3) The Extended Yale Cropped Face Dataset [61]:

The original extended Yale Face Dataset B [61] contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions. In these experiments we have used a cropped version of this dataset reported in [77]. This dataset contains grayscale images that are manually aligned, cropped and then resized to  $168 \times 192$  pixels. For this dataset, we have used 2242 images of 38 subjects.

#### 4) Georgia Tech. Face Dataset [62]:

Dataset contains images of 50 different human subjects with 15 RGB images for each of the subject. These face images vary in size, facial expression, illumination and rotation. The average size of the faces in these images is  $150 \times 150$  pixels.

#### 5) Stirling Face Dataset [63]:

Stirling face dataset is comprised of 312 images of 35 subjects (18 female, 17 male). These monochrome images with a spatial resolution of  $269 \times 369$  vary in pose and facial expression.

#### 6) Indian Face Dataset [64]:

This database contains images of 55 distinct subjects (22 female, 33 male) with eleven different poses for each individual. In addition to the variation in pose, images with four emotions - neutral, smile, laughter, sad/disgust - are also included for every individual. As this dataset separate sets for female and male subjects, so in our experimentation we have used them as separate datasets. The size of each image is  $640 \times 480$  pixels, with 256 grey levels per pixel.

#### 7) FEI Face Dataset [16]:

FEI face dataset comes in four different parts. Each part contains 50 subjects with 14 RGB images per subject. 10 out of these 14 images smoothly cover a profile rotation of upto  $180^\circ$ , whereas remaining 4 images contain illumination and expression variation. Size of each image is  $640 \times 480$  pixels. Again, as these images are provided in four different parts, so each part is used in our experimentation as a separate dataset.

Before performing actual experimentation, face information is extracted from a given image. For this purpose, images of all the datasets are manually cropped to remove the background information as shown in figure 4.3.

### 4.2.2 Face Recognition (FR) Systems

Four different face recognition (FR) systems are used in our experimentation to verify the effectiveness of our proposed VM. As the purpose of this work is not to provide a comparison between different face recognition system and we are also not proposing any state-of-the-art face recognition system, therefore to prove our point, we have chosen four such appearance based face recognition approaches that can easily be implemented and are relatively simple. A brief description of these approaches is given below:

#### 1) Eigenfaces:

The “Eigenfaces”, introduced by Turk and Pentland [54], is one of the most thoroughly investigated approaches to face recognition [78-80]. Eigenfaces are the eigenvectors that characterize the variation across different face images of training dataset. Each  $N$ -dimensional face image is a linear combination of these eigenvectors and can be best approximated using only a few  $M$  ( $M \ll N$ ) ‘the best’ eigenvectors or principal components (PCs) having the largest corresponding eigenvalues and together containing  $P$  percent of overall training data variance. Normally, the value of  $P$  is kept in the range of 90 – 95 here  $P = 95$ . Face images from both training and testing datasets are projected on a subspace, also called as facespace, spanned by these  $M$  Eigenfaces. Then recognition is performed in the facespace by calculating the distance between known points (i.e. training data) and unknown points (i.e. testing data).

#### 2) Fisherfaces:

The second face recognition technique, we used in our experiments, is the well-known “Fisherfaces”. The Fisherfaces approach [80], is based on a two-stage strategy. In first stage, principal component analysis (PCA) is performed, the same way as discussed in Eigenfaces approach, to reduce the face image dimension, and then linear discrimination analysis (LDA)

is used to extract discriminative information out of these reduced dimensional features. The major aim of LDA is to provide such basis vectors that best describe the variation among different classes by maximizing the between-class variation and minimizing the within-class variation. Fisherfaces have been heavily investigated and modified to yield a number of different face recognition systems [81-84].

### 3) PCA + Multi-Class SVM:

In the third face recognition technique, PCA is used as a preprocessing step for dimensionality reduction and then the well-known Support Vector Machine (SVM) is used in multi-class mode to classify these reduced dimensional feature vectors.

SVM, originally introduced for binary classification by Vapnik and Cortes [56], is normally extended for a multi-class problem by using two basic strategies, that are i) One-versus-One and ii) One-versus-All [85]. The basic difference in both strategies is the number of classifiers trained. In One-versus-One strategy, one classifier for each pair of classes is trained so for  $N$  classes,  $\frac{N(N-1)}{2}$  classifiers are needed to be trained. During classification process, for every test sample, each classifier votes for one of the two classes and the class with maximum votes is selected for that test sample. In case of One-versus-All, one classifier per class is built and trained to classify between each class and rest of the classes, in this way for  $N$  classes we get  $N$  classifiers. We have used One-versus-All approach as it is computationally less expensive due to the use of a smaller number of classifiers.

SVM works on the principle of finding an optimal linear hyperplane that separates two classes from each other. In most real-world applications, with face recognition being one of them, linear separation is not a feasible solution for classifying data. Thus SVM is modified to act as a non-linear classifier using kernel technique. The purpose of such kernels is to transform data to a higher dimensional space where it can be linearly classified.

In our experimentation, the radial basis function (RBF) kernel is used. RBF kernel is based on a Gaussian kernel and is dependent on two parameters, one is called kernel parameter  $\sigma$  and the other is known as penalty factor  $C$ . For each dataset, a number of experiments with different pairs of these parameters have been performed and the best pair was selected on basis of maximum system classification performance.

#### 4) Normalized Cross-Correlation:

In the last face recognition technique, face image classification is done on the basis of maximum Normalized Cross-Correlation (NCC) between input test face image and the training images. Before calculating NCC, both test and training images are first normalized to zero mean and unit variance.

### 4.2.3 Results and Discussion

The effectiveness of the VM computed over a certain face image dataset, to be used as a means in predicting the performance of FR systems operating on the same image dataset is considered in this section, see figure 4.4 test architecture. Experimentation is done in two phases; the first phase involves computation of VM and actual recognition performance for all the datasets using the above four listed FR systems. Furthermore and for all datasets, FR systems are separately designed for delivering maximum performance.

Performance of each FR system is evaluated using the k-fold approach; k is equal to the number of images per subject in a particular face image dataset, and each fold contains one image per subject. For a k-fold cross validation test, k experiments are performed and in each experimental run, (k-1)-folds are used to train the classifier whereas the remaining fold is used for testing.

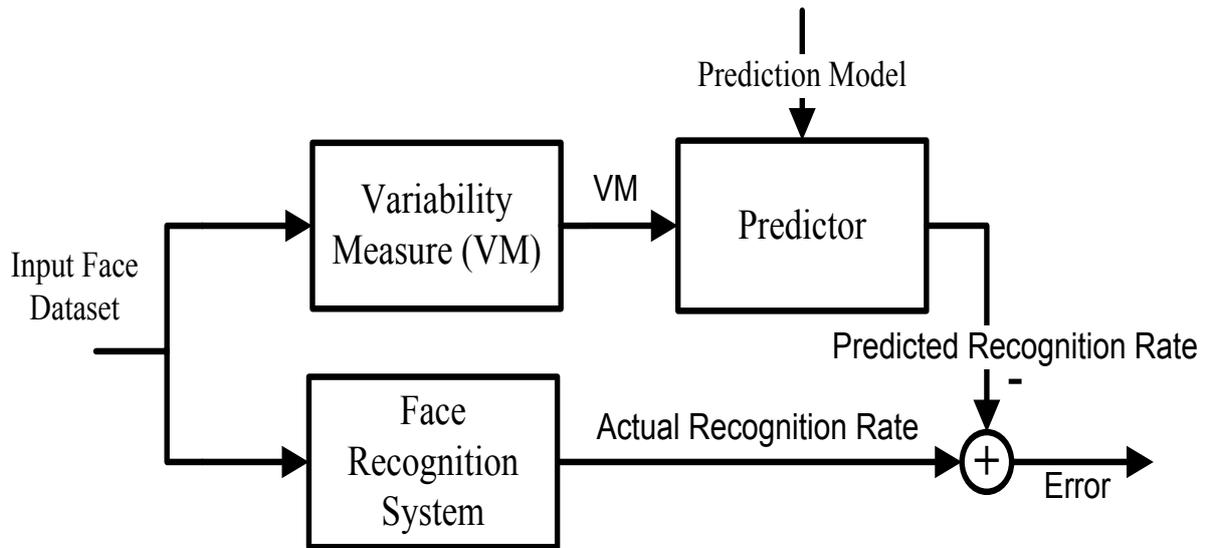


Figure 4.4: Experimental framework for evaluating the ability of VM to predict FR performance.

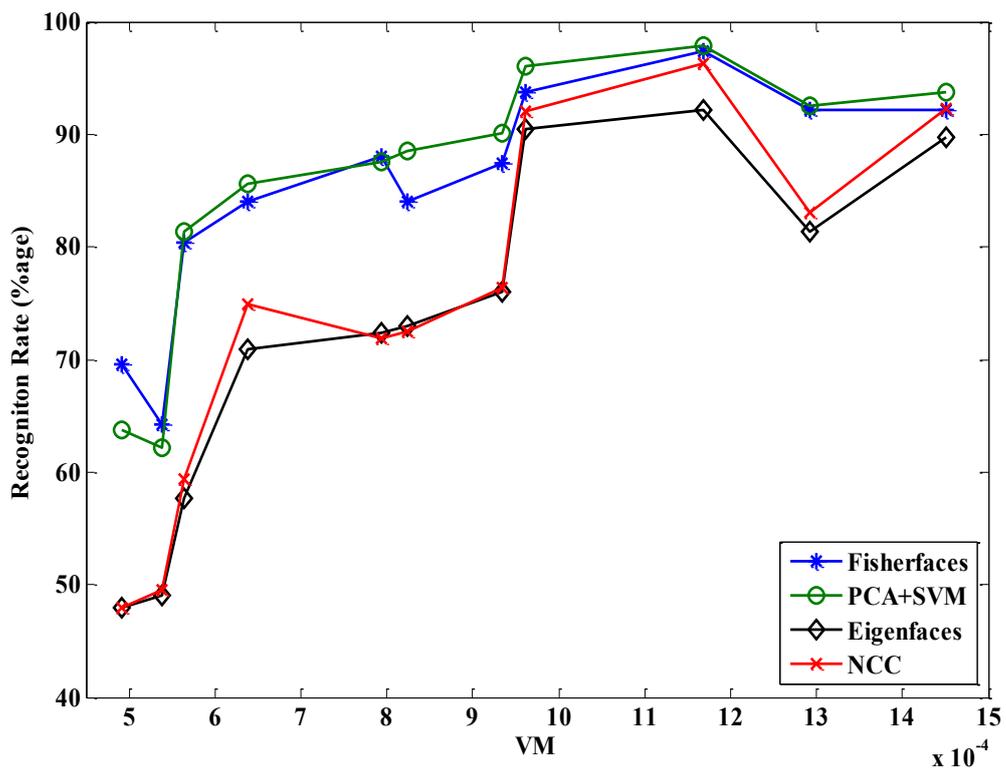


Figure 4.5: Recognition Rate Vs VM curves corresponding to different FR systems: A general increasing trend can be noticed.

At the end, an average recognition rate across all folds is calculated. Recognition performance versus VM curves for all FR systems are shown in figure 4.5. Curve points are obtained from different face image datasets. A general increasing trend in all curves shows that system classification performance improves with increasing variability measure (VM) values. Note however that this relationship is not monotonic.

The second phase of experimentation involved the polynomial modeling/approximation of these system performance-v-VM curves and was performed in two stages. In first stage, a polynomial model, that can best fit this relationship, is selected on basis of two ‘goodness of fit’ parameters: i) R-squared ( $R^2$ ) and ii) adjusted R-squared ( $\overline{R^2}$ ).  $R^2$ , generally known as coefficient of determination, is defined as the ratio of the sum of squares of the regression and the total sum of squares.  $\overline{R^2}$  is the modified version of  $R^2$  that has been adjusted for the number of terms (variables).

The value of  $R^2$  always increases with an increase in number of terms, even if new terms have no significance in improving the model. Note that  $\overline{R^2}$ , even being positively biased, is more consistent and only increases if the new term improves the model. Therefore, it is believed that a model would be preferred if and only if values for both  $R^2$  and  $\overline{R^2}$  are higher and the difference between the two is minimum [86]. Mathematically, both  $R^2$  and  $\overline{R^2}$  are, respectively, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

where

$$\bar{y} = \sum_{i=1}^n y_i, \quad (4.8)$$

and

$$\overline{R^2} = 1 - (1 - R^2) \frac{n-1}{n-p-1},$$

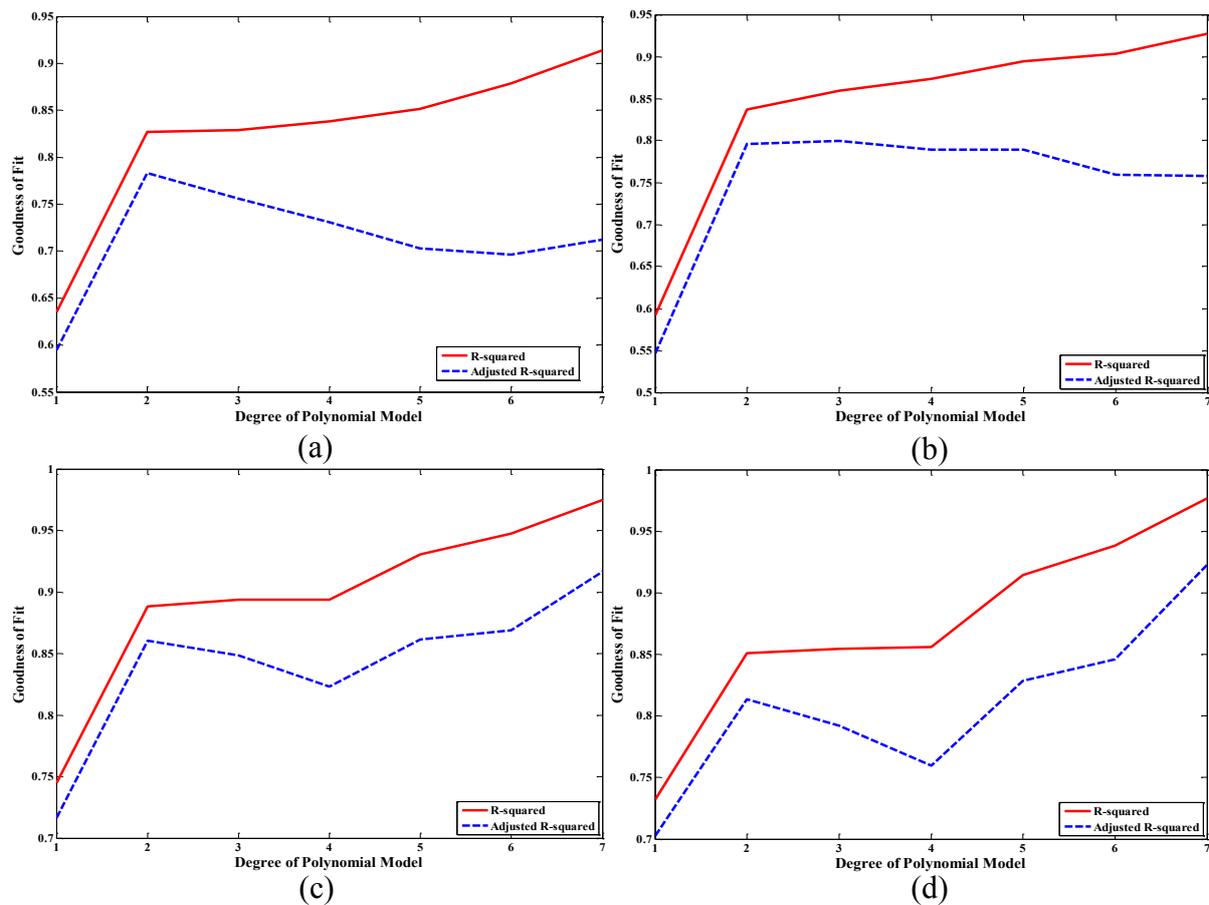


Figure 4.6:  $R^2$  and  $\overline{R^2}$  Vs degrees of Polynomial; a) Fisherfaces, b) PCA+SVM, c) Eigenfaces, and d) NCC. Degree  $d = 2$ , in all cases, is the maximum value for which the difference between  $R^2$  and  $\overline{R^2}$  is minimum.

where  $n$  and  $p$  are the number of sample points and number of variables (without constant term) and  $\hat{y}_i$  are the predicted by the model values.

The curves in figure 4.6 show  $R^2$  and  $\overline{R^2}$  values against different polynomial degrees for all recognition schemes. It can be noticed that as the polynomial degree ( $d$ ) is increased more than  $d = 2$ , the rate of increase of  $\overline{R^2}$  values gets smaller than that of  $R^2$ , and in some cases it is negative, which shows that for the available sample data polynomial degrees greater than 2 cause over-fitting. Therefore and in order to avoid over-fitting  $d = 2$  is chosen for all the recognition schemes. The resulting approximation models for all classifiers are shown in

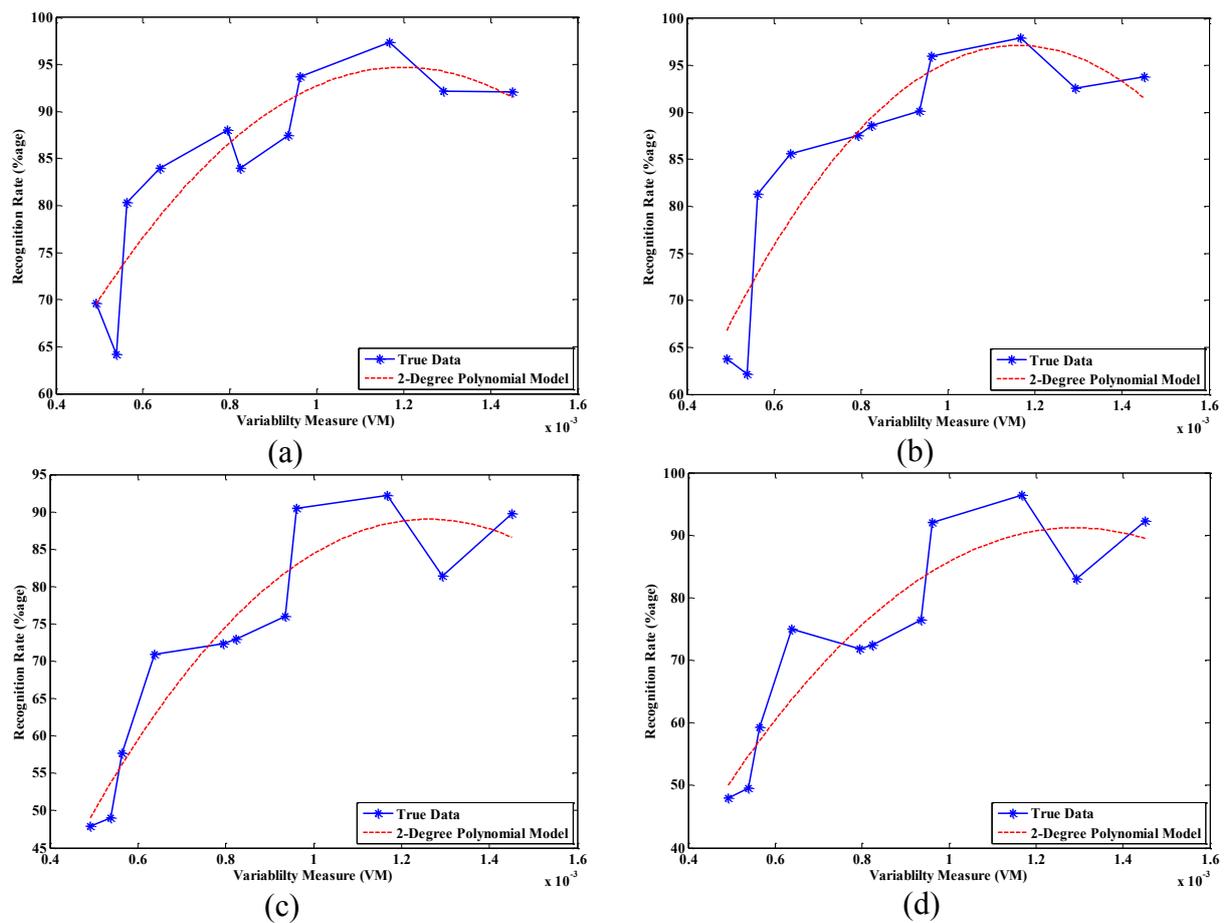


Figure 4.7: Recognition Rate Vs VM a) Fisherfaces, b) PCA+SVM, c) Eigenfaces, and d) NCC. 2nd degree polynomial shown as dotted line is following the trend of original data.

Table 4.4: Parameters for 2nd-degree Polynomial Model

FR Systems	$R^2$	$\overline{R^2}$	Avg. AE (%age)	Error Range (%age)
Fisherfaces	0.827	0.783	3.27	0.13-8.5
PCA+SVM	0.837	0.796	3.63	0.41-8.7
Eigenfaces	0.889	0.860	4.40	1.08-8.1
NCC	0.8510	0.814	5.47	2.08-11.1

figure 4.7 along with their corresponding original data curves, whereas their corresponding values of  $R^2$ ,  $\overline{R^2}$ , the average Absolute Error (Avg. AE), and Error Range are given in table 4.4. The Avg. AE between actual  $R_{Ac}$  and predicted recognition rate  $R_{Pr}$  is calculated as:

$$Avg. AE = \frac{1}{N} \sum_{i=1}^N |R_{Ac}^i - R_{Pr}^i|, \quad (4.9)$$

where N is the total number of face image datasets. The graphs in figure 4.7 and the data in table 4.4 are showing that the 2<sup>nd</sup>-degree polynomial provides a quite good fit to the available data and suggests a useful relationship between data variability and recognition performance.

The mathematical equations of the models for all four FR systems are given as:

Fisherfaces:-

$$y = p_1 x^2 + p_2 x + p_3$$

$x$  is normalized by mean 0.000878 and std 0.0003201.

Coefficients (with 95% confidence bounds).

(4.10)

$$p_1 = -5.188 (-9.211, -1.165)$$

$$p_2 = 10.38 (6.5, 14.25)$$

$$p_3 = 89.53 (84.6, 94.46)$$

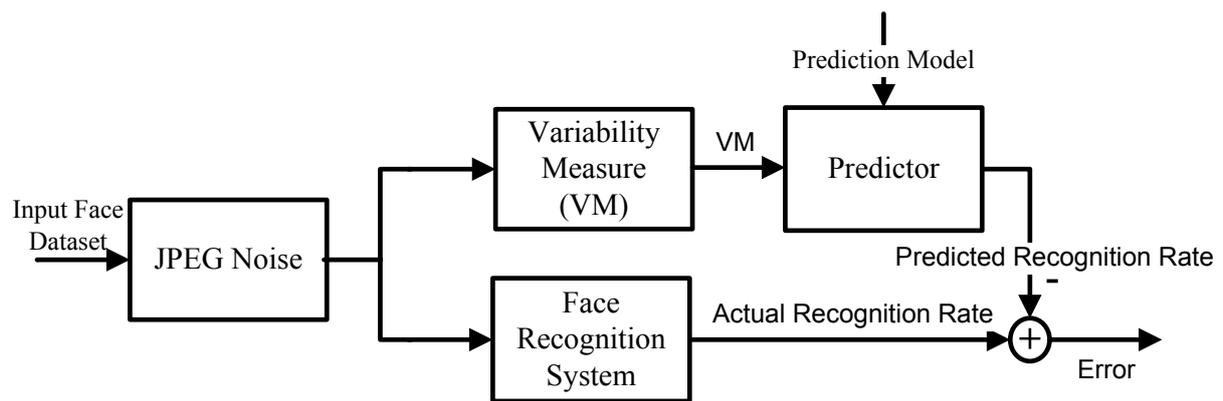


Figure 4.8: Experimental framework for evaluating proposed VM using noisy data.

PCA+SVM:-

$$y = p_1x^2 + p_2x + p_3$$

$x$  is normalized by mean 0.000878 and std 0.0003201.

Coefficients (with 95% confidence bounds).

(4.11)

$$p_1 = -6.926 (-11.54, -2.31)$$

$$p_2 = 12.28 (7.827, 16.72)$$

$$p_3 = 91.69 (86.04, 97.35)$$

Eigenfaces:-

$$y = p_1x^2 + p_2x + p_3$$

$x$  is normalized by mean 0.000878 and std 0.0003201.

(4.12)

Coefficients (with 95% confidence bounds).

$$p_1 = -6.919 (-11.9, -1.933)$$

$$p_2 = 16.58 (11.78, 21.39)$$

$$p_3 = 79.08 (72.97, 85.18)$$

NCC:-

$$y = p_1x^2 + p_2x + p_3$$

$x$  is normalized by mean 0.000878 and std 0.0003201.

Coefficients (with 95% confidence bounds).

(4.13)

$$p_1 = -6.632 (-12.69, -0.5738)$$

$$p_2 = 17.05 (11.21, 22.89)$$

$$p_3 = 80.23 (72.81, 87.65)$$

Note:- These models have been created using MATLAB R20011b simulations.

Next and given the VM value of an “unseen” image face dataset, the effectiveness of this FR system performance approximation approach is of course of interest. For the assessment of our proposed VM and the model, we have used same face datasets with noise of different PSNR levels. Furthermore and since in real life FR applications some type of compression coding is used prior to FR, which results in images of lower than the original image quality, experimentation was also performed using JPEG coded face image datasets. This introduces a block type of noise/distortion, and is reflected in the experimental set up of figure 4.8 by the inclusion of a “noise” box. This introduction of noise in input sample images can answer an important question i.e. How does the above system performance models which have been trained using noise-free image data, react to noisy input data and thus does the relationship

defined between VM and system recognition performance also holds for noisy images? To answer this question, further experimentation was performed involving one of the four recognition schemes (i.e. Fisherfaces) and JPEG coding noise at four the four average PSNR values of 55.57, 33.48, 26.86, 23.98.

In these experiments VMs and Recognition Rates were computed for all eleven face datasets corresponding to each average PSNR value as shown in figure 4.9. The increasing trend shown in case of noise-free datasets can also be noticed here even for the low average PSNR values. Moreover, a shift of curve to bottom-left corner with an increase in noise level is also an evidence for the robustness of our proposed VM/FR modeling. The downward shift indicates that coding noise has suppressed the facial variation across different subjects, which in turn caused a decrease in recognition performance for some facial datasets, and a left shift shows that a simultaneous decrease in VM has successfully kept the relationship intact. In addition, models derived from clean/un-coded data were employed to predict the recognition performance of systems operating on JPEG coded image data.

The histograms of absolute prediction error (%age) for all the datasets and Average Absolute Error (*Avg. AE*) values corresponding to each average PSNR level are shown in figure 4.10 and table 4.5, respectively. It is obvious from these values that in spite of introducing moderate image quality coding degradation in the input face images, model error ranges are approximately the same with those derived from noise-free data. This is indicative of the relative robustness of the proposed VM/ FR system performance relationship with respect to coding distortion.

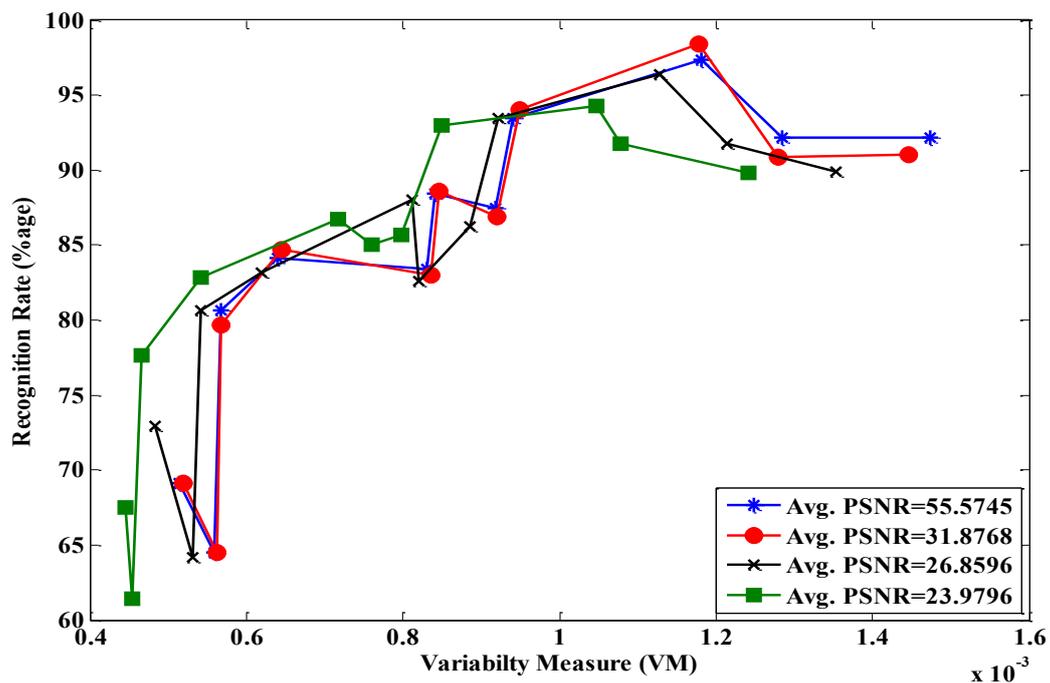


Figure 4.9: Recognition Rate Vs VM curves corresponding to Fisherfaces FR system: same overall increasing trend as seen in figure 4.5 can also be noticed in case of noise with all PSNR values.

**Table 4.5: Avg. Absolute Error (Avg. AE) at different Avg. PSNR values.**

Avg. PSNR	Avg. AE (%age)	Error Range (%age)
55.57	3.54	0.17-9.59
33.48	3.82	0.13-9.85
26.86	4.16	0.86-8.01
23.98	3.95	0.001-10.9

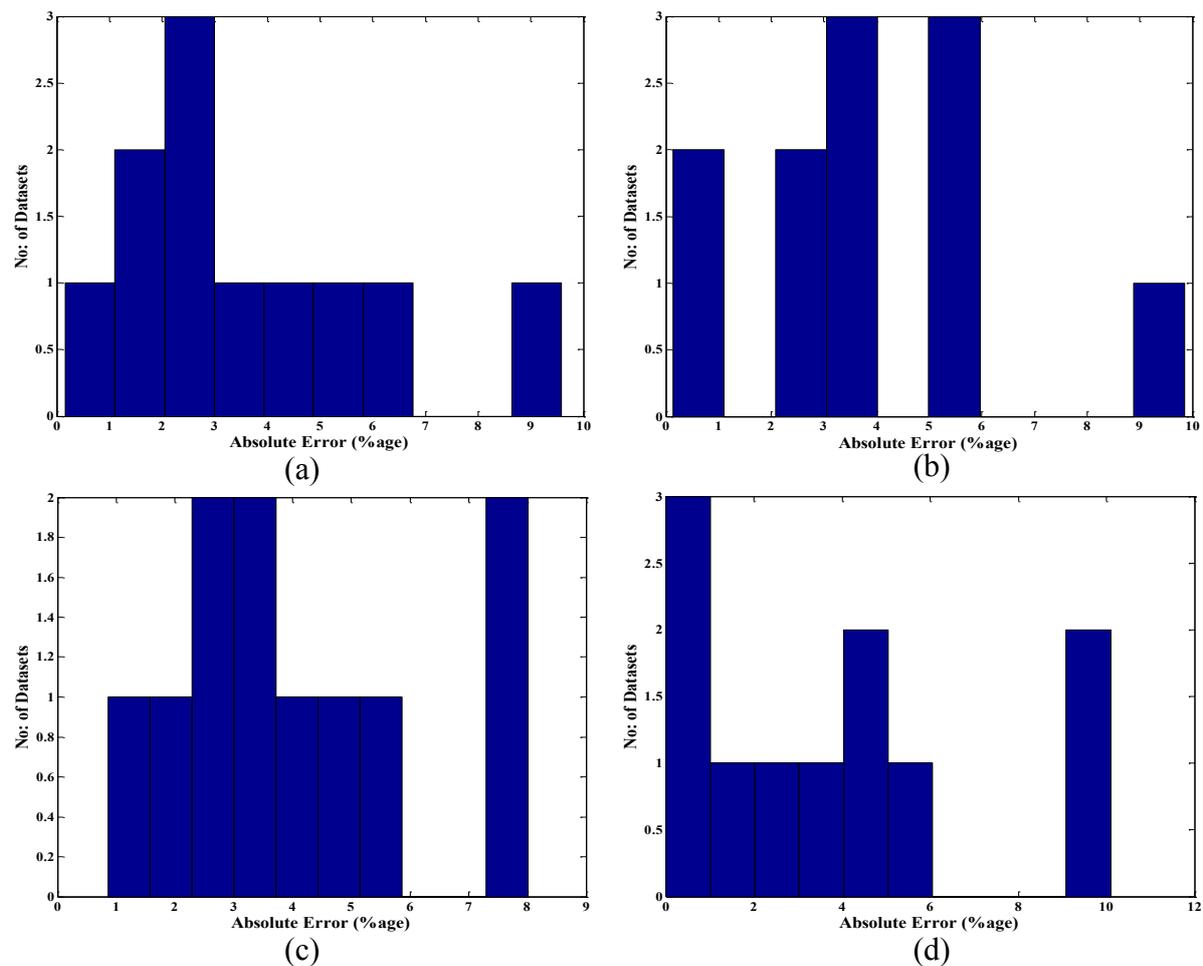


Figure 4.10: Histograms of Absolute Errors for the datasets with noise at four different PSNR values: a) 55.57, b) 33.48, c) 26.86, and d) 23.98.

### 4.3 Conclusion

The classification performance of face image based automatic age estimation (FI-AAE) systems is heavily dependent on the variation among the images of input dataset. Now, if somehow this dependence/relationship could be modeled, classification system performance could be predicted. With this as motivation work reported in this chapter investigated the issue of face image data variability and came up with a measure that can be used to represent/rank a dataset according to the amount of variation it contains. However, due to

lack of availability of many age-based face image datasets, the work presented in this chapter is applied on face recognition (FR) systems. Thus a new variability measure (VM) that characterizes overall image face data variability has been proposed and applied in FR systems. In addition, relationships between such VM values and the performance of four FR systems have been determined experimentally and also modeled using second order polynomials. Note that the proposed VM takes into account both the inter- and intra-class correlation characteristics of a given image dataset, i.e. VM-interSC and VM-intraSC.

Thus computer simulation results involving 11 publically available face image datasets show VM/FR performance prediction errors of less than 10%, for all four FR systems, with minimum Avg. AE of 3.27% for Fisherfaces and maximum Avg. AE 5.47% in case of NCC. Note that currently models are trained using eleven face image datasets. An increase in the number of training datasets should yield further improvement in modeling accuracy.

Furthermore, the prediction accuracy of the above VM/FR performance models is also assessed using noisy i.e. JPEG coded, image data at different PSNR values. Prediction errors (i.e. Avg. AE) corresponding to face image data for each PSNR value were obtained and show that models kept their error range approximately the same with those produced by of noise-free input data. Moreover and in case of noisy data, the VM/FR performance curves show the same increasing trend as it was in case of noise-free data. The robustness of our proposed VM/FR performance models is also evident from the left-downward shift of VM/FR performance curves of noisy data, which indicates that if recognition performance decreases due to noise the proposed VM also decreases and keeps the general relationship intact.

At the end, as the proposed formulation of VM is originally based on inter-class and intra-class similarities/correlation characteristics of a face dataset, we expect that in the future and

depending on the availability of large face image datasets, the above proposed variability assessment framework can be easily adapted to FI-AAE applications.

## **Chapter 5**

### **Face Image based Automatic Age Estimation**

Automatic age estimation (AAE), or more precisely, face image based automatic age estimation (FI-AAE) that aims to determine the specific age year or age range of a person based on a given facial image, is a challenging yet attractive topic due to its roots in numerous real-life applications. These application areas include surveillance and security control (for example, an accurate age estimation system can, prevent minors from entering bars or wine shops, stop underage smokers from purchasing cigarettes from vending machines, refuse elderly people to try a roller coaster in an amusement park, and deny children access to adult websites or restricted movies, etc.), law enforcement (for example, a good automatic age estimation module can be helpful in filtering out the potential suspects more efficiently and accurately from a database), health care (for example, a personalized Avatar can be selected automatically from the custom-built Avatar database to interact with patients from different age groups with particular preferences). Moreover AAE can be used to develop age-specific human-computer interaction (HCI) systems, etc. [87]

An FI-AAE system based on facial images is generally composed of two major components (see figure 1.1); i) feature extraction and ii) age estimation. Feature extraction is a process of extracting facial features that are affected by aging process. In general, these features are categorised as shape-based and texture-based features. Shape-based features represent the

information related to facial geometry, whereas texture-based features keep track of skin aging that includes skin and facial hair colour, wrinkles, age spots, etc. As discussed earlier in Chapter-1, the human aging process affects face geometry (shape) and facial texture differently during different periods of life (i.e. *birth-to-adulthood* and *adulthood-to-old age*) and hence an FI-AAE system cannot depend upon only on one of these features for accurate age estimation. This makes the feature extraction process very important, as the extracted features can significantly affect the age classification. Due to this very reason, a great amount of effort has been directed towards the extraction of such discriminating aging features that represent both shape and texture information, simultaneously. Examples of such research work include Gabor wavelet transform (GWT) [5, 6], subspace features using image intensity [7, 8], Active Appearance Models (AAM) and Image Frequency [9]. Among these, the Active Appearance Model (AAM) [15] is considered as the most popular among researchers [10-14] due to its ability to represent both facial shape and texture information with a small set of parameters that can be subsequently used to recreate the face image.

AAM involves a two-stage Principal Component Analysis (PCA) process to generate a low dimensional hybrid set of parameters that represents both face shape and texture. However as suggested in [2], this dimensionality reduction process, while retaining the individual identity characteristics, becomes a major cause of losing some important aging information such as wrinkles and skin texture (see figure 1.5) that affects the overall age estimation process. Even though some techniques [2, 16, 17] have been proposed to overcome this problem by introducing separate models for missing features, their performance gains are negligible in comparison to the overhead complexity introduced to the overall system.

In order to overcome the above mentioned problem, the research work presented in this chapter adopts a simpler strategy. Shape and texture information are used separately and the

outcomes from both processing streams are later fused to reach a final decision. Note here that shape is represented by a set of landmark coordinate points of different facial components boundaries, whereas the texture corresponds to the set of intensity values enclosed by these landmark points. Furthermore and since we are not performing fusion at feature level, the possibility of losing discriminative information related to each feature, is reduced.

Given a feature, the next step is age estimation. As discussed earlier in Chapter-1, the existing age estimation techniques, generally, operate in one of the two types of system architectures; i) Single-level age estimation [7, 11, 13, 14, 16] and ii) Hierarchical age estimation [2, 11, 20-23]. Among these, the hierarchical age estimation methods, despite being prone to error propagation, perform better as they have ability to effectively exploit group-specific characteristics in order to reach the final decision. Moreover training of classifiers, using relatively smaller groups, helps in reducing overall computational complexity.

In extant hierarchical age estimation techniques [11, 21, 22], a major cause of classification error is the hard boundaries between the images of two adjacent age groups. Since aging is a slow process, some images at the boundaries of two adjacent groups are often quite similar which makes it very hard for a classifier to correctly classify them. This classification error propagates within the hierarchy and adversely affects overall age estimation performance. To overcome this problem, authors in [2, 20] have proposed solutions using age groups with soft overlapping boundaries. These solutions reduce errors at the coarse levels of hierarchy; however, errors at finer levels are still quite large, which makes the resulting overall improvement in age estimation rather insignificant. There is, therefore need for a hierarchical age estimation framework that mitigates classification errors while operating on age partitions with larger inter-group variation.

This chapter presents such a novel Multi-Level Age Estimation (ML-AE) framework that minimizes age classification errors significantly by overcoming performance limitations introduced by the smooth variation of images belonging to adjacent age groups. This is achieved by a novel method used in the selection of training data so that large inter-class variations are generated and used during training operations.

The proposed ML-AE approach operates in an unbalanced ternary tree structure that produces three children nodes for each parent at a specific tree level. Like any other classification system, the ML-AE is also composed of two phases; i) System Training and ii) System Testing.

During the training process, a given dataset that contains facial features (shape/texture), is first partitioned into four age groups at a particular level of hierarchy, then sets of facial features (shape/texture) corresponding to alternative age groups are used to train four binary classifiers, two per feature type i.e. shape or texture. Afterwards, these four age groups are used to form three parent datasets for the next level by merging first age group with second, second with third and third with fourth age group. Note: the use of alternative age groups which are employed in binary classification, instead of adjacent groups, provides two major advantages; i) the adjacent groups suffer with low inter-group variation, whereas alternative groups contain large inter-group variation that can facilitate the classification process, ii) a multi-class classification problem is broken into two simpler binary classification problems.

During testing phase, facial features of a given input face image are classified using binary classifiers that were trained as discussed earlier. The four decision outcomes of the classifiers are fused in a way (which is explained in detail in a subsequent section) to decide which path to traverse next in the decision tree.

Computer simulation based experiments cover different aspects of proposed ML-AE performance. An initial experiment studied the impact that the fusion of shape and texture decisions has on overall system performance. Results showed that shape or texture alone, as a facial feature, is not sufficient for the representation of facial aging. A second investigation compared ML-AE system performance using two different classifiers i.e. Support Vector Machines (SVMs) with Radial Basis Function (RBF) and Linear Discriminant Analysis (LDA). Although the overall performance of SVM based ML-AE-SVM is better than that of LDA based systems, ML-AE-LDA produced smaller mean absolute errors (MAE) over some age ranges. These results are indicative of further research work to be directed towards systems with multiple classifiers and the use of an overall fusion of decisions at the very end. Finally, a third study was performed which compared ML-AE system performance with that obtained from some existing age estimation systems. Experimental results showed the effectiveness and potential of the proposed Age Estimation system.

This chapter is organized as follows: Section 5.1 presents, in detail, the design/structure of the proposed Multi-level Age Estimation (ML-AE) framework, it also covers the algorithmic implementation of the proposed framework. Section 5.2 explains the experimental setup used to produce computer simulation results and provides an in depth discussion on the experimental results. Finally, concluding remarks are given in Section 5.3.

## **5.1 Multi-Level Age Estimation (ML-AE) Framework**

The proposed Multi-Level Age Estimation (ML-AE) framework is comprised of two major modules; i) System Training and ii) System Testing, as depicted in the high level system diagram shown in figure 5.1. The system training module is based on an offline process and runs in an unbalanced ternary tree structure in which each parent node gives birth to three

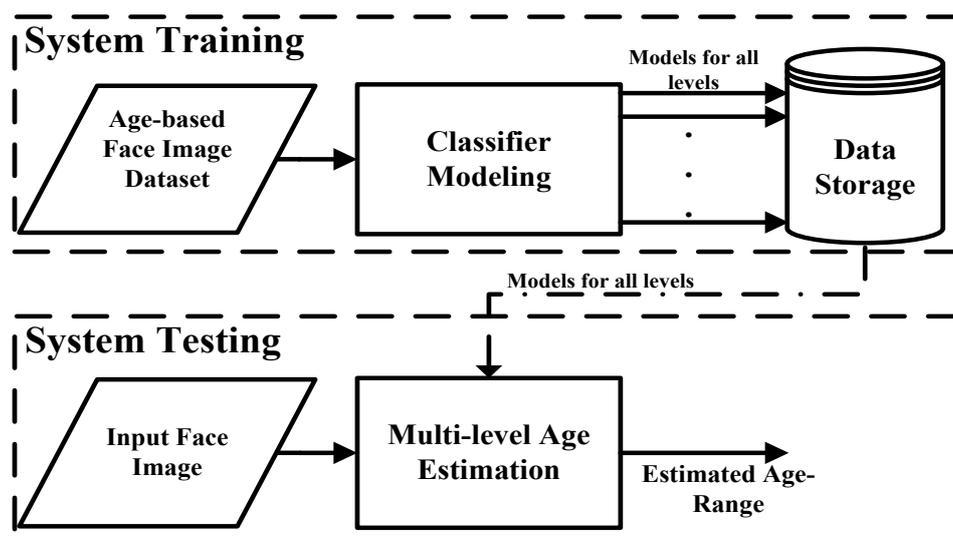


Figure 5.1: High-level system block diagram of our proposed ML-AE framework.

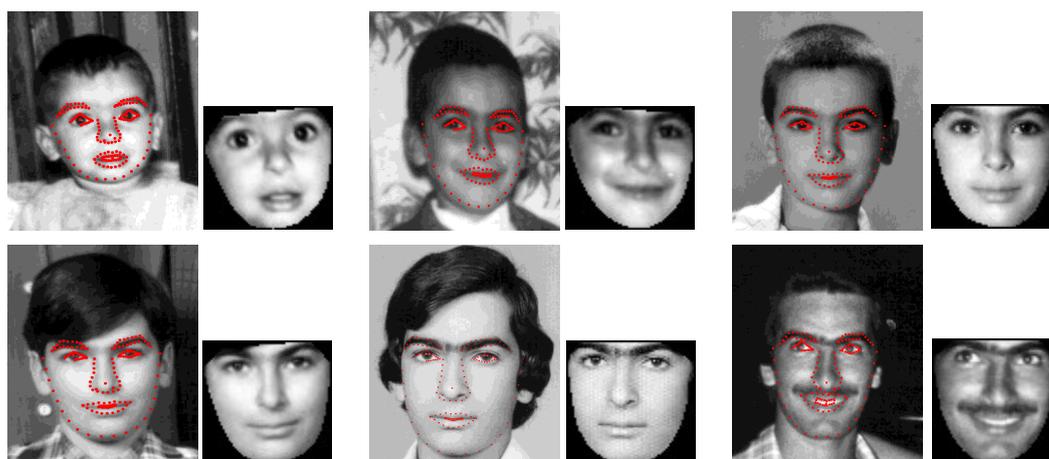


Figure 5.2: Sample face images with their superimposed shape coordinates (red dots) and extracted facial texture.

children. On each tree level, four binary classifiers are trained; two for shape and two for texture image data.

System testing is performed in an online setting in which an input face image (both shape and texture) is traversed through the ternary tree structure to obtain a final estimate of the age

year. At each tree level during system testing, decisions from all four classifiers are fused to decide for the next path to traverse. A detailed description of the two modules is provided in subsequent sub-sections.

### 5.1.1 System Training

The overall ML-AE system training is based on the following two major components (see figure 5.1).

#### 1) Age-based Face Image Dataset

The training module of ML-AE contains an original age-based face image dataset  $\mathbf{F} = \{\mathbf{S}|\mathbf{G}\}$  whose data is presented to the Classifier modeling module; here  $\mathbf{S}$  is a set that contains shape information of all the training images in the form of the coordinates of landmarks points, whereas set  $\mathbf{G}$  contains face texture, i.e. all the pixels (intensity values) enclosed by shape coordinates, (see Chapter 3, Eqs. 3.1 and 3.2). Some example face images with their superimposed shape coordinates and extracted texture profiles are shown in the figure 5.2. Note: as the extracted face textures of different images vary in sizes, therefore they are, first, made equal in size by using bicubic interpolation before doing further processing.

#### 2) Training Operations

The multi-level training of ML-AE operates in an off-line mode. On each level, it takes an age-based face image dataset as input in the form described in previous section. This dataset, which satisfies a pre-defined criteria (see below for details), is passed through the, *Classifier Modeling* process, which in turn returns three subsets of the input dataset. These serve as parent sets for next level, see figure 5.3. *Classifier Modeling* is comprised of two sub-processes, i) Split and ii) Merge, as shown in figure 5.4. The ‘Split’ process partitions a parent dataset into four subsets on basis of four equal partitions of the age-range associated

with the parent dataset. Note here that age-range means all the age years from minimum to maximum for which face images are available in the dataset. These four subsets are then used to train four binary classifiers at that level to produce two models one for shape and another for texture data. On the other hand, the ‘Merge’ process combines these four output datasets produced from the ‘Split’ process, in a particular way, and thus forms three new datasets that are the outputs of the *Classifier Modeling* and serve as the parent sets for next level of a tree type hierarchical architecture. The whole process for one level can be elaborated using the following example.

Consider that at some particular tree level, face images in a training dataset cover an age-range from year  $A$  to year  $B$ . This dataset goes through following steps:

- In the first step, a parent dataset is checked against two criteria, i) number of images  $L$  in the dataset must be greater than a pre-specified minimum number of images  $LTh$ , so that for subsequent classifier training, each class contains enough number of image samples, ii) the age-range corresponding to parent dataset must contain a number of age years  $R$  which are greater than  $RTh$  i.e. a lower bound for this age-range. If a parent dataset satisfies both of these criteria then we proceed to the next step otherwise this tree branch is terminated at this point.
- In the second step, the ‘Split’ process makes four equal partitions of the age-range as  $p_1 = A \rightarrow \frac{B}{4}$ ,  $p_2 = \frac{B}{4} + 1 \rightarrow \frac{B}{2}$ ,  $p_3 = \frac{B}{2} + 1 \rightarrow \frac{3 \times B}{4}$ , and  $p_4 = \frac{3 \times B}{4} + 1 \rightarrow B$  and thus four dataset  $D_1, D_2, D_3$ , and  $D_4$  out of initial input face dataset are formed that contain face images (both shape and texture) for their corresponding age-ranges  $p_1, p_2, p_3$ , and  $p_4$ , respectively.

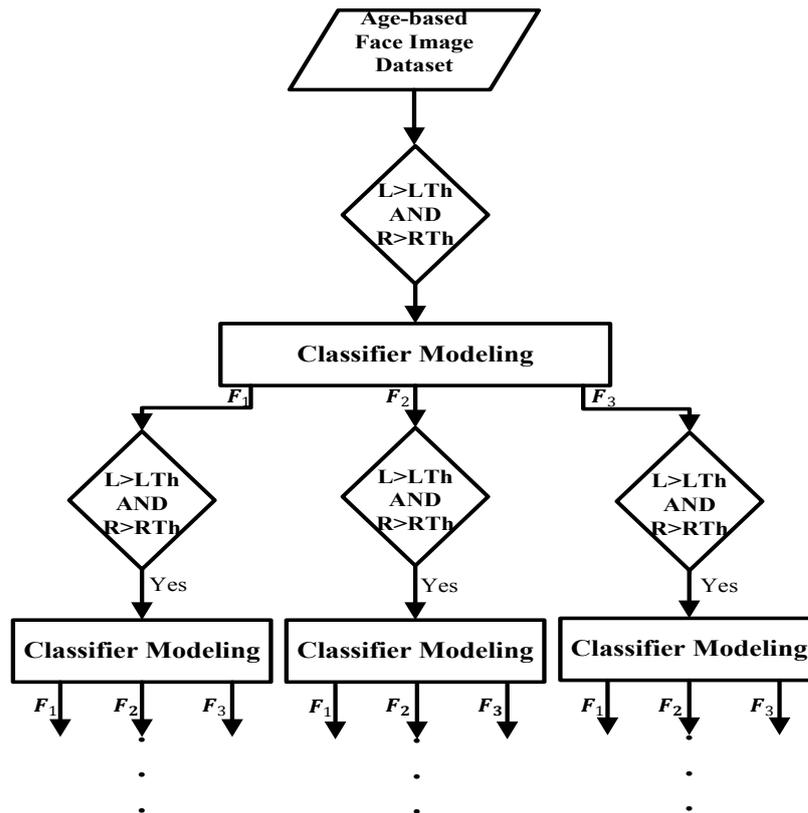


Figure 5.3: Flow diagram of ML-AE system training. Here  $L$  and  $R$  are the lower bounds for number of images and number of age years covered by the age-range of corresponding parent dataset.  $LTh$  and  $RTh$  are two pre-specified thresholds. Internal structure of Classifier Modelling block is explained in figure 5.4.

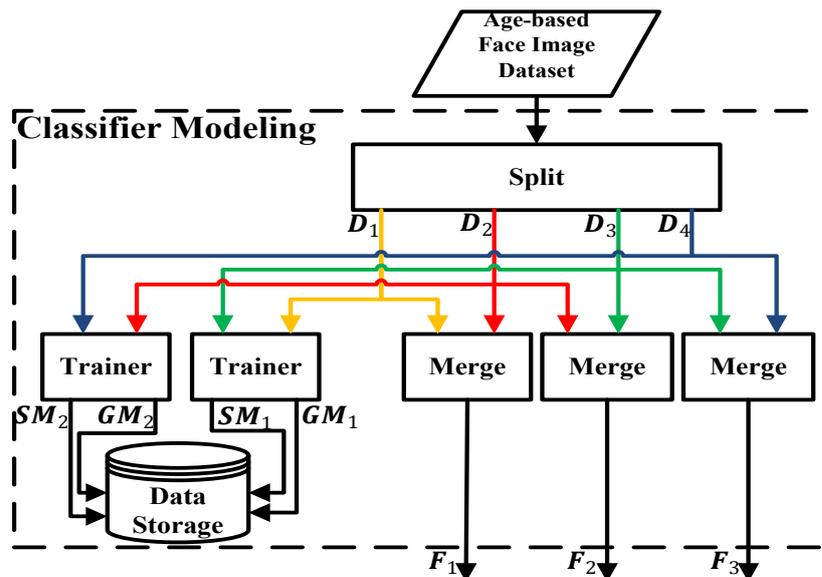


Figure 5.4: Internal structure of Classifier Modelling block shown in figure 5.3.

- Thirdly, shapes and textures of the images, from two datasets  $D_1$  and  $D_3$ , are passed separately to a ‘Trainer’. In this ‘Trainer’ block, Principal Component Analysis (PCA) [54] is separately applied to shape and texture of all the training images available to produce corresponding reduced dimensional feature sets. These feature sets are, separately, used to train two binary classifiers (e.g. in our case SVM [56] or LDA) that produce classification models  $SM_1$  and  $GM_1$  for shape and texture, respectively. Similarly, we get two models  $SM_2$  and  $GM_2$  for datasets  $D_2$  and  $D_4$ . As discussed earlier, the purpose of using alternative data partitions ( $D_1$  with  $D_3$  and  $D_2$  with  $D_4$ ), instead of adjacent partitions, to train binary classifiers is the relatively large inter-class variation which facilitates the classification training process to correctly distinguish between images belonging to two age clusters.
- In the last step, the ‘Merge’ is used to form three subsets  $F_1$ ,  $F_2$  and  $F_3$  of parent face dataset by combining  $D_1$  with  $D_2$ ,  $D_2$  with  $D_3$  and  $D_3$  with  $D_4$ . Each of these three subsets becomes the parent node for the next level.

The purpose of creating these subsets with overlapping boundaries is to further reduce the possibility of miss-classification that is caused due to hard boundaries.

The last component of the training system, shown in figure 5.1, is a storage device where classification models obtained at each training level, are stored to be used later during System Testing.

**Algorithm 5.1: Classifier Modeling**

```

recursive function: ClfModeling(Data, RTh, LTh)
Input: Data           A data structure with three elements
       Data.F        Face images dataset containing both shape and texture
       Data.L        Number of images in the dataset
       Data.R        Age-range i.e. minimum year and maximum year
       LTh           Lower bound on number of images in the dataset
       RTh           Lower bound on number of years in the age-range

Output: Models      a data structure containing all the classification models

If
  Data.L < LTh OR Data.R < RTh
  retrun Null
Else
  [D1, D2, D3, D4] = Split(Data)
  //Call Split function, see Algorithm-2
  [Models.SM1, Models.GM1] = Trainer(merge(D1, D3))
  [Models.SM2, Models.GM2] = Trainer(merge(D2, D4))
  //Call merge and Trainer function, see Algorithm-3 for merge
  Models.Path1 = ClfModelling(merge(D1, D2), RTh, LTh)
  Models.Path2 = ClfModelling(merge(D2, D3), RTh, LTh)
  Models.Path3 = ClfModelling(merge(D3, D4), RTh, LTh)
  retrun Models
End

```

A more detailed description of the algorithmic Classifier Modeling implementation is given below.

**Classifier Modeling Algorithmic Implementation**

The multi-level system training is implemented using a recursive function ‘ClfModeling’, see Algorithm 5.1. The function, on each call, takes three parameters i.e. *Data*, *RTh* and *LTh*, as input. Here, the parameter *Data* represents a data-structure with three elements *Data.F* i.e. face image dataset containing both shape and texture information for all the images, *Data.L* i.e. number of images in the dataset and *Data.R* i.e. age-range in years covered by all the images. The second input parameter *RTh* is the minimum number of years an age-range can have. Third and final input parameter *LTh* is the lower bound for the number of images in a dataset. The later two input parameters serve as the stopping criteria for the recursion.

**Algorithm 5.2: Splitting Dataset into Subsets**

<b>function:</b> Split( <i>Data</i> )	
<i>Input: Data</i>	A data structure with three elements
<i>Data.F</i>	Face images dataset containing both shape and texture
<i>Data.L</i>	Number of images in the dataset
<i>Data.R</i>	Age-range i.e. minimum year and maximum year
<i>Output:</i> $D_1, D_2, D_3, D_4$ Four subsets of input dataset	
$D_1.R \leftarrow \min(Data.R)$	To $(\max(Data.R)/4)$
$D_1.F \leftarrow$	All the image data corresponding to age-range $D_1.R$
$D_1.L \leftarrow$	number of images in $D_1.F$
$D_2.R \leftarrow (\max(Data.R)/4)+1$	To $(\max(Data.R)/2)$
$D_2.F \leftarrow$	All the image data corresponding to age-range $D_2.R$
$D_2.L \leftarrow$	number of images in $D_2.F$
$D_3.R \leftarrow (\max(Data.R)/2)+1$	To $(3 \times \max(Data.R)/4)$
$D_3.F \leftarrow$	All the image data corresponding to age-range $D_3.R$
$D_3.L \leftarrow$	number of images in $D_3.F$
$D_4.R \leftarrow (3 \times \max(Data.R)/4)+1$	To $\max(Data.R)$
$D_4.F \leftarrow$	All the image data corresponding to age-range 4. $R$
$D_4.L \leftarrow$	number of images in $D_4.F$
<b>return</b> $D_1, D_2, D_3, D_4$	

The recursive function ‘ClfModeling’ returns a data structure *Models* that contains, for all levels of tree structure, classification models for both shape and texture.

During the process, Algorithm 5.1 calls three more functions ‘Split’, ‘Merge’ and ‘Trainer’. ‘Split’ function, see Algorithm 5.2, partitions the input data into four subsets as described before. Here each subset is a data structure that contains face images (both shape and Texture), corresponding age-range and number of images in appropriate data structure fields.

In Algorithm 5.2, the functions min() and max() are used to find minimum and maximum value of the age-range in a dataset, respectively. The ‘Merge’ function, see Algorithm 5.3, combines two subsets to form one dataset. The ‘Trainer’ function represents training phase of a binary classifier (e.g. SVM or LDA) that takes datasets with labels to generate training classification models. The block diagram, shown in figure 5.3, explains the recursive calls to function ‘ClfModeling’, whereas block diagram shown in figure 5.4 illustrates the internal working of ‘ClfModeling’ and calls to other functions.

**Algorithm 5.3: Merging of Two Datasets**

<b>function:</b> Merge( $D_1, D_2$ ) <i>Input:</i> $D_1, D_2$ Two data structures with three elements  <i>Output:</i> $D$ Combined dataset $D.R \leftarrow \min(D_1.R)$ To $\max(D_2.R)$ $D.F \leftarrow$ All the image data corresponding to age-range $D.R$ from both sets $D.L \leftarrow$ number of images in $D.F$  <b>Return</b> $D$
--

**Table 5.1: Look-up Table used for Decision Fusion**

Decision Fusion Strategy	Next branch to Traverse (Decision)
If at least 3 out of 4 classifiers return label '0'	$F_1$
If $TClf_1$ and $SClf_1$ return label '1' and $TClf_2$ and $SClf_2$ return label '0'	$F_2$
If at least 3 out of 4 classifiers return label '1'	$F_3$
Otherwise	Fail! Return age-range of parent dataset

**5.1.2 System Testing**

Once all the classifier models are obtained using offline training process, they are used in an online evaluation system testing phase. In fact, ML-AE system testing is a tree-traversal process for a given face image  $f^t = \{s^t | g^t\}$  using the best possible path. This is determined by fusion of the four classification decisions at each tree-level.

System testing involves one major component, *Age Estimation*, as shown in figure 5.1. The working of the *Age Estimation* component is described in the following steps:

**Step-1:** Read input face image  $f^t = \{s^t | g^t\}$ .

**Step-2:** Project the shape information  $s^t$  on Eigen space, generated by PCA using the shape information of corresponding training data, to get appropriate reduced dimensionality shape feature set. This feature set is passed through two classifiers  $SClf_1$  and  $SClf_2$  that were trained earlier using training models  $SM_1$  and  $SM_2$ , respectively.  $SClf_1$  results in one class

label  $RS_1 = 0$  Or  $1$  that represent classes  $D_1$  and  $D_3$ , respectively.  $SClf_2$  results in class label  $RS_2 = 0$  Or  $1$  that corresponds to classes  $D_2$  and  $D_4$ , respectively.

**Step-3:** Similarly, project the texture information  $g^t$  on its corresponding Eigen space to produce a new texture feature set, which is then passed through two classifiers  $TClf_1$  and  $TClf_2$ , that were trained earlier using training models  $TM_1$  and  $TM_2$ , respectively.  $TClf_1$  results in one class label  $RT_1 = 0$  Or  $1$  that represents two classes  $D_1$  and  $D_3$ , respectively.  $TClf_2$  results in class label  $RT_2 = 0$  Or  $1$  that corresponds to classes  $D_2$  and  $D_4$ , respectively.

**Step-4:** Fuse four classifiers decisions, obtained in **Step-3** and **Step-4**, using the look-up of table 5.1.

**Step-5:** Proceed to the next level with  $F_1$  or  $F_2$  or  $F_3$  as new parent datasets and repeat **Step-1** to **Step-5** till the last level, or stop here and return the mean age year of the parent dataset as final estimated age.

The decision fusion strategy, given in table 5.1, is mainly based on the majority rule; therefore an output label '0' from at least three out of the four classifiers means that the given input face image should be a possible member of first two adjacent data partitions. Similarly a label '1' declares the input face image as the member of last two adjacent partitions. However, the exception is when both classifiers of first pair return label '1' and both classifiers of second pair return '0' as the output class label. In this case, the input face image is declared as the possible member of second and third partition.

## 5.2 Experiments

### 5.2.1 Input Dataset

Experimentation was performed using computer simulation of the ML-AE framework and involved the FG-NET dataset [24]. As explained earlier, FG-NET dataset contains 1002 colour or grey-scale face images of 82 persons (12 images per person on average) with an age range from infant to 69 years. The data distribution of FG-NET dataset according to age is shown in figure 5.5. Furthermore each image comes with 68 hand labeled shape points. Some FG-Net sample images are shown in figure 5.6.

Thus the shape information of training face images consists of 68 landmark points see figure 5.2. For the test data, these 68 landmark points, which are available with the FG-NET dataset, can be used as they are or they can be automatically obtained by an MM-AAM fitting process [28]. In order to concentrate on age estimation performance without having to account for the effects of an MM-AAM fitting process errors, all the landmark points used in the experimental part of this work, are those made available by FG-NET.

### 5.2.2 Experimental Setup

Experimentation was performed on the basis of the Leave-One-Person-Out (LOPO) strategy. LOPO is used mainly to prevent the same person from being included in both the training and the test dataset [2, 7, 12, 88]. Thus in each fold, the images of one person were used as the test set and those of the others were used as the training set. After 82 folds, each subject had been used as test set once, and the final average performance results were calculated from all estimates.

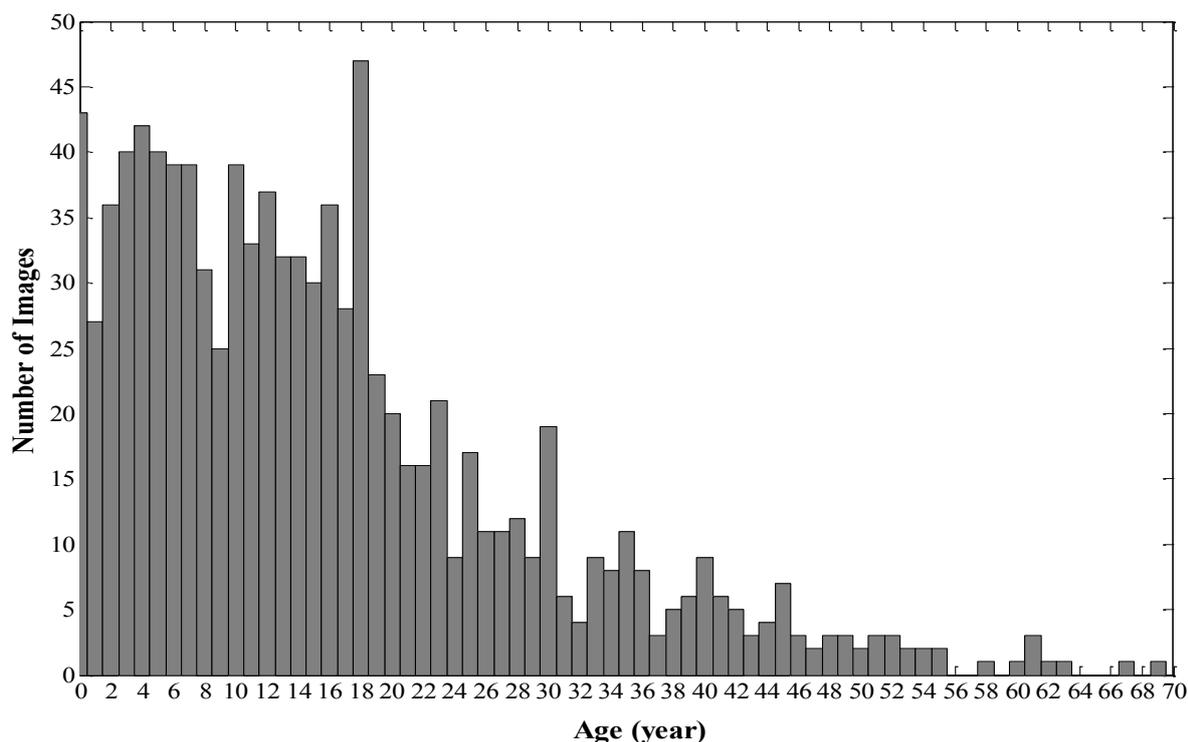


Figure 5.5: Image distribution of FG-Net face dataset according to age.

As most of the images in FG-NET represent younger ages and the number of images corresponding to older age years is quite small (see figure 5.5), training of ML-AE started with four non-uniform data partitions based on age years with an aim to have enough number of images per partition. The four non-uniform age-based partitions are  $p_1 = 0 \rightarrow 10$ ,  $p_2 = 10 \rightarrow 20$ ,  $p_3 = 20 \rightarrow 35$  and  $p_4 = 35 \rightarrow 69$ . This non-uniform partitioning is just to start the training process of ML-AE and actual age is estimated, during the testing process, after a complete tree traversal. After the first tree level, data is uniformly partitioned in to four classes.  $RTh$  and  $LTh$ , the two threshold values to terminate the training process, are set as 2 years and 5 images, respectively. During the training of classifiers at each level of ML-AE, reduced dimensional feature sets for both shape and texture are obtained using only those Eigen vectors of PCA that correspond to the largest Eigen values while 95% of cumulative signal energy is captured.



Figure 5.6: Sample images of a person taken from the FG-Net aging database.

Furthermore, ML-AE performance is evaluated by mean absolute errors (MAE) and cumulative scores (CS) [2, 7, 12, 88]. MAE is defined as the mean of the absolute difference between the estimated age and the ground-truth age i.e.

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{A}_i - A_i|, \quad (5.1)$$

where  $N$  is the total number of test images,  $\hat{A}_i$  is the estimated age of  $i$ th test image and  $A_i$  is the ground-truth age of the test image  $i$ . The cumulative score (CS) of an age difference  $t$  is defined as the percentage of estimations that have an estimation error of less than or equal to  $t$  years, mathematically it can be given as

$$CS(t) = \frac{n_{|\hat{A}-A|\leq t}}{N} \times 100, \quad (5.2)$$

where  $n_{|\hat{A}-A|\leq t}$  is the number of images whose estimation error ( $|\hat{A} - A|$ ) is less than or equal to  $t$  years and  $N$  is the total number of test images.

### 5.2.3 Experimental Results

This subsection provides ML-AE system performance evaluation results in terms of the MAE and CS using the FG-Net aging database.

**Table 5.2: MAE and standard deviation of error comparison of ML-AE using fusion of shape and texture classifications with ML-AE schemes using shape or texture classifications.**

Method	Mean Absolute Error (years)	Standard Deviation of Error (years)
ML-AE (Shape)	5.1437	5.7794
ML-AE (Texture)	4.2825	5.5737
ML-AE (Shape + Texture)	4.0406	4.9559

### 1) Performance Evaluation of Shape- and Texture-based Decision Fusion

The aim of this experiment was to investigate the impact of decision fusion of two features i.e. Shape and Texture. Therefore, here the performance of ML-AE system which employ feature (i.e. Shape and Texture) decision fusion was compared with that of ML-AE systems without decision fusion. Note: in these experiments binary SVMs with RBF kernel are used for classification purpose. The overall MAE for ML-AE with fusion of shape and texture decisions is 4.0406. This represents an improvement of 21.45% and 5.65% over ML-AE systems using only shape or texture, respectively (see table 5.2).

Figure 5.7 shows Mean Absolute Errors (MAEs) at different age years for the above three ML-AE systems. It is obvious from these curves that the fusion of classification decisions made with respect to shape and texture (blue curve) offers better overall performance than that of the other two systems see green curve and red curves. These curves also explain the algorithms' dependency on the number of images per year available for training; since the number of images for senior ages (40 – 60) is relatively small, classification results deteriorate significantly as compared to MAE values achieved for younger ages. It could be argued that, given the availability of properly populated input training datasets, most of ML-AE absolute errors should be less than 5 and not more than 10 years. Also cumulative scores (see figure 5.8) moved up on an average of 0.5389 and 5.7916 with respect to texture and shape, respectively, by the proposed fusion of these two features.

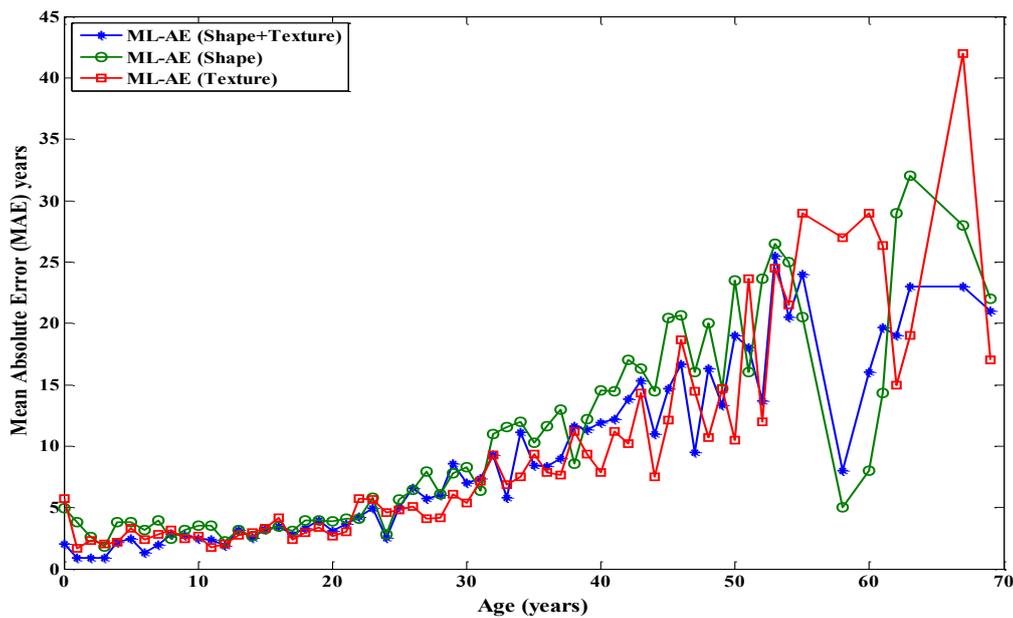


Figure 5.7: Mean Absolute Errors at different age years for FG-NET dataset; blue curve is representing MAEs for ML-AE based on both shape and texture, whereas green and red curves are for ML-AE schemes based on shape or texture, respectively.

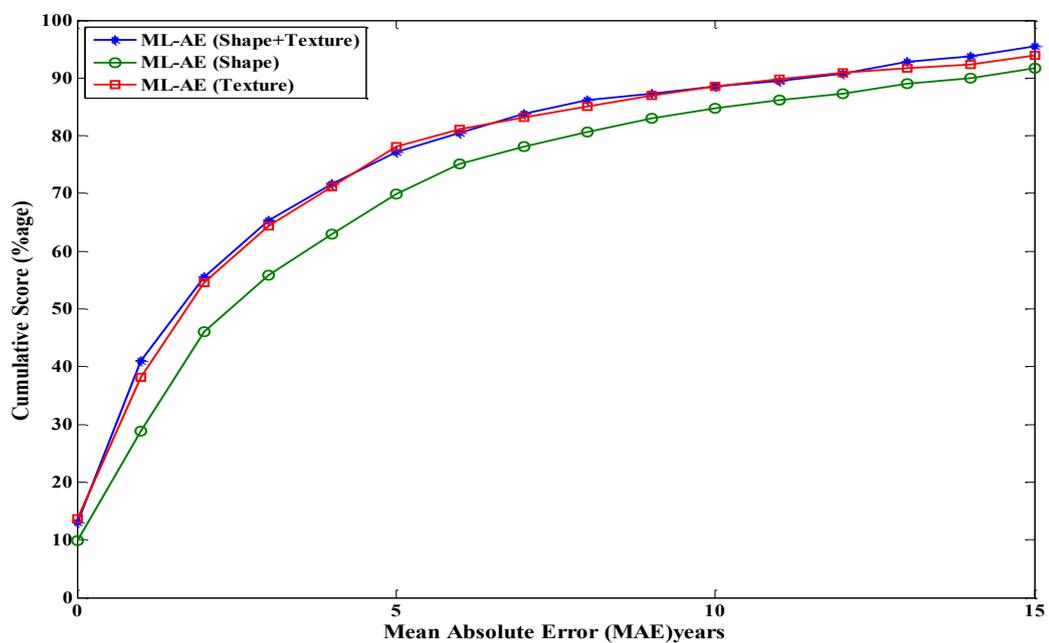


Figure 5.8: Cumulative scores comparing the proposed Multi-level Age Estimation with SVM classifier (ML-AE-SVM) framework, based on fusion of both shape and texture classifications, and ML-AE-SVMs based on shape or texture on their own.

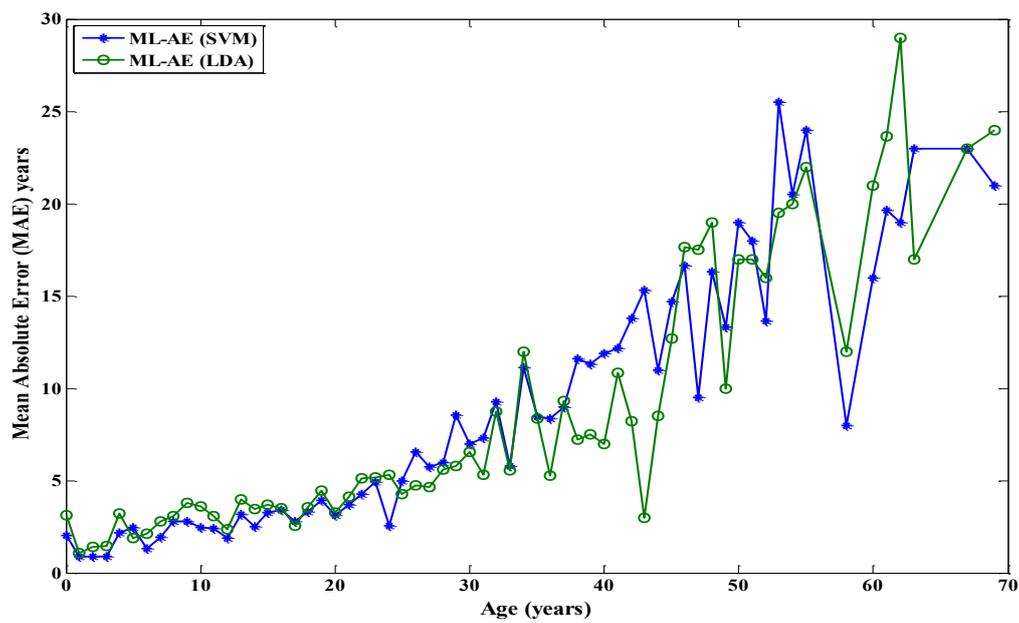


Figure 5.9: Mean Absolute Errors at different age years for FG-NET dataset.

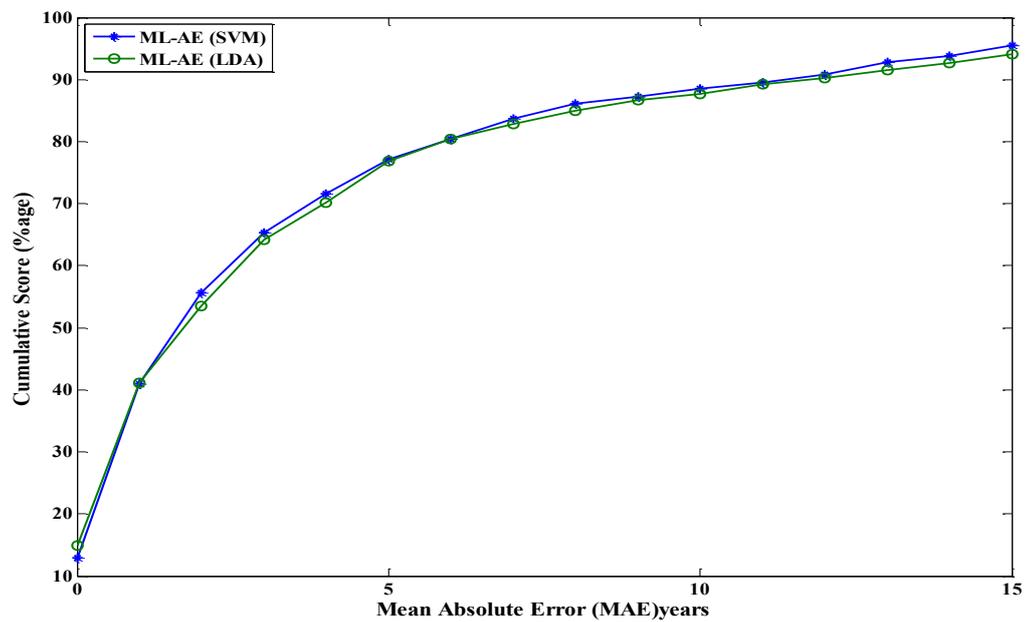


Figure 5.10: Cumulative scores comparing the proposed Multi-level Age Estimation framework with SVM (ML-AE-SVM) and with LDA (ML-AE-LDA).

## 2) Impact of Classifier

In this experiment the impact of the type of classifier used in ML-AE is highlighted. For this purpose, the performance of ML-AE, using Support Vector Machine (SVM) type classification is compared to that of ML-AE using Linear Discriminant Analysis (LDA) classification.

Computer simulation results show that ML-AE-SVM yields **4.16%** less MAE than that ML-AE-LDA (see table 5.3). Note that these are average values and in fact ML-AE-LDA gives slightly better performance over certain age years, see figure 5.9. Therefore and due to the lack of sufficient training data for older persons, a definite judgement in favour of one of the two systems cannot be secured, see also figure 5.10.

**Table 5.3: Comparison of ML-AE-SVM with ML-AE-LDA in terms of the MAE and standard deviation of error**

Method	Mean Absolute Error (years)	Standard Deviation of Error (years)
ML-AE-LDA	4.2161	5.2722
ML-AE-SVM	4.0406	4.9559

## 3) Comparison of ML-AE with Existing AE Systems.

Here ML-AE performance is compared against that of some state-of-the-art methods. Furthermore and for this purpose, only those schemes, which have evaluated (and their performance reported) using the FG-NET face dataset are included.

Mean Absolute Errors (MAEs) and the Cumulative Scores (CSs), for different Age Estimation methods, are presented in table 5.4. The cumulative scores, given in table, reflect correct age estimations within a 5-year absolute error.

Table 5.4 shows that even the best known method for FG-NET dataset with 4.1 years MAE and 73% cumulative score, reported in [89], is outperformed by ML-AE with 1.4390% improvement in terms of MAE and 4.11% improvement in terms of CS. These results are indicative of ML-AE's effectiveness and potential.

**Table 5.4: Comparison of AE methods with respect to Mean Absolute Error and Cumulative Score (MAE / CS).**

Method	Feature Extraction	MAE / CS (%age)
Thukral et al. [23]	Landmark based hierarchical approach	6.2 / NA
Han et al. [20]	Component and holistic biologically inspired features (BIF)	4.6 / 74.8%
Geng et al. [12]	Holistic appearance, principal component analysis (PCA)	6.8 / 65%
Suo et al. [17]	Holistic and local topology, 2D shape, colour, and gradient	6.0 / 55% Approx.
Guo et al. [44]	Holistic BIF	4.8 / 47%
Choi et al. [2]	Holistic appearance, Gabor, LBP	4.7 / 73%
Chao et al. [40]	Label-sensitive relevant component analysis	4.4 / NA
Han et al. [90]	BIF	4.5 / NA
Khryashchev et al. [91]	Local Binary Patterns (LBP)	7.47 / 37% Approx.
Chang et al. [92]	Ordinal hyper-plane ranking	4.5 / 74.7%
Wu et al. [93]	Grassmann manifold of Facial shape	5.9 / 62%
Luu et al. [89]	Holistic contourlet appearance mode	4.1 / 73% Approx.
ML-AE-SVM	Shape and Texture features Decision Fusion	4.041 / 77.11%

### 5.3 Conclusion

The proposed Multi-level Age Estimation (ML-AE) framework aims at the automatic and accurate estimation of human age on the basis of a given input face image. ML-AE operates in an unbalanced ternary tree structure to reach a final age estimation decision and yields better performance compared to different existing age estimation systems.

At each level of the tree ML-AE partitions a given face dataset, containing both shape and texture-based features for all the images, into four age-groups. Afterwards and for each feature type, two binary classifiers are trained on the basis of two different pairs of age-groups. During the testing phase, the best path to traverse such a tree is selected on the basis decision fusion of four classifiers i.e. two per type of features.

The novelty of the proposed framework stems from the notion that age classification errors can be reduced, if a given classifier is trained using age-groups with large inter-class variation. For this purpose, classifiers at each level of hierarchy are trained on alternative age-groups instead of adjacent groups, an action that is beneficial to the subsequent age classification process.

In conclusion computer simulation based experiments, performed to evaluate the proposed ML-AE approach, highlight the following points:

- ML-AE, based on the decision fusion of the two types of features i.e. shape and texture, has outperformed the ML-AE systems based on shape and texture alone, by yielding 21.45% and 5.65% less MAE respectively. Similarly, ML-AE with decision fusion has increased cumulative scores by an average of 0.5389% and 5.7916% with respect to texture and shape. These results suggest that neither shape nor texture when used alone as a discriminating feature can provide adequate age estimation performance.

- Secondly, the proposed ML-AE-SVM has shown signs of superiority over ML-AE-LDA but unfortunately this cannot be supported with certainty due to lack of training image data for older subjects.
- Finally, the performance gains, obtained using the proposed Multi-level Age Estimation framework, over a range of existing systems operating on the same input dataset, are indicative of ML-AE's potential.

## **Chapter 6**

### **Concluding Remarks and Future Work**

This chapter outlines some of the important conclusions derived from the research work undertaken as part of this thesis. Some possible extensions of proposed methodologies and their incorporations in various real life applications are also discussed.

#### **6.1 Conclusions**

This thesis explored the problem of face image based automatic age estimation (FI-AAE), an important process in many real life applications requiring age estimates (or age ranges) given the facial image of a person. Thus we proposed and applied a novel Multi-Level Age Estimation (ML-AE) framework; a hierarchical classification scheme which maximizes and then exploits inter-class variation among different age groups at each level of the hierarchy. Furthermore, the proposed scheme exploits age based discriminating information taken from two different cues (i.e. facial shape and texture) at the decision level which improves age estimation results.

During the process of achieving our main objective of age estimation, we have contributed to image analysis knowledge in general and to the following three face image processing areas in particular: i) Face image modeling and synthesis; a process of representing face image data with a low dimensional set of parameters. It is considered as the precursor to every face image based age estimation system, ii) Face image data variability and iii) Face image based

age estimation. Some key observations associated with each of these three studies are as follows:

### 6.1.1 Face Image Modeling and Synthesis

- The conventional Active Appearance Model (AAM) scheme for face image modeling and synthesis is not generic in nature and hence is not capable for representing unseen face images due to large variation of pose, expression and illumination. Moreover, during the process of dimensionality reduction, AAM tends to lose important discriminating age related information that adversely affects the overall age estimation system performance. The MM-AAM and MC/MM-AAM schemes for face image modeling and synthesis, proposed in Chapter-3, strive to cope with the above mentioned challenges associated with AAM. MM-AAM operates on the whole face image whereas MC/MM-AAM exploits local information by using each face component separately. The idea of the proposed work is based on the fact that face data variability (due to pose, expression, illumination etc.) can be reduced by splitting data into smaller groups of similar characteristics. For this purpose, both MM-AAM and MC/MM-AAM cluster face image data into small groups on the basis of shape similarities and yield multiple models, out of which the best one is selected for the synthesis of the unknown test face image. Similar AAM based existing techniques that operate on small groups of face data only cater for the pose variations, whereas the clustering method employed in our proposed schemes cope with all types of variation and hence yield better face representation/synthesis performance gains in terms of both seen and unseen data.
- MC/MM-AAM that operates on local facial information by using each face component (i.e. eyes, nose, mouth, etc.) separately has performed better in terms of synthesizing facial shape and texture, in comparison with MM-AAM. However, decomposing a given face

image into its components and then applying separate AAM operations for each component is computationally quite expensive job that made the MC/MM-AAM performance improvement over MM-AAM an undesirable option for applications such as face recognition and face image based age estimation.

- In Chapter-3, we also examined the applicability of MM-AAM under the application of face recognition and proposed a FR-MM-AAM framework. This proposed FR framework effectively outperformed the face recognition system based on conventional AAM. Although, MM-AAM has performed significantly better in synthesizing shape and texture of unseen face images and outperformed AAM, texture synthesis errors are still quite large as compared to shape synthesis errors. Furthermore since face recognition operates on texture information and large texture synthesis errors can affect recognition performance, we have utilized the synthesized shape to extract the original texture from the face image which is then used for the recognition purposes. Computer simulation experimental results have shown the effectiveness and considerable potential of this new FR framework.

### 6.1.2 Face image data variability

- The second contribution of this research work is quantifying face image data variability and proposes a single image variability measure VM that represents the overall variation of a face dataset. In this work and due to a lack of availability of age based face datasets, we have used face image datasets that are available for the purpose of face recognition. Furthermore, the relationship between the proposed VM and system recognition performance is also modeled using a 2<sup>nd</sup> degree polynomial. This can be useful in predicting the recognition performance of a given classifier, given a particular face image dataset. The polynomial model is selected on the basis of two ‘goodness of fit’ parameters: R-squared ( $R^2$ ) and adjusted R-squared ( $\overline{R^2}$ ).

- The proposed formulation of VM benefits from the correlation among images to, separately, form inter- and intra-subject variability measures, i.e. VM-interSC and VM-intraSC, which are then combined to produce a single measure VM to represent overall variability.
- Computer simulation experiments based on both noise-free and noisy (coded) data showed the effectiveness of this variability measure VM to characterize equally well FR system performance while operating on noise free face image or noisy/coded data.

### **6.1.3 Face image based Automatic Age Estimation (FI-AAE)**

- In general, existing hierarchical face image based age estimation systems with hard boundaries between adjacent age groups suffer inaccuracies due to smooth age variation among images located on boundaries of age groups. The multi-level age estimation ML-AE framework, proposed in Chapter-5, operates on a novel method of age based data partitioning that maximizes inter-group variation to effectively overcome the challenge of this smooth variation and as a result outperforms existing age estimation systems. The proposed ML-AE, at each level of its hierarchy, splits face image clusters into four age-based partitions and then performs classification using alternative age groups with larger inter-group variation.
- Secondly, the proposed ML-AE operates separately on facial shape and texture and at each level fusion of decisions takes place from both streams in order to decide the next path to traverse. This proposed method of using shape and texture separately helps in avoiding the loss of the important discriminating features that is a major limitation in the case of conventional AAM based systems with hybrid feature sets.
- Computer simulation based experiments highlight different aspects of the proposed ML-AE framework by i) showing the positive impact of decision fusion of the two features,

ii) comparing the performance of the systems based on two different classification techniques i.e. LDA and SVM and finally iii) comparison with other existing techniques further shows the effectiveness of this proposed ML-AE framework.

## 6.2 Future Work

Although this research work offers a number of novel contributions, it deals only with face image based age estimation and therefore can be extended to face based age synthesis, a process of generating a face image for a given age year. Face image based age synthesis has become an interesting topic these days because of emerging new applications. For example in forensics, age synthesis is used to modify and enhance photographs for the purpose of identifying suspects, victim and lost persons. In real life applications such as passport renewal and border security, a face recognition system equipped with an age synthesis module can recognize faces after a gap of several years by dynamically tuning facial shape and texture model parameters. By enhancing the face image modeling and synthesizing capability of our proposed MM-AAM, a face image based age synthesis system can be built and effectively used in the many real life applications. One way to achieve this is to obtain model parameter sets for each cluster during a MM-AAM modeling phase. Thus an  $n$ -degree polynomial model can be generated for each of the parameter to provide the evolution over time of that parameter. Therefore, consider if we have  $u$  parameters in one vector it means we will have  $u$  number of models. During synthesis, once we get the best model parameters for the input face image using MM-AAM, a closest parameter is found out of all training model parameters for each of these input parameters. Then polynomial model created for each of these closest parameters will be used to obtain the estimated parameter for the target age. Once all the estimated parameters for the target age are obtained, they will be used to synthesis the face image corresponding to the target age. However, for this purpose we need a

quite large age-based face dataset so that we can have enough number of images, corresponding to different age years, in each cluster, that can be obtained either by combining different age-based face datasets or creating a new dataset using images available on internet.

Furthermore, one more useful application of this research work can be low bandwidth data transmission. For example, sometimes it is required to transmit human facial images for further processing using bandwidth restricted links, in this case the proposed MM-AAM or MC/MM-AAM can be used to represent face image with a small set of parameters that can be efficiently transmitted and used on receiver side to regenerate the face image.

Finally, some suggestions to improve, in the future, the performance of proposed age estimation system ML-AE are listed below:

- Since age progression affects faces of male and female persons differently, making ML-AE gender aware, should further improve age estimation performance. This means, the ML-AE system is trained using only male face images to estimate the age of a male person, whereas for female faces ML-AE is trained using only female faces. Other types of subdivision of input data use for training can also be exploited, e.g. ethnicity.
- ML-AE framework is currently based on only one classifier i.e. SVM or LDA, this work can be extended and enhanced to accommodate operate on a bank of state-of-the-art classifiers with a following fusion of decisions process and can yield final results based on fusion of decisions from all the streams.

## Bibliography

- [1] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, pp. 387-394.
- [2] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognition*, vol. 44, pp. 1262-1281, 2011.
- [3] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, pp. 1955-1976, 2010.
- [4] L. S. Mark, J. B. Pittenger, H. Hines, C. Carello, R. E. Shaw, and J. T. Todd, "Wrinkling and head shape as coordinated sources of age-level information," *Perception & Psychophysics*, vol. 27, pp. 117-124, 1980.
- [5] H.-C. Lian and B.-L. Lu, "Age estimation using a min-max modular support vector machine," in *Twelfth International Conference on Neural Information Processing*, 2005, pp. 83-88.
- [6] F. Gao and H. Ai, "Face age classification on consumer images with gabor feature and fuzzy lda method," in *Advances in biometrics*, ed: Springer, 2009, pp. 132-141.

- [7] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *Image Processing, IEEE Transactions on*, vol. 17, pp. 1178-1188, 2008.
- [8] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *Multimedia, IEEE Transactions on*, vol. 10, pp. 578-584, 2008.
- [9] H. Fukai, H. Takimoto, Y. Mitsukura, and M. Fukumi, "An apparent age estimation system using the evolutionary algorithm," in *2007 International Conference on Control, Automation and Systems*, 2007.
- [10] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," in *Proceedings of the 14th annual ACM international conference on Multimedia*, 2006, pp. 307-316.
- [11] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 34, pp. 621-628, 2004.
- [12] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 2234-2240, 2007.
- [13] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.

- [14] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 442-455, 2002.
- [15] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 681-685, 2001.
- [16] J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen, and X. Gao, "Design sparse features for age estimation using hierarchical face model," in *Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on*, 2008, pp. 1-6.
- [17] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, pp. 385-401, 2010.
- [18] W.-B. Horng, C.-P. Lee, and C.-W. Chen, "Classification of age groups based on facial features," *Tamkang Journal of Science and Engineering*, vol. 4, pp. 183-192, 2001.
- [19] A. Günay and V. V. Nabiyev, "Automatic age classification with LBP," in *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, 2008, pp. 1-4.
- [20] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *Biometrics (ICB), 2013 International Conference on*, 2013, pp. 1-8.

- [21] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, "A study on automatic age estimation using a large database," in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1986-1991.
- [22] K. Luu, K. Ricanek Jr, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, 2009, pp. 1-5.
- [23] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 1529-1532.
- [24] T. Cootes and A. Lanitis, "The FG-NET aging database," ed, 2008.
- [25] K. Ricanek Jr and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 341-345.
- [26] MORPH Face Database [Online]. Available: <http://faceaginggroup.com/>
- [27] O. Aghazadeh and S. Carlsson, "Properties of datasets predict the performance of classifiers," in *British Machine Vision Conference*, 2013, pp. 22-31.

- [28] M. A. Khan, C. Xydeas, and H. Ahmed, "Multi-model AAM framework for face image modeling," in *Digital Signal Processing (DSP), 2013 18th International Conference on*, 2013, pp. 1-5.
- [29] M. A. Khan, C. Xydeas, and H. Ahmed, "Multi-Component/Multi-Model AAM framework for face image modeling," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 2124-2128.
- [30] M. A. Khan, C. Xydeas, and H. Ahmed, "On the application of AAM-based systems in face recognition," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, 2014, pp. 2445-2449.
- [31] Y. H. Kwon and N. D. V. Lobo, "Age classification from facial images," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, 1994, pp. 762-767.
- [32] L. G. Farkas, *Anthropometry of the Head and Face*: Raven Pr, 1994.
- [33] A. Günay and V. V. Nabiyevev, "Automatic detection of anthropometric features from facial images," in *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, 2007, pp. 1-4.
- [34] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *Image Processing, IEEE Transactions on*, vol. 7, pp. 359-369, 1998.

- [35] K. Ricanek Jr, Y. Wang, C. Chen, and S. J. Simmons, "Generalized multi-ethnic face age-estimation," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*, 2009, pp. 1-6.
- [36] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics*, vol. 32, pp. 407-499, 2004.
- [37] A. Sethuram, E. Patterson, K. Ricanek, and A. Rawls, "Improvements and performance evaluation concerning synthetic age progression and face recognition affected by adult aging," in *Advances in Biometrics*, ed: Springer, 2009, pp. 62-71.
- [38] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, pp. 2037-2041, 2006.
- [39] Y.-L. Chen and C.-T. Hsu, "Subspace learning for facial age estimation via pairwise age ranking," *Information Forensics and Security, IEEE Transactions on*, vol. 8, pp. 2164-2176, 2013.
- [40] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognition*, vol. 46, pp. 628-641, 2013.
- [41] T. F. Cootes and C. J. Taylor, "Statistical models of appearance for computer vision," ed: Technical report, University of Manchester, 2004.

- [42] R. Gross, I. Matthews, and S. Baker, "Generic vs. person specific active appearance models," *Image and Vision Computing*, vol. 23, pp. 1080-1093, 2005.
- [43] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *Image processing, IEEE Transactions on*, vol. 11, pp. 467-476, 2002.
- [44] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 112-119.
- [45] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 285-339, 1991.
- [46] T. F. Cootes, G. V. Wheeler, K. N. Walker, and C. J. Taylor, "View-based active appearance models," *Image and vision computing*, vol. 20, pp. 657-664, 2002.
- [47] L. Teijeiro-Mosquera and J. L. Alba-Castro, "Performance of active appearance model-based pose-robust face recognition," *Computer Vision, IET*, vol. 5, pp. 348-357, 2011.
- [48] C. Zhang and F. S. Cohen, "Component-based active appearance models for face modelling," in *Advances in Biometrics*, ed: Springer, 2005, pp. 206-212.
- [49] R. C. Luo, C. Y. Huang, and P. H. Lin, "Alignment and tracking of facial features with component-based active appearance models and optical flow," in *Advanced*

- Intelligent Mechatronics (AIM), 2011 IEEE/ASME International Conference on*, 2011, pp. 1058-1063.
- [50] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *Communications, IEEE Transactions on*, vol. 28, pp. 84-95, 1980.
- [51] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann, "The IMM face database-an annotated dataset of 240 face images," Technical University of Denmark, DTU Informatics, Building 3212004.
- [52] X. Gao, Y. Su, X. Li, and D. Tao, "A review of active appearance models," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 40, pp. 145-158, 2010.
- [53] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognition Letters*, vol. 28, pp. 1885-1906, 2007.
- [54] M. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, 1991, pp. 586-591.
- [55] G. McLachlan, *Discriminant analysis and statistical pattern recognition* vol. 544: John Wiley & Sons, 2004.
- [56] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.

- [57] V. Franc and V. Hlavác, "Statistical pattern recognition toolbox for Matlab," *Prague, Czech: Center for Machine Perception, Czech Technical University*, 2004.
- [58] S. Z. Li, *Encyclopedia of Biometrics: I-Z* vol. 1: Springer Science & Business Media, 2009.
- [59] S. J. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 970-984, 2008.
- [60] AT&T face database [Online]. Available: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
- [61] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Generative models for recognition under variable pose and illumination," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 277-284.
- [62] Georgia Tech Face Database [Online]. Available: <ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/>.
- [63] Stirling Face Database [Online]. Available: <http://pics.stir.ac.uk>
- [64] A. M. Vedit Jain. The Indian Face Database [Online]. Available: <http://www.cs.umass.edu/~vidit/IndianFaceDatabase/>

- [65] FEI Face Database [Online]. Available: <http://fei.edu.br/~cet/facedatabase.html>
- [66] K. Messer, J. Matas, J. Kittler, J. Luetin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Second international conference on audio and video-based biometric person authentication*, 1999, pp. 965-966.
- [67] D. B. Graham and N. M. Allinson, "Characterising virtual eigensignatures for general purpose face recognition," in *Face Recognition*, ed: Springer, 1998, pp. 446-456.
- [68] Bern Univ. Face Database [Online]. Available: <ftp://iamftp.unibe.ch/pub/Images/FaceImages/>
- [69] R. Singh, M. Vatsa, A. Ross, and A. Noore, "A mosaicing scheme for pose-invariant face recognition," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 37, pp. 1212-1225, 2007.
- [70] D. Beymer and T. Poggio, "Face recognition from one example view," in *Computer Vision, 1995. Proceedings., Fifth International Conference on*, 1995, pp. 500-507.
- [71] S. Chen and B. C. Lovell, "Illumination and expression invariant face recognition with one sample image," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 300-303.
- [72] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image and vision computing*, vol. 16, pp. 295-306, 1998.

- [73] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, 2002, pp. 46-51.
- [74] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *Image Processing, IEEE Transactions on*, vol. 16, pp. 1716-1725, 2007.
- [75] T. Kanade and A. Yamada, "Multi-subregion based probabilistic approach toward pose-invariant face recognition," in *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, 2003, pp. 954-959.
- [76] J. Lewis, "Fast normalized cross-correlation," in *Vision interface*, 1995, pp. 120-123.
- [77] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 684-698, 2005.
- [78] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society*, vol. 40, pp. 1170-1175, 2010.
- [79] T. Shkunaga and K. Shigenari, "Decomposed eigenface for face recognition under various lighting conditions," in *Computer Vision and Pattern Recognition, 2001.*

- CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. I-864-I-871 vol. 1.
- [80] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 711-720, 1997.
- [81] N. Kwak and J. Oh, "Feature extraction for one-class classification problems: Enhancements to biased discriminant analysis," *Pattern Recognition*, vol. 42, pp. 17-26, 2009.
- [82] Z. Liang, Y. Li, and P. Shi, "A note on two-dimensional linear discriminant analysis," *Pattern Recognition Letters*, vol. 29, pp. 2122-2128, 2008.
- [83] L. Rueda and M. Herrera, "Linear dimensionality reduction by maximizing the Chernoff distance in the transformed space," *Pattern Recognition*, vol. 41, pp. 3138-3152, 2008.
- [84] Y. Yan and Y.-J. Zhang, "A novel class-dependence feature analysis method for face recognition," *Pattern Recognition Letters*, vol. 29, pp. 1907-1914, 2008.
- [85] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *Neural Networks, IEEE Transactions on*, vol. 13, pp. 415-425, 2002.

- [86] E. Gayawan and R. A. Ipinyomi, "A comparison of Akaike, Schwarz and R square criteria for model selection using some fertility models," *Australian Journal of Basic and Applied Sciences*, vol. 3, pp. 3524-3530, 2009.
- [87] A. K. Jain and S. Z. Li, *Handbook of face recognition* vol. 1: Springer, 2005.
- [88] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, pp. 2401-2412, 2013.
- [89] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *Biometrics (IJCB), 2011 International Joint Conference on*, 2011, pp. 1-8.
- [90] H. Han and A. Jain, "Age, gender and race estimation from unconstrained face images," *Dept. Comput. Sci. Eng., Michigan State Univ., East Lansing, MI, USA, MSU Tech. Rep.(MSU-CSE-14-5)*, 2014.
- [91] V. Khryashchev, A. Ganin, O. Stepanova, and A. Lebedev, "Age estimation from face images: challenging problem for audience measurement systems," in *Open Innovations Association (FRUCT16), 2014 16th Conference of*, 2014, pp. 31-37.
- [92] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 585-592.

- [93] T. Wu, P. Turaga, and R. Chellappa, "Age estimation and face verification across aging using landmarks," *Information Forensics and Security, IEEE Transactions on*, vol. 7, pp. 1780-1788, 2012.