Chapter

INTRODUCTION

1.1 General Face Recognition Structure

Face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding over the last decade. Because of the nature of the problem, not only computer science researchers are interested in it, but neuroscientists and psychologists also.

A general definition of the face recognition in computer vision can be given as follows: Given still or video images of a scene, identify or verify one or more persons in the scene by comparing facial features from the test image and facial database.

Face recognition has a wide range of applications, including law enforcement, civil applications, and surveillance systems. Face recognition applications have also been extended to smart home systems where the recognition of the human face and expression is used for better interactive communications between human and machines.

Face recognition generally divided into three steps: Face Detection, Feature Extraction, and Face Recognition as shown below in figure.

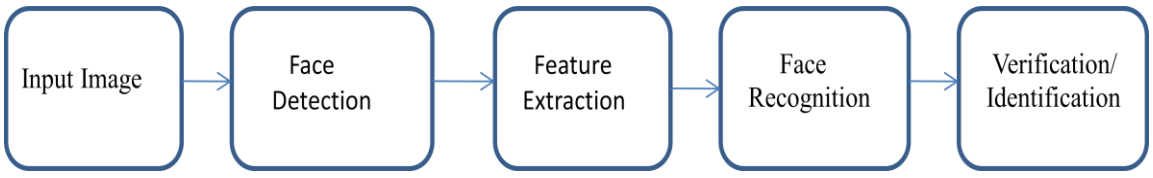


Figure 1.1: General Face Recognition Structure

1.1.1 Face Detection

For face recognition the first step is face detection [1][2] so further processing on face can be done.

Some of the well-known face detection approaches can be categorized as :-

i) color based [3] ii) template based [4] and iii) feature based [5]

Color based approaches learn the statistical model of skin color and use it to segment face candidates in an image. Template based approaches use templates that represent the general face appearance, and use cross correlation based methods to find face candidates. State-of-the-art face detection methods are based on local features and machine learning based binary classification (e.g., face versus non-face) methods.

1.1.2 Feature Extraction / Representation

After the face detection step, we have to represent the face and for that features extractions techniques are used in literature. As the faces may contain different expression, illuminations, occlusions, and clutter under various camera alignment. So to overcome these problems in face recognition we have to select that feature extraction technique which are robust to these drawbacks and should be invariant to poses, illuminations, occlusions etc.

Most face recognition techniques use one of two representation approaches:

i) Local feature based [6] [7] [8] or ii) holistic based [9] [10] [11].

Local feature based approaches identify local features (e.g., eyes, nose, mouth, or skin irregularities) in a face and generate a representation based on their geometric configuration.

Holistic approaches localize the face and use the entire face region in the sensed image to generate a representation.

1.1.3 Face Recognition

After the representation of each face, the next step is to recognize the identities of these faces. In order to achieve automatic recognition, a face database is required to build. For each subject, many images are taken and feature extraction is performed on them and extracted features are stored in the database. Then when we input the test face image, we again perform face detection and feature extraction, and extracted features of test image face are compared with feature of each face class stored in the database.

1.1.4 Verification / Identification

Now after features comparison of test face with the stored features of faces we can verify or indentify the test face. Face recognition includes both face identification and face verification. Face verification means validating a claimed identity based on the image of a face, and either accepting or rejecting the identity claim (one-to-one matching). On the other hand, the goal of face identification is to identify a person based on the image of a face. This face image has to be compared with all the registered persons (one-to-many matching).

1.2 Issues In Face Recognition

Table 1.1 List of problems in Face Recognition

Pose Variations	Pose variation results from performance degradation in image acquisition process due to different angles and location during imaging of face. It falls into two categories: (i) intra-user variation (two different viewpoints of the same subject) (ii) inter-user variation (same view point for two different subjects).
Lighting Variation	Since the face is a 3D object, different lighting sources can generate various illumination conditions and shadings. So to compensate lighting variations it is better to have the knowledge of its sources.
Expression Variation	Facial expression is an internal variation that causes large intra-class variation as different persons uses different facial expressions to express their feelings. These variation results in the facial feature shape change.
Occlusion	In surveillance videos face images may often get occluded by other objects or by the face itself (i.e., self-occlusion) makes the recognition of subject difficult as the facial features becomes unobserved.
RST Variation	During the image acquisition process rotations, scaling and translation variations (RST) get encountered in image.
Age Variation	It is the most difficult challenge in automatic face recognition because with the age progress facial appearance changes due to shape variations such as weight loss and gain and texture variations such as wrinkles and speckles with the progression of age.

Figure 1.2: Example of Pose Variations

(a) glasses (b) sunglasses (c) hat (d) scarf

Figure 1.3: Examples of Occlusions

Figure 1.4: Examples of lighting variations

Figure 1.5: Examples of Expression Variations

Figure 1.6: Examples of Aging variation as age progress

1.3 Problem Statement of Age invariant Face Recognition

Here the age invariant face recognition means recognizing a face irrespective of age. As already mentioned in issue of face recognition section aging variation is one of major problem in automatic face recognition. Designing age invariant face recognition system has many application such as checking whether the same person has been issued multiple government documents (e.g., passports and driver license), missing children identification etc. Further related work on age invariant face recognition is very limited so this has been the active research area from past decade which motivated me to find robust method to compensate of aging variation in face recognition.

I studied related work and found that local descriptors for image are more robust [12] and greater recognition rate than global one as they possess spatial locality and orientation selectivity. Thus these properties make the local features to be robust to aging, illumination, and expression variation.

As age invariant database FG-NET has pose variation , illumination, rotation and scale variations so in our approach we use SIFT[13] and LBP[14] descriptors as both of them are good representation of images[12].

Then combining both the features and learning through multiclass SVM [15] makes method more efficient to give more recognition rate.

Now in the preprocessing steps to make our method more robust we are using AAM models [16] to make posture correction as discriminative model for age invariant face recognition purposed by Zhifeng Li [17] was vulnerable to pose changes.

1.4 Literature Survey on Age Invariant Face Recognition

Lanitis et al. [18], proposed a model based face recognition in which the age progressive training set is projected onto a model space. The face model thus formed contained 50 model parameters and is a combination of shape and intensity model. The variation caused due to aging effect is isolated using an aging function (linear, quadratic or cubic) which operated on the model parameters to produce the actual age of the image. After the estimation of age, using appropriate ageing function, the model parameters are converted into new set of parameters corresponding to the target age. The test and training image parameters are obtained at target age which is the mean age of training set. Thus obtained parameters are used as feature vectors for recognition.

Sethuram et al.[19], face aging model built by him based on AAMs, support vector machines (SVMs) and Monte-Carlo simulation reported high accuracy. He performed two experiments. In experiment 1, they showed that when we probe faces age then there is a decrease in accuracy of face recognition. In experiment 2, they use face aging model on the probe faces to artificially age the faces to the same age of the gallery to make the face recognition algorithm more accurate then experiment 1.

Wang, Shang, Su and Lin [20]**,** presented an age simulation model which transforms image to a target image for recognition. For age simulation, first using ASM model, shape features are extracted and is spanned to texture image by triangle based affine transformation. Shape Eigen face and texture Eigen face are acquired by applying PCA on shape and texture image, which are further combined to form facial feature vector. A polynomial aging function along with K-means classification of aging way is used for estimating age. This estimated age along with typical vector creating function was used to generate feature vector at target age. Finally the shape and texture vector were reconstructed in Eigen spaces and combined to produce facial image at the target age which was further used for recognition.

Ramanathan & Chellappa [21], proposed Bayesian age difference classifier that is built on probabilistic Eigen space framework. To remove irregular illumination effects, better illuminated half of the image was extracted using mean and optimal mean intensity curves, under the assumption of bilateral symmetry. These half faces thus obtained were given the name of Point Five faces. An age difference classifier was developed in which difference of given pair of Point Five faces were grouped as Intra personal or Extra personal image difference by computing its Posteriori probability using Bayes rule. Those pairs which came out to be Intra personal were further classified into four groups representing their age difference by selecting the maximum of their Posteriori probabilities.

Ling, Soatto, Ramanathan and Jacobs [23], proposed using SVM classifier and GOP descriptor for efficient face recognition across aging. For any two given images, it is first mapped onto feature space using feature extraction function and then classified into two categories as intra personal or extra personal using support vector machine. The feature extraction function used is gradient orientations at different scales which forms a pyramid hierarchal representation. At each scale, the image values of previous scale are convolved with Gaussian kernel with 0.5 as standard deviation. Thus obtained gradient vectors are normalized to form gradient orientation at each scale. Thus for a given pair of images, the feature vector formed is the concatenation of cosines of difference of gradient orientations at each scale and pixel position. This SVM+GOP approach were compared to several other techniques to find it as the best technique for face recognition which includes aging effect. Also an empirical study on aging process was performed which showed that after an age difference greater than 4, the recognition rate saturates.

Park, Tong and Jain [24], Park et al has proposed a 3D deformable model that can accompany any pose and lighting invariant 3D model for age invariance. For converting a 2D face into 3D, a face mesh having 81 feature point vertices is formed and PCA is applied on these shape vectors to form shape space. The 3D shape space is a matrix of M number of rows and N number of columns, where M is number of distinct ages and N is number of distinct subjects. Each element of the space is a vector representing 3D shape of the face. For a given probe shape at age x, a weighted sum of the shapes at that age can be generated. Using these weights, a new artificial face can be generated at target age y. In the same fashion, a texture vector and space is also created. The model is then tested on FaceVACS software and a slight increase in recognition rate is identified.

Mahalingham and Kambhanmettu [25], proposed an aging model combined with graphical representation of faces for constructing a more efficient face recognition system. The facial image is constructed as a graph with feature points as vertices. Each vertex is labeled by its descriptor i.e. texture information. To extract feature points LFA (Local Feature Analysis) is used which constructs n kernels, where n is the number of pixels in the image. To reduce this dimensionality of the representation, Fishers Linear Discriminant method is used to choose a subset of kernels which has higher fisher score. These kernels correspond to the feature points in the image. After that, a uniform LBP (Local Binary Pattern) operator is used to extract feature descriptor of each feature point. For testing the model, two-step process is used. First, a graph is constructed using feature points as vertices and corresponding descriptors and likelihood score is calculated with each training image. The ones with highest scores are chosen for the next step of matching. The graph of the probe image is matched with training images and recognition result is calculated.

Hsieh, Pan and Hu [26], proposed a facial aging synthesis method which comprised of ASM model to detect facial landmarks and proper alignment of them, Log Gabor wavelet to analyze the landmarks and finally synthesizing age using decomposition maps to construct the image at target age. ASM is used to detect location information of important facial features and to attain geometric invariance, the distance between inner corner of eyes are made horizontal and the distance between chin and nose is made vertical. These two halves of the face are thus combined and philtrum of test and training image is matched. To obtained skin topographies, Log-Gabor wavelet decomposition maps are convolved with the face image. The higher frequency portion of the target image is placed on higher frequency portion of test image and thus formed decomposition map yields the final target image. The proposed method got 100% accuracy in aligning eyes, nose, mouth and all three in test and training set. Age synthesis was achieved successfully.

Luefei-Xu, Luu, Savvides1, Bui, and Suen [27], proposed the method of using per ocular region of human face for adding age invariance factor to face recognition system, as per ocular regions undergo least amount of changes as the face ages. Walsh-Hadamard Transform Encoded Binary Pattern (WLBP) based feature extraction technique is used and Unsupervised Discriminant Projection (UDP) is used to build subspaces on WLBP image. It is a fusion of Walsh-Hadamard Transform and LBP, in which LBP is not applied directly to raw image pixels but after some transformations. Walsh masks are used as convolution filter for faster location of local image characteristics. Walsh function is used to extract samples at integer points to produce 2D basis images. After Walsh coefficients are obtained, LBP is applied on them to get the feature points of the image. UDP has the potential to minimize local and maximize non-local characteristics of the image concurrently by formulating scatter matrix for both. Normalized cosine distance measurement is taken on to compute similarity matrix between training and test image. The results showed better recognition rate as compared to using PCA or LPP as classifier.

Li, Park and Jain [17], proposed a discriminative model that has densely sampled location feature description scheme with Scale Invariant Feature Transform (SIFT) and Multi Scale Local Binary Pattern (MLBP) as descriptors and Multi Feature discriminant Analysis (MFDA) as classifier. For feature extraction, the image to first normalized and divided into set of overlapping patches and each patch is represented by 88 and 408 dimensional SIFT and MLBP feature vector. Since, the resulting dimensionality is very high, MFDA framework is proposed to reduce dimensions and resolve over fitting problem. The feature sets of training images are divided into slices of same feature. PCA is applied on each slice to construct 10 random PCA subspaces and calculate within class scatter matrix for each subspace. Then each subspace is whitened to remove intra personal variations. Five different between class scatter matrix is constructed using bagging technique and thus 5 LDA classifiers are constructed for each of 10 subspaces. Thus each slice has 50 different classifiers. In testing phase, slices are made for each test image and classification outputs are calculated for each slice. Using the min-max score normalization scheme, these outputs are normalized and score sum based fusion rule is used for the final decision.

Dihong Gong, Zhifeng Li and Dahua Lin[28], purposed new approach " Hidden Factor Analysis for Age Invariant Face Recognition", motivated by the belief that the facial image of a person can be expressed as combination of two components: an identity-specific component that is stable over the aging process, and the other component that reflects the aging effect.

In particular, they introduce two latent factors: an identity factor and an age factor, which respectively govern the generation of these two components. Intuitively, each person is associated with a distinct identity factor, which is largely invariant over the aging process and thus can be used as a stable feature for face recognition; while the age factor changes as the person grows.

For computational simplicity, they assume a linear model, where the identity components and the age components lie on two different subspaces. In this way, the problem of separating identity and age factors naturally reduces to a problem of learning the basis of these subspaces.

As both the subspaces and the latent factors are unknown in the training stage, then they derive an algorithm that can jointly estimate both from a set of training image, based on an Expectation-Maximization process. In this process, the latent factors and the model parameters are iteratively updated to maximize a unified objective. In the testing, given a pair of face images with unknown ages, we compute the match score between them by inferring and comparing the posterior mean of their identity factors.

D.Sungatullina[29] propose a new multi-view discriminative learning (MDL) method for age invariant face recognition, in which three different local feature descriptors scale invariant feature transform (SIFT), local binary patterns (LBP) and gradient orientation pyramid (GOP) for each face image are used to exploit the discriminative information. Then, a discriminative learning method with multi-view feature representations, called MDL, is used to project different types of local features into a latent discriminative subspace where the intra-class variation of each feature is minimized, the interclass variation of each feature and the correlation of different features of the same person are maximized.

1.5 Organization of Thesis

Chapter 2: It discusses the Age Invariant Face Recognition existing approaches in detail such as Generative approaches & Discriminative approaches.

Chapter 3: It presents the purposed approach for Age Invariant Face Recognition.

Chapter 4: In this section, all results have been conducted and Comparison of the results obtained with the existing results has been shown.

Chapter 5: Conclusion of the thesis and the future scope of the implemented work are presented.

References: This section gives the reference details of the thesis.

Chapter 2

AGE INVARIANT FACE RECOGNITION APPROACHES

Existing methods on Age Invariant Face Recognition are limited and can be categorized into two approaches: (i) Generative approaches (ii) Discriminative approaches

2.1 Generative Approaches

It is one of the successful method in which face images are transformed to match the target age before recognition. It used the concept of building 2D or 3D generative model for face aging for compensation of aging process in face matching.

Linitis et al. [18] proposed a statistical method to capture the facial shape and texture variation over age progression.

Ramanathan and Chellapa [30] presented a face growing model for face verification across age for people under the age of 18 years old.

Wang et al.[20] also presented an age simulation method by transforming facial textures and shapes of source age to the target age for face verification in age invariant face recognition.

Park et al. [24] presented a 3D aging modeling technique to compensate for the age variation due to shape of face and its texture as age progress to improve the face recognition performance. This method shows better result in generative approaches. So it is discussed below in detail.

2.1.1 3D Aging Modeling

Park et al. [24], first overcome the pose variations using pose correction step and then he consider the aging process as 3D process, thus he further model the aging pattern in 3D domain easily. For that he converts a 2D face aging database to 3D aging model because there is no 3D aging database is available. In this modeling aging patterns are defines as an array of face models from a single subject indexed with age

He purposed separate shape modeling only, separate shape and texture modeling, and combined shape and texture modeling. Then he shows that the separate modeling is better than combined modeling on the FG-NET database.

2.1.2 Shape Aging Pattern

The internal shape changes variations and face size variations are captured by shape pattern space. For constructing the shape aging pattern space a preprocessed pose corrected stage is used.

Then PCA is performed on all the 3D shapes pose corrected images, A_i^j. Projection of all the mean subtracted A_i^j onto the subspace spanned by the columns of U_a to obtain a_i^j is written as

$$a_i^j = U_a^T(\mathrm{A}_i^j - \bar{\mathrm{A}}) \quad \text{.....(2.1)}$$

which is a $L_s \times 1$ vector.

Where j^{th} row is age j and the i^{th} column is subject i, thus the entry at (j,i) is a_i^j, U_a^T is transformation matrix, $\bar{\mathrm{A}}$ is mean.

The shape aging pattern space is the space which contains all linear combinations of the patterns which are expressed in PCA basis.

$$a_{w_a}^j = \bar{a}^j + \sum_{i=1}^{n}\left(a_i^j - \bar{a}^j\right)w_{a,i}; 0 \le j \le m-1 \quad \text{.....(2.2)}$$

As the weight w_a in the linear combination above is not unique for the same aging pattern. Now to overcome this problem, regularization term can be used in the aging simulation.

With the mean shape $\bar{A}$, transformation matrix U_a and complete shape pattern space, the shape aging model with weight w_a is defined as:

$$A_{w_a}^j = \bar{A} + U_a a_{w_a}^j ; 0 \le j \le m-1 \quad \text{.....(2.3)}$$

2.1.3 Texture Aging Pattern

To obtain texture pattern B_i^j for subject i at age j, first map the original face image to frontal projection of the mean shape $\bar{B}$ and then take column-wise concatenation of the image pixels.

Transformation matrix U_b and the projected texture t_i^j is calculated by applying PCA on B_i^j Now to construct the complete texture pattern space using t_i^j, the same procedure is used as we have used in shape pattern space.

Thus a new texture $B_{w_b}^j$ obtained, at age j and with a set of weights w_b, is given as:

$$b_{w_b}^j = \bar{b}^j + \sum_{i=1}^{n} \left(b_i^j - \bar{b}^j \right) w_{b,i} \quad \text{.....(2.4)}$$

$$B_w^j = \bar{B} + U_b b_{w_b}^j ; 0 \le j \le m-1 \quad \text{.....(2.5)}$$

2.1.4 Aging Simulation

First obtain the 3D shape and the texture for an image at age x , A_{new}^{x} and B_{new}^{x} respectively. After obtaining 3D shape and texture we project them to reduced spaces and b_{new}^{x} respectively. Then the next step is to calculate weighting vector, w_a , for a reduced space a_{new}^{x} that generates the closest possible weighted sum of the shapes at age x as

$$\hat{w}_a = \arg\min_{c_- \le w_a \le c_+} \| a_{new}^{x} - a_{w_a}^{x} \|^2 + r_a \| w_a \|^2 \quad \text{.....(2.6)}$$

Where r_a is the weight of a regularizer which is used to handle the multiple solutions or when the solution has a large condition number. We constrain each element of weight vector, $w_{a,i}$,within $[c_-, c_+]$ to avoid strong domination by a few shape basis vectors.

Now by using $\hat{w}_a$, we can calculate age-adjusted shape at age y by carrying $\hat{w}_a$ over to the shapes at age y and transforming the shape descriptor back to the original shape space as

$$A_{new}^{y} = A_{\hat{w}_a}^{y} = \bar{A} + U_a a_{\hat{w}_a}^{y} \quad \text{.....(2.7)}$$

The same procedure is used for the texture simulation process by calculating $\hat{w}_b$ as below:

$$\hat{w}_b = \arg\min_{c_- \le w_b \le c_+} \| b_{new}^{x} - b_{w_b}^{x} \|^2 + r_b \| w_b \|^2 \quad \text{.....(2.8)}$$

And then projecting the $\hat{w}_b$ to the target age y followed by the back projection to get

$$B_{new}^{y} = B_{\hat{w}_b}^{y} = \bar{B} + U_b b_{\hat{w}_b}^{y} \quad \text{.....(2.9)}$$

The aging simulation process is illustrated in Fig2.1 given below.

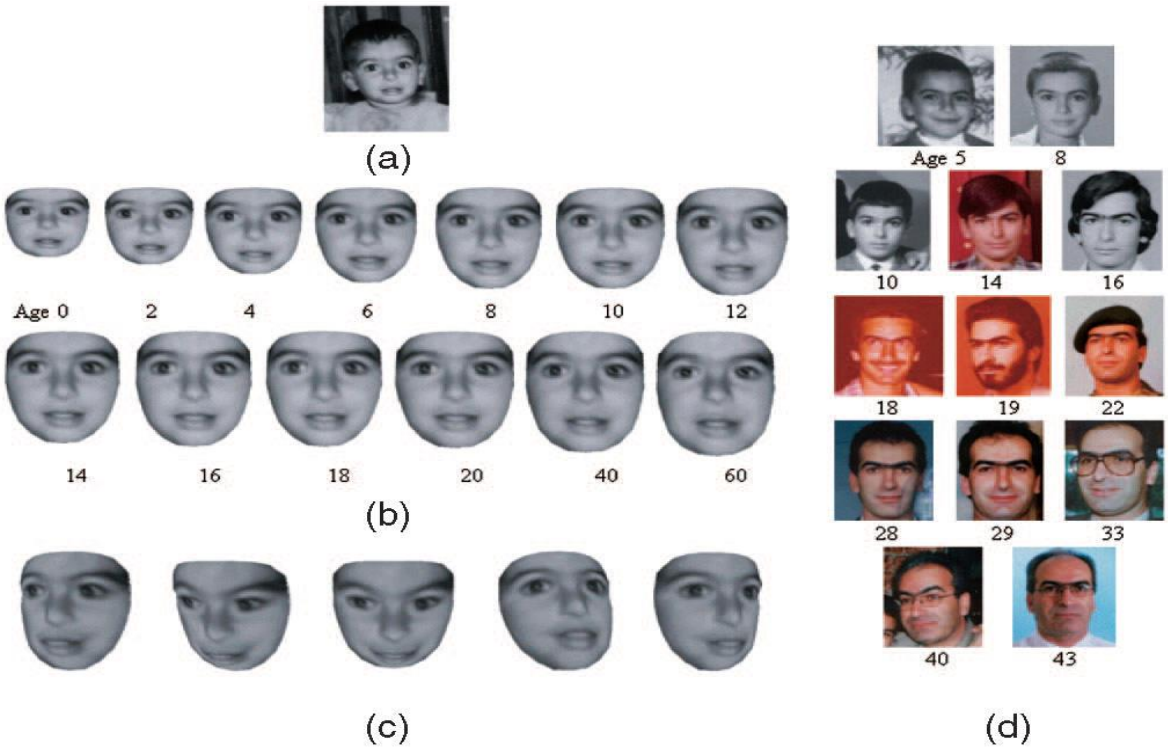


Figure 2.1: Aging simulation Example

(a) Given image at age 2

(b) Aging-simulated images 0 to 60.

(c) Face images at five different poses from the aging-simulated image at age 20.

(d) Ground truth images from database

2.2 Discriminative Approaches

Before discussing discriminative approaches in detail there are some limitation of generative approaches which leads to the concept of discriminative method for age invariant face recognition.

2.2.1 Limitations of Generative Approaches

(i) There is difficulty in face models construction in generative methods and when training sample size is limited then these approaches are not suitable of representing the aging process accurately.

(ii) For construction of the aging model, strong parametric assumptions are needed, which lead to unrealistic synthesis results and high complexity in computation.

(iii) Also in construction process of aging model some additional information is need such as location of landmarks on each face and true ages of the training sample images.

(iv) Further constraint is that images used in aging model should be captured under frontal pose, normal illumination, and neutral expression etc.

2.2.2 Discriminative Methods History and Basic

Therefore to overcome the above mentioned problems, approaches based on discriminative models are popular in aging problem. In these approaches robust feature descriptors and discriminative learning methods are usually applied to reduce the gap between face images collected at different ages.

For example good works of discriminative models is purposed by Ling et al [23], in which gradient orientation pyramid (GOP) is used for feature representation, and support vector machine is used for verifying faces across age progression.

More recently Z.Li, U.Park and A.jain [17], presented a discriminative age-invariant face recognition framework by combining multiple local features, such as local binary pattern (LBP) and SIFT. It is capable of handling aging variations, and can also handle intra-user variations such as pose, illumination, expression etc. This model consists of two components: densely sampled local description and multi-feature discriminant analysis (MFDA).

2.2.2.1 Densely Sampled Local Feature Description

Face images can be represented by local features effectively at diverse scales and orientations Local features representations are more robust to illumination and geometric distortions

Hence, local image descriptor techniques are used for face representation in which the each face image is divided into a set of overlapping patches and then local image descriptors is applied to each patch. Then features from these patches are extracted which further combined together to form a feature vector which has large dimensionality.

For example for a face image of $U \times V$ size, we can divide it into a set of $r \times r$ overlapping patches that overlap by $'s'$ pixels. Then the number of horizontal patches (H), vertical patches (W) are given as:

$$H = (V - r)/s + 1 \quad \text{.....(2.10)}$$

$$W = (U - r)/s + 1 \quad \text{.....(2.11)}$$

For each patch of size $H \times W$, a d–dimensional feature vector is calculated. Then these each patch feature vectors is combined to form a single $H \times W \times d$ -dimensional feature vector for a given face image.

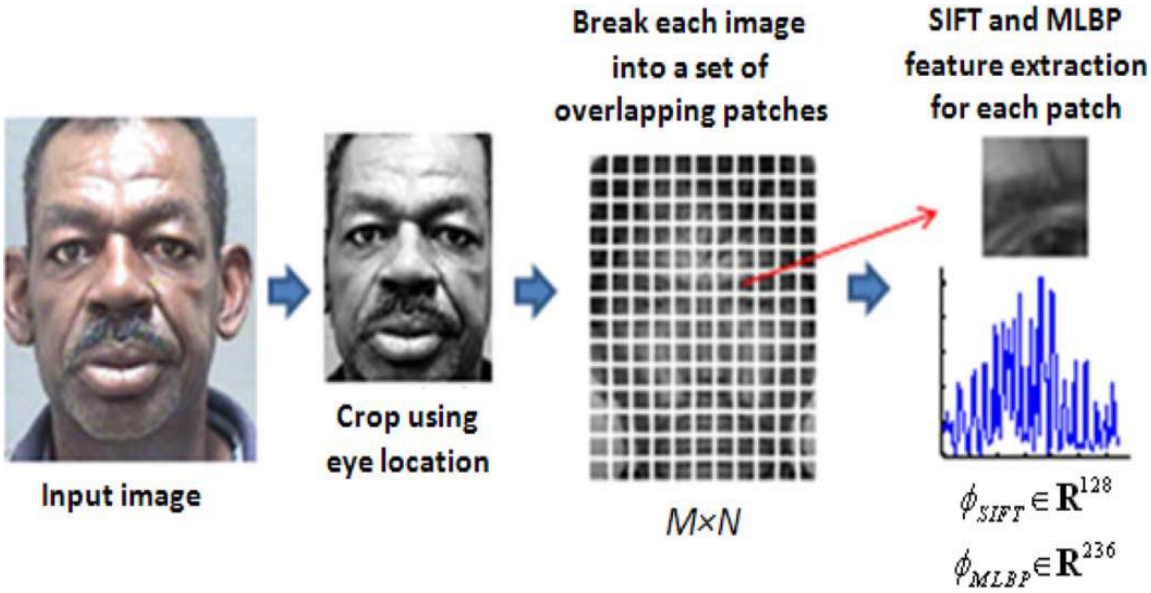


Figure 2.2: Local feature representation of a face image

Here SIFT [13], is used for spatial location and orientation within image patch, and then histogram is calculated in where each bin show orientation and spatial location.

Further extended version of one more local feature descriptor namely LBP is MLBP [32], has been used to describe the face at multiple scales, by computing the LBP descriptors computed at four different radii $\{1,3,5,7\}$.

Each face is first normalized to 150×200 pixels, and then entire face is divided into either 88 overlapping patches where each patch has size of 32×32 or 408 overlapping patches where each patch has size of 16×16. Then SIFT and MLBP features are extracted from each patch which is then further combined to form a single feature vector.

Thus now each patch is represented by a 128-dimensional SIFT feature vector or a 236-dimensional MLBP feature vector. Here the resulting feature dimensionality is very high and thus it is necessary to reduce the dimensionality. A direct approach can be to apply LDA (Linear Discriminant Analysis) to the SIFT and MLBP features separately and then the results from classifier based on SIFT and the other classifier based on MLBP can be fused. But, this straightforward approach has some disadvantages.

Thus to overcome these limitations of LDA, the multi-feature discriminant analysis (MFDA) framework is purposed. It is an extension and improvement in the LDA by using multiple features combined with two different random sampling methods in feature and sample spaces, as explained in the following sections.

2.2.2.2 Multi-Feature Discriminant Analysis (MFDA)

The LDA [33], is very much popular discriminant analysis method for face recognition.

The LDA uses the within-class scatter matrix and the between-class scatter matrix concept to define a function which is further used to measure the class separability. The within-class and between-class scatter matrices are defined as:

$$S_w = \sum_{i=1}^{c} \sum_{X_j \in C_i} \left(X_j - \mu_i \right) \left(X_j - \mu_i \right)^T \quad \text{.....(2.12)}$$

$$S_b = \sum_{i=1}^{c} \left(\mu - \mu_i \right) \left(\mu - \mu_i \right)^T \quad \text{.....(2.13)}$$

Where μ_i is mean of the class C_i, μ is defined as the total mean and c gives the number of classes. The main aim of LDA is to find the optimal projection W_{opt}, which is calculated for the maximum ratio of the determinant of the between-class matrix to that of the within-class matrix, and can be defined as:

$$W_{opt} = \arg\max_{w} \frac{|W^T S_b W|}{|W^T S_w W|} \quad \text{.....(2.14)}$$

Mathematically, it is equivalent to computing the leading eigenvectors of $S_w^{-1} S_b$.

If we directly use the LDA then the following problems occurs:

(i) When there is small size of the aging training data and the feature vector space has very high dimensionality then the accuracy and stability of S_w gets drastically reduced.

(ii) If only class means are used for calculating S_b, then LDA fails to capture the boundary structure of the classes

Now to overcome the above problems and to improve performance of LDA, random sampling techniques are used. In literature there are two popular random sampling methods:

i) Random subspace [34] and ii) Bagging [35]

In the random subspace method, we randomly sampled the feature space to construct multiple classifiers. Then the decisions of these individual classifiers are fused to gives the final decision to make the improvement in classification.

In the bagging method, we randomly sampled the training set to generate multiple training subsets. Then for each generated training subset, we construct a classifier and the outputs of these multiple classifiers are combined.

Now the combination of random subspace and bagging techniques, a MFDA framework is developed which is based on random sampling as shown below:

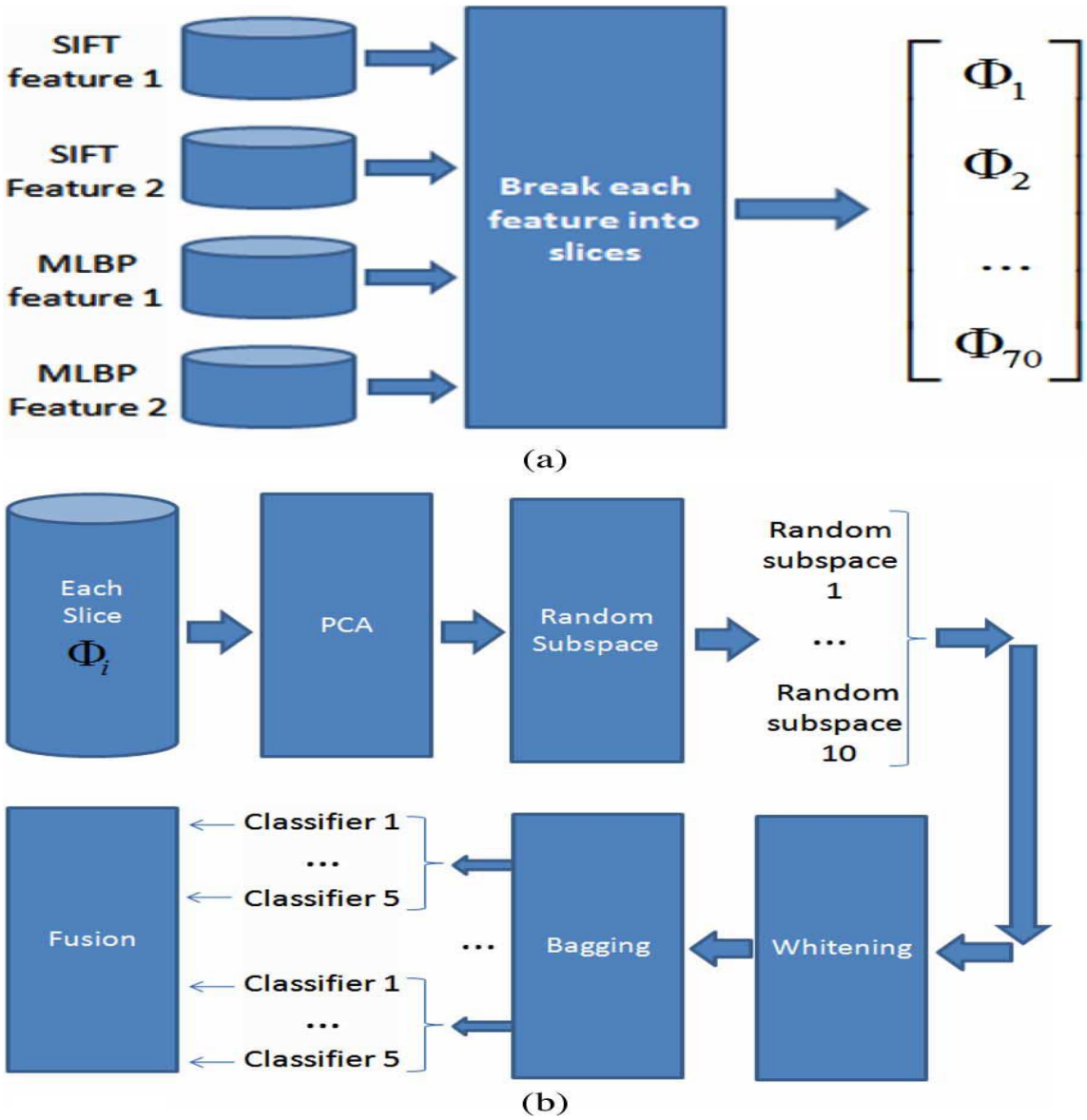


Figure 2.3: Block diagram of MFDA
(a) Breaking local features into slices.
(b) Each slice for fusion is train by 50 different classifiers.

Chapter 3

PURPOSED APPROACH FOR AGE INVARIANT FACE RECOGNITION

This chapter gives a brief description of purposed method for age invariant face recognition. As discussed in problem statement of this thesis in chapter 1, SIFT and Gabor Wavelet transform are used as image representation for the faces. Before that in preprocessing step AAM model is used for pose correction. For classification of test images SVM is used.

3.1 Flow Chart for Purposed Method

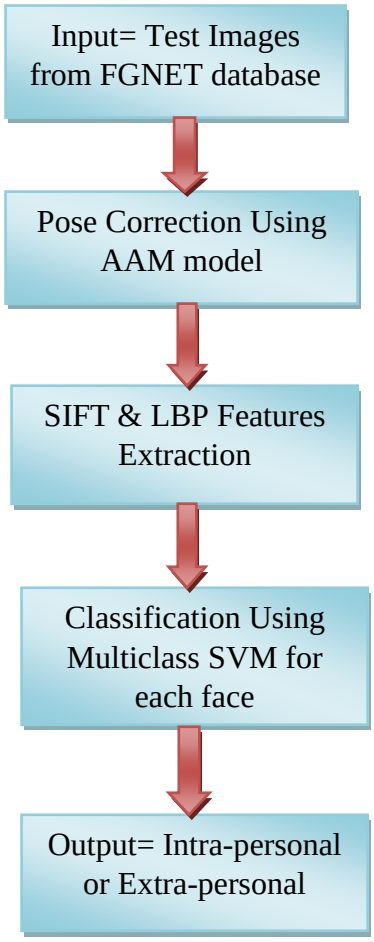


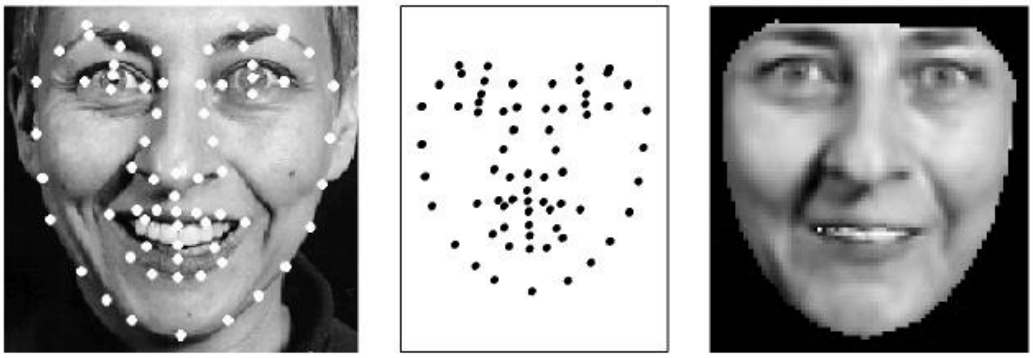
Figure 3.1: Purposed method block diagram

3.2 Pose Correction Using AAM Model

As to formulate a model to "interpret" face images has received considerable attention [36][37]. To this context the first well known approach was ASM[38] which was fast, simplistic but not robust when new images are introduced and further it does not incorporate gray-level information(i.e., texture). Thus to take advantage of all the available information AAM was introduced by Edwards, Cootes and Taylor [39] for matching a statistical model of appearance to images.

This model is actually a combination of model of shape variation and texture variation. Here texture means intensities pattern across an image patch. Model is then build by providing a set of training images, together with coordinates of landmarks that defines the main features in all of the images.

Then these landmarks points are represented as vector X and PCA is performed and build a statistical shape model [38]. Then warping of ach training image is done to match with mean shape to obtain 'shape free patch' as shown below in figure.



(a) Labelled Image (b) Points (c) Shape-free Patch

Figure 3.2: Labeled training image gives points with shape-free patch

Further eigen-analysis is used for texture model and in final step correlations between shape and texture model are learned to obtain combined appearance model.

Now for matching to new images this algorithm uses the optimization process in which difference between the current estimate of appearance and the target image is used in the least square sense technique.

3.3 Scale-Invariant Feature Transform (SIFT)

For "feature description" of face interesting points from the face can be extracted. When test image with multiple faces is given these descriptors can be used to identify the face. For effective recognition the extracted features should be detectable under various varying conditions. They should be scale and illumination invariant and should work under noisy situations.

SIFT algorithm published by David Lowe [13] can robustly identify objects as SIFT feature descriptor has advantages such as scale invariability, invariant to orientation, and partially invariant to affine distortion and illumination invariability.

From a database of training image these descriptors are extracted and saved. These descriptors are called features now on onwards. For testing purpose when given an image, features are extracted from the image and these are compared with saved feature set. Due to various reasons negative matches can also be found they can be eliminated with the aid of some post processing.

SIFT algorithm has following steps:

3.3.1 Scale-space Extrema Detection

This stage is used for extracting key-points used in SIFT framework. The image $I(x,y)$ is smoothed using the Gaussian filter. Filter with different scales is used. It can be represented as,

$$L(x,y,k\sigma) = G(x,y,k\sigma) * I(x,y) \quad \text{.....(3.1)}$$

Here $G(x,y,k\sigma)$ is Gaussian convolution kernel which can be expressed as

$$G(x,y,\sigma) = 1/2\pi\sigma \exp\left(\frac{x^2+y^2}{2\sigma^2}\right) \quad \text{.....(3.2)}$$

And $L(x,y,k\sigma)$ is caller spatial scale images, σ is scale factor and with the change of σ scales images are obtained.

Now Difference of Gaussian-blurred images (DOG) at multiple scales is obtained as:

$$D(x,y,\sigma) = L(x,y,k_i\sigma) - L(x,y,k_j\sigma) \quad \text{.....(3.3)}$$

Hence $D(x,y,\sigma)$ can be interpreted as difference of images which are convolved with Gaussian filter with scale $k_i\sigma$ and $k_j\sigma$ respectively.

The maxima/minima among different scales of DOG images are considered as key points. To get key points, around a pixel a small neighborhood is considered at all scales. If there are three different scales and if we assume a 3x3 neighborhood there would be 26 pixel values to compare. If the considering pixel is maxima/minima among them then it is considered as key point.

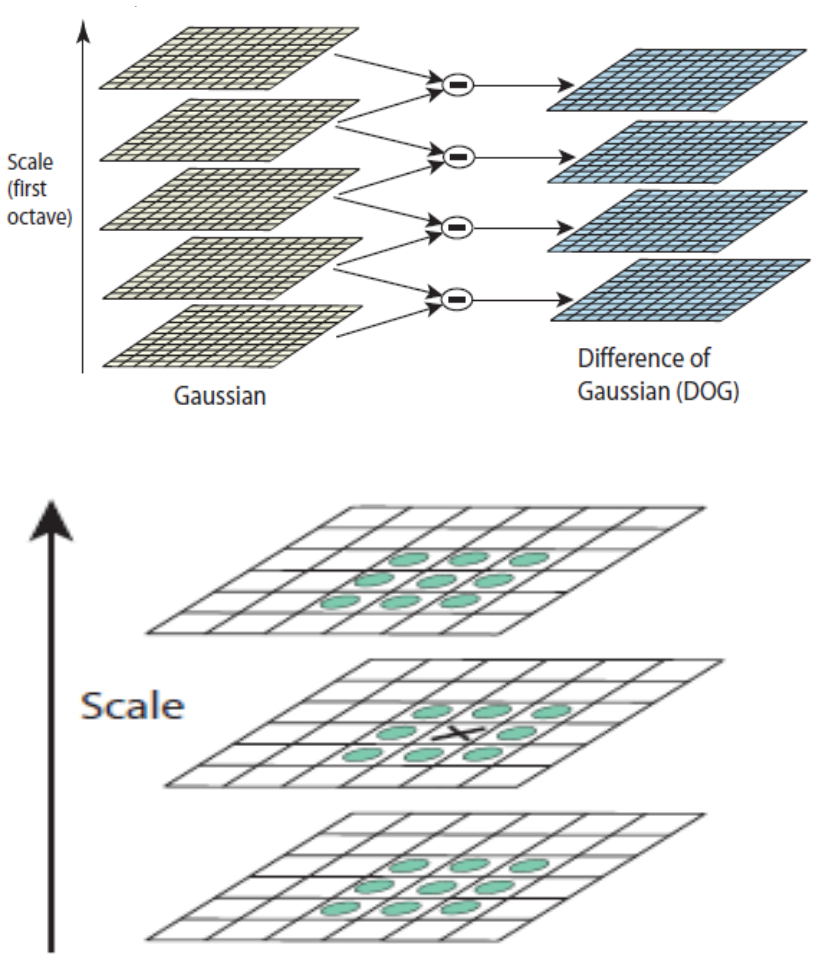


Figure 3.3: Key-point Detection

3.3.2 Key-point Localization

Scale-space extrema detection results in lots of key points, out of which some are sensitive to noise or have no edge effect. Thus in this step these low contrast key-points can be rejected. Now this has following stages:

(i) Interpolation of nearby data for accurate position

To determine accurate position of each candidate key-point interpolation of nearby data is used. The interpolation is done using the quadratic Taylor expansion of the DOG image, with the candidate key-point as the origin. This Taylor expansion can be written as:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad \text{.....(3.4)}$$

At the candidate key-point D and its derivatives are evaluated, $X = (x, y, \sigma)$ is the offset from this point. By taking the derivative of the above expression with respect to X and making it equal to zero, the extremum location X can be found. A threshold is used to decide the aureate position of key point. Offset X values greater than threshold implies the extremum is closer to some other key point candidate. Then the above process is repeat with respect to that point. The present key point candidate is considered and the accurate position is calculated by adding the Offset X , when the offset values are less than threshold. The threshold is taken as 0.5.

(ii) Discarding Low-contrast Key-points

The low contrast key points are also unwanted. At the offset X ,second-order Taylor expansion $D(X)$ is computed. This value is used to eliminated those key points with low contrast. They are discarded if a value of less than 0.03 occurs.

(iii) Eliminating edge responses

Such a point has large principal curvature across the edge but a small one in the perpendicular direction. The principal curvatures can be calculated from a Hessian function:

$$H = \begin{pmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{pmatrix}$$

The eigen values (α, β) of H are proportional to the principal curvatures, so two eigen values shouldn't differ too much this can be explained below:

Trace of H is sum of diagonal values given as:

$$Tr(H) = D_{xx} + D_{yy} \quad \text{.....(3.5)}$$

And determinant of H is given as:

$$Det(H) = D_{xx}D_{yy} - D_{xy}{}^2 \quad \text{.....(3.6)}$$

And $R = \frac{Tr(H)^2}{Det(H)} = \frac{(r+1)^2}{r}$ where $r = \frac{\alpha}{\beta}$ .Thus R is minimum if eigen values equal to each other and will be higher if difference between them is larger. Now for some threshold say r_{th} R is calculated and candidate key-point is rejected if $R > \frac{(r_{th}+1)^2}{r_{th}}$

3.3.3 Orientation Assignment

In this step, after position and scale of the key-point to achieve rotation invariance magnitude and direction of key-point is assigned. For example:

For an image sample $L(x,y)$ at scale σ, the gradient magnitude, $m(x,y)$, and orientation, $\theta(x,y)$, are precomputed using pixel differences:

$$m(x,y)=\sqrt{\left\{\left[L(x+1,y)-L(x-1,y)\right]^2+\left[L(x,y+1)-L(x,y-1)\right]^2\right\}} \quad \text{.....(3.7)}$$

$$\theta(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \quad \text{.....(3.8)}$$

In Gaussian-blurred image, for every pixel in a neighborhood region of key-point the magnitude and orientation for the gradient are calculated.

A histogram is formed by quantizing the orientations into 36 bins, with each bin covering 10 degrees. The peak in the histogram corresponds to dominant orientations. Once the histogram is filled, the orientations corresponding to the highest peak and local peaks that are within 80% of the highest peaks are assigned to the key-point. For the same scale and location, there can be multiple key-points with different orientations.

3.3.4 Key-point Descriptor

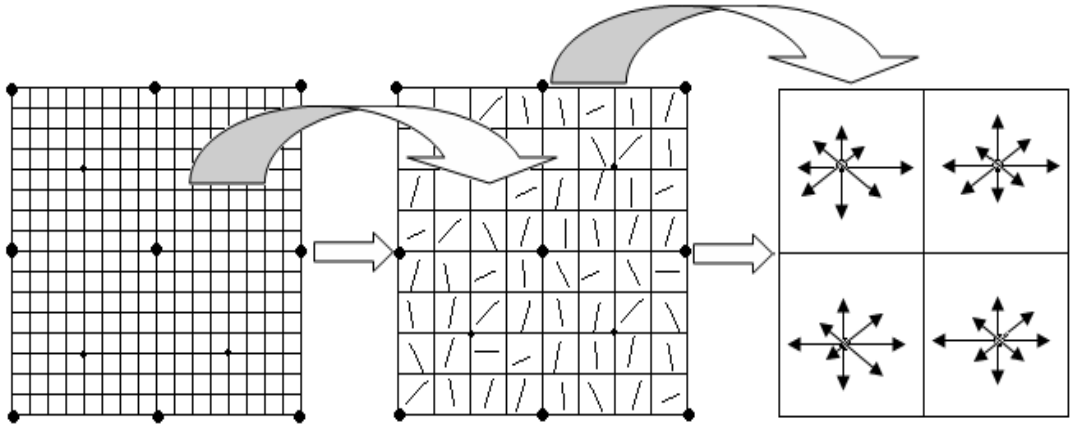
To key make the descriptors illumination variant the following process is followed.

Around a key point is considered a 16x16 neighborhood is considered, and this 16x16 cell is divided into 16 blocks of 4x4. In this 4x4 blocks according to the orientation they are divided into 8 bins by dividing 0 to $360^0$with a gap of 45^0.

The same thing is done for all the 16, 4x4 blocks. So a total of 16*8=128, i.e., a descriptor of length 128 to represent a key point is found

The feature description is obtained by connecting the direction descriptions of all subfields; the total of the direction descriptions is 16, so the length of the feature description is 128=16×8;

In order to ensure the illumination invariance, the feature description should be normalized;



(a) Neighborhood (b) Direction (c) Descriptor of points

Figure 3.4: Feature Description of Key-points

3.4 Local Binary Pattern (LBP)

The LBP operator [11] is used as texture descriptors in computer vision applications. We are using it in age invariant face recognition because of highly discriminative nature and have advantages, such as invariance to gray-level changes as we have texture variations in our FGNET database.

.

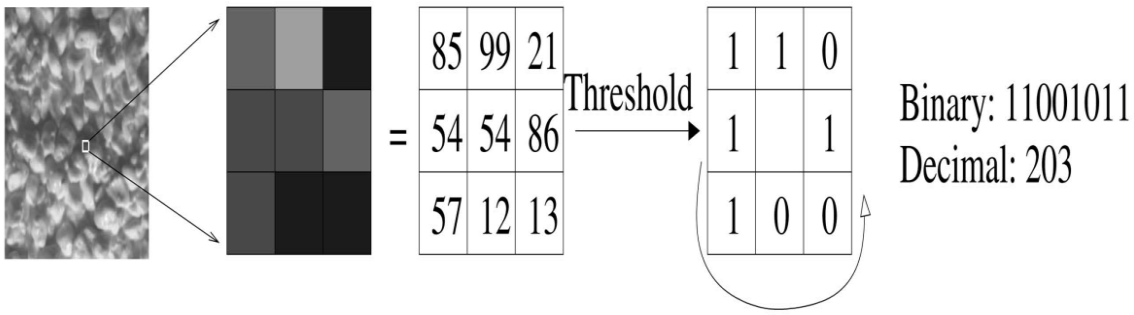


Figure 3.5: Basic LBP Operation

The LBP feature is extracted as follows:

- First divide the window which we are examining into cells (e.g. 16x16 pixels for each cell).
- Now in a cell formed for each pixel we are comparing the pixel to its 8 neighbors, pixels are follow along circle as clockwise or anticlockwise.
- When the center pixel's value is greater than the neighbor's value, we write "1". Otherwise, "0". This gives an 8-digit binary number which is further converted to decimal number.
- Now histogram is computed, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center).
- Then normalize the histogram.
- In final step concatenation of normalized histograms of all cells is done which gives the feature vector for the examined window.

The feature vector further processed using the SVM.

3.4.1 Face Description with LBP

FGNET face images can be considered as a composition of micro-patterns which having texture variations along age progression, thus it can be effectively detected by the LBP operator.

The LBP histograms extracted for each sub-region are concatenated into a single, spatially enhanced feature histogram feature vector which is defined as:

$$H_{i,j} = \sum_{x,y} I\left(f_l\left(x,y\right) = i \right) I\left(x,y\right) \in R_j \quad \text{.....(3.9)}$$

where i=0,……L-1 and j=0,…..M-1.

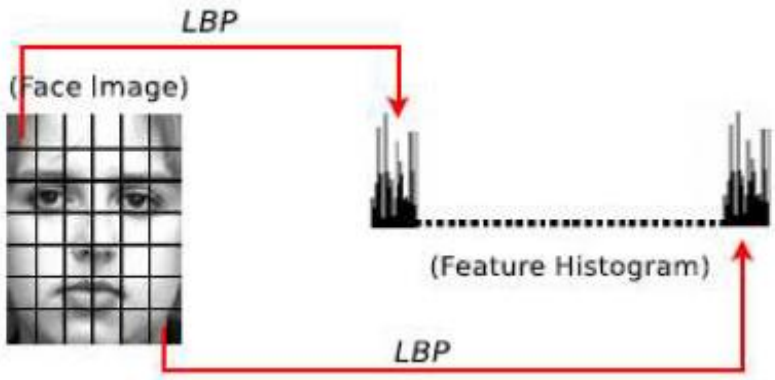


Figure 3.6: LBP representation of face.

3.5 Classification using Multiclass SVM

Support vector machine (SVM) [15] is an example of supervised learning classification in which there are predetermined classes. Statistical processes (i.e. based on an a priori knowledge of probability distribution functions) or distribution free processes can be used to extract class descriptors.

The classifier is first trained using a set of training examples. The training examples are pre marked as belonging to one of the two categories and based on these examples, the SVM classifier builds a model that assigns new examples (here images to be tested) their suitable classes. The training examples are represented as points in space and are mapped such that there is a clear gap which divides the examples belonging to separate classes. The new examples are then mapped into the same space by analyzing to which of the two classes they suit better.

Thus after features representation the next task is verification in which one must determine whether two images come from the same person. Therefore we model face verification as a two class classification problem. Given an input image pair $I1$ and $I2$, the task is to assign the pair as either **intra-personal** (i.e. $I1$ and $I2$ from the same people) or **extra-personal** (i.e. $I1$ and $I2$ from different individuals).Thus the case becomes linearly separable date as binary classification.

3.5.1 Linear SVM

Here the goal is written as- To find the hyper-plane (i.e. decision boundary) linearly separating our classes. Our boundary will have equation: $W^T X + b = 0$.

Anything above the decision boundary should have label 1.

For X_i if $W^T X_i + b > 0$ will have corresponding $y_i = 1$

Similarly, anything below the decision boundary should have label -1

For X_i if $W^T X_i + b < 0$ will have corresponding $y_i = -1$.

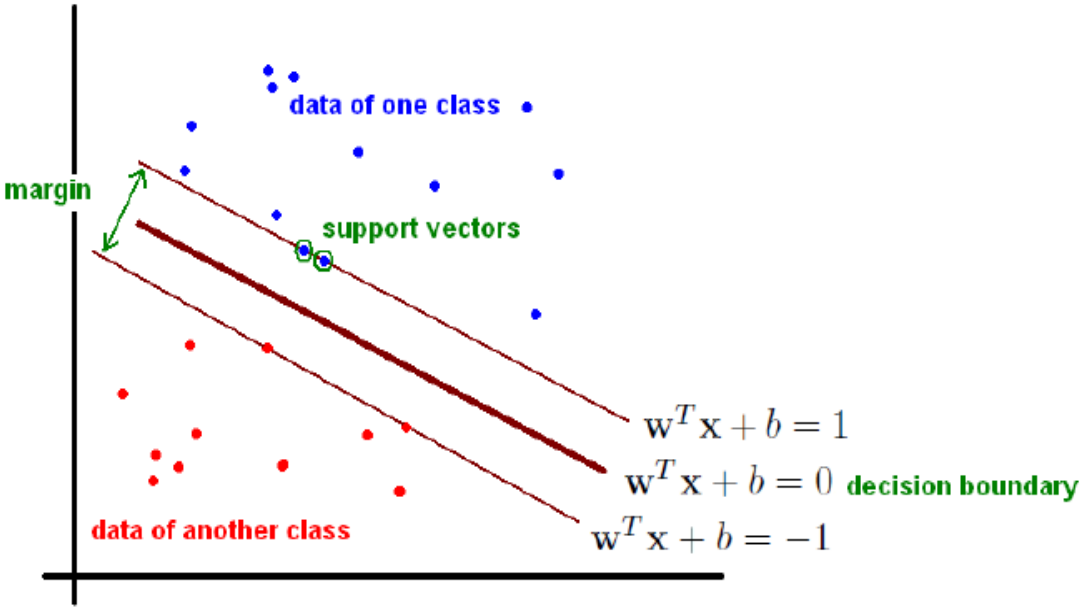


Figure 3.7: SVM hyper-plane with support vectors

3.5.2 Nonlinear SVM

Nonlinear SVM is applied in cases where the data sets have nonlinear decision boundaries. The trick applied here is to transform the data from its original coordinate space to a new space where a linear decision boundary can be used to separate the instances in the transformed space. Transformation of space may suffer from dimensionality problem which is associated with high dimensional data. Moreover, it is not always easy to find out what mapping must be used to ensure that a linear decision boundary can be constructed in the transformed space. This problem can be solved by 'Kernel' trick. It is a method for computing similarity in the transformed space using the original attribute set. The measure of similarity used is the dot product of the two input vectors. The similarity function computed in the original attribute space is called the 'kernel function'. The kernel trick eliminates the need to know the exact form of the mapping function.

3.5.3 Multiclass SVM

Multi class SVM classifies data into more than one class. For multiclass case, we transform the problem into multiple binary classification problems. Here in this thesis we are using Multiclass SVM for classification as there is multiple users with each having intra-class variation and inter-class variation between other so we are taking each of them as separate class. There are following types of multiclass SVM:

(i) One-against-all

For the given N-class problems where $N > 2$, N two-class SVM classifiers may be constructed to finally build a multiclass classifier. When we train the i^{th} SVM classifier the i^{th} class samples are considered as positive examples while all the rest samples of other classes are considered as negative examples. For the recognition part, one can give a test sample as input to N SVMs classifiers and output class is assigned according to the maximum of N classifier output. The drawback of this method is that when the training samples of classes is large then training becomes very difficult.

(ii) One-against-one

Here one can build $N(N-1)/2$ two-class classifiers. All classifiers used in SVM are the binary pair wise combinations of the N classes. While training the each classifier, the data points of the first class are considered as positive examples and the data points of the second class are considered as negative examples. When we combine these $N\ (N-1)\ /2$ classifiers, we adopt Maximum Wins algorithm to find the output class by voting the classes according to the results of each of the classifier and then finding the most voted class. The disadvantage of this method is that every test sample has to be presented to the large number of classifiers $N\ (N-1)\ /2$. This results in faster training but slower testing, especially when the number of the classes in the problem is big.

So to construct a multi-class classifier, one way is to construct N separate SVMs. Here the N^{th} model is trained from the N^{th} class as the positive examples and all the remaining $N-1$classes are treated as the negative examples.

The other way to build a multi-class classifier is to decompose an N-class problem into a number of two-class problems, in which one-against-all the most commonly used implementations [2]. The results from the multiple classifiers are used in the final decision as $D_i(\mathrm{x}) \geq -1$.

Let the maximum margin output that separates i[th] class from the remaining classes, is given by

$$D_i(\mathrm{x}) = w_i^T \Phi(\mathrm{x}) + \mathrm{b}_i \quad \text{.....(3.10)}$$

Where w_i is the l-D vector, $\Phi(x)$ is known as the mapping function that maps the data points into the l -D feature space, and $\mathrm{b_i}$ is known bias [8]. The optimal hyper plane is selected for classification, and if it is possible to separate classes then the training data belonging to class satisfy $D_i(\mathrm{x}) = 0$ and for the remaining classes satisfy $D_i(\mathrm{x}) \leq 1$. If it is not possible to the data-set into classes then unbounded support vectors satisfy $|D_i(\mathrm{x})| = 1$ and bounded support vectors belonging to class satisfy $D_i(\mathrm{x}) \leq 1$ and those belonging to a class other than class satisfy $D_i(\mathrm{x}) \geq -1$. Data points from the different classes, *x* is classified into the class is given by:

$$\arg \max_{i=1,.....n} D_i(\mathrm{x}) \text{.....(3.11)}$$

How to effectively cast the multiclass problems is still an on-going issue.

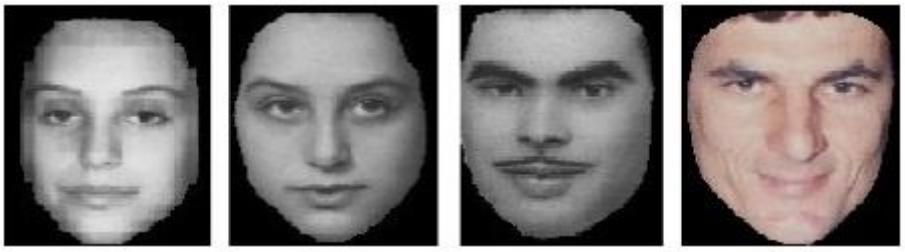
Chapter 4

RESULTS AND COMPARISON

4.1 Examples of Pose Correction Using AAM

As discussed in chapter 3 about AAM technique where we use 68 annonated landmarks point on face and then using them for pose correction are shown below and cropped face area using these points is also shown on which we further applying feature extraction technique.

(a): Original FGNET database images



(b): Pose Corrected FGNET database images

Figure 4.1: Pose Correction Using AAM model

4.2 SIFT Feature Extraction on FGNET Images

Now after pose correction SIFT & LBP features are extracted which are shown below:

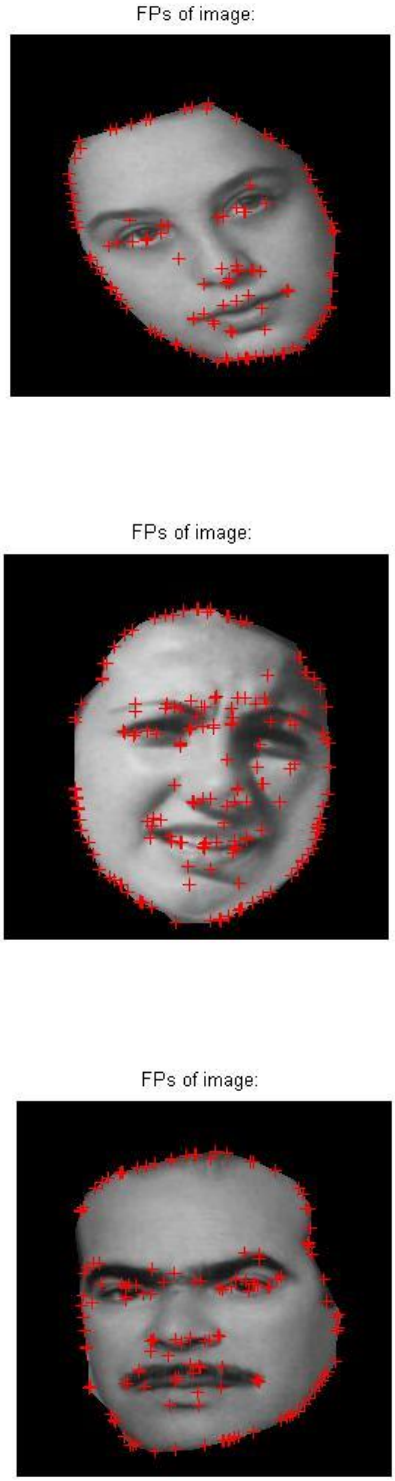


Figure 4.2: SIFT Feature Detection

4.3 Example of Test Image and Recognized Image

Figure 4.3: Test & Recognized Images

4.4 Performance Measures

To see performance of our purposed method we are calculating various performance parameters which tell how much our method is efficient in recognizing aging variation in face recognition and also we can compare purposed method with other pre-existed methods using these parameter. These parameters are: FAR, FRR, EER, ROC Curve, and CMC Curve.

4.4.1 FAR

False Acceptance (FA), which occurs when the system accepts an impostor face (in our case it is extra-personal pair which means input image pair are from the different subject), and the False Acceptance Ratio thus defined as below:

FAR = number of false acceptance / number of imposter face representations

In our experiment on FGNET database we get FAR=0.421983

4.4.2 FRR

False Rejection (FR), which occurs when the system refuses a true face (in our case it is intra-personal pair which means input image pair are from the same subject), and False Rejection Ration thus defined as below:

FRR = number of false rejection / number of true face representations

In our experiment on FGNET database we get FRR=0.422917

4.4.3 EER

Equal error rate (EER), defined as the error rate when a solution has the same CAR and CRR, is frequently used to measure verification performance.

In our experiment on FGNET database we get EER=0. 422450

4.4.4 ROC Curve

Receiver Operating Characteristic curve (ROC curve) of a verification system, expresses the quality of a 1:1 matcher. The ROC plots the False Accept Rate (FAR) of a 1:1 matcher versus the False Reject Rate (FRR).

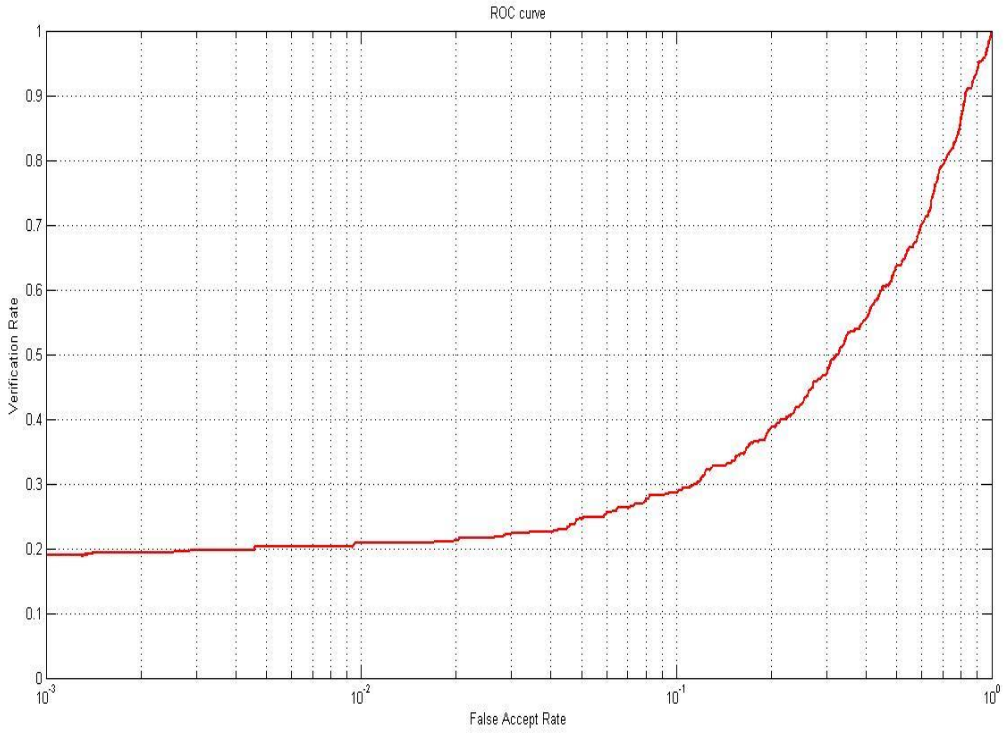


Figure 4.4: ROC Curve

4.4.5 CMC Curve

The Cumulative Match Curve (CMC) is used as a measure of 1:m identification system performance. It judges the ranking capabilities of an identification system. To estimate the CMC, the match scores between a query sample and the *m* biometric samples in the database are sorted. The lower the rank of the genuine matching biometric in the enrollment database, the better the 1: m identification system.

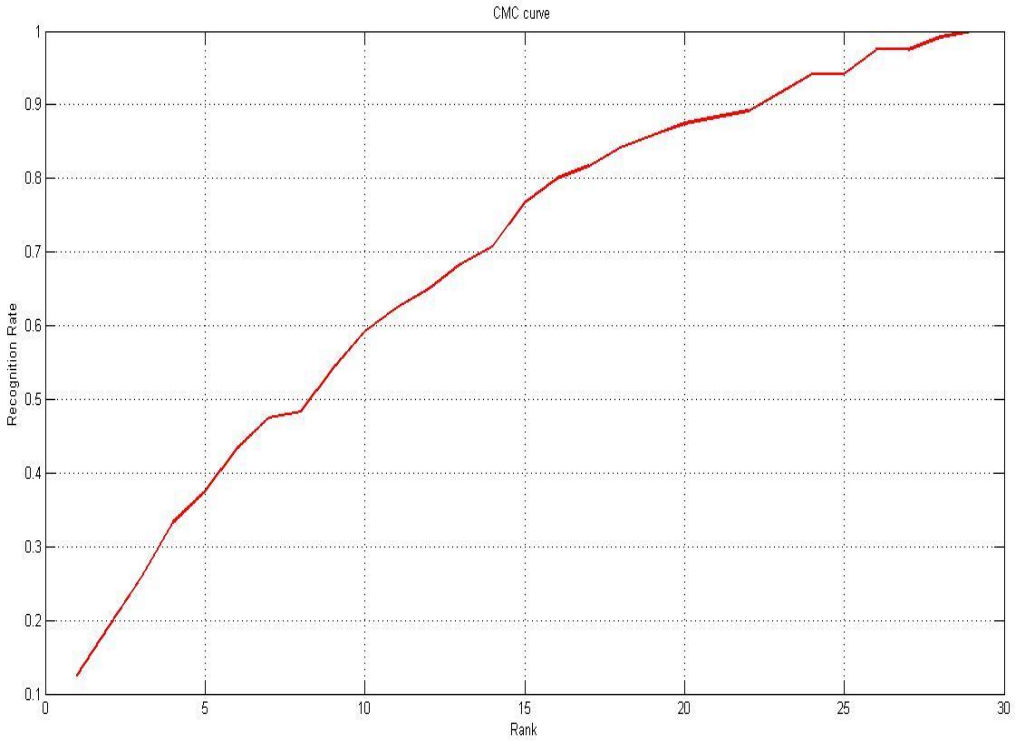


Figure 4.5: CMC Curve

4.5 Comparison of Purposed Approach with others

Table 4.1: Comparison of age invariant face recognition methods on FG-NET database

(Only Global Approaches included)

Author	Technique Used	Rank-1 Recognition Rate
Geng et al (2007) [22]	Learn aging pattern on concatenated PCA coefficients of shape and texture across a series of ages	38.1 %
Park et al (2010) [24]	Learn aging pattern based on PCA coefficients in separate 3D shape and texture spaces from the given 2D database.	37.4 %
Gayathri Mahalingam et al (2010) [25]	GMM + Graph Matching	50 %
Zhifeng Li et al (2011) [17]	Multi-feature discriminant analysis (MFDA) to process SIFT and MLBP-based local feature spaces of each face in a unified framework.	47.5 %
Diana Sungatullina et al (2013) [29]	Multiview Discriminative Learning	65.2%
Purposed method	Multiclass SVM+Fusion of SFIT and LBP based local feature with pose correction.	76.6%

Chapter 5

CONCLUSION AND FUTURE SCOPE OF WORK

5.1 Conclusion

In this thesis we purposed a discriminative model with pose correction for age invariant face recognition. This method overcomes the problem in generative methods as in that there is need of a training set of subjects with minimum variations in pose and illuminations.

Further our Rank 1 recognition accuracy (for age > 18) comes to be larger than other reported method in literature as shown in table of comparison in result and comparison chapter.

5.2 Future Work

As our purposed method is best suited for aging process (age > 18) and shows good results. But for the age <18 our method performance is slow as it growth and development stage where facial appearance gets many changes thus it is difficult to recognize in age range <18.

Therefore in future more discriminative method which is capable of recognizing in age <18 can be studied and further also other pose correction methods can also be studied.

Further AAM model we are using sometimes fails for much pose variations as automatic landmarks detection fails in that subject thus in future pose correction method can also be studied which is more invariant to pose variations than AAM.