Factorial Discriminant Analysis for 3D Face Recognition System using SVM Classifier

P. S. Hiremath
Department of P. G. Studies and Research in Computer Science, Gulbarga University, Gulbarga-585106 Karnataka, India

Manjunatha Hiremath
Department of P. G. Studies and Research in Computer Science, Gulbarga University, Gulbarga-585106 Karnataka, India

ABSTRACT
In past three decades, two dimensional face recognition has been one of the most important and attractive research areas in computer vision. However, pose and illumination variations in the face images have been the dominant factors which have hindered many practical applications of two dimensional face recognition systems. In order to overcome these limitations and inherent drawbacks of two dimensional recognition system, many researchers have turned to 3D or surface facial information which is now commonly believed to have the potential to achieve greater recognition accuracy than just 2D. In this paper, a novel 3D face recognition approach based on radon transform and factorial discriminant analysis using SVM is proposed. The proposed approach has been tested on three publicly available databases, namely, Bosphorus, Texas and CASIA 3D face databases. The experimental results yielded 99.70% recognition accuracy using SVM classifier.

General Terms
Image Processing, Pattern Recognition, Classification

Keywords
Radon transform, Factorial Discriminant Analysis, KNN, SVM.

1. INTRODUCTION
The human face is considered to be the most commonly used biometric trait; we recognize each other and, in many cases, establish our identities based on faces. Hence, it has become a standard practice to incorporate face photographs in various tokens of authentication such as ID cards, passports, and driver’s licenses. Face recognition can be defined as the process of establishing a person’s identity based on their facial characteristics. In its simplest form, the problem of face recognition involves comparing two face images and determining if they are of the same person. While humans seem to be adept in determining the similarity between two face images acquired under diverse conditions, the process of automated face recognition is beset with several challenges. Face images of a person may have variations in age, pose, illumination, and facial expressions as well exhibit changes in appearance due to make-up, facial hair, or accessories (e.g., sunglasses). Training a machine to recognize face images exhibiting such unconstrained intra-user variations is a difficult task, especially since the exact cognitive and neural processes involved in humans for the task of face recognition (and recollection) is still not completely known. Techniques for automated face recognition have been developed for the purpose of person recognition from still 2-dimensional (2D) images, video (a sequence of 2D images), and 3D range (depth) images. Efforts in acquiring the face biometric in a 3D format have resulted in the development of 3D face capture systems. There are two types of 3D face capture systems: one is based on laser scanning and the other is based on stereographic reconstruction.

It is generally regarded that laser scanners provide more accurate 3D face models, while stereographic cameras provide near real-time capture capability with slight loss in accuracy. 3D face models are usually represented as a polygonal mesh structure (e.g., triangular or rectangular) for computational efficiency. The format of data produced by a 3D face sensor may depend on the technology used to obtain an image. A survey of literature on the research work focusing on various potential problems and challenges in the three dimensional (3D) face recognition can be found in [1-9]. The most prominent method in current 3D face recognition system is to use 3D point clouds to represent faces. In point cloud-based approaches, raw 3D point sets are used to register faces, and then features are extracted from registered faces[10]. Many recognition systems use depth or range images where, each pixel value represents the distance from the sensor to the facial surface. The 3D face recognition is then formulated as a problem of dimensionality reduction for planar images. The principal component analysis (PCA) based “Eigenfaces” can be used for dimensionality reduction [11]. The basis vectors are however typically holistic and of global support. The PCA can be combined with the linear discriminant analysis (LDA) to form “Fisherfaces” with enhanced class seperability properties [12,13]. Hiremath and Manjunath [28] have employed radon transform, PCA and LDA data analysis approach for 3D face recognition, which is further extended to symbolic data analysis framework [30] yielding recognition accuracy of 95.30% and 99.50%, respectively.

In this paper, the objective is to propose a new 3D face recognition method based on radon transform and factorial discriminant analysis using KNN and SVM classifier, which are applied on 3D facial range images. The experimentation is done using three publicly available databases, namely, Bosphorus, Texas and CASIA 3D face database. The experimental results demonstrate the effectiveness of the proposed method.

2. MATERIALS AND METHODS
For purpose of experimentation of the proposed methodology, the face images drawn from the following 3D face databases are considered: (i) Bosphorus 3D face database, (ii) Texas 3D face database, (iii) CASIA 3D face database.

2.1 Bosphorus 3D Face Database
The Bosphorus 3D face database consists of 105 subjects in various poses, expressions and occlusion conditions. The 18 subjects have beard/moustache and the 15 subjects have hair. The majority of the subjects are aged between 25 and 35. There are 60 men and 45 women in total, and most of the subjects are Caucasian. Two types of expressions have been considered in the Bosphorus database. In the first set, the expressions are based on action units. In the second set, facial expressions corresponding to certain emotional expressions are collected. These are: happiness, surprise, fear, sadness, anger and disgust. The facial data are acquired using Inspeck Mega Capturor II 3D, which is a commercial structured-light based
3D digitizer device. The sensor resolution in x, y & z (depth) dimensions are 0.3mm, 0.3mm and 0.4mm respectively, and colour texture images are high resolution (1600x1200 pixels). It is able to capture a face in less than a second. Subjects were made to sit at a distance of about 1.5 meters away from the 3D digitizer. A 1000W halogen lamp was used in a dark room to obtain homogeneous lighting. However, due to the strong lighting of this lamp and the device’s projector, usually specular reflections occur on the face. This does not only affect the texture image of the face but can also cause noise in the 3D data. To prevent it, a special powder which does not change the skin colour is applied to the subject’s face. Moreover, during acquisition, each subject wore a band to keep his/her hair above the forehead to prevent hair occlusion, and also to simplify the face segmentation task. The propriety software of the scanner is used for acquisition and 3D model reconstruction[19].

2.2 Texas 3D Face Database
The Texas 3D Face Recognition (Texas 3DFR) database is a collection of 1149 pairs of facial color and range images of 105 adult human subjects. These images were acquired using a stereo imaging system manufactured by 3Q Technologies (Atlanta, GA) at a very high spatial resolution of 0.32 mm along the x, y, and z dimensions. During each acquisition, the color and range images were captured simultaneously and thus the two are perfectly registered to each other. This large database of two 2D and 3D facial models was acquired at the company Advanced Digital Imaging Research (ADIR), LLC (Friendswood, TX), formerly a subsidiary of Iris International, Inc. (Chatsworth, CA), with assistance from research students and faculty from the Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin. This project was sponsored by the Advanced Technology Program of the National Institute of Standards and Technology (NIST). Texas 3DFR was created to develop and test 3D face recognition algorithms intended to operate in environments with co-operative subjects, wherein, the faces are imaged in a relatively fixed position and distance from the camera[16-18].

2.3 CASIA 3D Face Database
CASIA 3D Face Database consisting of 4624 scans of 123 persons using the non-contact 3D digitizer, Minolta Vivid 910. During building the database, not only the single variations of poses, but also expressions and illuminations are considered [20].

3. PROPOSED METHODOLOGY
The proposed methodology employs the following: (i) Radon transform(RT) and (ii) Factorial Discriminant Analysis (FDA), which are described in the following sections.

3.1 Radon Transform
The Radon Transform (RT) is a fundamental tool in many areas. The 3D Radon Transform is defined using 1D projections of a 3D object \( f(x,y,z) \) where these projections are obtained by integrating \( f(x,y,z) \) on a plane, whose orientation can be described by a unit vector \( \vec{a} \). Geometrically, the continuous 3D Radon transform maps a function \( f(x) \) into the set of its plane integrals in \( \mathbb{R}^3 \). Given a 3D function \( f(x,y,z) \) and a plane whose representation is given using the normal \( \vec{a} \) and the distance \( s \) of the plane from the origin, the 3D continuous Radon Transform of \( f(x,y,z) \) is defined by

\[
Rf(\vec{a}, s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x \sin \theta \cos \phi + y \sin \theta \sin \phi + z \cos \theta - s) dx dy dz
\]

where \( \vec{a} = [x, y, z]^T \),
\( \phi = \sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta \)^T,
and \( \delta \) is Dirac’s delta function defined by \( \delta(x) = 0, x \neq 0, \delta(x) = 1 \) for \( x = 0 \). The Radon transform maps the spatial domain \( (x,y,z) \) to the range domain \( (\vec{a}, s) \).

3.2 Factorial Discriminant Analysis (FDA)
Factorial Discriminant Analysis (FDA) is performed on a 3D face data set \( E = \{1,...,n\} \) on \( n \) individuals (training set) which are each characterized by a vector of \( p \) quantitative predictor variables \( Y = (Y_1,Y_2,...,Y_p) \). Each element \( k \) of \( E \) belongs to one of \( m \) classes \( \Pi_1,\Pi_2, ..., \Pi_m \), and their membership is known a priori. It is described by a nominal variable \( c \) on \( E \) with \( m \) categories; \( c(k) = i \) if \( k \) belongs to \( \Pi_i \) [21].

FDA deals with two aims: a descriptive one that consists in finding \( p \) new discriminant variables, also called canonical variables (with \( p \leq p \)), as linear combinations of the \( p \) predictors, such that the projections of the \( m \) classes are well separated in the space of canonical variables. The second objective of FDA is the decision-oriented one, which deals with the specification of a classification rule (both geometrical and probabilistic) in order to assign new individuals, for which the same variables \( Y(j = 1,..., p) \) have been observed to one of the given classes \( \Pi_i(i = 1,...,m) \).

FDA starts from a classical data matrix \( X = (x_{ij})_{n \times p} \) with \( x_{ij} = Y(j) \), where the \( j \)th column corresponds to the \( j \)th explicative variable \( Y_j \). Let \( C(c_{ik}) \) denote the \( n \times m \) indicator matrix associated to the classificatory variable \( c \) such that \( c_{ik} = 1 \) (or 0) iff \( c(k) = i \) (or not), i.e., if the \( k \)th individual belongs to the class \( \Pi_i \) (or not). The discriminant variables to be constructed should maximize the variance between the classes (in the training set \( E \)) and, at the same time, minimize the variance within the classes. Assuming \( X \) to be a centered matrix, the global empirical covariance matrix \( V \) of the data matrix \( X \) is given by

\[
V = \left( \frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \right)_{p \times p} = X' H X
\]

where \( H \) is the matrix of the weights of the individuals. Supposing all weights to be equal to \( \frac{1}{n} \), we get \( H = n^{-1} I_n \) (with \( I_n \) the identity matrix of dimension \( n \times n \)). Suppose that \( C \) is the sub-set of elements \( k \) of \( E \) (i.e., \( C \subset E \)), which belong to the class \( \Pi_i \), and \( n_i = |C| \) is the number of these elements. If the diagonal matrix \( Q \) of the weights \( n_i / n \) of the classes is denoted by

\[
Q = \text{diag} \left( \frac{n_1}{n}, ..., \frac{n_m}{n} \right) = C' H C,
\]

the centroids \( x_{Ci} \) of the \( m \) classes \( C_i \) given by:

\[
x_{Ci} = n_i^{-1} \sum_{x_i} x_i, \quad \text{for} \ i = 1,...,m,
\]
are just the rows of the matrix:

\[ \tilde{G} = Q^{-1}(C'HX) \]

With this notation, the $p \times p$ covariance matrix $B$ between the classes $C_1, \ldots, C_m$ is given by:

\[ B = \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij} x_{ij}^T = (X'HC)Q^{-1}(C'HX) \]

The between-class variance of the data vectors $x_1, \ldots, x_n$ is equal to $tr(B)$. Factorial discriminant analysis assumes that the underlying variance-covariance matrices in the $m$ classes $\Pi_i$ are all the same and therefore considers, as its estimate, the empirical variance-covariance matrix $W$ within the classes $C_1, \ldots, C_m$ is:

\[ W = \sum_{i=1}^{m} n_i W_i \]

which is obtained as a weighted sum of the empirical variance-covariance matrices $W_i$ of the classes $C_i$ given by:

\[ W_i = (n_i)^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T \]

$tr(W)$ is the within-class variance of the data vectors of $E$. The first step of a discriminant analysis is to look for a "discriminant axis" $\eta \in \mathbb{R}^p$ such that the $m$ groups $C_1, \ldots, C_m$ are distinguished as well as possible in the sense as to maximize the ratio of the between and within variances of the observed elements of the training set. Due to $V = W + B$, this solve the maximization problem:

\[ \max_{\eta \in \mathbb{R}^p} \frac{\eta^T B \eta}{\eta^T V \eta} \]

The optimum $\eta_i$ is obtained as solution of the eigen-equation:

\[ B \eta_i = \lambda \eta_i \]

$V\eta_i$ corresponding to the largest eigenvalue $\lambda_i$ of the matrix $V^{-1}B$, which represents a measure of the discriminating power of the first factorial axis $\eta_i$. The other factorial axes are obtained as the eigenvectors $\eta_2, \eta_3, \ldots$ associated to the next largest eigenvalues $\lambda_2 \geq \lambda_3 \geq \ldots$ of the matrix $V^{-1}B$ in decreasing order.

In the $s$-dimensional factorial plane, the $\alpha^k$ coordinate of the individual $k \in E$ is given by the $k^{th}$ element of the discriminant variable $Z_{\alpha} = X\eta_k$, for $\alpha = 1, \ldots, s$; here $s$ is the number of axes of the best discriminant subspace. Similarly, the projection of the centroids $\bar{x}_i$ on the factorial plane has, as its $\alpha^k$ coordinate, the $i^{th}$ element of the vector $(C'H\bar{X})^{-1}C'HZ_{\alpha}$.

3.3 Proposed Method

The proposed method comprises the following steps:

(i) Radon transform is applied to the input depth and intensity images of a 3D face, which yields binary images that are used to crop the facial areas in the corresponding images.

(ii) Factorial discriminant analysis is applied to the cropped facial images, to achieve dimensionality reduction and obtain subsampled feature vectors.

(iii) Lastly, the minimum distance classifier is used to perform face recognition based on subsampled feature vectors.

The Figure 1 shows the overview of proposed framework. The algorithms of the training phase and the testing phase of the proposed method are given below:

**Algorithm 1: Training Phase**

1. Input the 3D face image $I_i$ from the training set containing 3D face data set $E = \{1, 2, \ldots, n\}$ of $n$ individuals (training set) which are each characterized by a vector of $p$ quantitative predictor variables

\[ Y = (Y_1, Y_2, \ldots, Y_p) \]

Each element $k$ of $E$ belongs to one of $m$ classes $\Pi_1, \Pi_2, \ldots, \Pi_m$.

2. Apply Radon transform, from image $0^\circ$ to $180^\circ$ orientations (in steps of $h$), to the input range image $I_i$ yielding a binary image $I_{ij}$.

3. Superpose the binary image $I_{ij}$ obtained in the Step 2 on the input range image $I_i$ to obtain the cropped facial range image $I'_{ij}$.

4. Repeat the Steps 1 to 3 for all the $M$ facial range images including subclasses in the training set.

5. Apply PCA to the set of cropped facial range images obtained in the Step 4 and obtain eigenclasses.

6. Compute the weights $w_1, w_2, \ldots, w_M$ for each training face image, where $s < M$ is the dimension of feature subspace on which the training face image is projected.

7. After computing the weights perform FDA on feature subspace.

8. Store the weights $w_1, w_2, \ldots, w_M$ for each training image as its facial features in the feature library of the face database.

**Algorithm 2: Testing Phase**

1. Input the m 3D face images $Z_i, (i = 1, \ldots, m)$ of a subject (individual), one image for each of $m$ subclass.

2. Apply Radon transform, for $Z_i, (i = 1, \ldots, m)$ to the input range images $Z_i$ yielding a binary images $Z_{si}$.

3. Superimpose the binary images $Z_{si}$ on $Z_i$ to obtain the cropped facial image $Z_{si}$.

4. Compute the weights $w_{i}^{m}, i = 1, 2, \ldots, m$ , for the test images $Z_i$, by projecting the test images on the feature subspace of dimension $s$.

5. For classification, apply SVM on feature vectors $w_{i}^{m}$ and the feature vectors $w_{i}$ stored in the feature library.

6. The face image in the face database classified by the SVM classifier in the Step 5 is the recognized face. Output the texture face image corresponding to the recognized facial range image.

The above algorithm for testing phase is modified to apply Minimum Distance Classifier and K-NN classifier to the feature set in the Step 6. The classification performances of these different classifiers are compared with SVM.
4. EXPERIMENTAL RESULTS

As in typical biometric systems, the proposed method includes two phases: the training phase and the testing phase as illustrated in the Figure 1. The training phase has 300 subjects (classes) with three subsets (i.e., \( n=300 \), \( m=3 \)), namely, normal subset, expression variation subset and pose variation subset for each subject and each subset contains 3 face images of that subject. For each class (subject) the images in a subset are divided into subclasses, such that each subclass consists of three consecutive images with variation. The radon transform is applied for each 3D face image as a preprocessing step and extracted facial region for all images for the training set. The outcome of the preprocessing step is superimposed with original 3D face images of the training set respectively and applied PCA for the whole set. The weights \( w_1, w_2, \ldots, w_m \) for each training face image is computed, where \( m < M \) is the dimension of feature subspace on which the training face image is projected. After computing the weights factorial discriminant analysis is performed on feature subspace and the weights for each training image as its facial features are stored in the feature library of the face database. For testing phase, three face images are taken from each subject, one image being drawn from each of the three subclasses.

The proposed method is implemented using Intel Core 2 Quad processor @ 2.66 GHz machine and MATLAB 2012b. The 4000 images of three databases, namely, Bhosphorus 3D face database, Texas and CASIA 3D face database, are divided into two subsets, which are the training set, and probe set. The sample training set of 3D face images used for experimentation are shown in the Figure 2. The sample testing set of 3D face images used for experimentation are shown in the Figure 3. The Table 1 shows performance comparison of the RT+Symbolic PCA and RT+Symbolic LDA methods with the proposed method in terms of recognition rates and the Table 2, shows the performance comparison of the proposed method with other methods. The Figure 4 shows a Receiver operating curve (ROC) space, which is defined by FAR versus FRR as x and y axes respectively, which depicts relative trade-offs between true positive and false positive for the FDA based face recognition for 3 databases, namely, Bhosphorus 3D face database, Texas 3D face database and CASIA 3D face database, with equal error rates (ERR) 12.2876, 10.4253 and 9.1877 respectively. The reason for lower ERR for CASIA 3D database is due to the fact that it contains more sample images with variations in pose, expression and illumination as compared to the other two databases, which is responsible for better training in case of CASIA 3D face database.

![Figure 1. Overview of proposed framework](image1)

![Figure 2. Sample training 3D face images used for experimentation](image2)
Figure 3. Sample testing 3D face images used for experimentation

Table 1. Performance comparison of the proposed with the RT+PCA+LDA, RT+ Symbolic PCA and RT + Symbolic LDA, in terms of recognition accuracy.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>61.60</td>
<td>62.00</td>
<td>64.46</td>
<td>64.46</td>
<td>65.86</td>
</tr>
<tr>
<td>10</td>
<td>77.90</td>
<td>78.00</td>
<td>81.46</td>
<td>80.86</td>
<td>81.66</td>
</tr>
<tr>
<td>15</td>
<td>85.10</td>
<td>85.50</td>
<td>87.96</td>
<td>87.76</td>
<td>88.96</td>
</tr>
<tr>
<td>20</td>
<td>91.20</td>
<td>91.00</td>
<td>93.96</td>
<td>93.86</td>
<td>94.66</td>
</tr>
<tr>
<td>25</td>
<td>94.20</td>
<td>95.00</td>
<td>97.16</td>
<td>96.90</td>
<td>97.16</td>
</tr>
<tr>
<td>30</td>
<td>95.91</td>
<td>96.00</td>
<td>98.96</td>
<td>97.66</td>
<td>98.56</td>
</tr>
<tr>
<td>35</td>
<td>97.90</td>
<td>96.50</td>
<td>99.26</td>
<td>98.56</td>
<td>98.99</td>
</tr>
<tr>
<td>40</td>
<td>99.16</td>
<td>97.00</td>
<td>99.60</td>
<td>99.66</td>
<td>99.70</td>
</tr>
</tbody>
</table>

Table 2. The performance comparison of proposed method with other methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Recognition Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faltiemier et al. [23] (Score-based Fusion)</td>
<td>94.90%</td>
</tr>
<tr>
<td>Maurer et al. [24] (Iterative Closest Point (ICP))</td>
<td>95.80%</td>
</tr>
<tr>
<td>Hiremath et al. [28] (RT+Symbolic PCA)</td>
<td>97.00%</td>
</tr>
<tr>
<td>Kakadiaris et al. [26] (Annotated model, ICP)</td>
<td>97.00%</td>
</tr>
<tr>
<td>H. usken et al. [27] (Hierarchical graph matching (HGM))</td>
<td>97.30%</td>
</tr>
<tr>
<td>Hiremath et al. [25] (RT + PCA + LDA)</td>
<td>99.16%</td>
</tr>
<tr>
<td>Mian et al.[29] (Spherical face representation (SFR), ICP)</td>
<td>99.30%</td>
</tr>
<tr>
<td>Hiremath et al. [30] (RT + Symbolic LDA)</td>
<td>99.50%</td>
</tr>
<tr>
<td>Proposed Method (RT + FDA, SVM)</td>
<td>99.70%</td>
</tr>
</tbody>
</table>

Figure 4. Receiver operating characteristic (ROC) curves for the proposed method, experimented with Bhosphorus, Texas and CASIA 3D face databases yielding equal error rates 12.2876, 10.4253 and 9.1877 respectively.
5. CONCLUSION

In this paper, we have proposed a novel method for three-dimensional (3D) face recognition using Radon transform and Factorial Discriminant Analysis (FDA) based features of 3D range face images using KNN and SVM. In this method, the FDA based feature computation takes into account of 3D face image variations to a larger extent and has advantage of dimensionality reduction. The experimental results have yielded 99.70% recognition performance using SVM classifier with reduced complexity, which compares well with other state-of-the-art methods. The experimental results demonstrate the efficacy and the robustness of the method to illumination and pose variations. The recognition accuracy can be further improved by considering a larger training set and a better classifier.

6. ACKNOWLEDGMENTS

The authors are indebted to the University Grants Commission, New Delhi, for the financial support for this research work under UGC-MRP F.No.39-124/2010 (SR).

7. REFERENCES


8. AUTHOR’S PROFILE

Dr. P. S. Hiremath was born in May 1952, and has obtained Ph.D. (1978) in Applied Mathematics and M.Sc. (1973) in Applied Mathematics, from Karnataka University, Dharwad, Karnataka, India. He had been in the Faculty of Mathematics and Computer Science of various institutions in India, namely, National Institute of Technology, Surathkal (1977-79), Coimbatore Institute of Technology, Coimbatore (1979-80), National Institute of Technology, Tiruchirappalli (1980-86), Karnataka University, Dharwad (1986-1993) and has been presently working as Professor of Computer Science in Gulbarga University, Gulbarga (1993 onwards). His research areas of interest are Computational Fluid Dynamics, Optimization Techniques, Image Processing and Pattern Recognition, and Computer Networks. He has published 156 research papers in peer reviewed International Journals and Proceedings of International Conferences.

Manjunatha Hiremath was born in July 1984, and has obtained M.Phil (2010) in Computer Science and M.Sc. (2008) in Computer Science from Gulbarga University, Gulbarga. Presently, he is working as Project Fellow in UGC Major Research Project since February 2011. His area of research interest is Image Processing and Pattern Recognition. He has published 21 research papers in peer reviewed International Journals and Proceedings of International Conferences.