August 2 0 0 9

Master's Thesis

Performance Evaluation of Principal Component Analysis and Linear Discriminant Analysis for Human Teeth Recognition

Graduate School of Chosun University

Department of Information and Communication Engineering

Poudel Santosh

Performance Evaluation of Principal Component Analysis and Linear Discriminant Analysis for Human Teeth Recognition

August 2009

Graduate School of Chosun University

Department of Information and Communication Engineering

Poudel Santosh

Performance Evaluation of Principal Component Analysis and Linear Discriminant Analysis for Human Teeth Recognition

Director: Youngsuk Shin, Ph. D

Thesis submitted for the Degree of Master of Engineering

April 2009

Graduate School of Chosun University

Department of Information and Communication Engineering

Poudel Santosh

Poudel Santosh의 석사학위 논문을 인준함

위 원 장	박 종 안	(인)
위 원	변 재 영	(인)
위 원	신 영 숙	(인)

2009 년 6 월

조 선 대 학 교 대 학 원

Table of Contents (목차)

Abstract ··· iii
List of Figures ··· iv
List of Tables ··· v
I. Introduction ··· 1
A. Background ··· 1
B. Pattern Recognition ··· 3
C. Feature Analysis ··· 3
D. Pattern Classification ··· 4
II. Previous Works ··· 6
A. A General Algorithm ··· 7
B. Why Study These Subspaces? ··· 9
III. Independent Feature Extraction ··· 10
A. Linear Feature Extraction Formulation ··· 10
B. Principal Component Analysis ··· 10
1. A Brief History of PCA ··· 10
2. Definition and Derivation of PCA ··· 11
3. PCA for Feature Dimensionality Reduction in Classification ··· 14
4. Eigenvector Selection ··· 15
5. Similarity & Distance Measures ··· 16
C. Linear Discriminant Analysis ··· 16
1. Fisher's Linear Discriminates ··· 16
IV. Experiments Using Teeth Images ··· 20
A. Utilized Teeth Database ··· 20
1. Database ··· 20
2. Preprocessing Techniques ··· 21
B. Eignespace Projection ··· 24
1. Recognizing Images Using Eignenspace ··· 25
2. PCA for Feature Extraction of Teeth ··· 27
3. Classification ··· 29

C. Fisher Discriminates....31
1. Fisher Discriminates Tutorial(Original method) 32
2. Fisher Discriminates Tutorial(Orthonormal Basis Method) 34
3. LDA based Teeth Classifier....34
D. Experimental Results 36
1. Experiment I.... 37
2. Experiment II....42
3. Experiment III 44
V. Conclusion 46
References 47

ABSTRACT

Performance Evaluation of Principal Component Analysis and Linear Discriminant Analysis for Human Teeth Recognition

Poudel, Santosh
Advisor: Youngsuk Shin, Ph. D
Department of Information and Communication Engineering
Graduate School of Chosun University

Biometric identification methods have been proved to be very efficient, natural, and easier for users than traditional methods of human identification. Biometric is defined as the science of recognizing a person based on certain physiological (fingerprints, face and voice) traits which possess low discriminating contents; these change over time for each individual. Thus, these biometrics show lower performance as compared to the strong biometrics (eg. fingerprints, iris, retina, etc.). Among various physiological biometrics, teeth biometrics has been found to be interesting and promising in the biometrics field. In this thesis, for the performance evaluation of appearance-based statistical methods, both Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are tested and compared for the recognition of human teeth images. In the transformed space, euclidean distance classifier is employed. Teeth were acquired using a simple low-cost setup consisting of a digital camera. Three sets of experiments are conducted for relative performance evaluations. In the first set of experiments, the recognition performances of PCA and LDA are demonstrated. The effect of illumination variations is evaluated in the second set, whereas teeth images are anterior and posterior occlusion in the third set of experiments.

The goal of this thesis is to present an independent and comparative study of two most popular appearance-based teeth recognition algorithms (PCA and LDA) in various conditions.

List of Figures

Fig. 1.1 Conventional pattern recognition systems....3
Fig. 1.2 Block diagram of the system 8
Fig. 4.1 Original picture of teeth images....20
Fig. 4.2 Cropping region area from original teeth image....21
Fig. 4.3 Cropped region area of teeth image....21
Fig. 4.4 Gray scale training images....22
Fig. 4.5 Gray scale testing images....22
Fig. 4.6 Normalization of training teeth images 23
Fig. 4.7 Sample training teeth image modalities (a) Low light image
(b) Light reflected image (c) Dark image....23
Fig. 4.8 Sample testing teeth image modalities (a) low light image
(b) Light reflected image (c) Dark image....23
Fig. 4.9 Illustration of average teeth image and eigenteeth....26
Fig. 4.10 Illustration of sample teeth and normalizing teeth images
(a) Training teeth image
(b) Normalization of training teeth images....27
Fig. 4.11 Average teeth image 28
Fig. 4.12 Eigenspace of teeth images 29
Fig. 4.13 Example of test and reconstruction teeth image
(a) Test teeth image (b) Reconstruction teeth image....29
Fig. 4.14 PCA approach for teeth recognition....30
Fig. 4.15 Example of 2-dimensional space for linear projection
(a) Points in 2-dimensional space
(b) Points mixed when projected onto a line
(c) Points separated when projected onto a line....31
Fig. 4.16 A comparison of principal component analysis (PCA) and
Fisher's linear discriminant (FLD) for a two class problem
where data for each class lies near a linear subspace 32

Fig. 4.17 LDA approach for teeth recognition 36
Fig. 4.18 Summary of the experimental results in low dimensional space
(a) Performance curves for PCA and LDA
(b) Comparison PCA and LDA curve
(c) Recognition rate of PCA and LDA..40
Fig. 4.19 The recognition rate on data set of high dimensional case
(a) Recognition rate of PCA and LDA
(b) Performance curves of PCA and LDA 41
Fig. 4.20 Teeth image in illumination variation 42
Fig. 4.21 Example of occlusion images
(a) Posterior teeth image
(b) Anterior teeth image...44

List of Tables

Table 4.1 Experiment I results: different training and testing subsets for the value of the dimensionality parameters....................................38
Table 4.2 High-dimensional spaces...38
Table 4.3 Experiment II results: estimations with illumination variation teeth...42
Table 4.4 Experiment III results: estimations with posterior and anterior teeth 45

I. Introduction

Pattern recognition has been an active research area over the last 30 years. It has been studied by scientists from different areas of psychophysical sciences and from different areas of computer sciences. Psychologists and neuro-scientists mainly focus with the human perception aspect of the topic, whereas engineers studying on machine recognition of human body parts deal with the computational aspects of pattern recognition.

In this thesis, proposed method is based on human teeth recognition system under the different testing methodologies. The performance of purposed system is evaluated by using two algorithms (PCA and LDA) through the various testing conditions. Experiments are conducted by utilizing newly constructed training and test human teeth database for this teeth recognition system.

A. Background

Biometrics is a method to automatically verify or identify individuals using their physiological or behavioral characteristics.

Biometric technologies have essential some requirements in order to be utilized in real applications. They are reliable, easy to use, easy to implement and cost effective. Iris identification requires a complicity of the data collection. Face can be deformed by expressions of a user. Fingerprint can be contaminated with materials such as sweat or dust. Voice can be changed by catch a cold. Human teeth are not generally deformed at the moment of image acquisition because of rigidity. Furthermore, in teeth biometric there is no need to touch any device, hence the user feels more comfortable. In addition, teeth identification does not require high resolution images. Thus data collection can use the digital cameras with low cost. Teeth, which concern recognizing individual, are a relatively new biometrics.

Training the teeth recognition system with images from the known individuals and classifying the newly received test images into one of the classes is the main aspect of the teeth recognition systems. The topic seems to be easy for a human, where limited memory can be a main problem; whereas the problems in machine recognition are manifold. Some of possible problems for a machine teeth recognition system are mainly;

Teeth expression change: Teeth expression (Anterior and Posterior) can affect teeth recognition system significantly.

Aging: Images that taken some time apart varying from 5 minutes to 5 years changes the system accuracy.

Rotation: Rotation of the individual's head clockwise or counter clockwise (even if the image stays frontal with respect to the camera) affects the performance of the system.

What biological measurements qualify to be a biometric? Any human physiological and/or behavioral characteristic can be used as a biometric characteristic as long as it satisfies the following requirements:

- *Universality*: each person should have the characteristic.
- *Distinctiveness*: any two persons should be sufficiently different in terms of the characteristic.
- *Permanence*: the characteristic should be sufficiently invariant (with respect to the matching criterion) over a period of time.
- *Collectability*: the characteristic can be measured quantitatively.

A practical biometric system should meet the specified recognition accuracy, speed, and resource requirements, be harmless to the users, be accepted by the intended population, and be sufficiently robust to various fraudulent methods and attacks to the system.

The rest of this thesis is organized as follows. Section two gives a description of related works. Section three describes feature extraction algorithms employed by our system. Section four presents newly constructed training and test database for teeth recognition. A detailed description and experimental results of three different modalities are given in section four. Finally, in section five, conclusion drawn from our experiments are discussed.

B. Pattern Recognition

Pattern recognition deals with mathematical and technical aspects of classifying different objects thorough their observable information, such as grey levels of pixels for an image, energy levels infrequency domain for a wave form and the percentage of certain contents in a product. It is normally beyond the scope of the study of pattern recognition. Thus, a typical pattern recognition system consists of two component: feature analysis, which includes parameter extraction and/ or feature extraction, and pattern classification. The structure of a conventional pattern recognition system is show in Fig. 1.1

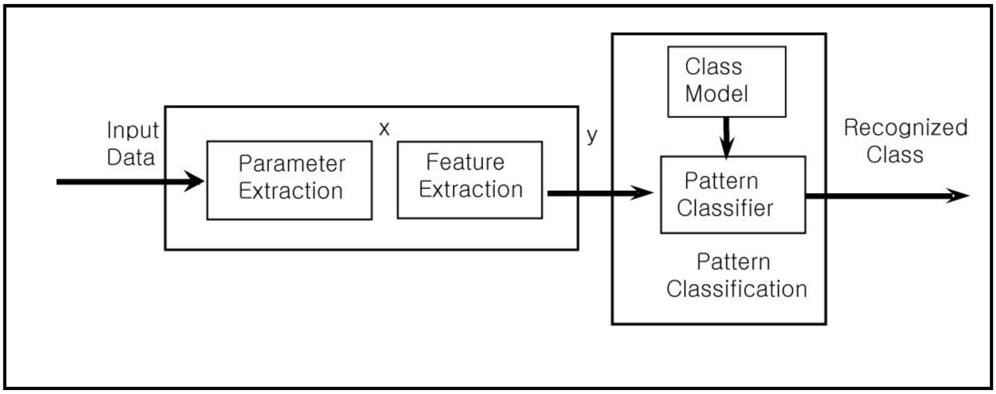


Fig. 1.1 Conventional pattern recognition systems

C. Feature Analysis

Feature analysis is achieved in two steps: parameter extraction and or/feature extraction. In the parameter extraction step, information relevant to pattern classification extracted from the input data in the form of a p-dimensional parameter vector. In the feature extraction step, the parameter vector x is transformed to a feature vector, which has a dimensionality m $(m<=p)$. If the parameter extractor is properly designed so that the parameter vector x is matched to the pattern classifier and its dimensionality is low, then there is no necessity for the feature extraction step. However in practice, parameter vectors are not suitable for pattern classifiers. For example, parameter vectors have to be de-correlated before applying them to a classifier based on Gaussian mixture modes (with diagonal variance matrices). Furthermore, the dimensionality of parameter vectors is normally very high and needs to be reduced for the sake of less computational cost and system complexity. Due to these reasons, feature extraction has been an important part in pattern recognition tasks.

Feature extraction can be conducted independently or jointly with either parameter extraction or classification. Independent feature extraction method is a well developed area of research. A number of independent feature extraction algorithms have been proposed [1, 2, 3].

Among them, LDA and PCA are two popular independent feature extraction methods. Both of them extract feature by projection the original parameter vectors onto a new feature space through a linear transformation matrix. But they optimize the transformation matrix with different intentions. PCA optimizes the transformation matrix by finding the largest variations in the original feature space [2, 3]. LDA pursues the largest ratio of between-class variation and within-class variation when projecting the original feature to a subspace [4]. The drawback of independent feature extraction algorithms is that their optimization criteria are different from the classifier's minimum classification error criterion, which may cause inconsistency between feature extraction and the classification stages of a pattern recognizer and consequently, degrade the performance of classifiers [5].

D. Pattern Classification

The objective of pattern classification is to assign an input feature vectors to one of K existing classes based on a classification measure. Conventional classification measures include distance (Mahalanobis or Euclidean distance), Likelihood and Bayesian a posteriori probability. These measures lead to linear classification methods, i.e., the decision boundaries they generate are linear. Linear method, however, has the limitation that they have little computational flexibility and are unable to handle complex non linear decision boundaries. SVM is a developed pattern classification algorithm with non-linear formulation. It is based on the idea that the classification that affords dot-products can be computed efficiently in higher dimensional feature spaces [6, 7, 8]. The classes which are not linearly separable in the original parametric space can be linearly separated in the higher dimensional feature space. Because of this, SVM has the advantage that it can handle the classes with complex non-linear decision boundaries. SVM has now evolved into an active area of research [9].

II. Previous Works

Projecting images into Eigenspace is a standard procedure for many appearance-based object recognition algorithms. Michael Kirby was the first to introduce the idea of the low dimensional characterization of faces. Examples of his use of eigenspace projection can be found in [10, 11, 12]. Turk & Pentland worked with eigenspace projection for face recognition [13]. More recently Shree Nayar used eigenspace projection to identify objects using a turntable to view objects at different angles as explained in [14]. R.A. Fisher developed Fisher's linear discriminant in the 1930's [15]. Not until recently have Fisher discriminates been utilized for object recognition. An explanation of Fisher discriminates can be found in [16]. Swets and Weng used Fisher discriminates to cluster images for the purpose of identification in 1996. Most studies on teeth identification had been used in postmortem identification and location missing and unidentified persons. Jain and Chen [17, 18] utilized dental radio-graphs to identify victims. Ammar and Nassar [19] analyzed radio-graphs to utilize underlying image structure that are often difficult to be assessed merely by visual examination. Zhou and Abdel-Mottaleb Mahoor and abdel-Mottaleb extracted the teeth contours and used shape representation based on extraction since poor quality images [20][21]. The dental radio graphs have a challenging problem of the shape extraction since poor quality images.

A couple of researches on using teeth for personal identification have been reported. Prajuabklang, K., Kumhom, P., Maneewarn, T. and Chamnongthai, K. [22], proposed Real-time Personal Identification from Teeth-image using Modified PCA. They applied the Principle Component Analysis (PCA). In this method, the eigenvectors and their corresponding eigenvalues were found and matched with the vectors of possible teeth images in the database. Shin presented gender identification on the dental image using geometric features [23]. Tae-Woo Kim, Tae-Kyung Cho, Byoung-Soo Park and Myung-Wook Lee [24] proposed a personal

identification method using teeth images. The method is composed of teeth image acquisition and teeth recognition in which there are teeth region extraction and pattern recognition procedure in a sequential step. In the teeth recognition, an input pattern is compared with each pattern of the teeth database in which each class has feature vectors for teeth set of a person. The method uses teeth images for anterior and posterior occlusion state. For pattern recognition, they used LDA method which is popular in appearance-based face recognition and a nearest neighbor (NN) algorithm. However, it is not a comparison of algorithms for human teeth.
In this thesis, comparison of PCA and LDA techniques based on human teeth recognition. Experiments show that PCA performance has better than LDA techniques. These experiments assure that teeth biometrics will play vital role in multimodal biometrics. This research will be able to draw the attention who are involving in biometrics fields.

A. A General Algorithm

An image may be viewed as a vector of pixels where the value of each entry in the vector is the gray-scale value of the corresponding pixel. For example, an 8 x 8 image may be unwrapped and treated as a vector of length 64. The image is said to sit in dimensional space, where N is the number of pixels (and the length of the vector). This vector representation of the image is considered to be the original space of the image. The original space of an image is just one of infinitely many spaces in which the image can be examined. Two specific subspaces are the subspace created by the eigenvectors of the covariance matrix of the training data and the basis vectors calculated by Fisher discriminants. The majority of subspaces, including eigenspace, do not optimize discrimination characteristics. Eigenspace optimizes variance among the images. The exceptions to this statement is Fisher discriminants, which does optimize discrimination characteristics. Although some of the details may vary, there is a basic algorithm for identifying images by projecting them into a subspace. First one selects a subspace on which to project the images. Once this subspace is selected, all training images are projected into this subspace. Next each test image is projected into this subspace. Each test image is compared to all the training images by a similarity or distance measure, the training image found to be most similar or closest to the test image is used to identify the test image.

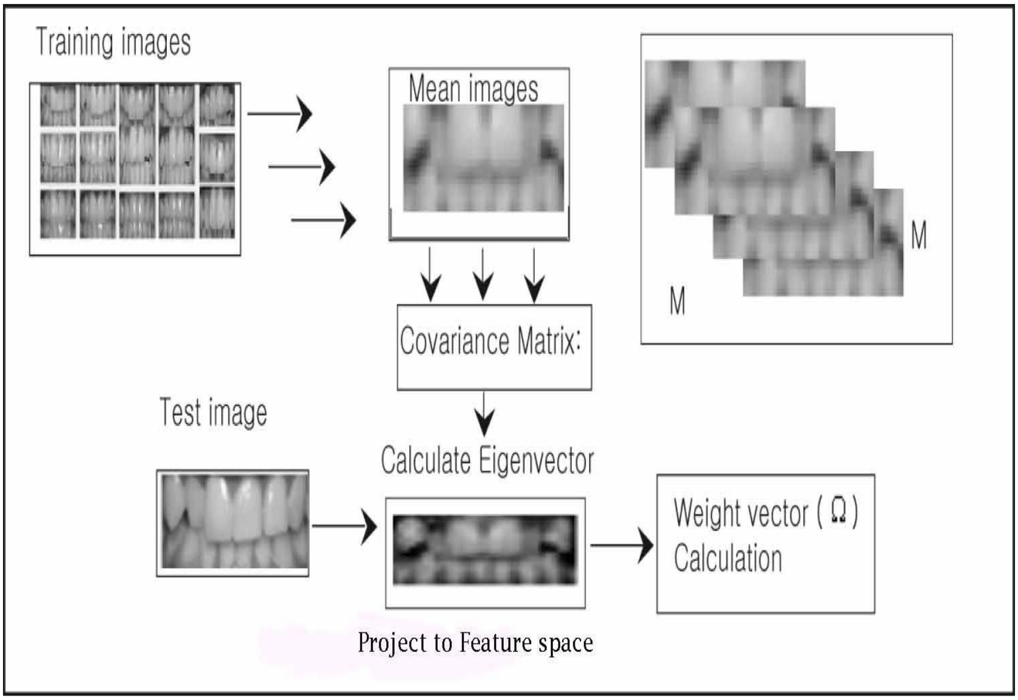


Fig. 1.2 Block diagram of the system

B. Why Study These Subspaces?

Projecting images into subspaces has been studied for many years as discussed in the previous work section. The research into these subspaces has helped to revolutionize image recognition algorithms, teeth recognition. When studying these subspaces an interesting question arises: under what conditions does projecting an image into a subspace improve performance. The answer to this question is not an easy one. What specific subspace (if any at all) improves performance depends on the specific problem? Furthermore, variations within the subspace also affect performance. For example, the selection of vectors to create the subspace and measures to decide which images are a closest match, both effect performances.

III. Independent Feature Extraction

A. Linear Feature Extraction Formulation

Linear feature extraction method is the most basis way of extraction feature vectors. It projects parameter vectors form parametric space onto feature space through a linear transformation matrix T. Suppose the input observation vector $\boldsymbol{x}$ be a $\boldsymbol{p}$-dimensional vector and $\mathbf{T}$ be a $p \times m\,(p \geq m)$ matrix. The extracted feature vector y *is:*

$$y = T^{T} x \tag{3.1}$$

The difference between linear feature extraction algorithms is that they optimize T by different criteria. A number of algorithms have been proposed to seek the optimized T. PCA and LDA are the most popular ones among them. Briefly speaking, PCA obtains T by searching for the directions that have the largest variations; LDA optimizes T by maximizing the ratio of between-class variation and within-class variation. In the following subsections, a detailed discussion of each of them will be given.

B. Principal Component Analysis

1. A Brief History of PCA

The earliest descriptions of PCA appear to be proposed by Pearson in 1901 [25] and Hotelling in 1933 [26]. In Pearson's paper, the main concern was to find lines and planes which best fit a set of points in a p-dimensional space and the geometric optimization problems considered lead to principal components (PCs). It seems that little relevant work has been published in the 32 years between Pearson's and Hotelling's papers.

Hotellings motivation is that are may be a smaller 'fundamental set of independent variables' which determines the values of the original p variables. The term components was introduced and they were chosen to maximize their successive contributions to the total of the variances of the original variables. Hotelling called the components derived in this was the 'principal components' and the analysis to find these components was then christened the 'method of principal components'. Hotelling derived the PCs by the power method.

In 1939, Girshick [27] investigated the asymptotic sampling distributions of the coefficients and variances of PCs. But apart from Girshick's work there appears to be little work on the development of different applications of PCA during nearly three decades following the publication of Hotelling's paper. Not until 1963, based on the earlier work by Girshick (1939), Anderson (1963) discussed the asymptotic sampling distributions of the coefficients and variances of the samples PCs which has built up the fundamental framework of PCA [28]. Rao (1964) provided a large number of new ideas concerning uses, interpretations and extensions of PCA [29]. Gower (1966) discussed some links between PCA and various other statistical techniques and provided a number of geometric insights [30].

Despite the simplicity of the technique, much research is still being carried out in the general are of PCA. Apart from being used basically as a dimensionality reduction tool, PCA is also widely used for feature extraction, data compression and preprocessing for pattern recognition etc.

2. Definition and Derivation of PCA

The central idea of PCA is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible the variation present in the data set.

Suppose x is a p-dimensional random vector. PCA first looks for a linear function $a_1^T x$ of x which has maximum variance, where $a_1 = \left\{ a_{11}, a_{12}, \ldots\ldots, a_{1p} \right\}$ is a p-dimensional vector and $$a_1^T x = a_{11}x_1 + a_{12}x_2 + ... + a_{1p}x_p = \sum_{i=1}^{p} a_{1i}x_i \quad (3.2)$$

Then, it looks for a second linear function $a_2^T x$ which is un-correlated with $a_1^T x$ and has the second maximum variance. Repeat this procedure until the desired kth linear function $a_k^T x$ is found. These k variables $a_1^T x, a_2^T x, ..., a_k^T x,$ are called k principle components (PCs). In general, up to p PCs can be found. The mathematical expressions of constraint on $a_i = (i = 1, 2, ..., p)$ are:

$$a_i^T a_j = \begin{Bmatrix} 1, \ if \ \ i = j \\ 0, \ if \ \ i \leq j \end{Bmatrix} . \quad (3.3)$$

Consider the first PC, $a_1^T x$. a_1 maximizes $\text{var}\ [a_1^T x] = a_1^T \sum a_1$ subject to $a_1^T a_1 = 1$. Use lagrange multiplier, we have:

$$\alpha_1^T \Sigma \alpha_1 - \lambda(\alpha_1^T \alpha_1 - 1) \quad , \quad (3.4)$$

where λ_1 is a lagrange multiplier. Differentiation (3.4) with respect to a_1 gives:

$$(\Sigma - \lambda_1 I_p)a_1 = 0 \quad , \quad (3.5)$$

where I_p is the $p \times p$ identity matrix. Thus, λ_1 is the eigenvalue of $\sum$ and a_1 is the corresponding eigenvector. Note the quantity to be maximized is:

$$a_1^T \sum a_1 = a_1^T \lambda_1 a_1 = \lambda_1 a_1^T a_1 = \lambda_1 \quad (3.6)$$

Thus, λ_1 must be the largest eigenvalue and a_1 is corresponding eigenvectors.

Consider the second PC, $a_2^T x$ maximizes $a_2^T x a_2$ subject to being un-correlated with the first PC, $a_1^T x$ that is:

$$\operatorname{cov}[\ a_1^T x, a_2^T x] = 0 \tag{3.7}$$

If choosing $a_2^T a = 0$ to specify the relationship in (3.5), the quantity to maximize is:

$$a_2^T \sum a_2 = \lambda_2 (a_2^T a_2 - 1) - \phi a_2^T a_1 \tag{3.8}$$

where λ_2 and ϕ are lagrange multiplier. Differentiation of (3.8) with respect to a_2 gives:

$$\sum a_2 - \lambda_2 a_2 - \phi a_2^T a_1 \tag{3.9}$$

Eq. (3.9) can be reduced to:

$$\sum a_2 - \lambda\, alpha_2 = 0 \tag{3.10}$$

$$\phi = 0$$

Again $\lambda_2 = a_2^T \sum a_2$, therefore, λ_2 is the second largest eigenvalue and a_2 is the corresponding eigenvector.

By using the same strategy, it can be shown that the coefficient vector a_k *of* *kth* PC(k=1,2,……,p) is the eigenvector corresponding to the k*th* largest eigenvalue of $\sum$.

3. PCA for Feature dimensionality Reduction in Classification

Given p-dimensional data set X, the m principal axes $T_1, T_2, \ldots\ldots T_m$, where $1 \le m \le p$, are orthogonal axes onto which the retained variance is maximum in the projected space. Generally $T_1, T_2, \ldots\ldots T_m$, can be given by the m leading eigenvectors of the sample covariance matrix $S = \frac{1}{2}\sum_{i=1}^{N}(x_i - \mu)^T(x_i - \mu)$, Where $x_i \in X$, μ the sample is mean and N is number of samples, so that:

$$ST_i = \lambda_i T_i \qquad i \in 1, \ldots\ldots, m \tag{3.11}$$

where λ_i is the ith largest eigenvalue of S. Then principal components of a given observation vector $x \in X$, are given by:

$$y = [y_1, \ldots\ldots y_m] = [T_1^T x, \ldots\ldots T_m^T x] = T^T x \tag{3.12}$$

The m principal components of x are the non-correlated in the projected space . In multi-class problems, the variations of data are determined on a global basis [35], the principal axes are derived from a global covariance matrix:

$$\hat{S} = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N_j}(x_{ji} - \hat{\mu})(x_{ji} - \hat{\mu})^T \quad , \tag{3.13}$$

where $\hat{\mu}$ is the global mean of all the sample, K is the number of classes, N_j is the number of samples in class j, N= $\sum_{j=1}^{K} N_j$ and x_{ji} represents the ith observation from class j. The principal axes $T_1, T_2, \ldots\ldots T_m$, are therefore the m leading eigenvectors of $\hat{S}$:

$$\hat{S}T_i = \hat{\lambda}_i T_i \qquad i \in 1,\dots\dots,m \quad . \tag{3.14}$$

Where, $\hat{\lambda}_i$ is the ith largest eigen value of $\hat{S}$.

An assumption made for dimensionality reduction by PCA is that most information of the observation vectors is contained in the subspace spanned by the first m principal axes, where m<p. Therefore, each original data vector can be represented by its principal component vector:

$$y = T^T x \tag{3.15}$$

where $T=[T_1,\cdots\cdots T_m]$ is a $p \times m$ matrix.

The merit of PCA is that the extracted features have the minimum correlation along the principal axes. On the other hand there are some defects that reside in PCA. First, as mentioned in [33], PCA is a scale-sensitive method, i.e, the principal component may be dominated by the elements with large variance. Another problem with PCA is that the direction of maximum variance is not necessarily the directions of maximum discrimination since there is no attempt to use the class information, such the between-class scatter and within-class scatter.

4. Eigenvector Selection

Until this point, when creating a subspace using eigenspace projection we use all eigenvectors associated with non-zero eigenvalues. The computation time of eigenspace projection is directly proportional to the number of eigenvectors used to create the eigenspace. Therefore, by removing some portion of the eigenvectors computation time is decrease. Furthermore, by removing additional eigenvectors that do not contribute to the classification of the image, performance can be improved.

5. Similarity & Distance Measures

Once images are projected into a subspace, there is the task of determining which images are most like one another. There are two ways in general to determine how alike images are. One is to measure the distance between the images in N-dimensional space. The second way is to measure how similar the two images are. When measuring distance, one has to minimize distance, so the two images that are alike produce a small distance. When measuring similarity, one wishes to maximize similarity, so that two like images produce a high similarity value.

C. Linear Discriminant Analysis

1. Fisher's Linear Discriminant

The goal of Fisher's linear discriminant is well separate the class by projection classes, samples from p-dimension space onto a finely orientated line. For a K-class problem c=min(K-1,p) different lines will be involved. Thus, the projection is from a p-dimensional space to a c-dimensional space [35]. Suppose we have K classes, $X_1,X_2,\cdots\cdots X_K$. Let the *ith* observation vector from the X_j be X_j where j=1, ⋯⋯K, i=1 ⋯⋯N_j is the number of observations from class *j*. The sample mean vector μ_j and the covariance matrix S_j of class *j* are given by:

$$\mu_j = \frac{1}{N_j}\sum_{i=1}^{N_j} x_{ji} \tag{3.16}$$

and

$$S_j = \frac{1}{N_j}\sum_{i=1}^{N_j}(x_{ji}-\mu_j)(x_{ji}-\mu_j)^T \tag{3.17}$$

The within-class covariance matrix S_w is given by:

$$S_W = \sum_{j=1}^{K} S_j \tag{3.18}$$

Define the overall mean μ and the total covariance matrix S_T as:

$$\mu = \frac{1}{N}\sum_{j=1}^{K}\sum_{i=1}^{N_u} x_{ji} = \frac{1}{N}\sum_{j=1}^{K} N_j \mu_j \tag{3.19}$$

and

$$S_T = \sum_{j=1}^{K}\sum_{i=1}^{N_u} (x_{ji} - \mu)(x_{ji} - \mu)^T \tag{3.20}$$

where N= $\sum_{j=1}^{K} N_j$. Then it follows that:

$$S_T = \sum_{j=1}^{K} \sum_{i=1}^{N_j} (x_{ji} - \mu_j + \mu_j - \mu)^T$$

$$= \sum_{j=1}^{K} \sum_{i=1}^{N_j} (x_{ji} - \mu_j)(x_{ji} - \mu_j)^T + \sum_{j=1}^{K} \sum_{i=1}^{N_j} (\mu_j - \mu)(\mu_j - \mu)^T$$

$$= S_W + \sum_{j=1}^{K} N_j (\mu_j - \mu)(\mu_j - \mu) \tag{3.21}$$

It is natural to define the second term in Eq.(3.20) the between-class covariance matrix, so that we have:

$$S_B = \sum_{j=1}^{K} N_j (\mu_j - \mu)(\mu_j - \mu)^T \tag{3.22}$$

and

$$S_T = S_W + S_B \tag{3.23}$$

The projection from a p-dimensional space to an m-dimensional space is accomplished by m-discriminant function:

$$y_i = w_I^t x \qquad i = 1,2,.........,m. \tag{3.24}$$

Eq.(3.24) can be re-written in matrix form:

$$y = W^t x \tag{3.25}$$

Then, corresponding mean and covariance matrix of y are defined as:

$$\tilde{\mu}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ji} \tag{3.26}$$

$$\tilde{\mu} = \frac{1}{N} \sum_{j=1}^{K} N_j \tilde{\mu}_j \tag{3.27}$$

$$\tilde{S}_W = \sum_{j=1}^{K} \sum_{i=1}^{N_j} (y_{ji} - \tilde{\mu}_j)(y_{ji} - \tilde{\mu}_j)^T \tag{3.28}$$

and

$$\tilde{S}_B = \sum_{j=1}^{K} N_j (\tilde{\mu}_j - \tilde{\mu})(\tilde{\mu}_j - \tilde{\mu})^T \tag{3.29}$$

It is straightforward to show that:

$$\tilde{S}_W = W^T S_W W \tag{3.30}$$

and

$$\tilde{S}_B = W^T S_B W \tag{3.31}$$

Fisher's linear discriminant is then defined as the linear function $W^T x$ for which the criterion function

$$J(W) = \frac{|\tilde{S}_B|}{|\tilde{S}_W|} = \frac{W^T S_B W}{W^T S_W W} \tag{3.32}$$

is maximum.

It can be shown that the solution of Eq.(3.32) is that the *i*th column of an optimal **W** is the generalized eigenvector corresponding to the *i*th largest eigenvalue of matrix $S_W^{-1} S_B$.

IV. Experiments using Teeth Images

A. Utilized Teeth Database

In this thesis, experiments have done for pattern recognition based on three methodologies for utilizing human teeth images.

1. Database

The newly constructed teeth images database was used for the experiments. It was created by members of human computer interaction laboratory at Chosun university. Teeth dataset were collected during this thesis study, in order to test the systems performances in a real-life application. Images were first taken from university students. Finally, these images were manually cropped in order the image to contain the teeth region.

Fig. 4.1 Original picture of teeth images

There are 75 individuals in this database each having 6 frontal teeth images, it can be suggested to use the first three images as the training set and the other images taken after the four weeks as a test sets. All these images are taken from normal digital camera.

2. Preprocessing Techniques

The main purpose of the preprocessing is to format the images in the test image set to be suitable for the algorithm to use. There are many kinds of preprocessing techniques such as Resizing, rotation correction, cropping, histogram equalization, and masking.

In this thesis, mainly the effect of utilize cropping preprocessing techniques by cropped teeth image manually. When the area of an image is much larger compared to that of a teeth, the region of the image where the teeth is located is cut out from the image and only this area is used in the process of teeth recognition. In this study, the teeth area is determined.

Preprocessing is done in the following steps:

a. Images are cropped to contain only the subject's teeth as shown in the Fig. 4.2.

Fig. 4.2 Cropping region area from original teeth image

- Left and right borders are determined by using top frontal six teeth
- Top and bottom borders are determined using the half of the distance between the top teeth and down teeth vertical position.

Fig. 4.3 Cropped region area of teeth image

b. Images were converted in to 8-bit gray scale intensity image
c. Images were resized in to 15x30 pixels.

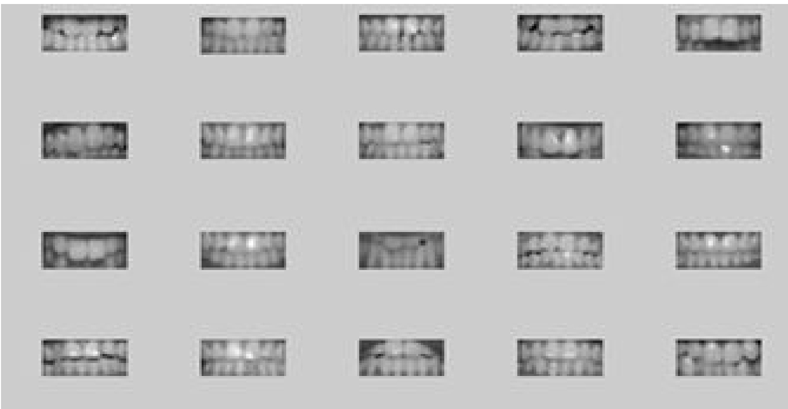


Fig. 4.4 Gray scale training images

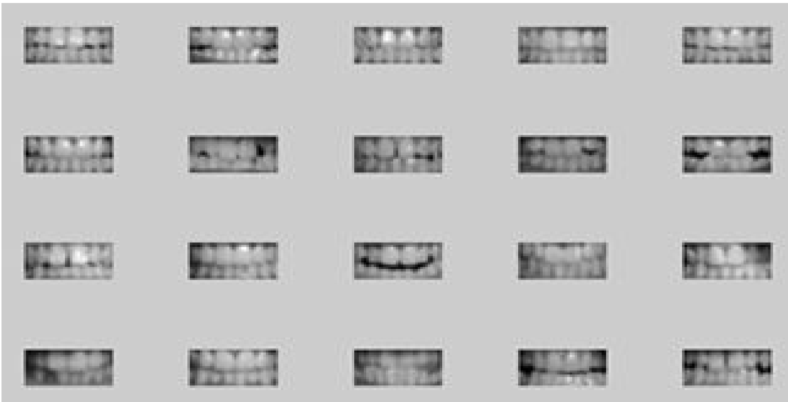


Fig. 4.5 Gray scale testing images

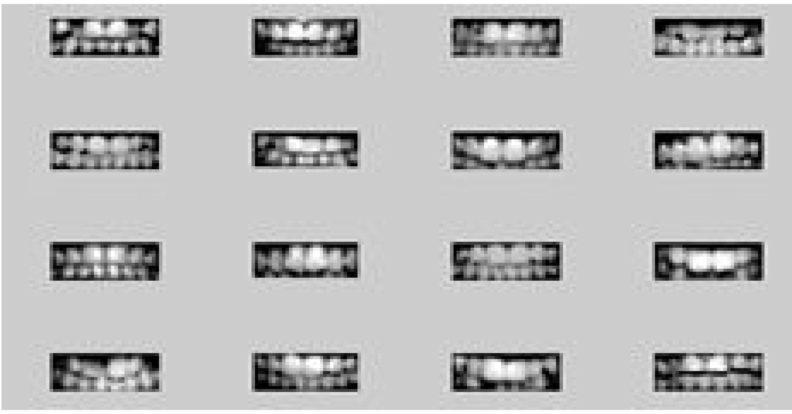


Fig. 4.6 Normalization of training teeth images

The database consists of 120 images (6 images for each of 20 subjects) in jpg format. The teeth images have been cropped, but they are not rotated and brightened. Example images are shown in Fig 4.7 and 4.8.

i. Training set

ii. Testing set.

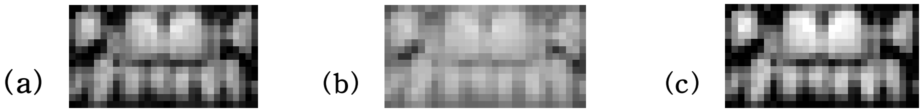


Fig. 4.7 Sample training teeth image modalities (a) Low light image (b) Light reflected image (c) Dark image

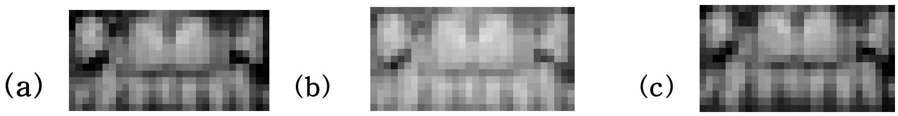


Fig. 4.8 Sample testing teeth image modalities (a) low light image (b) Light reflected image (c) Dark image

B. Eigenspace Projection

Eigenspace is calculated by identifying the eigenvectors of the covariance matrix derived from a set of training images. The eigenvectors corresponding to non-zero eigenvalues of the covariance matrix form an orthonormal basis that rotates and/or reflects the images in the N-dimensional space. Specifically, each image is stored in a vector of size N.

$$x^{i} = [x_{1}^{i} \ldots\ldots\ldots \; x_{N}^{i}]^{T} \tag{4.1}$$

The images are mean centered by subtracting the mean image from each image vector.

$$x^{-1} = x^{i} - m, where \quad m = \frac{1}{p}\sum_{i=1}^{P} x^{i} \tag{4.2}$$

These vectors are combined, side-by-side, to create a data matrix of size N x P (where P is the number of images).

$$\overline{X} = [\, \overline{x}^{1} \mid \quad \overline{x}^{2} \mid \ldots\ldots\ldots \quad \mid \overline{x}^{P} \,] \tag{4.3}$$

The data matrix **X** is multiplied by its transpose to calculate the covariance matrix.

$$\Omega = \overline{X}\overline{X}^{T} \tag{4.4}$$

This covariance matrix has up to P eigenvectors associated with non-zero eigenvalues, assuming P<N. The eigenvectors are sorted, high to low, according to their associated eigenvalues. The eigenvector associated with the largest eigenvalue is the eigenvector that finds the greatest variance in the images. The eigenvector associated with the second largest eigenvalue is the eigenvector that finds the second most variance in the images. This trend continues until the smallest eigenvalue is associated with the eigenvector that finds the least variance in the images.

1. Recognizing Images Using Eigenspace

$$\Omega V = \Lambda V \tag{4.5}$$

Here, V is the set of eigenvectors associated with the eigenvalues Λ .

Order eigenvectors: Order the eigenvectors $v_i \varepsilon V$ according to their corresponding eigenvalues $\lambda_i \varepsilon \Lambda$ from high to low. Keep only the eigenvectors associated with non-zero eigen values. This matrix of eigenvectors is the eigenspace V , where each column of V is an eigenvector.

$$V = [\, v_1 \mid v_2 \mid \ldots\ldots\ldots \mid v_P \,] \tag{4.6}$$

Project training images: These are projected into the eigenspace. To project an image into the eigenspace, calculate the dot product of the image with each of the ordered eigenvectors.

$$\tilde{x}^t = V^T \bar{x}^t \tag{4.7}$$

Therefore, the dot product of the image and the first eigenvector will be the first value in the new vector. The new vector of the projected image will contain as many values as eigenvectors.

Identify test images: Each test image is first mean centered by subtracting the mean image, and is then projected into the same eigenspace defined by V.

$$\bar{y}^t = y^t - m\,, \; where \quad m = \frac{1}{P}\sum_{i=1}^{P} x^t \tag{4.8}$$

$$\tilde{y}^t = V^T \tilde{y}^t \tag{4.9}$$

The projected test image is compared to every projected training image and the training image that is found to be closest to the test image is used to identify the training image. The images can be compared using any number of similarity measures; the most common is the 2l norm.

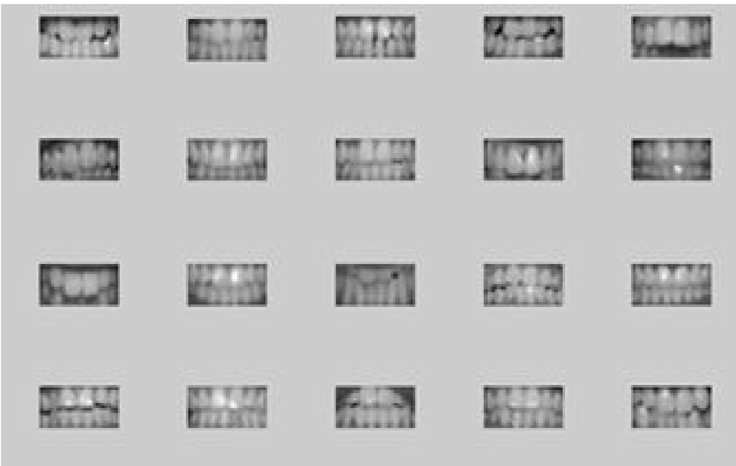
Fig. 4.9 Illustration of average teeth image and eigenteeth

In the language of information theory, we want to extract the relevant information in a teeth image, encoding with a database of model encoded similarly. A simple approach to extraction the information contained in an image of the teeth are to somehow capture the variation in a collection of teeth images, independent of any judgement of features, and use this information to encode and compare individual teeth images.

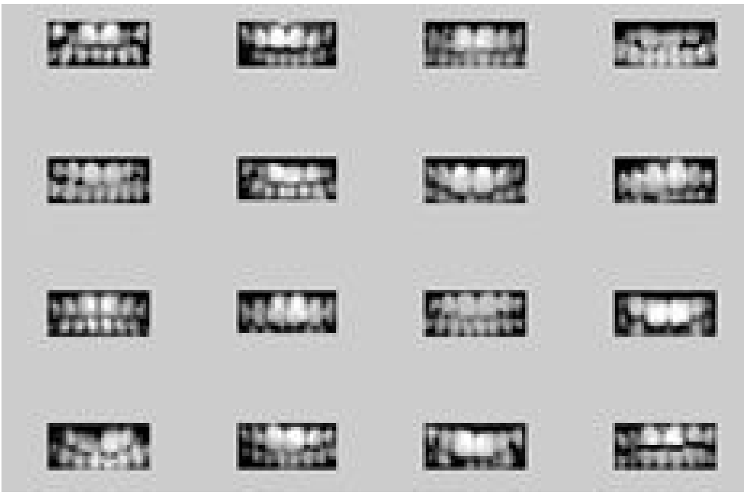
Each teeth image in the training set can represented exactly in terms of a linear combination of the eigenteeth. The number of possible eigenteeth are equal to the number of teeth images in the training set. However the teeth can also be approximated using only the best eigenteeth those that have the largest eigenvaluses. and which therefore account for the most variance within the set of teeth images.

2. PCA for feature extraction of teeth

Image normalization refers to eliminating image variations (such as noise or illumination). Image normalization can be a useful preprocessing stage to improve significantly the accuracy of recognition.



(a)



(b)

Fig. 4.10 Illustration of sample teeth and normalizing teeth images (a) Training teeth image (b) Normalization of training teeth images

Let the training set of teeth images be Γ1, Γ2, ⋯⋯, ΓM, then the average of the set is defined by eq. (4.10).

$$\psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \tag{4.10}$$

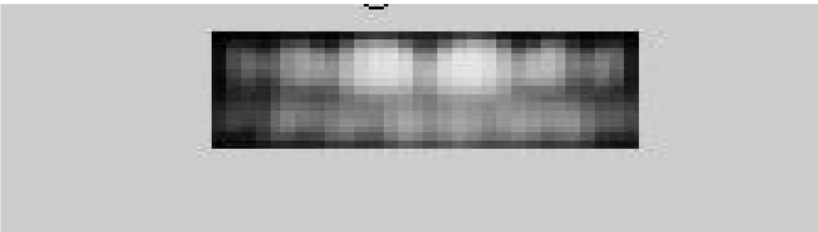


Fig. 4.11 Average teeth image

Each teeth differs from the average by the vector Γ_i and the average ψ is determined by.

$$\Phi_i = \Gamma_i - \Psi \quad . \tag{4.11}$$

This set of very large vectors is subjected to principle component analysis which seeks a set of K orthonormal vectors *υk*, k = 1,⋯⋯⋯.,K and their associated eigenvalues λk which best describe the distribution of data. The vectors *υk* and scalars λk are the eigenvectors and eigenvalues of the covariance matrix.

$$C = \frac{1}{M}\sum_{n=1}^{M}\Phi_n\Phi_n^T = AA^T \quad , \tag{4.12}$$

where is matrix $A = [\Phi_1\,\Phi_2\,\Phi_3 \ldots\ldots\, \Phi_M]$. Finding the eigenvectors of matrix C is computationally intensive. The eigenvectors of in our system can be determined by principal components for best performance and taking a linear combination of the resulting vectors.

$$U_l = \sum_{k=1}^{M} v_{lk}\Phi_k,\ l = 1,\ldots\ldots\ldots M \tag{4.13}$$

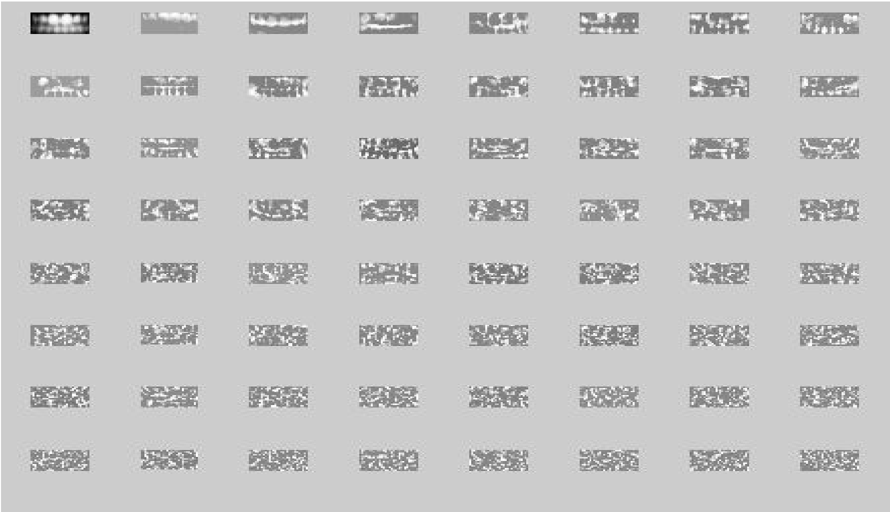
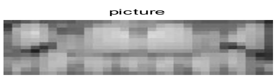


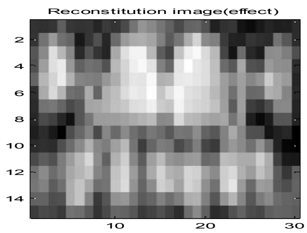
Fig. 4.12. Eigenspace of teeth images

3. Classification

A new teeth image (Γ) is transformed into its eigenteeth components by a simple operation.

$$w_k = U_k^T(\Gamma - \Psi) \tag{4.14}$$





(a) (b)

Fig. 4.13 Example of test and reconstruction teeth image
(a) Test teeth image (b) Reconstruction teeth image

For k=1,⋯.M. The weights form a projection vector,

$$\Omega^{T} = [w_1 w_2 \ldots\ldots w_{M'}] \tag{4.15}$$

Given a set of teeth classes ε_k and corresponding feature vectors Ω_k ,the simplest method for determining which teeth class provides the best description of test teeth image Γ.The projection vector is then used in a pattern recognition algorithm to identify which of a number of predefined teeth classes.

This comparison is based on Euclidean distance between the training teeth classes and the test teeth image. This is given in below eq. (4.16). The idea is to find the teeth class k that minimizes the Euclidean distance.

$$\varepsilon_k = \left\| \Omega - \Omega_k \right\| \tag{4.16}$$

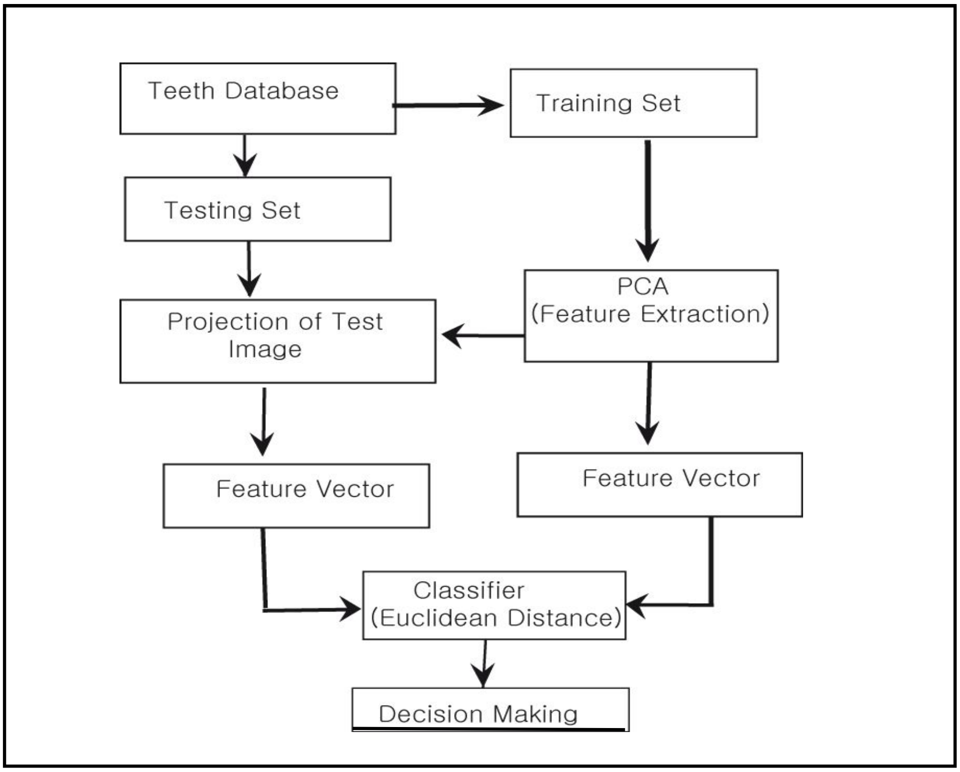


Fig. 4.14 PCA approach for teeth recognition

C. Fisher Discriminates

Linear Discriminant Analysis has been shown to be a powerful tool for pattern recognition in general and for face recognition in particular. In the previous chapter and [31] we have shown that the normalized correlation outperforms the simple Euclidean metric score. In this chapter the issue of matching score in the LDA space is revisited. The reason behind the success of the normalized correlation will be established. The understanding gained about the role of metric will then naturally lead to a novel way of measuring the distance between a probe image and a representative of the hypothesized class.

Fisher discriminates group images of the same class and separates images of different classes. Images are projected from N-dimensional space (where N is the number of pixels in the image) to C-1 dimensional space (where C is the number of classes of images). For example, consider two sets of points in 2-dimensional space that are projected onto a single line Fig. 4.15 (b). Depending on the direction of the line, the points can either be mixed together or separated Fig. 4.15(c). Fisher discriminates find the line that best separates the points. To identify a test image, the projected test image is compared to each projected training image, and the test image is identified as the closest training image.

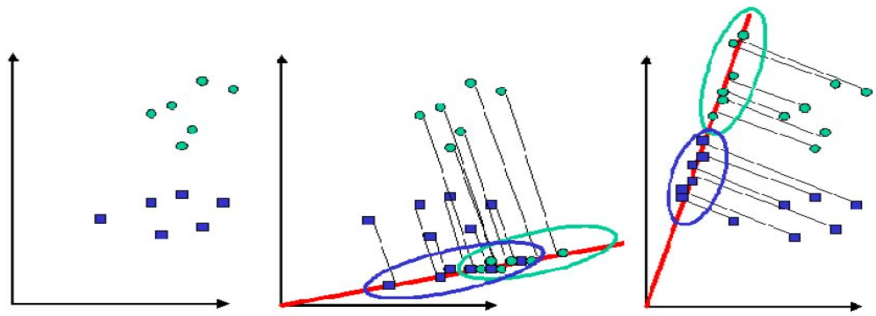


Fig. 4.15 Example of 2-dimensional space for linear projection (a) Points in 2-dimensional space (b) Points mixed when projected onto a line (c) Points separated when projected onto a line

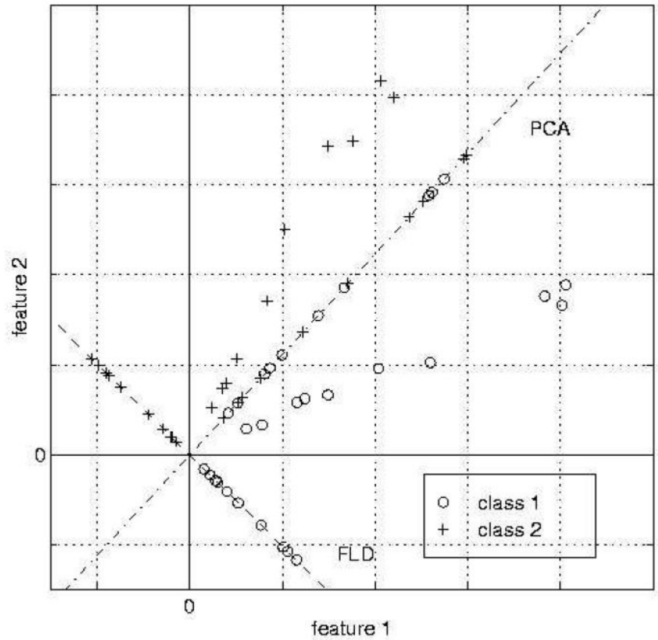


Fig. 4.16 A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace

1. Fisher Discriminants Tutorial (Original Method)

As with eigenspace projection, training images are projected into a subspace. The test images are projected into the same subspace and identified using a similarity measure. What differs is how the subspace is calculated. Following are the steps to follow to find the Fisher discriminants for a set of images.

Calculate the within class scatter matrix: The within class scatter matrix measures the amount of scatter between items in the same class. For the *ith* class, a scatter matrix (S_i) is calculated as the sum of the covariance matrices of the centered images in that class.

$$S_i = \sum_{x \in X_i} (x - m_i)(x - m_i)^T \tag{4.17}$$

where m_i is the mean of the images in the class. The within class scatter matrix S_w is the sum of all the scatter matrices.

$$S_W = \sum_{t=1}^{C} S_t \tag{4.18}$$

.

Calculate the between class scatter matrix: The between class scatter matrix S_B measures the amount of scatter between classes. It is calculated as the sum of the covariance matrices of the difference between the total mean and the mean of each class.

$$S_B = \sum_{t=1}^{C} n_t (m_t - m)(m_t - m)^T \tag{4.19}$$

where n_t is the number of images in the class, m_i is the mean of the images in the class and m is the mean of all the images.

Solve the generalized eigenvalue problem: Solve for the generalized eigenvectors (V) and eigenvalues (Λ) of the within class and between class scatter matrices.

$$S_B V = \Lambda S_W V \tag{4.20}$$

Keep first C-1 eigenvectors: Sort the eigenvectors by their associated eigenvalues from high to low and keep the first *C-1* eigenvectors. These eigenvectors form the Fisher basis vectors.

Project images onto Fisher basis vectors: Project all the original (i.e. not centered) images onto the Fisher basis vectors by calculating the dot product of the image with each of the Fisher basis vectors. The original images are projected onto this line because these are the points that the line has been created to discriminate, not the centered images.

2. Fisher Discriminates Tutorial (Orthonormal Basis Method):

Two problems arise when using Fisher discriminant. Firstly, the matrices needed for computation are very large, causing slow computation time and possible problems with numeric precision. Second, since there are fewer training images than pixels, the data matrix is rank deficient. It is possible to solve the eigenvectors and eigenvalues of a rank deficient matrix by using generalize singular value decomposition routine, but a simpler solution exists. A simpler solution is to project the data matrix of training images into an orthonormal basis of size PxP (where P is the number of training images). This projection produces a data matrix of full rank that is much smaller and therefore decreases computation time. The projection also preserves information so the final outcome of Fisher discriminants is not affected. Following are the steps to follow to find the Fisher discriminants of a set of images by first projecting the images into any orthonormal basis.

3. LDA-based Teeth Classifier

A two-dimensional teeth image is considered as a vector, by concatenating each row (or column) of the image. Let $X = (x_1, x_2, \cdots\cdots x_N)$ denote the data matrix, where N is the number of teeth images in the training set. Each x_i is a teeth vector of dimension $\boldsymbol{n}$, concatenated from a $P \times P$ teeth image, where n represents the total number of pixels in the teeth image and $n = P \times P$. The Linear Discriminant Analysis (LDA) [17, 18] representation is a linear transformation from the original image vector to a projection feature vector, i.e.

$$Y = W^T_{LDA}X \qquad (4.21)$$

where Y is the $d \times N$ feature vector matrix, d is the dimension of the feature vector, $d<=n$ and W_{LDA} is the transformation matrix derived by

$$W_{LDA} = \frac{|W^T S_B W|}{|W^T S_w W|} \quad , \tag{4.22}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix shown as:

$$S_B = \sum_{i=1}^{c} N_i (x_i - m)(x_i - m)^T , \tag{4.23}$$

and

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad . \tag{4.24}$$

In the above expression, N_i is the number of training samples in classes i; c is the number of distinct classes; m is the mean vector of all the samples, i.e., $m = \sum_{i=1}^{N} x_i : m_i$ is the mean vector of samples belonging to class i and Xi represents the set of samples belonging to class i. In the teeth recognition problem, if the within-class scatter matrix S_W is singular, due to the facts that the rank of S_W is at most (N-c) and the number of training samples is generally less than the dimensionality of the teeth image (number of pixels), [19] PCA can be used to reduce the dimensionality of the original teeth image space prior to applying LDA. LDA derives alow dimensional representation of a high dimensional teeth feature vector space. The teeth vector is projected by the transformation matrix W_{LDA}. The projection coefficients are used as the feature representation of each teeth image. Testing was carried out by using the nearest-neighbor algorithm using the standard L_2-norm for the euclidean distance.

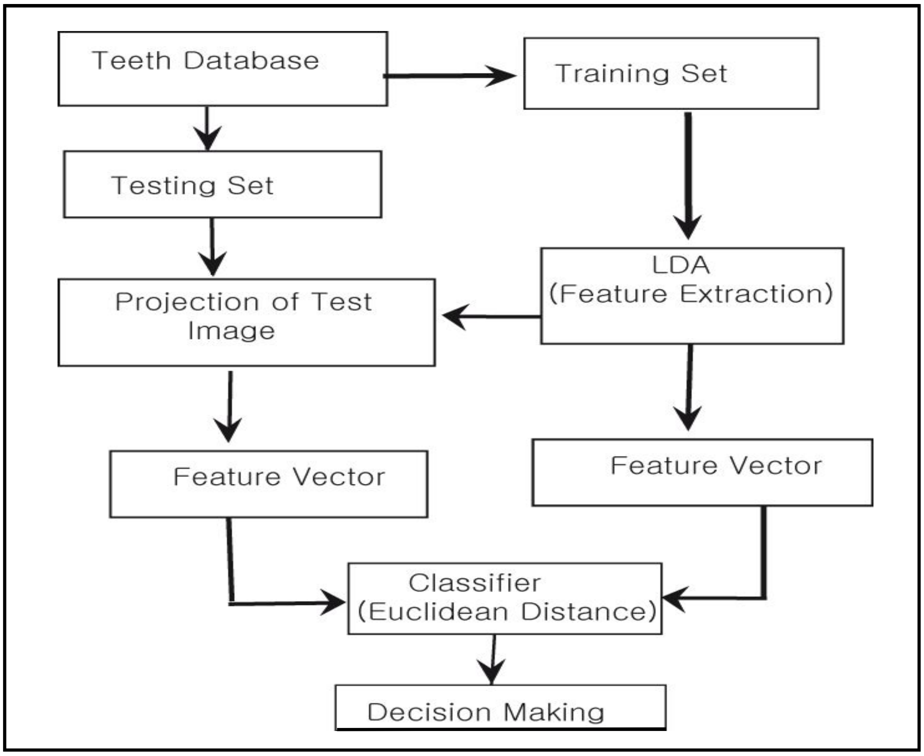


Fig. 4.17 LDA approach for teeth recognition

C. Experimental Results

The results to be presented in this section were obtained using the teeth database which was prepared in the HCI laboratory of Chosun University. This database consists of over 120 images of the frontal images of teeth of 20 subjects. There are 6 different images for each subject. For each subject, these images were recorded in one month, each session consisting of 3 images. For illustration, these images for one subject are shown in Fig.4.1. All images were taken by the same camera under tightly controlled conditions of illumination and viewpoint. Each image in the database consists of a 15x30 array of pixels. For the experiments reported in this section were randomly selected from this database.

1. Experiment I

Small training data set

To simulate the effects of a small training data set, our results here use two teeth images per person for training and two for testing. For example, subject in Fig.4.7 and Fig.4.8. There are only four unconcluded images for each subject, there are obviously many different ways a total of ten different ways of separating the data into the training and the testing parts for the results show below.

To each of the 10 different training and testing dataset created in the manner described above, these method are applied i. PCA and LDA algorithm using the standard L_2 - norm for the euclidean distance. The dataset were indexed 1,2,⋯10, and the test results for the i^{th} dataset were represented by **test#i**. In Fig.4.18, we have shown the results for **test#3**, **test#5**, and **test#8.**

We choose **test#4** and **test#9** for displaying Fig. 4.19 because each represents a different type of comparative performance from the two algorithms tested. The performance curves for **test#4** are typical of the data sets for which PCA outperformed LDA. The performance curves for **test#9** are typical for those data sets for which PCA proved to be superior to LDA for some values of the dimensionality and inferior for PCA.

In Fig.4.18 experiments have focused on only low-dimensional spaces because it has to make a comparison of the most discriminant features for the LDA case with the most descriptive (in the sense of pacing the most 'energy') features for the PCA case.

The dimensionality of LDA is upper-bounded by c-1, where c is the number of classes, since that is the rank the $S_w^{-1}S_b$ matrix. Since we used 20 classes, this gives us an upper bound of 19 for the dimensionality of the LDA space.

The dimensionality of the underlying PCA space cannot be allowed to exceed N-c where N is the total number of samples available. This is to prevent S_w from becoming singular. Since we used 40 samples and since we have 20 classes, the dimensionality of the underlying PCA space cannot be allowed to exceed 20.

Since it makes no sense to extract a 19 dimensional LDA subspace out of a 20 dimensional PCA space, we arbitrarily hard limited the dimensionality of the LDA space to 14.

Table 4.1 summarizes the result for all 10 cases of training and testing data sets for the case of low dimensionality. And table 2 does the same for the case of high-dimensionality.

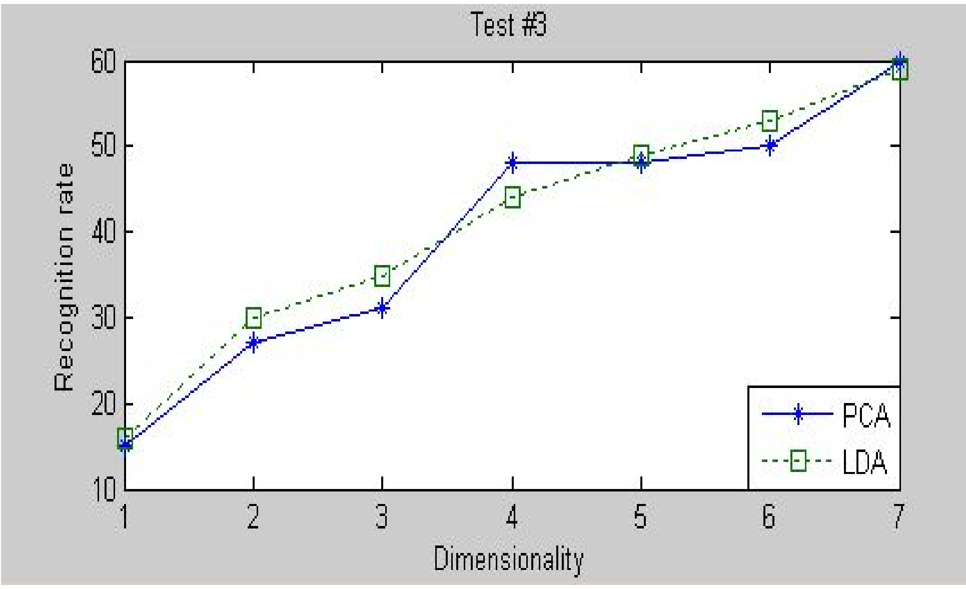
Table 4.1 Experiment I results: different training and testing subsets for the value of the dimensionality parameters

Method	f=2	f=3	f=5	f=6	f=7
PCA	3	4	5	7	6
LDA	7	6	5	3	4

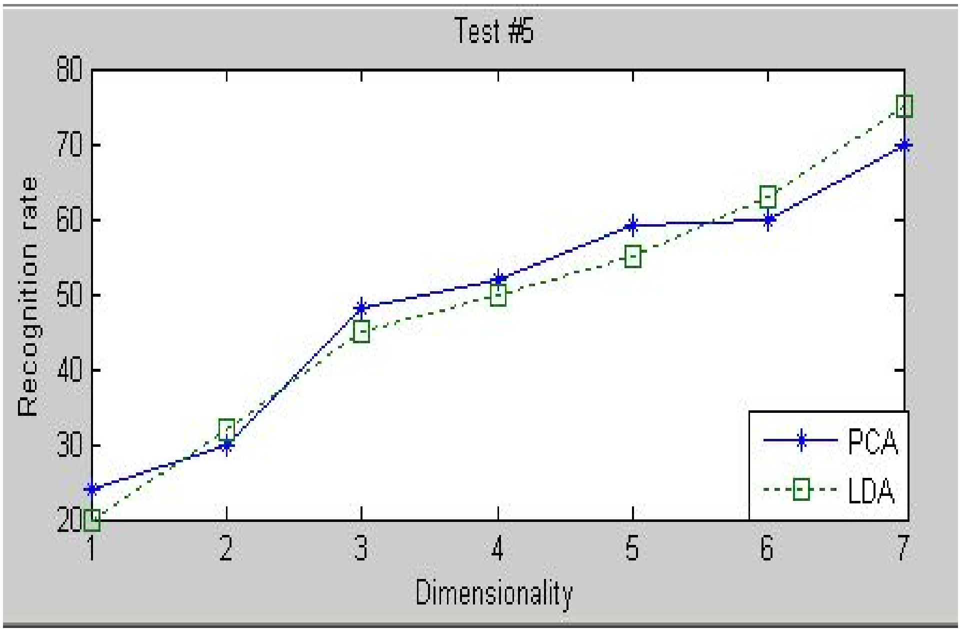
Table 4.2 High-dimensional spaces

Method	f=12	f=13	f=14
PCA	3	3	2
LDA	7	7	8

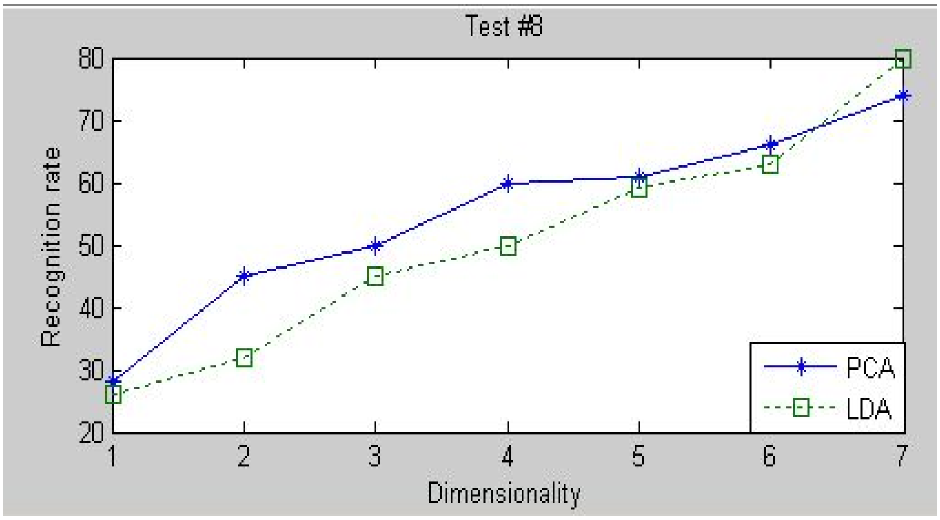
The small training dataset has to do with the relative behavior of PCA and LDA as the dimensionality parameter becomes larger.


Test #3
Recognition rate
Dimensionality
PCA
LDA
60
50
40
30
20
10
1
2
3
4
5
6
7

(a)

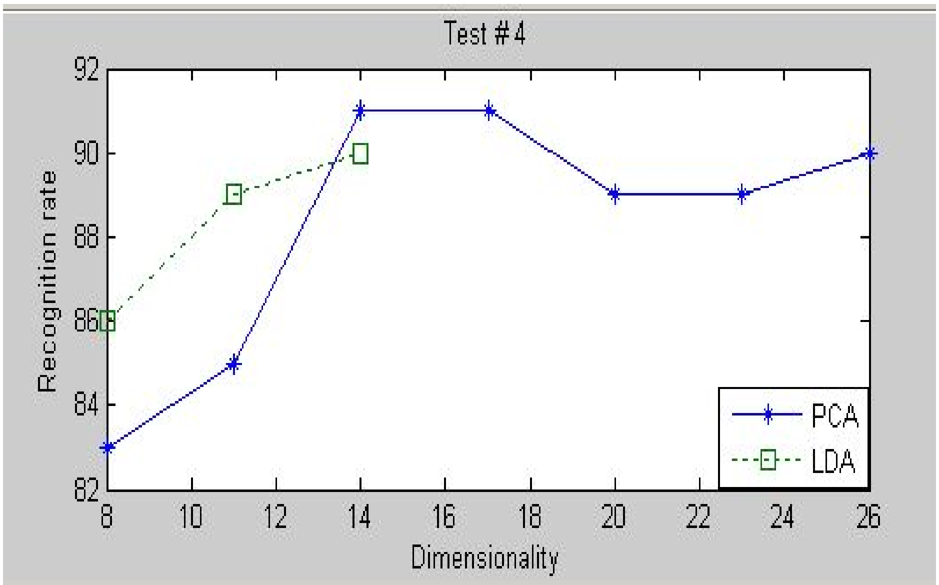

Test #5
Recognition rate
Dimensionality
PCA
LDA
80
70
60
50
40
30
20
1
2
3
4
5
6
7

(b)

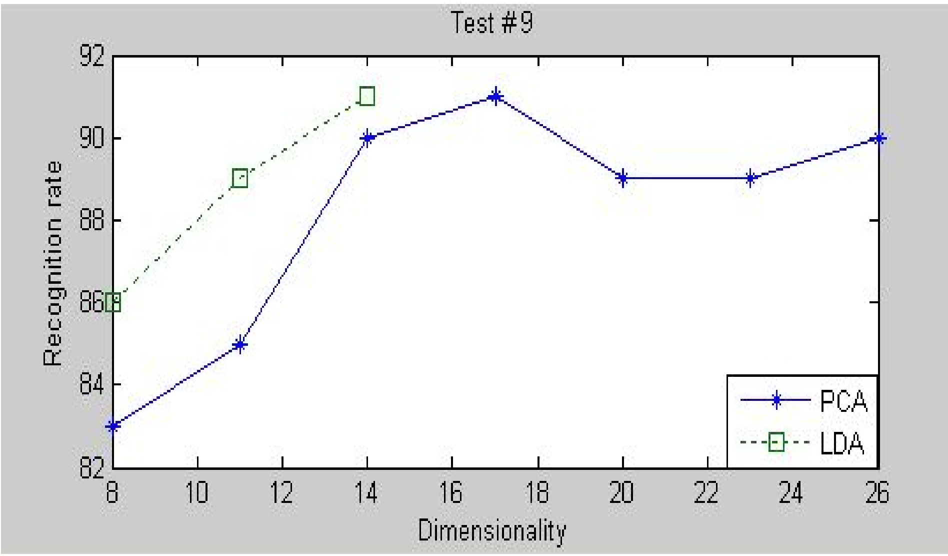


(c)

Fig. 4.18 Summary of the experimental results in low dimensional space.
(a) Performance curves for PCA and LDA
(b) Comparison PCA and LDA curve.
(c) Recognition rate of PCA and LDA



(a)



(b)

Fig. 4.19 The recognition rate on data set of high dimensional case
(a) Recognition rate of PCA and LDA (b) Performance curves of PCA and LDA

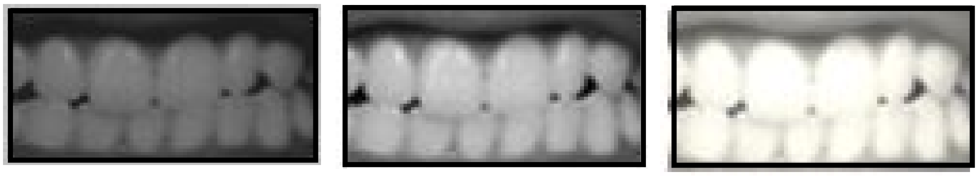
For high-dimensional space, we can draw comparable conclusions, except, that LDA has a greater chance of outperforming PCA for our data set. But note that this conclusion applied only to the specific data set used by us for the experiments reported here. One may end up with an entirely different conclusion for a different data set.

Small training data sets has to do with the relative behaviour of PCA and LDA as dimensionality parameter f becomes larger. The performance of both transforms gets better as the value of f increases.

2. Experiment II

Illumination variation test methodology

In second category of experiments, different illumination variation of the teeth images, PCA and LDA algorithm were used in the variation in illumination of teeth images. Teeth images were obtained for three different cases. The training teeth images were variation in illumination. Three different databases were formed with these training images. Three different illumination applied are shown in Fig. 4.20.



(A) (B) (C)

Fig. 4.20 Teeth image in illumination variation

Case A: Dark teeth image

Case B: Normal teeth image

Case C: Bright teeth image

Table 4.3 Experiment II results: estimation with illumination variation teeth images

Method	No. of Training teeth images	Case A (Classification Accuracy)	Case B (Classification Accuracy)	Case C (Classification Accuracy)
PCA	60	79	86	93
	120	83	85	91
LDA	60	75	82	85
	120	80	82	88

The experiments were done with 60 and 120 images for each case using PCA and LDA. The recognition results, using the closest distance image measure are given in Table 4.3. In most of the cases, PCA estimated near the correct teeth image, but LDA is more sensitive and it may not even estimate the correct teeth images. Whenever the number of training images in increased to 120, PCA success rates are decreased but according to LDA, PCA is still good result than LDA. PCA's success rates are better than that of LDA on illumination varying teeth images.

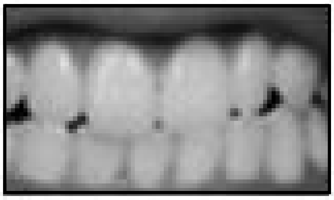
3. Experiment III

Posterior and Anterior based test methodology

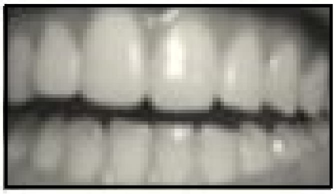
In the third category of experiments, different teeth images at anterior and posterior occlusion expression. Two different databases were formed with these training images Teeth images were obtained for the two different cases.

i. Anterior: Anterior teeth are those located in the front of the mouth, the incisors, and the cupids. Normally, these are the teeth that are visible when a person smiles.

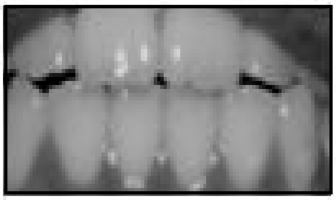
ii. Posterior: The posterior teeth are those located in the back of the mouth-the bicuspids and molars



Normal image



(a) Anterior image



(b) Posterior image

Fig. 4.21 Example of occlusion images (a) Posterior teeth image (b) Anterior teeth image

Table 4.4 Experiment III results: estimation with posterior and anterior teeth

Method	No. of training Teeth images	Case A (Classification Accuracy)	Case B (Classification Accuracy)
PCA	20	79	91
	40	83	89
LDA	20	75	82
	40	78	80

Experiments were done with 20 and 40 training images for each case using PCA and LDA. The recognition results, using the closest distance image measure are given in Table 4.4. Using 20 training image with Posterior teeth show that PCA is not very sensitive than LDA. In most of the cases PCA estimates the correct teeth image, but LDA is more sensitive to Posterior teeth images. However the anterior of training images have less sensitive. PCA is still less sensitive to anterior images. In general, the sensitivity of the two algorithms on posterior and anterior whenever number of training images is increased. PCA's success rates are better than LDA. PCA is slightly better than LDA on posterior teeth images and anterior teeth images .

V. Conclusion

This thesis evaluated the performance of appearance-based statistical methods which are PCA and LDA. Both methods are tested and compared for the recognition of human teeth images. PCA and LDA extract features by projecting the parameter vectors into a new feature space through a linear transformation matrix. But both methods optimize the transformation matrix with different intentions. PCA optimizes the transformation matrix by finding the largest variation in the original feature space. LDA pursues the largest ratio of between-class variation and within-class variation when projecting the original feature space to a subspace. PCA and LDA are well established techniques especially for face recognition but in this study, these algorithms are first time ever used for performance evaluation using human teeth. These methods were applied on newly constructed teeth database throughout this study. Euclidean distance was employed to classify the teeth images using these features.

This experimental setup yielded three different methodologies which were conducted for relative performance evaluations. In the first set of experiments, the recognition performances of PCA and LDA are demonstrated. To simulate the effects of a small training data set, two teeth images per person for training and two for testing are used. LDA result shows less accuracy especially when number of training images per person is not adequate. In the second set of experiments, three illumination variations teeth images were utilized. Experimental results show that the PCA algorithm is better than LDA algorithm under different illumination variations. In the third set of experiments, teeth images are anterior and posterior occlusion. LDA is more sensitive than PCA on anterior and posterior occlusions based test methodology. PCA performance is better than LDA while performance on anterior and posterior occlusions.

The experimental results prove that PCA shows more than 90% accuracy result, whereas LDA shows only 83% under the same testing condition.

References

[1] N.A.Campbell, Shrunken, “Estimators in Discriminate and Canonical Variate Analysis”, Applied Statistics, Vol. 29, No. 1, pp. 5-14, 1980

[2] C.R Rao, “The use and interpretation of principal component analysis in applied research”, Sankhya A, 26, pp 329-358, 1964.

[3] T.J Hastie, R.Tibshirani, “Flexible Discriminant Analysis by Optimal Scoring”, AT and T Bell Labs Technical Report, December, 1993

[4] Xuechuan, Wang and Kuldip K. Paliwal “Feature extraction and dimensionality reduction algorithms and their application in vowel recognition” pattern Recognition, Vol. 36 pp. 2429-2439, 2003

[5] W.L poston and D.J Marchette, “Recursive Dimensionality Reduction Using Fisher's Linear Discriminant”, Pattern Recognition, Vol. 31, No. 7 pp. 881-888, 1998

[6] B.H Juang and S.Katagiri, “Discriminative Learning for Minum Error Classification”, IEEE Transactions on Signal processing, Vol 40, No. 12, December, 1992

[7] V.Vapnik, The nature of statistical Learning Theory, Springer, N.Y, 1995

[8] V.Roth and V.steinhage, “Nonlinear discriminant analysis using kernel functions”, Technical Report, Nr, IAI-TR-99-7, ISSN 0944-8535, University Bonn, 1999

[9] B.Scholkopf, C. Gurges and V.VApnik, “Incorporating invariance in support vector learning machines”, International Conference on Artificial Neural Networks-ICANN 96, pp. 47-52, Berlin, 1996

[10] Turk M. A. and Pentland P., “Face Recognition Using Eigenfaces”, Proc. Of IEEE Conference on Computer Visionand Pattern Recognition, pp. 586-591, 1991.

[11] Kirby M., “Dimensionally of Reduction and Pattern Analysis an empirical approach.” Under contract with Wiley, 2000.

[12] Kirby M. and Sirovich L., “Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces,” *IEEE Trans. PAMI*, vol. 12, no.1, pp. 103-108, 1990.

[13] Sirovich L. and Kirby M., “A low-dimensional procedure for the characterization of human faces,” The Journal of the Optical Society of

America, vol. 4, pp. 519-524, 1987.
[14] Yambor W., Draper B., and Beveridge J.R., “Analysis of PCA-based Face Recognition Algorithms: Eigenvector Selection and Distance Measures,” Second Workshop on Empirical Evaluation Methods in Computer Vision, 2000
[15] Nayar S., Nene S., and Murasr H., “Real-Time 100 Object Recognition System,” Proceedings of ARPA Image Understanding Workshop. 1996.
[16] Fisher, R.A., “The use of Multiple Measures in Taxonomic Problems,” Ann. Eugenics, vol.7, pp. 179-188, 1936.
[17] Duda R. and Hart P., “Pattern Classification and Scene Analysis”, NewYork: JohnWiley & Sons, 1973.
[18] Jain, A.K., Chen, H., “Matching of dental X-ray images for human identification,” Pattern Recognition vol. 37, no. 7, 2004.
[19] Chen, H., Jain, A.K., “Teeth contour extraction for matching dental radiographs,” proc. 17th ICPR (3) Cambridge, UK, August, pp. 522-525, 2004.
[20] Ammar, H.Hany., Nassar, Diaa Eldin M., “A neural network system for dental radiograph comparison,” Proc. of the end. IEEE international symposium on signal processing and information technology, Marrakwesh-Moroco, Dec, 2002.
[21] Zhou, J., Abdel-Mottaleb, M., “Automatic human identification based on dental X-ray images,” Proc. Of the SPIE technologies for homeland security and low enforcement, the biometric technology for human identification conference, Orlando, FL, Gesture Recognition, pp. 336-341, 2005.
[22] Prajuabklang, K., Kumhom, P., Maneewarn, T. and Chamnongthai,K., 2004, “Real-time Personal Identification from Teeth-image using Modified PCA,” Proceeding, the 4th information and computer Engineering Postgraduate Workshop, Vol. 4, No. 1, pp. 172-175
[23] Shin, Young-suk., “Gender Identification on the Teeth Based on Principal Component Analysis Representation,” Springer 4th International Conference, AMDO 2006.
[24] Kim, Tae-Woo., Cho, Tae-Kyung., “Teeth image recognition for Biometrics”, IEICE Trans D, Information, March 2006.
[25] K. pearson., “On lines and planes of closet fit to systems of points in space”, Hill, mag, No 6, Vol 2, pp. 559-572, 1901
[26] M. Hotelling, “Analysis of a complex of statistical Variables onto Principal Components”, Jornal of Edvcational Psychology, 24, pp. 498-520, 1933

[27] M. A, Girshick, “On the sampling theory of roots of determinant equation”, Ann. Math. Statist. 10, pp. 203-224, 1939

[28] T.W. Anderson, “Asymptotic theory for principal component analysis”, Ann. Statist, Section, 3 pp. 77-95, 1963

[29] C.R. Rao, “The use and interpretation of principal component analysis in applied research”, Sankhya, A, 26, pp. 329-358, 1964.

[30] V.Vapnik, “The nature of statistical Learning Theory”, Springer, N.Y, 1995

[31] V.Roth and V.steinhage, “Nonlinear discriminant analysis using kernel functions”, Technical Report, Nr, IAI-TR-99-7, ISSN 0944-8535, University Bonn, 1999

[32] B.Scholkopf, C. Gurges and V.VApnik, “Incorporating invariance in support vector learning machines”, International Conference on Artificial Neural Networks-ICANN 96, pp. 47-52, Berlin, 1996

[33] J. C. Gower, “Some distance properties of latent root and vector methods used in multivariate analysis”, Biometrika, 53, pp. 325-338, 1966

[34] R.O dula, P.E hart and D. Gstork “Pattern Classification and scene Analysis” John Wiley and Sons press, New york, 1973

[35] W.J. Krsanowski, “Principal component analysis in the presence of group structure”, Applied statistics, 33, pp. 164-168, 1984