İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

EAR BIOMETRICS IN PERSONAL IDENTIFICATION

M. Sc. Thesis by

Bahattin KOCAMAN, B.Sc.

Department : Electronics and Communication Engineering

Programme : Electronics Engineering

FEBRUARY 2008

İSTANBUL TECHNICAL UNIVERSITY ★ INSTITUTE OF SCIENCE AND TECHNOLOGY

EAR BIOMETRICS IN PERSONAL IDENTIFICATION

M. Sc. Thesis by
Bahattin KOCAMAN, B.Sc.
504051203

Date of submission : 24 December 2007
Date of defence examination: 28 January 2008

Supervisor: Assoc. Prof. Dr. Mürvet KIRCI

Members of the Examining Committee Prof. Dr. Bilge GÜNSEL

Assoc. Prof. Dr. Zehra ÇATALTEPE

FEBRUARY 2008

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

KİMLİK TESPİTİNDE KULAK BİYOMETRİĞİ

YÜKSEK LİSANS TEZİ

Müh. Bahattin KOCAMAN

504051203

Tezin Enstitüye Verildiği Tarih : 24 Aralık 2007

Tezin Savunulduğu Tarih : 28 Ocak 2008

Tez Danışmanı : Doç. Dr. Mürvet KIRCI

Diğer Jüri Üyeleri Prof. Dr. Bilge GÜNSEL

Doç. Dr. Zehra ÇATALTEPE

ŞUBAT 2008

ACKNOWLEDGEMENTS

First and foremost I would like to thank Assoc. Prof. Dr. Mürvet KIRCI, my supervisor, who helped me to complete this thesis and was available whenever I need help. Also I would like to thank Prof. Dr. Ece Olcay GÜNEŞ for her helps.

Special thanks to my father Bahaettin KOCAMAN and my mother Firdes KOCAMAN. Their continuous support that brought me to this point.

December 2007 **Bahattin KOCAMAN**

CONTENTS

ABBREVIATIONS v
LIST OF TABLES vi
LIST OF FIGURES vii
ÖZET viii
SUMMARY ix

1. INTRODUCTION 1

1.1 Applications Of Ear Recognition 2
1.2 Stages In Ear Recognition 2

2. WHY EAR RECOGNITION 4

2.1 History of Ear Recognition 5
2.2 Ear Recognition and Ear Detection 6

3. EAR RECOGNITION ALGORITHMS 8

3.1 Principal Component Analysis (PCA) 8
3.1.1 Theory of PCA 8
3.1.2 How to Use PCA on Ear Images 11
3.1.2.1 Ear Is Viewed As A Vector 12
3.1.2.2 Statistical PCA 15
3.2 Fisher's Linear Discriminant (FLD) 21
3.2.1 Implementation of FLD 27
3.2.2 Multiple Discriminant Analysis 30
3.3 Small Sample Size Problem 35
3.3.1 Problems with PCA 35
3.3.2 Problems with FLD 35
3.4 Discriminative Common Vector Approach (DCVA) 36
3.4.1 Obtaining The Discriminative Common Vectors By Using The Range Space Of S_W 37
3.5 Laplacianears 41
3.5.1 Learning a Locality Preserving Subspace 43
3.5.2 Statistical View of LPP 45
3.5.3 Learning Laplacianears for Representation 47

4. EXPERIMENTS USING İTU-E DATABASE 49

4.1 Testing PCA by Random Subsampling 50
4.2 Testing PCA by K-Fold Cross-validation 51
4.3 Testing FLD by Random Subsampling 52
4.4 Testing FLD by K-Fold Cross-validation 53
4.5 Testing DCVA by Random Subsampling 54
4.6 Testing DCVA by K-Fold Cross-validation 54

4.7 Testing Laplacianears by Random Subsampling 55
4.8 Testing Laplacianears by K-Fold Cross-validation 56
4.9 Error Rate vs. Dimension Number 57
4.10 Effects of Cropping 65
5. CONCLUSION **73**
REFERENCES **74**
RESUME **77**

ABBREVIATIONS

PCA	: Principal Component Analysis
FLD	: Fisher's Linear Discriminant
LDA	: Linear Discriminant Analysis
DCVA	: Discriminative Common Vector Approach
LPP	: Locality Preserving Projections
LLE	: Locally Linear Embedding
ERT	: Ear Recognition Techniques
İTU-E	: İstanbul Technical University Ear Database

LIST OF TABLES

Page Number

Table 3.1 Dataset Divided into 2 Classes 27
Table 4.1 Random Subsampling Hit Rates of PCA 50
Table 4.2 K-Fold Cross-validation Hit Rates of PCA 51
Table 4.3 Random Subsampling Hit Rates of FLD 52
Table 4.4 K-Fold Cross-validation Hit Rates of FLD 53
Table 4.5 Random Subsampling Hit Rates of DCVA 54
Table 4.6 K-Fold Cross-validation Hit Rates of DCVA 54
Table 4.7 Random Subsampling Hit Rates of Laplacianears 55
Table 4.8 K-Fold Cross-validation Hit Rates of Laplacianears 56
Table 4.9 Hit Rates of PCA Over Cropped Left Ear Images 66
Table 4.10 Hit Rates of PCA Over Masked Left Ear Images 67
Table 4.11 Hit Rates of PCA Over Cropped Right Ear Images 67
Table 4.12 Hit Rates of PCA Over Masked Right Ear Images 68
Table 4.13 Hit Rates of FLD Over Cropped Left Ear Images 68
Table 4.14 Hit Rates of FLD Over Masked Left Ear Images 69
Table 4.15 Hit Rates of FLD Over Cropped Right Ear Images 69
Table 4.16 Hit Rates of FLD Over Masked Right Ear Images 70
Table 4.17 Hit Rates of DCVA Over Ear Images 70
Table 4.18 Hit Rates of LPP Over Cropped Left Ear Images 71
Table 4.19 Hit Rates of LLP Over Masked Left Ear Images 71
Table 4.20 Hit Rates of LPP Over Cropped Right Ear Images 72
Table 4.21 Hit Rates of LPP Over Masked Right Ear Images 72

LIST OF FIGURES

Page Number

Figure 3.1: Construction of ear vector [1]..... 11
Figure 3.2: Formation of the vector of ear from the image of ear 13
Figure 3.3: Mapping NxN image into N^2 vector [1]..... 13
Figure 3.4: Image (ear) space [1]..... 14
Figure 3.5: Image space and ear cluster [1] 14
Figure 3.6: High-level functioning principle of the eigenear-based ear recognition algorithm 17
Figure 3.7: Eigenear generation process [1] 19
Figure 3.8: The Way that PCA Handles the Dataset [5]..... 21
Figure 3.9: Projection of the Same Set of Samples onto Two Different Lines in the Direction Marked w [5] 22
Figure 3.10: Shows the Projected Point x_i [5]..... 22
Figure 3.11: Stand-alone $\left|\tilde{\mu}_1 - \tilde{\mu}_2\right|$ is not always a Good Measure of Separation [5] 23
Figure 3.12: The Plot of Dataset. "x" donates class 1, "o" donates class 2..... 28
Figure 3.13: Projection Direction of FLD..... 29
Figure 3.14: Shows the Best Possible Projection of Dataset for Classification..... 30
Figure 3.15: The Projection of Three-Dimension onto Two-Dimension [2]..... 33
Figure 4.1: Error Rates of PCA at Left Ear... 57
Figure 4.2: Error Rates of PCA at Right Ear 58
Figure 4.3: Error Rates of FLD at Left Ear..... 59
Figure 4.4: Error Rates of FLD at Right Ear..... 60
Figure 4.5: Error Rates of DCVA at Left Ear..... 61
Figure 4.6: Error Rates of DCVA at Right Ear 62
Figure 4.7: Error Rates of Laplacianears at Left Ear..... 63
Figure 4.8: Error Rates of Laplacianears at Right Ear 64
Figure 4.9: Test images are cropped and renormalized to 128*128 pixels. (a): uncropped. (b): cropped from left. (c): cropped from right. (d): cropped from top. (e): cropped from bottom. (f): cropped from all sides..... 65
Figure 4.10: Black mask applied to test images. (a): unmasked. (b): masked from left. (c): masked from right. (d): masked from top. (e): masked from bottom. (f): masked from all sides.....66

KİMLİK TESPİTİNDE KULAK BİYOMETRİĞİ

ÖZET

Bu tez biyometrik tabanlı, kimlik tespitine dayalı güvenlik sistemi geliştirme fikriyle başlayan çalışmanın bir parçasıdır. Günümüzde, biyometrik tabanlı ve türleri içinde en yüksek doğruluk oranına sahip parmak izi ve iris tarama yöntemleri kriminal vakalarda ve yüksek güvenlik gerektiren tesislerde kullanılmaktatır. Yüz tanıma hala gelişmekte olan bir biyometrik yöntemidir, fakat yapılan literatür araştırmalarında ortam ışıklandırması, makyaj, verilen poz, duygusal ifadeler ve estetik operasyonlar gibi yüz görünümü üzerinde etkisi olan faktörlerin yüz tanıma probleminde doğrudan yöntemlerin başarımını azaltacak yönde etkili olduğu görülmüştür. Dolayısıyla, yüz gibi erişimi kolay fakat onun gibi gündelik hayatın makyaj, duygusal ifadeler, bıyık ve sakal bırakma gibi faktörlerinden etkilenmeyecek bir biyometrik gereksinimi ortaya çıkmıştır. Alternatif biyometriğin başarımının yüzle kıyaslanabilir mertebelerde olması gerektiği açıktır. Araştırmaların devamında, tek yumurta ikizlerinin birbirlerine ne kadar benzeseler de kulak yapılarının farklı olduğu, kulağın 3 boyutlu olsa da yüz kadar detay içermediği ve kulağın yapısı itibariyle duygusal açılımlar ifade edilirken biçimini değiştirmediği görülmüştür. Bunların ışığında, kulak yüze karşı güçlü bir alternatif biyometrik olarak ortaya çıkmaktadır. Bu çalışmada, literatürde önerilen yöntemler kulak resimleri üzerine uygulanmıştır. Bu yöntemler veri kümesi olarak 2 boyutlu resimleri kullanan ve veri kümesi üzerinde sınıflandırma yapan, lineer yöntemlerdir. Yapılan çalışma sonunda görülmüştür ki, PCA, FLD, FLD'nın geliştirilmesiyle oluşturulan DCVA ve LPP yöntemlerinin kulak tanımadaki başarımları yüz tanımadaki başarılarından daha yüksektir. Bu yöntemlerin kulak tanımadaki doğru eşleştirme oranları, literatürde bulunan, yüz tanımadaki eşleştirme oranlarıyla karşılaştırıldıklarında daha yüksektir. Yapılan bu çalışmanın sonuçları biyometrik tabanlı kimlik tesbit yöntemleri için kulağın yüzden daha iyi bir alternatif olduğunu göstermiştir.

EAR BIOMETRICS IN PERSONAL IDENTIFICATION

SUMMARY

This thesis is one of the parts of a biometric based identity verification security system development project. Today, the most successful biometric based identification technologies such as fingerprint and iris scan are used worldwide in both criminal investigations and high security facilities. Face recognition is one of the developing biometric methods; however illumination, makeup, posing, emotional expressions and face-lifting reduce the success of face recognition. A new biometric which is not effected by any of the factors above is needed. The new biometric should be as successful as face recognition. Twins are identical but their ears differ from each other, ear is also 3-dimensional but it is simpler than face and emotional expressions do not affect the ear. In the light of this, ear is a good alternative to face, as a biometric. In this study, the methods presented in the literature are tested on ear images. These methods are linear classification algorithms that work on 2D image databases. It is found out that, PCA, FLD, modified FLD which is also known as DCVA and LPP has better results at ear recognition than face recognition. Ear recognition has higher hit rates, when compared with face recognition researches that are presented in the literature previously. The results of this study proved that ear is the best alternative to face at personal identification tasks.

1. INTRODUCTION

Ear recognition is considered to be a part of pattern recognition technology. Ear recognition and recognition of moving people in natural scenes require a set of visual activities to be performed. This process consists of three tasks: detection, normalization and recognition. Detection means the detection and tracking of ear-like image patches in dynamic scenes. Normalization is the segmentation, alignment and normalization of the ear images and finally recognition is the representation and modeling of ear images as identities, and the association of novel ear images with known models [1].

A number of approaches for recognition and classification tasks have been proposed in the literature. These can be classified as Principal Components Analysis (PCA) [2,3,4], Fisher's Linear Discriminant (FLD) [2,5,6], Discriminative Common Vectors (DCV) [7] and Locality Preserving Projections (LPP) [8,9,10].

Techniques like PCA and FLD treat the ear image as a vector in a high-dimensional space and derive a lower dimensional representation (in the case of PCA) or a discriminatory representation (in the case of FLD). FLD provides a better performance but it is computationally more intensive compared to feature-based approaches. Also, the performance of data analysis techniques depends on the training data. Discriminative Common Vector approach (DCVA) reduces computational cost in recognition stage because a common vector is chosen from each class, instead of dealing with all ears in the dataset. LPP tries to preserve local structure of ear images for classification, however, PCA, FLD and DCVA focus on global structure of ear images [8,9].

Comparison of ear and face, using one or two recognition methods is the topic of previous research papers [11,12]. This study will compare the performance of four most well-known traditional and new recognition algorithms at ear recognition task, for the first time.

1.1 Applications of Ear Recognition

The applications of Ear Recognition Techniques (ERT) can be divided into commercial and law enforcement applications. ATMs, safe deposit boxes and ear recognition based security systems can be thought of commercial applications of ear recognition. There are surveillance cameras at all banks and ATMs. These cameras can be repositioned to record both the faces and ears of customers. In the case of robbery both ear and face images of suspects can be analyzed to identify them. These kinds of precautions make the security policies of banks stronger. Another example is an ear recognition system can be placed at the entrance of a parking lot and the gate opens if the driver is a known person. These kinds of applications can also be used at high-tech facilities where security is mandatory. These are all commercial examples.

If law enforcement applications are considered, there are many wanted people all over the world. If the records of surveillance cameras, which are placed at crowed places like squares, train stations and airports, are processed by the algorithms such as ear and face recognition simultaneously, detecting and arresting these wanted people become easier [13,14]. So world becomes a safer place.

1.2 Stages in Ear Recognition

The definition of ear recognition task is to identify one or more people in a scene using a stored database of ears. The solution of the general problem is divided into three different stages:

- Segmentation of ears from cluttered scenes.
- Extraction of features from the ear region.
- Decision.

Segmentation is usually achieved by the following algorithm. An edge map is created, and then edges are connected together using several heuristics and the edges are matched into an elliptical shape using a Hough transformation. If the input is composed of video images (moving objects), motion could be used for segmentation.

The second and most important stage is the extraction of features. There are two types of features: holistic features and partial features. Partial features techniques use crucial points in the ear for recognition, whereas holistic feature techniques always consider the ear as a whole. For example PCA is a holistic feature technique.

In the last stage, using the data collected in the previous stages a decision is made. There can be three types of decisions that can be made depending on the application: 1. identification: where labels for each individual must be obtained; 2. recognition of a person: where a decision is made based on the ear that the individual has already been seen, and; 3. categorization; in which the ear must be assigned to a certain category [1]. This study will focus on two stages of ear recognition: extraction of features and decision.

2. WHY EAR RECOGNITION

When there is a requirement for determining people's identity, the obvious question is what technology is best at supplying this information. There are many ways that humans can identify each other and it is the same for machines. There are many different identification methods available, many of which have been used in commercial applications for years. The most common type of application is password/PIN known as Personal Identification Number systems. The password/PIN applications such as credit cards are not unique. Swindlers can clone and use credit cards. The owner of credits card can forget his/her password and also lose his/her credit card. The credit card can be stolen. These are most possible and common problems of password/PIN applications. To solve these problems there has been considerable interest in biometric identification systems, which use pattern recognition methods to identify people using characteristic features.

There are many behavioural and physiological characteristics of humans that satisfy the definition of biometric. All of these biometrics researched to find the best solution for biometric based identification problem. Some of them are efficient enough to use in criminal investigations and some of them are still developing.

First research about head area biometrics is done by Bledsoe [15]. Kanade offered some models and face biometric ratios for face classification problem [16,17]. Pentland and Turk offered a 2D image based face recognition method for the first time in 1991 [4]. There are many researches about other biometrics. For example, Petejan developed a recognition system depending on lipreading, during his Ph. D. thesis researches in 1984 [18]. Nishida developed a speech recognition system in MIT in 1986 [19]. Daugman and Downing showed that iris can be the best biometric for human identification [20,21]. Nixon and Carter did research about gait identification [22]. Persuad and Sommerville did a research about human scent for identification [23,24]. There are also researches about hand writing and hand geometry. References [25,26,27] can be checked for more details about hand based biometrics.

There are two types of identification techniques that are active identification techniques and passive identification techniques. However, active identification

techniques are not easy to use and can be intrusive both physically and socially. For example, in bank transactions and entry into secure areas, the user must position the body relative to the sensor and then pause for a second to declare himself or herself since the quality of the image is important. While the pause and present interaction is useful in high-security situation, it is inefficient in other circumstances. So, popularity of passive identification techniques rises. For instance, a store that wishes to recognize its best customers or an information kiosk that remembers you or a house that knows the people who live there does not want to interrupt the individual's daily activities. Passive biometric identification systems have a natural place in these next generation smart environments, they are unobtrusive, do not restrict user movement, and are now both low power and inexpensive.

2.1 History of Ear Recognition

Ear recognition was established when machines started to become more intelligent and were able to fill in correct or help the lack of human abilities and senses. Face and ear recognition and computer vision subjects are both important because of the practical importance of the topic and theoretical interest from cognitive science. Face recognition is not the only method of recognizing images and humans because of some restrictions. Humans also use senses between each other to recognize images and others. Today machines are used for different recognition purposes such as fingerprinting or iris scanning [21,28]. These methods of identification are more accurate than face recognition but face recognition is more interesting for researchers because of its non-invasive nature and because it is the primary method used by humans for identifying people. However, as mentioned above, face recognition comes with some restrictions, such as sensitivity to make up, illumination, posing, the rotation angle of face to up/down/right/left directions and emotional expressions such as smiling and frowning brows, which need to be eliminating during the recognition process. As a result of these restrictions, most of common face image databases, which are used to test face recognition methods, developed under the same illumination conditions, no emotional expressions, no makeup and limited degree of face rotation. These kind of studies that are done under extremely ideal conditions, will have far less accuracy in real world applications. Thinking of this obvious situation directed researchers to other biometrics which naturally do not have emotional expressions, beard (for males), makeup, etc.... Ear is an option for recognition tasks. First, Iannarelli showed that ear is adequate for being a biometric

after early research on a manual approach [29]. Burge and Burger proved that ear has a similar performance like face at personal identification tasks [30,31]. Chang and Bowyer compared PCA technique on both face and ear images and showed that they have similar performance as biometrics [11,12]. Hurley, Nixon and Carter did a research using Forced Field Transformation and Forced Field Feature Extraction techniques to classify ear images [13,32]. Yan and Bowyer did a research using PCA and Hausdorff matching on 2D ear images and ICP on 3D ear images [33,34]. Middendorff, Bowyer and Yan also researched Multi-model biometrics to increase the performance of their previous work [35]. The aim of this thesis is to provide the knowledge of performance differences of 2D image based, linear methods, on same ear image database, to the literature.

At the beginning of the research, there were two main approaches in recognition technologies: 1. geometrical approach and 2. pictorial approach.

The geometrical approach uses the spatial configuration of ear features. That means the main geometrical features of the ear such as the helix, lobule, antihelix, concha, tragus, antitragus, etc... are first located and then ears are grouped or classified on the basis of various geometrical distances and angles between features. The pictorial approach uses templates of the features of ear. It uses templates of the major features and entire ear to perform recognition on frontal views of ears. Many of these studies that are based on these two approaches have some common extensions that handle different pose backgrounds. Different from these two techniques there are other recent template-based approaches, which form templates from the image gradient and the principal component analysis approach, which can be thought of a sub-optimal template approach.

2.2 Ear Recognition and Ear Detection

As discussed earlier, face recognition is a technique that is used for recognizing faces but it is not necessary to freeze the user to take a picture. The problem with recognizing a face arises when the pose of the face is different, but in particular, there is a limit on face rotations in depth, which include left and right and up and down rotations. Face recognition is a difficult task because it has to discriminate among similar objects. To be more specific, when two faces are similar, recognition is going to be a challenge. By adding pose to a face, the problem becomes more complex. The appearance of face changes under rotation since the face has a complex three dimensional structure. Switching face with a simpler biometric will

automatically eliminate some limitations such as pose, emotional expressions and effect of makeup and enhance the performance of recognition algorithms.

Ear recognition and ear detection are two distinct processes. The main difference is that ear recognition is a technique to detect ears and search through a dataset in order to find an exact match, but on the other hand ear detection is looking for any match and as soon as a match is found the search stops.

In ear recognition techniques, three visual cues are most important; motion, color and ear appearance. Most of recognition algorithms are not concerned with color, neither is this study. Grayscale images are used in this study.

Recognition is an area of research, which involves different fields such as biology, mathematics and computer science. Different recognition methods have been proposed in recent years resulting in interesting applications. The research presented here tries to bring a highly compatible alternative for human identification tasks. The results of this research prove that ear is a better option than face. The features of ear are simpler than the features of face which results in raising the performance of recognition methods.

3. EAR RECOGNITION ALGORITHMS

3.1 Principal Component Analysis (PCA)

3.1.1 Theory of PCA

It is needed to represent all of the vectors in a set of n d-dimensional samples $x_1,...,x_n$ by a single vector x_0. To be more specific, suppose that it is desired to find a vector x_0 such that the sum of the squared distances between x_0 and the various x_k is as small as possible. The squared-error criterion function $J_0(x_0)$ is defined by

$$J_0(x_0) = \sum_{k=1}^{n} \|x_0 - x_k\|^2 \quad \textbf{(3.1)}$$

and seek the value of x_0 that minimizes J_0. It is simple to show that the solution to this problem is given by $x_0 = \mu$, where μ is the sample mean,

$$\mu = \frac{1}{n}\sum_{k=1}^{n} x_k \quad \textbf{(3.2)}$$

This can be easily verified by writing

$$J_0(x_0) = \sum_{k=1}^{n} \|(x_0 - \mu) - (x_k - \mu)\|^2 \quad \textbf{(3.3a)}$$

$$J_0(x_0) = \sum_{k=1}^{n} \|x_0 - \mu\|^2 - 2\sum_{k=1}^{n} (x_0 - \mu)^T (x_k - \mu) + \sum_{k=1}^{n} \|x_k - \mu\|^2 \quad \textbf{(3.3b)}$$

$$J_0(x_0) = \sum_{k=1}^{n} \|x_0 - \mu\|^2 - 2(x_0 - \mu)^T \sum_{k=1}^{n} (x_k - \mu) + \sum_{k=1}^{n} \|x_k - \mu\|^2 \quad \textbf{(3.3c)}$$

$$J_0(x_0) = \sum_{k=1}^{n} \|x_0 - \mu\|^2 + \underbrace{\sum_{k=1}^{n} \|x_k - \mu\|^2}_{\text{independent of } x_0} \quad \textbf{(3.3d)}$$

Since the second sum is independent of x_0, this expression is obviously minimized by the choice $x_0 = \mu$.

The sample mean is a zero-dimensional representation of the data set. It is simple, but it does not reveal any of the variability in the data. More interesting, one-dimensional representation can be obtained by projecting the data onto a line running through the sample mean. Let e be a unit vector in the direction of the line. Then the equation of the line can be written as

$$x = \mu + ae \tag{3.4}$$

where the scalar a (which takes on any real value) corresponds to the distance of any point x from the mean μ. If x_k is presented by $\mu + a_k e$, an optimal set of coefficients a_k can be found by minimizing the squared-error criterion function

$$J_1(a_1,...,a_n,e) = \sum_{k=1}^{n} \left\| (\mu + a_k e) - x_k \right\|^2 = \sum_{k=1}^{n} \left\| a_k e - (x_k - \mu) \right\|^2 \tag{3.5a}$$

$$J_1(a_1,...,a_n,e) = \sum_{k=1}^{n} a_k^2 \left\| e \right\|^2 - 2\sum_{k=1}^{n} a_k e^T (x_k - \mu) + \sum_{k=1}^{n} \left\| x_k - \mu \right\|^2 \tag{3.5b}$$

Recognizing that $\|e\| = 1$, partially differentiating with respect to a_k, and setting the derivative to zero, the Eq. (3.6) is obtained

$$a_k = e^T (x_k - \mu) \tag{3.6}$$

Geometrically, this result merely says that a least-squares solution is obtained by projecting the vector x_k onto the line in the direction of e that passes through the sample mean.

This brings a more interesting problem of finding the best direction e for the line. The solution to this problem involves the so-called scatter matrix S defined by

$$S = \sum_{k=1}^{n} (x_k - \mu)(x_k - \mu)^T \tag{3.7}$$

The scatter matrix should look familiar, it is merely $n-1$ times the sample covariance matrix. It arises here when a_k found in Eq.(3.6) is substituted into Eq.(3.5) to obtain

$$J_1(e) = \sum_{k=1}^{n} a_k^2 - 2\sum_{k=1}^{n} a_k^2 + \sum_{k=1}^{n} \left\| x_k - \mu \right\|^2 \tag{3.8a}$$

$$J_1(e) = -\sum_{k=1}^{n} \left[e^T (x_k - \mu) \right]^2 + \sum_{k=1}^{n} \left\| x_k - \mu \right\|^2 \tag{3.8b}$$

$$J_1(e) = -\sum_{k=1}^{n} e^T (x_k - \mu)(x_k - \mu)^T e + \sum_{k=1}^{n} \|x_k - \mu\|^2 \quad \textbf{(3.8c)}$$

$$J_1(e) = -e^T S e + \sum_{k=1}^{n} \|x_k - \mu\|^2 \quad \textbf{(3.8d)}$$

Clearly, the vector e that minimizes J_1 also maximizes $e^T S e$. The method of Lagrange multipliers is used to maximize $e^T S e$ subject to the constraint that $\|e\| = 1$. Letting λ be the undetermined multiplier, Eq. (3.9) is differentiated

$$u = e^T S e - \lambda\left(e^T e - 1\right) \quad \textbf{(3.9)}$$

with respect to e to obtain

$$\frac{\partial u}{\partial e} = 2Se - 2\lambda e \quad \textbf{(3.10)}$$

Setting this gradient vector equal to zero, it is seen that e must be an eigenvector of the scatter matrix:

$$Se = \lambda e \quad \textbf{(3.11)}$$

In particular, because $e^T S e = \lambda e^T e = \lambda$, it follows that to maximizes $e^T S e$, it is needed to select the eigenvector corresponding to the largest eigenvalue of the scatter matrix. In other words, to find the best one-dimensional projection of the data (best in the least-sum-of-squared-error sense). The data is projected onto a line through the sample mean in the direction of the eigenvector of the scatter matrix having the largest eigenvalue

This result can be readily extended from a one-dimensional projection to a d'-dimensional projection. In place of Eq.(3.4), Eq. (3.12) is written

$$x = \mu + \sum_{i=1}^{d'} a_i e_i \quad \textbf{(3.12)}$$

where $d' \le d$. It is not difficult to show that the criterion function

$$J_{d'} = \sum_{k=1}^{n} \left\| \left(\mu + \sum_{i=1}^{d'} a_{ki} e_i \right) - x_k \right\|^2 \quad \textbf{(3.13)}$$

is minimized when the vectors $e_1, ..., e_{d'}$ are the d' eigenvectors of the scatter matrix having the largest eigenvalues. Because the scatter matrix is real and symmetric, these eigenvectors are orthogonal. They form a natural set of basis vectors for

representing any feature vector **x**. The coefficients a_i in Eq.(3.12) are the components of x in that basis, and are called the principal components. Geometrically, if the data points $x_1,...,x_n$ are considered to form a d-dimensional, hyperellipsoidally shaped cloud, then the eigenvectors of the scatter matrix are the principal axis of that hyperellipsoid. Principal component analysis reduces the dimensionality of feature space by restricting attention to those directions along which the scatter of the cloud is greatest [2].

3.1.2 How to Use PCA on Ear Images

There are patterns that occur in any input signal (image). Such patterns, which can be observed in all signals, could be - in the domain of ear recognition - the presence of some objects, such as helix, antihelix and ear lobe, in any ear as well as relative distances between these objects. These characteristic features are called eigenears (principal components) in ear recognition domain. They can be extracted out of the original image data by means of a mathematical tool called Principal Component Analysis (PCA).

By means of PCA each original image of the training set can be transformed into a corresponding eigenear. Furthermore, an important feature of PCA makes possible to recontruct any original image from the training set by combining the eigenears. Eigenears are the characteristic features of the ear, the original ear image can be reconstructed from eigenears if all the eigenears (features) are added up in the right proportions. Each eigenear represents certain features of the ear, which may or may not be present in the original image. If the feature is present in the original image to a higher degree, then the share or sum of the corresponding eigenear should be greater. If the particular feature is not (or almost not) present in the original image, then the corresponding eigenear should contribute a smaller (or not at all) part to the sum of eigenears. so, in order to reconstruct the original image from the eigenears, one has to build a kind of weighted sum of all eigenears. That is, the reconstructed original image is equal to a sum of all eigenears, with each eigenface having a certain weight. This weight specifies, to what degree the specific feature (eigenear) is present in the original image.

Eigenears that are extracted from original images can exactly reconstruct back these images. It is also possible to reconstruct original images approximately by using some of eigenears. Losses due to omitting some of the eigenears can be minimized by choosing the most important features (eigenears). The omission of eigenears is necessary due to scarcity of computational resources.

How does this relate to ear recognition? It is possible to extract the ear from eigenears that is given by a set of weights. It is also possible to extract the weights from the eigenears and recognize the ear. These weights tell the amount by which the ear in question differs from typical ears represented by the eigenears. Weights can determine two important things;

1. Determine if the image in question is a ear at all. If the weights of the image differ a large degree from the weights of ear images (ie. images that are ears for sure), the image probably is not a ear.

2. Similar ears (images) possess similar features (eigenears) to similar degrees (weights). If one extracts weights from all the images available, the images could be grouped into clusters. That is, all images having similar weights are likely to be similar ears.

3.1.2.1 Ear Is Viewed As A Vector

Consider ear as a 2D image. This image can be formed as a vector. Suppose that width of the image is w pixels and height of the image is h pixels. Thus the number of pixels for each vector is w*h. to construct the vector, the rows of the image are put beside each other as shown in Figure 3.1 and Figure 3.2

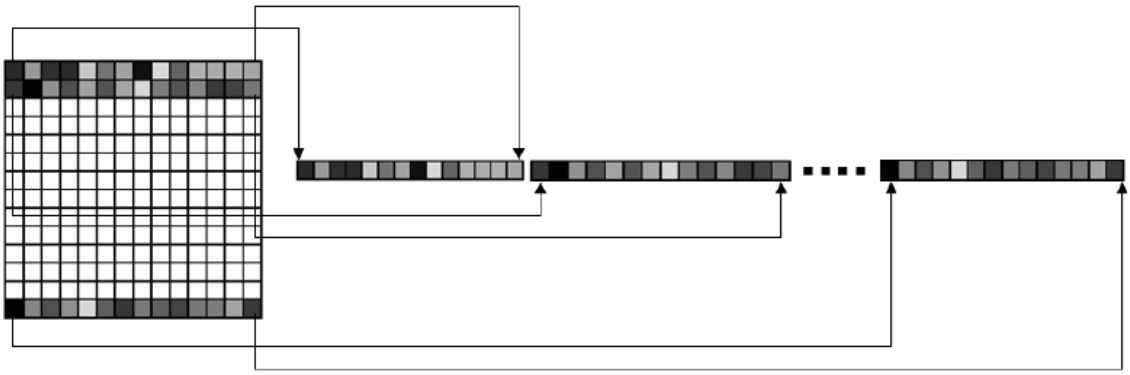


Figure 3.1: Construction of ear vector [1]

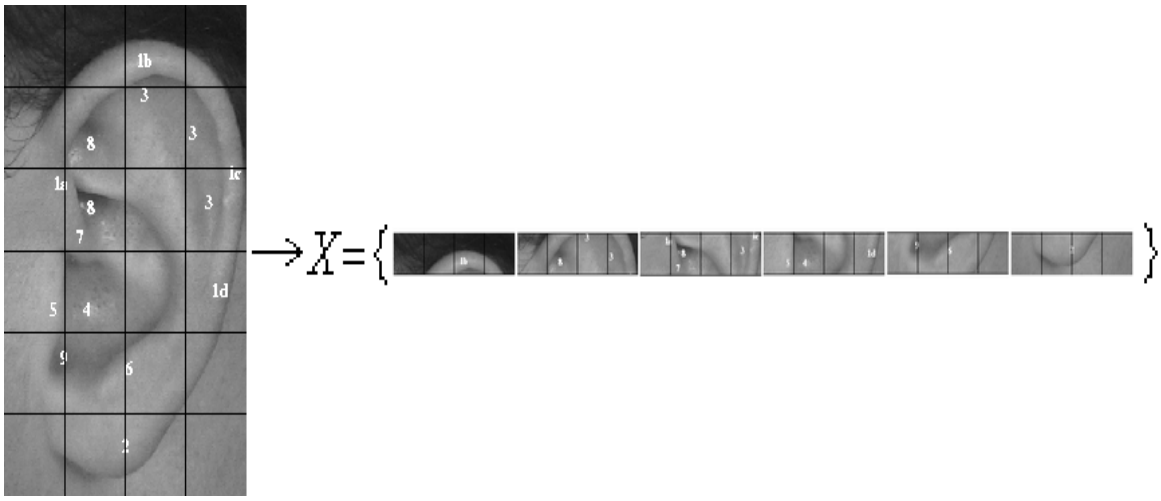


Figure 3.2: Formation of the vector of ear from the image of ear. 1 Helix, 2 Lobule, 3 Antihelix, 4 Concha, 5 Tragus, 6 Antitragus, 7 Crus of Helix, 8 Triangular Fossa, 9 Incisure Intertragica

By putting rows of the images beside each other, you make a vector as shown in Figure 3.3

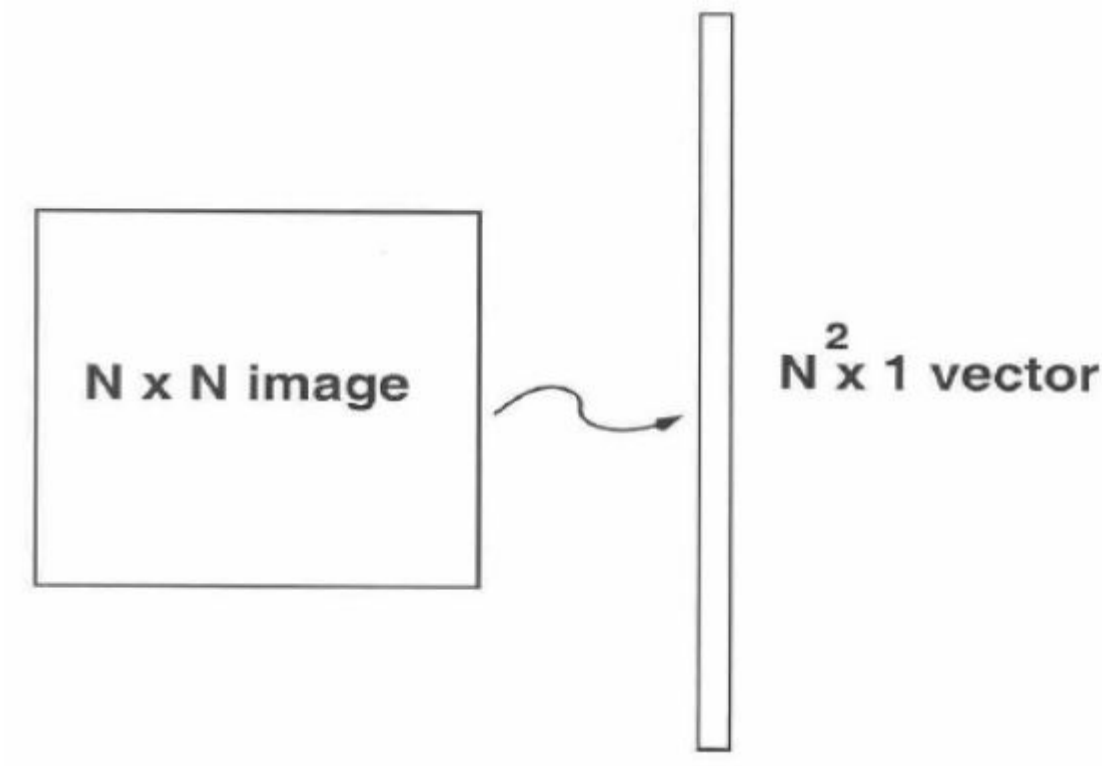


Figure 3.3: Mapping NxN image into N^2 vector [1]

The ear vector belongs to an ear space. This space is the image space, the space of all images whose dimensions are w by h pixels. The basis of the image space is composed of the following vectors;

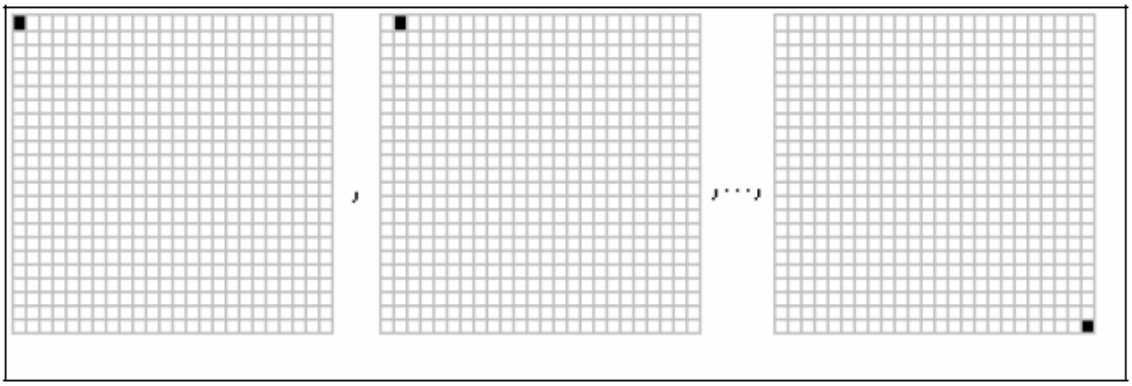


Figure 3.4: Image (ear) space [1]

All the ears look like earch other. They all have helix, antihelix, ear lobe, etc. located at the same place. There for, all the ear vectors are loceted in a very narrow cluster in the image space, as shown in Figure 3.5.

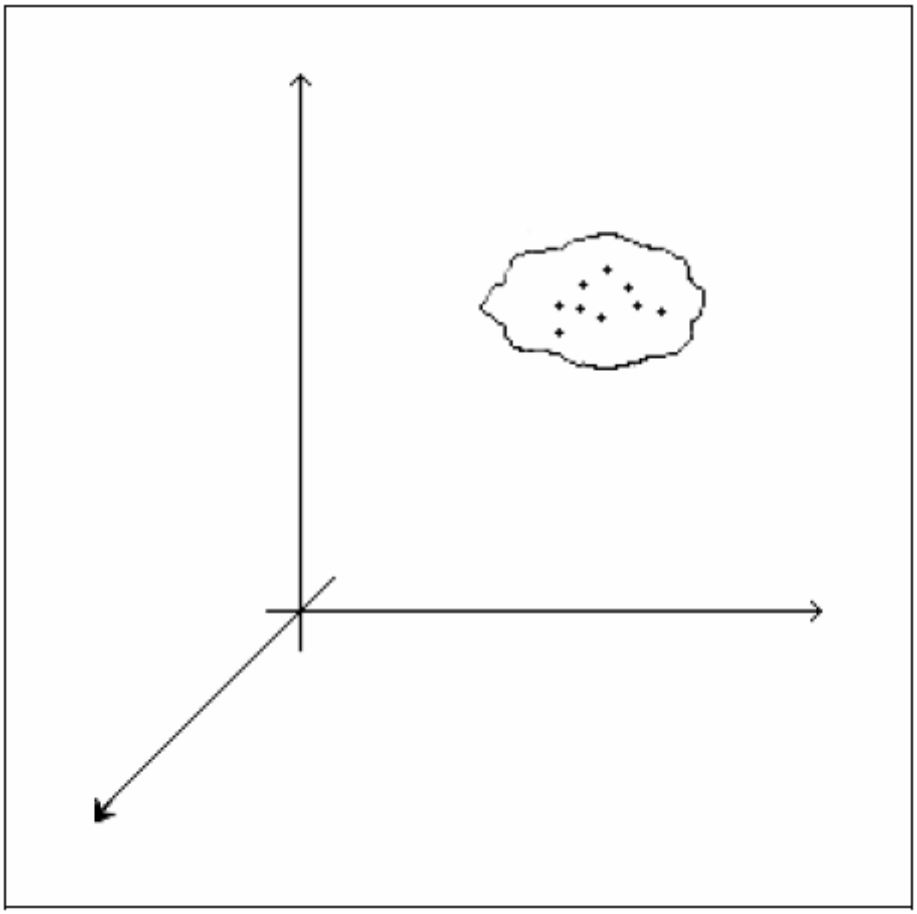


Figure 3.5: Image space and ear cluster [1]

The full image space is not an optimal space for ear description. The task presented here aims to build an ideal ear space that describes the ear better. The basis vectors of this ear space are so called eigenears (principal components) [4].

In the field of recognition one of the most-preferred, because of its simplicity and accuracy, methods is Principal Components Analysis. PCA is based on the Karhunen-Loeve (K-L) or Hotelling Transform, which is the optimal linear method for reducing redundancy, in the least mean squared reconstruction error sense. The idea of PCA is based on the identification of linear transformation of the co-ordinates in a system: The three axis of the new coordinate system coincide with the directions of the three largest spreads of the point distributions.

PCA uses the singular value decomposition to compute the principal components. A matrix whose rows consist of the eigenvectors of the input covariance matrix multiplies the input vector. This produces transformed input vectors whose components are ordered according to the magnitude of their variance. Those components, which contribute only a small amount to the total variance in the data set, are eliminated. It is assumed that the input data set has already been normalized so that it has a zero mean.

The most important components of each ear are located in a very narrow cluster. Thus the full image is not an optimal space for ear recognition and there are many redundant components that are not important for ear recognition. The purpose of PCA is to reduce the dimension of the set or the space. That means it aims to catch the total variation in the set of the training ears, and to explain this variation by few variables. Dealing with few variables is always more advantageous than dealing with huge numbers of variables, especially if there are huge number of ears to be processed. There are two approaches in PCA;

1) Statistical approach

2) Neural network approach

This study will focus on PCA using statistical approach.

3.1.2.2 Statistical PCA

In statistics, PCA is used to simplify the data set, more formally it is a linear transformation that chooses a new coordinate system for the data set so that the first axis of new coordinate system, which is called principal component, holds the greatest variance by any projection of the data set, the second axis holds the second greatest variance and so on. PCA can be used for reducing dimensionality in a data set while leaving out least effective principal components, however keeping most effective principal components which carry the information of significant variances that characterizes the data set. In this application, the characteristics are the most

important, but it is not always necessary. It is an application dependent case. The purpose of PCA is making recognition simpler and faster by reducing vector space dimensions because the actual image space is redundant.

Unlike other linear transformations, PCA does not have a fixed basis vectors. Its basis vectors depend on data set which makes it an optimal linear transformation that keeps the subspace of largest variance.

The aim of PCA when data set has zero mean is to find a projection vector or matrix that makes the variance maximum as defined by the formula below;

$$w_1 = \arg\max_{\|w\|=1} \operatorname{var}\left\{w^T x\right\} = \arg\max_{\|w\|=1} E\left\{\left(w^T x\right)^2\right\} \quad \textbf{(3.14)}$$

There are two stages in PCA;

1) The training stage.

2) The recognition stage.

In the training stage, the training image set is used to calculate PCA variables. In the recognition stage, calculated variables are used to recognize an unidentified image.

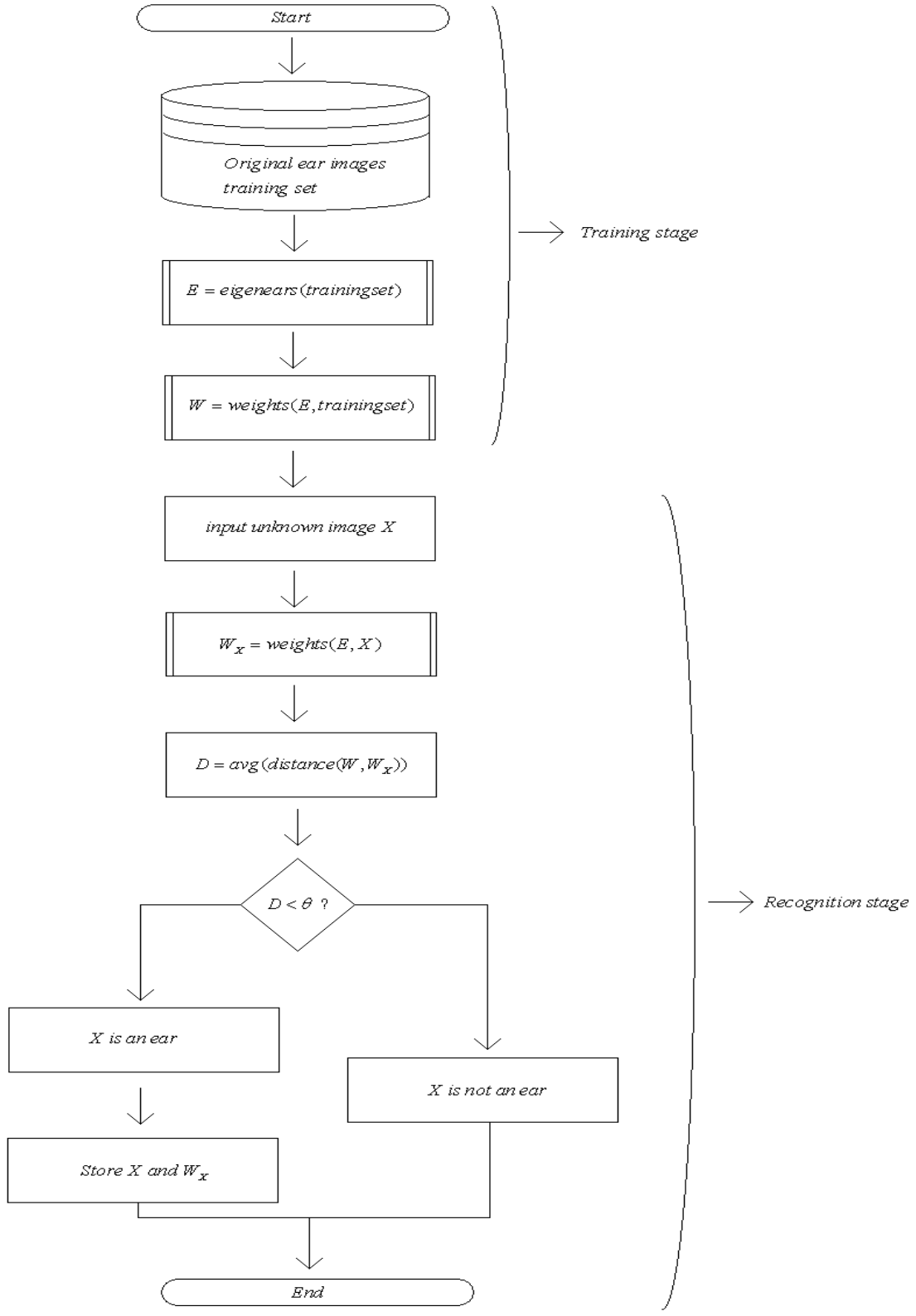


Figure 3.6: High-level functioning principle of the eigenear-based ear recognition algorithm

As shown in the algorithm diagram above in Figure 3.6 [1], the first step of training is transforming the training images to eigenears. Then the weights are calculated for

the each training image. In the recognition stage, weights for the unknown images are calculated, as well. In the last step of recognition step, the difference between the weights in training set and the weights for the unknown images are compared. The closest difference based on the threshold θ is considered as the recognized image.

Suppose that, each training image consists N pixels (N = w*h). Let $\{\Gamma_i \mid i = 1,...,M\}$ be the training set of images (Γ_i is the vector form of images and M is the number of images in training set). The average of these M images is calculated by the following;

$$\Psi = \frac{1}{n}\sum_{i=1}^{M}\Gamma_i \quad \textbf{(3.15)}$$

Then each ear Γ_i differs from the average ear Ψ by Φ_i.

$$\Phi_i = \Gamma_i - \Psi ; i = 1,...,M \quad \textbf{(3.16)}$$

A covariance matrix of the training images can be constructed as follows;

$$C = AA^T \quad \textbf{(3.17)}$$

Where $A = [\Phi_1\Phi_2...\Phi_M]$. The basis vectors, that are so called eigenears, of the ear space, are then the orthogonal eigenvectors of the covariance matrix C.

A is a NxM matrix and A^T is a MxN matrix so C will be a NxN matrix. Finding eigenvectors of a NxN matrix is a difficult task due to its size. Therefore, a simple calculation is required.

Since the number of training images is usually less than the number of pixels in an image, there will be only M-1, instead of N, meaningful eigenvectors. The eigenears are computed by first finding the eigenvectors, $\upsilon_l (l = 1,...,M)$, of the M by M matrix L.

$$L = A^T A \quad \textbf{(3.18)}$$

The eigenvectors, $u_l (l = 1,...,M)$, of the matrix C are then expressed by a linear combination of the difference ear images, $\Phi_i (i = 1,...,M)$, weighted by $\upsilon_l (l = 1,...,M)$;

$$L\upsilon = \lambda\upsilon \quad \textbf{(3.19a)}$$

$$A^T A\upsilon = \lambda\upsilon \quad \textbf{(3.19b)}$$

$$AA^T Av = A\lambda v \tag{3.19c}$$

$$C(Av) = \lambda(Av) \Rightarrow Cu = \lambda u \tag{3.19d}$$

$$u = Av \tag{3.19e}$$

$$u = [u_1, ..., u_M] = [\Phi_1, ..., \Phi_M][v_1, ..., v_M] = Av \tag{3.20}$$

Each eigenear can be viewed as a feature. When a particular ear is projected onto the ear space, its vector, which is made up of its weight values with respect to each eigenear, into the ear space describes the importance of each of those features in the ear.

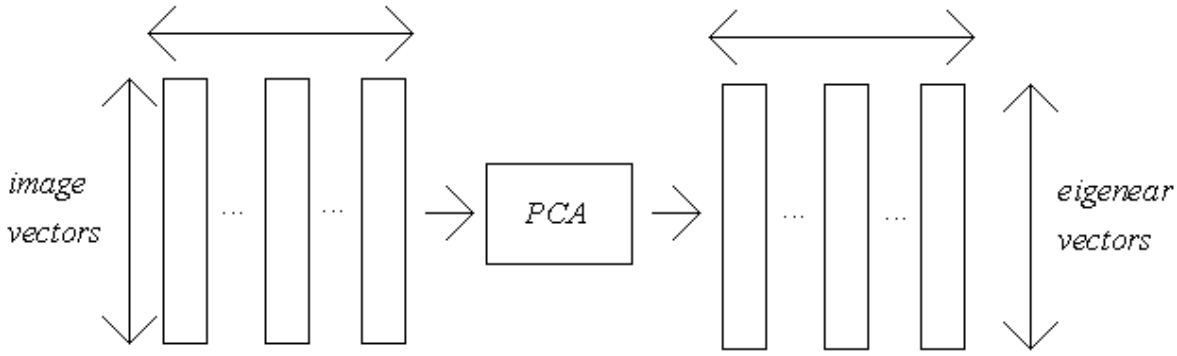


Figure 3.7: Eigenear generation process [1]

In practice, a smaller set of $M'(M' < M)$ eigenears is sufficient for ear identification. Hence, only M' significant eigenvectors of L, corresponding to the largest M' eigenvalues, are selected for the eigenear computation, resulting in a further data compression. M' is determined by a threshold, θ_λ, of the ratio of the eigenvalue summation.

$$M' = \min_r \left\{ r \mid \frac{\sum_{l=1}^{r} \lambda_l}{\sum_{l=1}^{M} \lambda_l} \right\} \tag{3.21}$$

In the training stage, the ear of each known individual, Γ_k, is projected into the ear space and an M'-dimensional vector, Ω_k, is obtained.

$$\Omega_k = U^T(\Gamma_k - \Psi); \quad k = 1, ..., M \tag{3.22}$$

For comparison, two methods are used to describe an ear class in the ear space. The first method, referred to as the averaging representation, calculates the class vector by averaging the projected vectors from the training images of the corresponding individual. The second method, the point-set representation, describes an ear class by a set of vectors projected from all the training images of an individual.

A distance threshold, θ_c, that defines the maximum allowable distance from an ear class as well as from the ear space, is set up by computing half the largest distance between any two ears;

$$\theta_c = \frac{1}{2}\max_{j,k}\left\{\left\|\Omega_j - \Omega_k\right\|\right\}; j,k = 1,...,M \quad \textbf{(3.23)}$$

once the eigenears have been computed, the ear space has to be populated with known ears. Usually these ears are taken from the training set. Each known ear is transformed into the face space and its components stored in memory.

at this stage the identification process can begin. An unknown ear is presented to the system. The system projects it onto the ear space and computes its distance from all the stores ears. the ear is identified as being the same individual as the ear space which is nearest to it in ear space. There are several methods of computing the distance between multidimensional vectors. In this study, a form of Euclidean distance is chosen.

In the recognition stage, a new image, Γ, is projected into the ear space to obtain a vector, Ω;

$$\Omega = U^T\left(\Gamma - \Psi\right) \quad \textbf{(3.24)}$$

The distance of Ω to each ear class is defined by

$$\varepsilon_k^2 = \left\|\Omega - \Omega_k\right\|^2; k = 1,...,M \quad \textbf{(3.25)}$$

For the purpose of discriminating between ear images and non-ear like images, the distance, ε, between the original image, Γ, and its reconstructed image from the eigenear space, Γ_f, is also computed;

$$\varepsilon^2 = \left\|\Gamma - \Gamma_k\right\|^2 \quad \textbf{(3.26)}$$

where

$$\Gamma_f = U\Omega + \Psi \quad \textbf{(3.27)}$$

These distances are compared with the threshold given in Eq.(3.23) and the input image is classified by the following rules [1,4];

IF $\varepsilon \geq \theta_c$ **(3.28)**

THEN input image is not an ear image;

IF $\varepsilon < \theta_c$ AND $\forall k, \varepsilon_k \geq \theta_c$ **(3.29)**

THEN input image contains an unknown ear;

IF $\varepsilon < \theta_c$ AND $\varepsilon_{k^*} = \min_k \{\varepsilon_k\} < \theta_c$ **(3.30)**

THEN input image contains the ear of individual k^*;

3.2 Fisher's Linear Discriminant (FLD)

The main idea of PCA is to find components that are useful for representing data, but it is not guaranteed that these components are useful for discriminating between data in different classes. In some cases, the directions that are discarded by PCA might be exactly the directions that are needed for distinguishing between classes. For example, if the data is uppercase letters O and Q, PCA might discover the gross features that characterize Os and Qs, but might ignore the tail that distinguishes an O from a Q. Where PCA seeks directions that are efficient for representation, in Figure 3.8, discriminant analysis seeks directions that are efficient for discrimination [2].

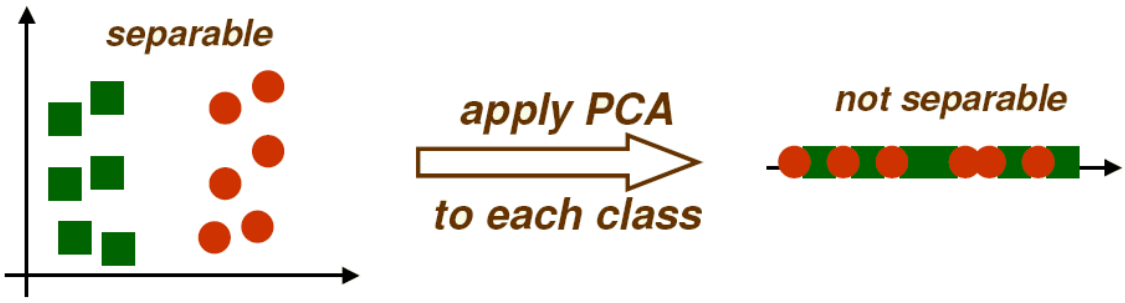


Figure 3.8: The Way that PCA Handles the Dataset [5]

Projecting data from d dimensions onto a line is a hard problem if you desire good recognition performance. Even if the samples formed well-separated, compact clusters in d-space, projection onto an arbitrary line will usually produce a confused mixture of samples from all of the classes and thus produce poor recognition performance. However, by moving the line around, we might be able to find an

orientation for which the projected samples are well separated, in Figure 3.9. This is exactly the goal of classical discriminant analysis.

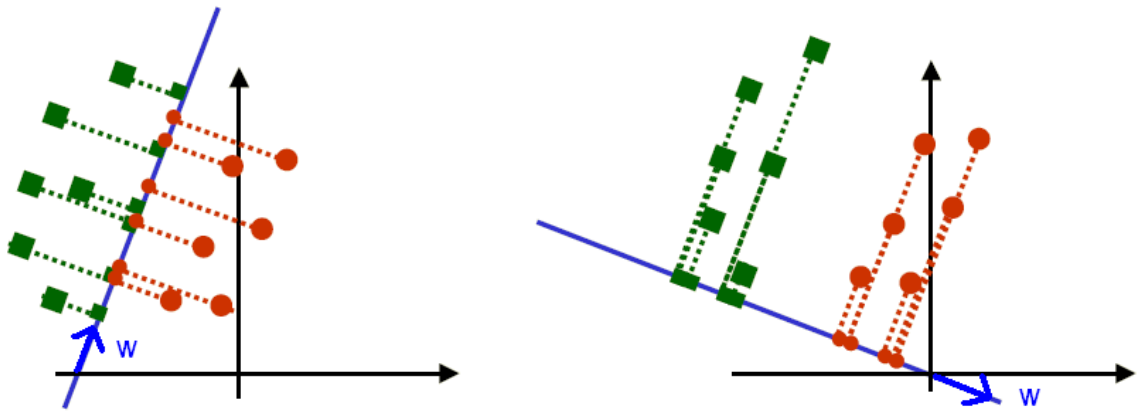


Figure 3.9: Projection of the Same Set of Samples onto Two Different Lines in the Direction Marked w [5]

In Figure 3.9, the figure on the right shows greater separation between the projected red and green points

Suppose that there is a set of n d-dimensional samples $x_1,...,x_n$, n_1 in the subset D_1 labeled ω_1 and n_2 in the subset D_2 labeled ω_2. If the linear combination of the components of x is formed, the scalar dot product, in Figure 3.10, is obtained;

$$y = w^T x \tag{3.31}$$

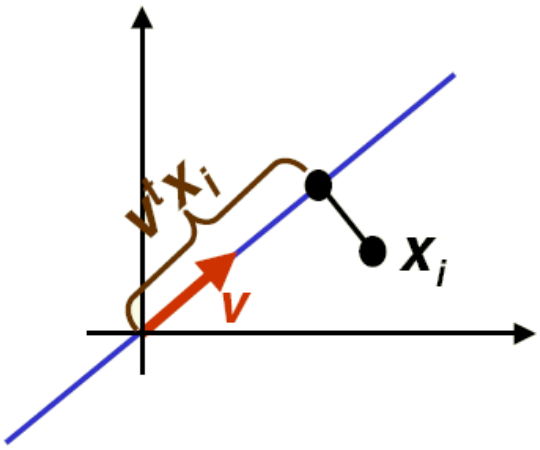


Figure 3.10: Shows the Projected Point x_i [5]

and a corresponding set of n samples $y_1,..., y_n$ divided in to the subsets Y_1 and Y_2. Geometrically, if $\|w\| = 1$, each y_i is the projection of the corresponding x_i onto a line in the direction of w. Actually, the magnitude of w is of no real significance, because it merely scales y. The direction of w is important, however, if we imagine

that the samples labeled ω_1 fall more or less into one cluster while those labeled ω_2 fall in another, we want the projections falling onto the line to be well separated, not thoroughly intermingled. It should be abundantly clear that if the original distributions are multimodal and highly overlapping, even the "best" w is unlikely to provide adequate separation, and thus this method will be of little use.

Now, it is turned to the matter of finding the best such direction w, one will enable accurate classification. A measure of the separation between the projected points is the difference of the sample means. If m_i is the d-dimensional sample mean given by

$$\mu_i = \frac{1}{n_i} \sum_{x \in D_i} x_i \tag{3.32}$$

then the sample mean for the projected points is given by

$$\tilde{\mu}_i = \frac{1}{n_i} \sum_{y \in Y_i} y \tag{3.33a}$$

$$\tilde{\mu}_i = \frac{1}{n_i} \sum_{x \in D_i} w^T x = w^T \mu_i \tag{3.33b}$$

and is simply the projection of μ_i.

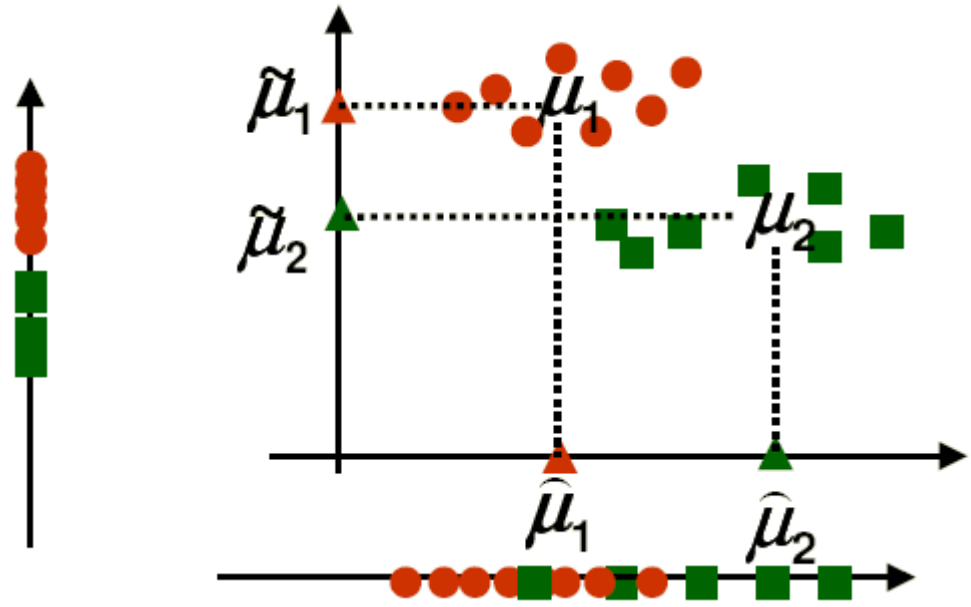


Figure 3.11: Stand-alone $|\tilde{\mu}_1 - \tilde{\mu}_2|$ is not always a Good Measure of Separation [5]

It follows that the distance between the projected means is

$$\left|\tilde{\mu}_1 - \tilde{\mu}_2\right| = \left|w^T(\mu_1 - \mu_2)\right| \tag{3.34}$$

and that we can make this difference as large as we wish merely by scaling w. Of course, to obtain good separation of the projected data we really want the difference between the means to be large relative to some measure of the standard deviations for each class, in Figure 3.11. Rather than forming sample variances, we define the scatter for projected samples labeled ω_i by

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \tilde{\mu}_i)^2 \tag{3.35}$$

Thus, $(1/n)\left(\tilde{s}_1^2 + \tilde{s}_2^2\right)$ is an estimate of the variance of the pooled data, and $\tilde{s}_1^2 + \tilde{s}_2^2$ is called the total within-class scatter of the projected samples. The Fisher's Linear Discriminant employs that linear function $w^T x$ for which the criterion function

$$J(w) = \frac{\left|\tilde{m}_1 - \tilde{m}_2\right|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} \tag{3.36}$$

is maximum (and independent of $\|w\|$). While the w maximizing $J(.)$ leads to the best separation between the two projected sets (in the sense just described), a threshold criterion will be needed before a true classifier is obtained. It is first considered how to find the optimal w, and later turn to the issue of thresholds.

To obtain $J(.)$ as an explicit function of w, we define the scatter matrices S_i and S_W by

$$S_i = \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T \tag{3.37}$$

and

$$S_W = S_1 + S_2 \tag{3.38}$$

then we can write

$$\tilde{s}_i^2 = \sum_{x \in D_i} (w^T x - w^T \mu_i)^2 \tag{3.39a}$$

$$\tilde{s}_i^2 = \sum_{x \in D_i} w^T (x - \mu_i)(x - \mu_i)^T w \tag{3.39b}$$

$$\tilde{s}_i^2 = w^T S_i w \tag{3.39c}$$

therefore the sum of these scatters can be written

$$\tilde{s}_1^2 + \tilde{s}_2^2 = w^T S_W w \tag{3.40}$$

Similarly, the separation of the projected means obeys

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = (w^T \mu_1 - w^T \mu_2)^2 \tag{3.41a}$$

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = w^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T w \tag{3.41b}$$

$$(\tilde{\mu}_1 - \tilde{\mu}_2)^2 = w^T S_B w \tag{3.41c}$$

where

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \tag{3.42}$$

We call S_W the within-class scatter matrix. It is proportional to the sample covariance matrix for the pooled d-dimensional data. It is symmetric and positive semidefinite, and it is usually nonsingular if $n > d$. Likewise, S_B is called the between-class scatter matrix. It is also symmetric and positive semidefinite, but because it is the outer product of two vectors, its rank is at most one. In particular, for any w, $S_B w$ is in the direction of $\mu_1 - \mu_2$, and S_B is quite singular.

In terms of S_B and S_W, the criterion function $J(.)$ can be written as

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \tag{3.43}$$

This expression is well known in mathematical physics as the generalized Rayleigh quotient. It is easy to show that a vector w that maximizes $J(.)$ must satisfy

$$S_B w = \lambda S_W w \tag{3.44}$$

for some constant λ, which is a generalized eigenvalue problem. This can also be seen informally by noting that at an extremum of $J(w)$ a small change in w in Eq. (3.43) should leave unchanged the ratio of the numerator to the denominator. If S_W is nonsingular we can obtain a conventional eigenvalue problem by writing

$$S_W^{-1} S_B w = \lambda w \tag{3.45}$$

In most cases, It is unnecessary to solve for the eigenvalues and eigenvectors of $S_W^{-1}S_B$ due to the fact that $S_B w$ is always in the direction of $\mu_1 - \mu_2$. Because the scale factor for w is immaterial, we can immediately write the solution for the w that optimizes $J(.)$:

$$w = S_W^{-1}(\mu_1 - \mu_2) \quad \textbf{(3.46)}$$

Thus, we have obtained w for Fisher's Linear Discriminant, the linear function yielding the maximum ratio of between-class scatter. Thus the classification has been converted from a d-dimensional problem to a hopefully more manageable one-dimensional one. This mapping is many-to-one, and in theory it cannot possibly reduce the minimum achievable error rate if we have a very large training set. In general, one is willing to sacrifice some of the theoretically attainable performance for the advantages of working in one-dimension. All that remains is to find the threshold, that is, the point along the one-dimensional subspace separating the projected points.

When the conditional densities $p(x \mid \omega_i)$ are multivariate normal with equal covariance matrices Σ, the threshold can be calculated directly. In that case the equation of optimal decision boundary is recalled

$$w^T x + w_0 = 0 \quad \textbf{(3.47)}$$

where

$$w = \Sigma^{-1}(\mu_1 - \mu_2) \quad \textbf{(3.48)}$$

and where w_0 is a constant involving w and the prior probabilities. If sample means and the sample covariance matrix is used to estimate μ_i and Σ, a vector in the same direction as w of Eq. (3.48) is obtained, that maximizes $J(.)$. Thus, for the normal, equal-covariance case, the optimal decision rule is merely to decide ω_1 if Fisher's Linear Discriminant exceeds some threshold, and to decide ω_2 otherwise. More generally, if the projected data is smoothed, or is fitted with a univariate Gaussion, then ω_0 where the posteriors in the one dimensional distributions are equal should be choosen.

The computational complexity of finding the optimal w for the Fisher linear discriminant is dominated by the calculation of the within-category total scatter and its inverse, an O (d^2n) calculation.

3.2.1 Implementation of FLD

In this section the implementation of FLD was done on a 2D random dataset. The purpose of this process is to show how FLD works.

Different from PCA, FLD handles dataset as devided into classes and tries to find the best direction for the good classification. The dataset is divided into 2 classes as shown in Table 3.7 and plotted as in Figure 3.12.

Table 3.1: Dataset Divided into 2 Classes

		X	Y
Class 1	P_2	1	2
	P_4	2	3
	P_7	3	3
	P_8	4	5
	P_{10}	5	5
Class 2	P_1	1	0
	P_3	2	1
	P_5	3	1
	P_6	3	2
	P_9	5	3
	P_{11}	6	5

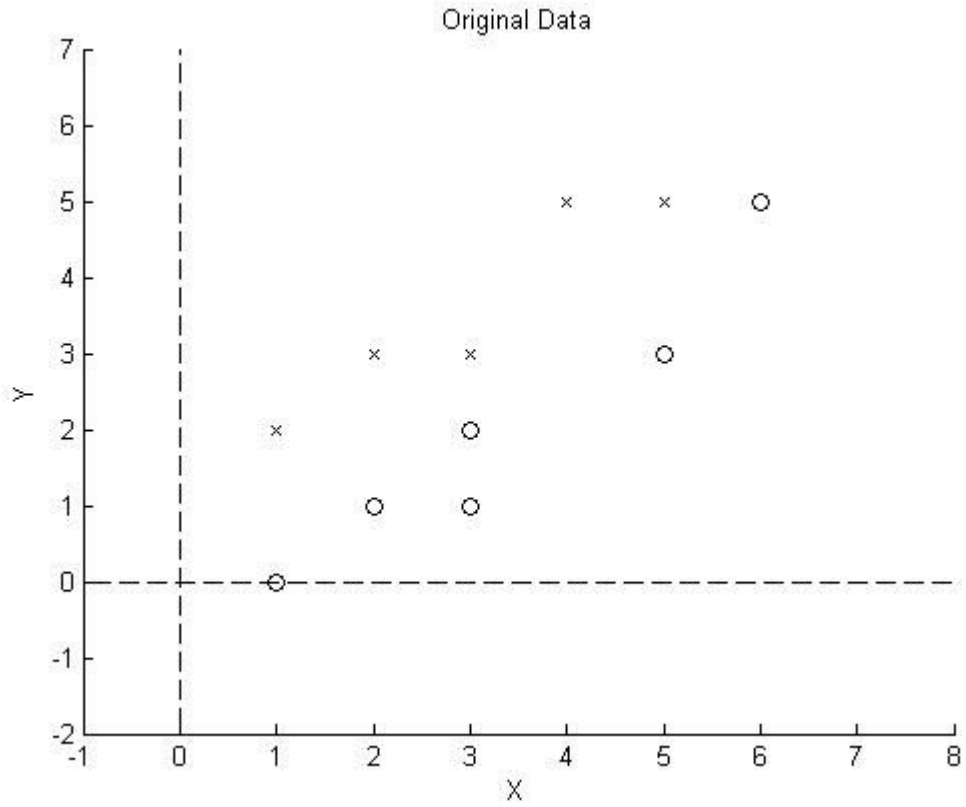


Figure 3.12: The Plot of Dataset. "x" donates class 1, "o" donates class 2

Mean of class μ_1 and μ_2 were calculated by the formula given in Eq. (3.32) and shown below;

$$\mu_1 = mean(c_1) = \begin{bmatrix} 3 & 3.6 \end{bmatrix}$$

$$\mu_2 = mean(c_2) = \begin{bmatrix} 3.3 & 2 \end{bmatrix}$$

Scatter matrices S_1 and S_2 for each class were calculated by the formula given in Eq. (3.37) and shown below;

$$S_1 = \begin{bmatrix} 10 & 8 \\ 8 & 7.2 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 17.3 & 16 \\ 16 & 16 \end{bmatrix}$$

Within the class scatter matrix S_W of dataset was calculated by the formula given in Eq. (3.38) and shown below;

$$S_W = S_1 + S_2 = \begin{bmatrix} 27.3 & 24 \\ 24 & 23.2 \end{bmatrix}$$

Between the class scatter matrix S_B of dataset was calculated by the formula given in Eq. (3.42) and shown below;

$$S_B = \begin{bmatrix} 0.09 & -0.48 \\ -0.48 & 2.56 \end{bmatrix}$$

After calculating the within the class scatter matrix S_W and between the class scatter matrix S_B, the generalized eigenvalue problem, given in Eq. (3.45), was solved and $c-1$ most significant eigenvector, which has the larger eigenvalues, was taken to form projection direction. In this case there are two classes so there is just one eigenvector, which is shown below;

$$w = \begin{bmatrix} -0.67 \\ 0.75 \end{bmatrix}$$

The plot of dataset and projection direction is given below, in Figure 3.20.

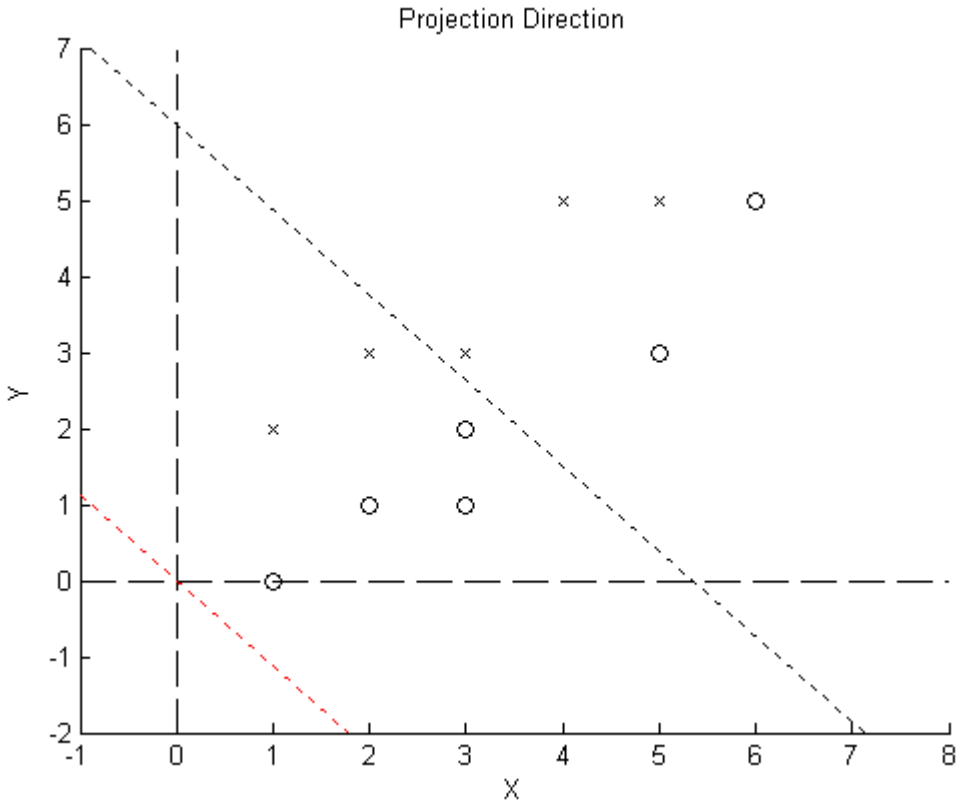


Figure 3.13: Projection Direction of FLD

In the Figure 3.13, the doted red line that passes through the origin shows the direction of significant eigenvector which is also the projection direction. The doted red line is in the form of $y = ax$. Here, the important thing is the slope of the line that

is a. The line can be shifted anywhere along the X axis, which can be mathematically expressed as $y = ax + b$. In the Figure 3.20, the doted black line presents the shifted projection direction. It does not matter if it is shifted or not, as long as the slope of line is not changed.

The projection of dataset on the projection direction is given in Figure 3.14.

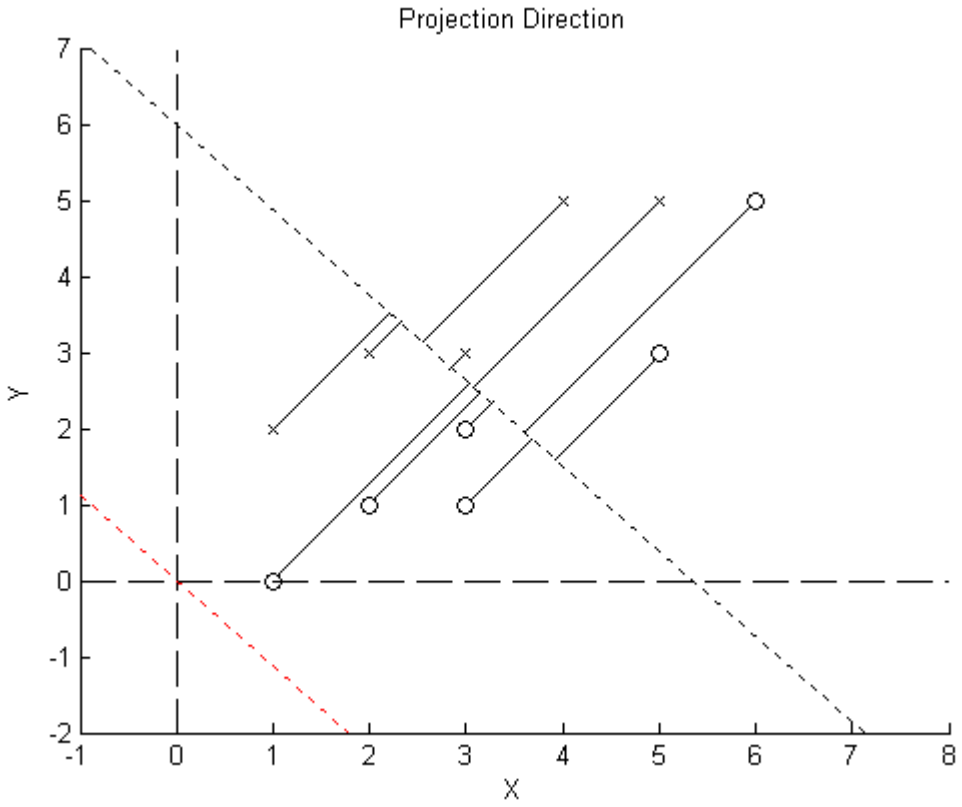


Figure 3.14: Shows the Best Possible Projection of Dataset for Classification.

The goal of PCA is to find the direction of greatest variance while FLD tries to find the best projection direction for classification [5].

3.2.2 Multiple Discriminant Analysis

For the c-class problem, the natural generalization of Fisher's linear discriminant involves $c-1$ discriminant functions. Thus, the projection is from a d-dimensional space to a $(c-1)$-dimensional space, and it is tacitly assumed that $d \geq c$. The generalization for the within-class scatter matrix is obvious:

$$S_W = \sum_{i=1}^{c} S_i \qquad \textbf{(3.49)}$$

where, as before,

$$S_i = \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T \tag{3.50}$$

and

$$\mu_i = \frac{1}{n_i} \sum_{x \in D_i} x_i \tag{3.51}$$

The proper generalization for S_B is not quite so obvious. Suppose that we define a total mean vector m and a total scatter matrix S_T by

$$\mu = \frac{1}{n} \sum_{x} x = \frac{1}{n} \sum_{i=1}^{c} n_i \mu_i \tag{3.52}$$

and

$$S_T = \sum_{x} (x - \mu)(x - \mu)^T \tag{3.53}$$

Than it follows that

$$S_T = \sum_{i=1}^{c} \sum_{x \in D_i} (x - \mu_i + \mu_i - \mu)(x - \mu_i + \mu_i - \mu)^T \tag{3.54a}$$

$$S_T = \sum_{i=1}^{c} \sum_{x \in D_i} (x - \mu_i)(x - \mu_i)^T + \sum_{i=1}^{c} \sum_{x \in D_i} (\mu_i - \mu)(\mu_i - \mu)^T \tag{3.54b}$$

$$S_T = S_W + \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{3.54c}$$

It is natural to define this second term as a general between-class scatter matrix, so that the total scatter is the sum of the within-class scatter and the between-class scatter:

$$S_B = \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^T \tag{3.55}$$

and

$$S_T = S_W + S_B \tag{3.56}$$

If we check the two-class case, we find that the resulting between-class scatter matrix is $n_1 n_2 / n$ times our previous definition.

The projection from a d-dimensional space to a $(c-1)$-dimensional space is accomplished by $c-1$ discriminant functions

$$y_i = w_i^T x \qquad i = 1,...,c-1 \tag{3.57}$$

If the y_i are viewed as components of a vector y and the weight vectors w_i are viewed as the columns of a d-by-$(c-1)$ matrix W, then the projection can be written as a single matrix equation

$$y = W^T x \tag{3.58}$$

The samples $x_1,...,x_n$ project to a corresponding set of samples $y_1,...,y_n$, which can be described by their own mean vectors and scatter matrices. Thus, if we define

$$\tilde{\mu}_i = \frac{1}{n_i} \sum_{y \in Y_i} y \tag{3.59}$$

$$\tilde{\mu} = \frac{1}{n_i} \sum_{i=1}^{c} n_i \tilde{\mu}_i \tag{3.60}$$

$$\tilde{S}_W = \sum_{i=1}^{c} \sum_{y \in Y_i} (y - \tilde{\mu}_i)(y - \tilde{\mu}_i)^T \tag{3.61}$$

and

$$\tilde{S}_B = \sum_{i=1}^{c} n_i (\tilde{\mu}_i - \tilde{\mu})(\tilde{\mu}_i - \tilde{\mu})^T \tag{3.62}$$

it is a straightforward matter to show that

$$\tilde{S}_W = W^T S_W W \tag{3.63}$$

and

$$\tilde{S}_B = W^T S_B W \tag{3.64}$$

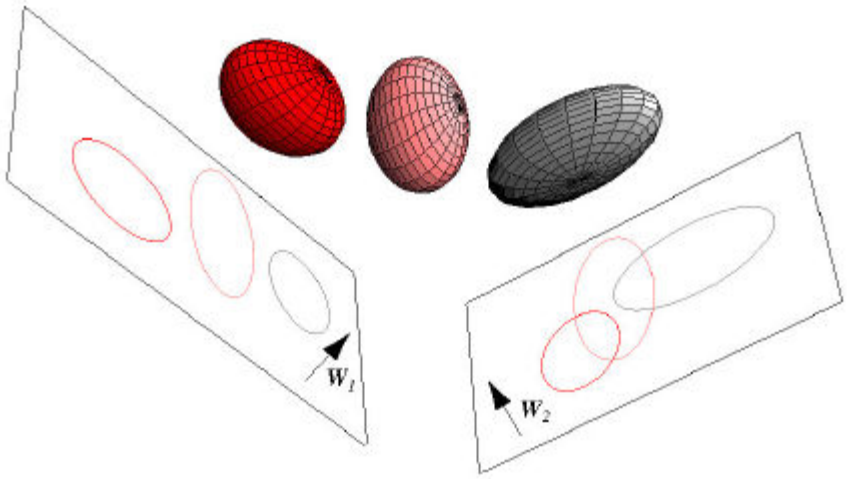


Figure 3.15: The Projection of Three-Dimension onto Two-Dimension [2]

Figure 3.15 shows three 3-dimensional distributions are projected onto 2-dimensional subspaces, described by a normal vectors W_1 and W_2. Informally, multiple discriminant methods seek the optimum such subspace, that is, the one with the greatest separation of the projected distributions for given total within-scatter matrix, here as associated with W_1.

These equations show how the within-class and between-class scatter matrices are transformed by the projection to the lower dimensional space, in Figure 3.15. What we seek is a transformation matrix W that in some sense maximizes the ratio of the between-class scatter to the within-class scatter. A simple scalar measure of scatter is the determinant of the scatter matrix. The determinant is the product of the eigenvalues, and hence is the product of the "variances" in the principal directions, thereby measuring the square of the hyperellipsoidal scattering volume. Using this measure, we obtain the criterion function.

$$J(W) = \frac{\left|\tilde{S}_B\right|}{\left|\tilde{S}_W\right|} = \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \qquad \textbf{(3.65)}$$

The problem of finding a rectangular matrix W that maximizes $J(.)$ is tricky, though fortunately it turns out that the solution is relatively simple. The columns of an optimal W are the generalized eigenvectors that correspond to the largest eigenvalues in

$$S_B w_i = \lambda_i S_W w_i \qquad \textbf{(3.66a)}$$

$$S_W^{-1} S_B w_i = \lambda_i w_i \qquad \textbf{(3.66b)}$$

A few observations about this solution are in order. First, if S_W is nonsingular, this can be converted to a conventional eigenvalue problem as before. However, this is actually undesirable, since it requires an unnecessary computation of the inverse of S_W. Instead of this unnecessary computation, one can find the eigenvalues as the roots of the characteristic polynomial

$$\left|S_B - \lambda_i S_W\right| = 0 \qquad \textbf{(3.67)}$$

and then solve

$$\left(S_B - \lambda_i S_W\right) w_i = 0 \qquad \textbf{(3.68)}$$

directly for the eigenvectors w_i. Because S_B is the sum of c matrices of rank one or less, and because only $c-1$ of these are independent, S_B is of rank $c-1$ or less. Thus, no more than $c-1$ of the eigenvalues are nonzero, and the desired weight vectors correspond to these nonzero eigenvalues. If the within-class scatter is isotropic, the eigenvectors are merely the eigenvectors of S_B, and the eigenvectors with nonzero eigenvalues span the space spanned by applying the Gram-Schmidt orthonormalization procedure to the $c-1$ vectors $\mu_i - \mu, i = 1,...,c-1$. Finally, we observe that in general the solution for W is not unique; the allowable transformations include rotating and scaling the axis in various ways. These are all linear transformations from a $(c-1)$-dimensional space to a $(c-1)$-dimensional space, however, and do not change things in any significant way; in particular, they leave the criterion function $J(W)$ invariant and the classifier unchanged.

If we have very little data, we would tend to project to a subspace of low dimension, while if there are more data, we can use a higher dimensions.

As in the two-class case, multiple discriminant analysis primarily provides a reasonable way to reducing the dimensionality of the problem. Parametric or nonparametric techniques that might not have been feasible in the original space may work well in the lower-dimensional space. In particular, it may be possible to estimate separate covariance matrices for each class and use the general multivariate normal assumption after the transformation where this could not be done with the original data. In general, if the transformation causes some unnecessary overlapping of the data and increases the theoretically achievable error rate, then the problem of classifying the data still remains [2.6].

3.3 Small Sample Size Problem

In ear recognition tasks, the dimension of the sample space is typically larger than the number of the samples in the training set. So the within-class scatter matrix is singular. This problem is known as the small sample size problem [1,2]. To overcome this problem, a new method is proposed called Discriminative Common Vector method, which is based on a variation of Fisher's Linear Discriminant Analysis for small sample size. This algorithm uses within-class scatter matrix to produce common vectors. Then the common vectors are used for classification of new ears. This method claims more accuracy, efficiency and stability comparing to traditional methods like PCA and FLD.

3.3.1 Problems with PCA

PCA is unsupervised since it does not consider the classes within the training set data. In choosing a criterion that maximizes the total scatter, this approach tends to model unwanted within-class variations such as those resulting from the differences in illumination and other factors. Also, because the criterion does not minimize the within-class variation, there could be overlap in the result compared to other methods. Thus, the projection vectors chosen for optimal reconstruction may obscure the existence of the separate classes.

3.3.2 Problems with FLD

FLD solves the limitations of the Eigenears method by applying Fisher's Linear Discriminant criterion as mentioned below;

$$J_{FLD}\left(W_{opt}\right) = \arg_W \max \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|} \tag{3.69}$$

where S_B is the between-class scatter-matrix and S_W is the within-class scatter matrix. By applying this method, the projection directions maximize the Euclidian distance between the ear images of different classes on the other hand and on the other minimize the distance between the ear images of the same class. The problem in this method is that it cannot be applied since the dimension of the sample space is typically large than the number of samples in the training set. Discriminative Common Vector approach is one of the algorithms those were proposed to fix this problem [7].

3.4 Discriminative Common Vector Approach (DCVA)

Since ear images have similar structure, the image vectors are correlated, and any image in the image space can be represented in a lower-dimensional subspace without losing a significant amount of information. The Eigenear method has been proposed for finding such a lower-dimensional subspace. The key idea behind the Eigenear method, which uses Principal Component Analysis (PCA), is to find the best set of projection directions in the sample space that will maximize the total scatter across all images such that

$$J_{PCA}\left(W_{opt}\right) = \arg\max_{W} \left|W^T S_T W\right| \quad \textbf{(3.70)}$$

is maximized. Here, S_T is the total scatter matrix of the training set samples, and W is the matrix whose columns are orthonormal projection vectors. The projection directions are also so called the eigenears. Any ear image in the image space can be approximated by a linear combination of the significant eigenears. The sum of the eigenvalues that correspond to the eigenears not used in reconstruction gives the mean square error of reconstruction. This method is an unsupervised technique since it does not consider the classes within the training set data. In choosing a criterion that maximizes the total scatter, this approach tends to model unwanted within-class variations such as those resulting from the differences in lighting, facial expression, and other factors. Additionally, since the criterion does not attempt to minimize the within-class variation, the resulting classes may tend to have more overlap than other approaches. Thus, the projection vectors chosen for optimal reconstruction may obscure the existence of the separate classes.

The Linear Discriminant Analysis (LDA) method overcomes the limitations of the eigenear method by applying the Fisher's Linear Discriminant criterion. This criterion tries to maximize the ratio of the function given in Eq. (3.69). Thus, by applying this method, we find the projection directions that on one hand maximize the Euclidean distance between the ear images of different classes and on the other hand minimize the distance between the ear images of the same class. The ratio is maximized when the column vectors of the projection matrix W are the eigenvectors of $S_W^{-1}S_B$. In ear recognition tasks, this method cannot be applied directly since the dimension of the sample space is typically larger than the number of samples in the training set. As a consequence, S_W is singular in this case. This problem is also known as the "small sample size problem" [1,2]. Numerous methods have been proposed to solve this problem, in the last decade.

The idea of common vectors was originally introduced for isolated word recognition problems in the case where the number of samples in each class was less than or equal to the dimensionality of the sample space. These approaches extract the common properties of classes in the training set by eliminating the differences of the samples in each class. A common vector for each individual class is obtained by removing all the features that are in the direction of the eigenvectors corresponding to the nonzero eigenvalues of the scatter matrix of its own class. The common vectors are then used for recognition. In our case, instead of using a given class's own scatter matrix, we use the within-class scatter matrix of all classes to obtain the common vectors. Then, a new set of vectors, called the discriminative common vectors, which will be used for classification are obtained from the common vectors [7].

3.4.1 Obtaining The Discriminative Common Vectors By Using The Range Space Of S_W

Let the training set be composed of C classes, where each class contains N samples, and let x_m^i be a d-dimensional column vector which donates the mth sample from the ith class. There will be a total of $M = NC$ samples in the training set. Suppose that $d > M - C$. In this case, S_W, S_B, and S_T are defined as,

$$S_W = \sum_{i=1}^{C} \sum_{m=1}^{N} \left(x_m^i - \mu_i\right)\left(x_m^i - \mu_i\right)^T \quad \textbf{(3.71)}$$

$$S_B = \sum_{i=1}^{C} N\left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T \quad \textbf{(3.72)}$$

and

$$S_T = \sum_{i=1}^{C} \sum_{m=1}^{N} \left(x_m^i - \mu\right)\left(x_m^i - \mu\right)^T = S_W + S_B \quad \textbf{(3.73)}$$

where μ is the mean of all samples, and μ_i is the mean of samples in the ith class.

In the special case where $\omega^T S_W \omega = 0$ and $\omega^T S_B \omega \neq 0$, for all $\omega \in R^d \setminus \{0\}$, the modified Fisher's Linear Discriminant criterion attains a maximum. However, a projection vector ω, satisfying the above conditions, does not necessarily maximize the between-class scatter. In this case, a better criterion is given below,

$$J\left(W_{opt}\right) = \underset{\left|W^T S_W W\right|=0}{\arg\max} \left|W^T S_B W\right| = \underset{\left|W^T S_W W\right|=0}{\arg\max} \left|W^T S_T W\right| \quad \textbf{(3.74)}$$

To find the optimal projection vectors ω in the null space of S_W, we project the ear samples onto the null space of S_W and then obtain the projection vectors by performing PCA. To do so, vectors that span the null space of S_W must first be computed. However, this task is computationally intractable since the dimension of this null space can be very large. A more efficient way to accomplish this task is by using the orthogonal complement of the null space of S_W, which typically is a significantly lower-dimensional space.

Let R^d be the original sample space, V be the range space of S_W, and $V^\perp$ be the null space of S_W. Equivalently,

$$V = span\{a_k \mid S_W a_k \neq 0, k = 1,...,r\} \quad \textbf{(3.75)}$$

and

$$V^\perp = span\{a_k \mid S_W a_k = 0, k = r+1,...,d\} \quad \textbf{(3.76)}$$

where $r < d$ is the rank of S_W, $\{a_1,...,a_d\}$ is an orthonormal set, and $\{a_1,...,a_r\}$ is the set of orthonormal eigenvectors corresponding to the nonzero eigenvalues of S_W.

Consider the matrices $Q = [a_1 \ ... \ a_r]$ and $\bar{Q} = [a_{r+1} \ ... \ a_d]$. Since $R^d = V \oplus V^\perp$, every ear image $x_m^i \in R^d$ has a unique decomposition of the form

$$x_m^i = y_m^i + z_m^i \quad \textbf{(3.77)}$$

where $y_m^i = Px_m^i = QQ^T x_m^i \in V$, $z_m^i = \bar{P}x_m^i = \bar{Q}\bar{Q}^T x_m^i \in V^\perp$ and P and $\bar{P}$ are the orthogonal projection operators onto V and $V^\perp$, respectively. Our goal is to compute

$$z_m^i = x_m^i - y_m^i = x_m^i - Px_m^i \quad \textbf{(3.78)}$$

To do this, we need to find a basis for V, which can be accomplished by an eigenanalysis of S_W. In particular, the normalized eigenvectors a_k corresponding to the nonzero eigenvalues of S_W will be an orthonormal basis for V. The eigenvectors can be obtained by calculating the eigenvectors of the smaller M by M matrix, $A^T A$, defined such that $S_W = AA^T$, where A is a d by M matrix of the form

$$A = \left[x_1^1 - \mu_1 \ \ ... \ \ x_N^1 - \mu_1 \ \ x_1^2 - \mu_2 \ \ ... \ \ x_N^C - \mu_C \right] \quad \textbf{(3.79)}$$

let λ_k and v_k be the kth nonzero eigenvalue and the corresponding eigenvector of $A^T A$, where $k \leq M - C$. Then, $a_k = Av_k$ will be the eigenvector that corresponds to

the kth nonzero eigenvalue of S_W. The sought-for projection onto $V^{\perp}$ is achieved by using Eq. (3.78). In this way, it turns out that we obtain the same unique vector for all samples of the same class,

$$x_{com}^{i} = x_{m}^{i} - QQ^{T}x_{m}^{i} = \overline{Q}\overline{Q}^{T}x_{m}^{i}, \quad m = 1,...,N, \;\; i = 1,...,C \tag{3.80}$$

i.e., the vector on the right-hand side of Eq. (3.80) is independent of the sample index m, we refer to the vectors x_{com}^{i} as the common vectors [36]. So it is enough to project a single sample from each class. This will greatly reduce the computational burden of the calculations.

After obtaining the common vectors x_{com}^{i}, optimal projection vectors will be those that maximize the total scatter of the common vectors,

$$J\left(W_{opt}\right) = \arg\max_{\left|W^{T}S_{W}W\right|=0} \left|W^{T}S_{B}W\right| = \arg\max_{\left|W^{T}S_{W}W\right|=0} \left|W^{T}S_{T}W\right| = \arg\max_{W} \left|W^{T}S_{com}W\right| \tag{3.81}$$

where W is a matrix whose columns are the orthonormal optimal projection vectors ω_k and S_{com} is the scatter matrix of the common vectors,

$$S_{com} = \sum_{i=1}^{C} \left(x_{com}^{i} - \mu_{com}\right)\left(x_{com}^{i} - \mu_{com}\right)^{T}, \quad i = 1,...,C \tag{3.82}$$

where μ_{com} is the mean of all common vectors,

$$\mu_{com} = \frac{1}{C}\sum_{i=1}^{C} x_{com}^{i} \tag{3.83}$$

In this case, optimal projection vectors ω_k can be found by an eigenanalysis of S_{com}. In particular, all eigenvectors corresponding to the nonzero eigenvalues of S_{com} will be the optimal projection vectors. S_{com} is typically a large d by d matrix and, thus, we can use the smaller matrix, $A_{com}^{T}A_{com}$, of size C by C, to find nonzero eigenvalues and the corresponding eigenvectors of $S_{com} = A_{com}A_{com}^{T}$, where A_{com} is the d by C matrix of the form

$$A_{com} = \left[\, x_{com}^{1} - \mu_{com} \;\; ... \;\; x_{com}^{C} - \mu_{com} \,\right] \tag{3.84}$$

There will be $C-1$ optimal projection vectors since the rank of S_{com} is $C-1$ if all common vectors are linearly independent. If two common vectors are identical, then the two classes, which are presented by this vector, cannot be distinguished. Since the optimal projection vectors ω_k belong to the null space of S_W, it follows that when the image samples x_m^i of the ith class are projected onto the linear span of the

projection vectors ω_k, the feature vector $\Omega_i = \left[\left\langle x_m^i, \omega_1 \right\rangle \ ... \ \left\langle x_m^i, \omega_{C-1} \right\rangle \right]^T$ of the projection coefficients $\left\langle x_m^i, \omega_k \right\rangle$ will also be independent of the sample index m. Thus, we have

$$\Omega_i = W^T x_m^i, \quad m = 1,...,N, \quad i = 1,...,C \tag{3.85}$$

We call the feature vectors Ω_i discriminative common vectors, and they will be used for classification of face images. The fact that Ω_i does not depend on the index m in Eq. (3.85).

To recognize a test image x_{test}, the feature vector of this test image is found by

$$\Omega_{test} = W^T x_{test} \tag{3.86}$$

which is then compared with the discriminative common vector Ω_i of each class using the Euclidian distance. The discriminative common vector found to be the closest to Ω_{test} is used to identify the test image.

Since Ω_{test} is only compared to a single vector for each class, the recognition is very efficient for real-time ear recognition tasks. In the Eigenear and the Fisherear methods, the test sample feature vector Ω_{test} is typically compared to all feature vectors of samples in the training set, making these methods impractical for real-time applications for large training sets.

The above method can be summarized as follows [7]:

- Step 1: Compute the nonzero eigenvalues and corresponding eigenvectors of S_W by using the matrix $A^T A$, where $S_W = AA^T$ and A is given by Eq. (3.79). Set $Q = \left[a_1 \ \ ... \ \ a_r \right]$, where r is the rank of S_W.
- Step 2: Choose any sample from each class and project it onto the null space of S_W to obtain the common vectors
  $x_{com}^i = x_m^i - QQ^T x_m^i, \quad m = 1,...,N, \quad i = 1,...,C$

Step 3: Compute the eigenvectors ω_k of S_{com}, corresponding to the nonzero eigenvalues, by using the matrix $A_{com}^T A_{com}$, where $S_{com} = A_{com} A_{com}^T$ and A_{com} is given in Eq. (3.84). There are at most $C-1$ eigenvectors that correspond to the nonzero eigenvalues. Use these eigenvectors to form the projection matrix $W = \left[\omega_1 \ \ ... \ \ \omega_{C-1} \right]$, which will be used to obtain feature vectors in Eq. (3.85) and Eq. (3.86).

3.5 Laplacianears

Many recognition techniques have been developed over the past few decades. One of the most successful and well-studied techniques to ear recognition is the appearance-based method. When using appearance-based methods, we usually present an image of size $n \times m$ pixels by a vector in an $n \times m$-dimensional space. In practice, however, these $n \times m$-dimensional spaces are too large to allow robust and fast ear recognition. A common way attempt to resolve this problem is to use dimensionality reduction techniques. Two of the most popular techniques for this purpose are Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

PCA is an eigenvector method designed to model linear variations in high-dimensional data. PCA performs dimensionality reduction by projecting the original n-dimensional data onto the $k(\square\ n)$-dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix. Its goal is to find a set of mutually orthogonal basis functions that capture the directions of maximum variance in the data and for which the coefficients are pairwise decorrelated. For linearly embedded manifolds, PCA is guaranteed to discover the dimensionality of the manifold and produces a compact representation.

LDA is a supervised learning algorithm. LDA searches for the project axis on which the data points of different classes are far from each other while requiring data points of the same class to be close to each other. Unlike PCA which encodes information in an orthogonal linear space, LDA encodes discriminating information in a linearly seperable space using bases that are not necessarily orthogonal. It is generally believed that algorithms based on LDA are superior to those based on PCA. However, some recent work shows that, when the training data set is small, PCA can outperform LDA, and also that PCA is less sensitive to different training data sets.

Recently, a number of research efforts have shown that the biometrics images possibly reside on a nonlinear submanifold. However, both PCA and LDA effectively see only the Euclidian structure. They fail to discover the underlying structure, if the ear images lie on a nonlinear submanifold hidden in the image space. Some nonlinear techniques have been proposed to discover the nonlinear structure of the manifold, e.g., Isomap, LLE and Laplacian Eigenmap. These nonlinear methods do yield impressive results on some benchmark artificial data sets. However, they yield maps that are defined only on the training data points and how to evaluate the maps on novel test data points remains unclear. Therefore, these nonlinear manifold

learning techniques might not be suitable for some computer vision tasks, such as ear recognition.

In the mean time, there has been some interest in the problem of developing low-dimensional representations through kernel based techniques for biometrics recognition. These methods can discover the nonlinear structure of the biometrics images. However, they are computationally expensive. Moreover, none of them explicitly considers the structure of the manifold on which the biometrics images possibly reside.

This part of my study focuses on a new approach to ear analysis (representation and recognition), which explicitly considers the manifold structure. To be specific, the manifold structure is modeled by a nearest-neighbor graph which preserves the local structure of the image space. An ear subspace is obtained by Locality Preserving Projections (LPP) [8]. Each ear image in the image space is mapped to a low-dimensional ear subspace, which is characterized by a set of feature images, called Laplacianears. The ear subspace preserves local structure and seems to have more discriminating power than the PCA approach for classification purpose.

Let highlight several aspects of the proposed approach below [8]:

1. While the Eigenears method aims to preserve the global structure of the image space, and the Fisherears method aims to preserve the discriminating information, Laplacianears method aims to preserve the local structure of the image space. In many real-world classification problems, the local manifold structure is more important than the global Euclidian structure, especially when nearest-neighbor like classifiers are used for classification. LPP seems to have discriminating power although it is unsupervised.
2. An efficient subspace learning algorithm for ear recognition should be able to discover the nonlinear manifold structure of the ear space. Laplacianears method explicitly considers the manifold structure which is modeled by an adjacency graph. Moreover, the Laplacianears are obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the ear manifold. They reflect the intrinsic face manifold structures.

LPP shares some similar properties to LLE, such as a locality preserving character. However, their objective functions are totally different. LPP is obtained by finding

the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. LPP is linear, while LLE is nonlinear. Moreover, LPP is defined everywhere, while LLE is defined only on the training data points and it is unclear how to evaluate the maps for new test points. In contrast, LPP may be simply applied to any new data point to locate it in the reduced representation space.

3.5.1 Learning a Locality Preserving Subspace

PCA and LDA aim to preserve the global structure. However, in many real-world applications, the local structure is more important. In this section, Locality Preserving Projection (LPP) is described, a new algorithm for learning a locality preserving subspace [9]. The complete derivation and theoretical justifications of LPP can be traced back to Eq. (3.90). LPP seeks to preserve the intrinsic geometry of the data and local structure. The objective function of LPP is as follows:

$$\min \sum_{ij} \left(y_i - y_j \right)^2 S_{ij} \tag{3.87}$$

where y_i is the one-dimensional representation of x_i and the matrix S is a similarity matrix. A possible way of defining S is follows:

$$S_{ij} = \begin{cases} \exp\left(-\left\| x_i - x_j \right\|^2 / t \right), & \left\| x_i - x_j \right\|^2 < \varepsilon \\ 0 & otherwise \end{cases} \tag{3.88}$$

or

$$S_{ij} = \begin{cases} \exp\left(-\left\| x_i - x_j \right\|^2 / t \right), & if\ x_i\ is\ among\ k\ nearest\ neighbors\ of\ x_j \\ & or\ x_j\ is\ among\ k\ nearest\ neighbors\ of\ x_i \\ 0 & otherwise \end{cases} \tag{3.89}$$

where ε is sufficiently small, and $\varepsilon > 0$. Here, ε defines the radius of the local neighborhood. In other words, ε defines the "locality". The objective function with our choice of symmetric weights $S_{ij} \left(S_{ij} = S_{ji} \right)$ incurs a heavy penalty if neighboring points x_i and x_j are mapped far apart, i.e., if $\left(y_i - y_j \right)^2$ is large. Therefore, minimizing it is an attempt to ensure that, if x_i and x_j are "close", then y_i and y_j are close as well. Following some simple algebraic steps, we see that

$$\frac{1}{2} \sum_{ij} \left(y_i - y_j \right)^2 S_{ij} = \frac{1}{2} \sum_{ij} \left(w^T x_i - w^T x_j \right)^2 S_{ij} \tag{3.90a}$$

$$= \sum_{ij} w^T x_i S_{ij} x_i^T w - \sum_{ij} w^T x_i S_{ij} x_j^T w \quad \textbf{(3.90b)}$$

$$= \sum_{i} w^T x_i D_{ii} x_i^T w - w^T XSX^T w \quad \textbf{(3.90c)}$$

$$= w^T XDX^T w - w^T XSX^T w \quad \textbf{(3.90d)}$$

$$= w^T X \left(D - S \right) X^T w \quad \textbf{(3.90e)}$$

$$= w^T XLX^T w \quad \textbf{(3.90f)}$$

where $X = \left[x_1, x_2, ..., x_n \right]$, and D is a diagonal matrix; its entries are column (or row since S is symmetric) sums of S, $D_{ii} = \sum_j S_{ji}$. $L = D - S$ is the Laplacian matrix. Matrix D provides a natural measure on the data points. The bigger the value D_{ii} (corresponding to y_i) is, the more "important" is y_i. Therefore, we impose a constraint as follows:

$$y^T Dy = 1 \quad \textbf{(3.91a)}$$

$$\Rightarrow w^T XDX^T w = 1 \quad \textbf{(3.91b)}$$

Finally, the minimization problem reduces to finding:

$$\underset{\substack{w \\ w^T XDX^T w = 1}}{\arg\min} \quad w^T XLX^T w \quad \textbf{(3.92)}$$

The transformation vector w that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$XLX^T w = \lambda XDX^T w \quad \textbf{(3.93)}$$

Note that the two matrices XLX^T and XDX^T are both symmetric and positive semidefined since the Laplacian matrix L and the diagonal matrix D are both symmetric and positive semidefinite.

The Laplacian matrix for finite graph is analogous to the Laplace Beltrami operator on compact Riemannian manifolds. While the Laplace Beltrami operator for a manifold is generated by the Riemannian metric, for a graph it comes from the adjacency relation. Belkin and Niyogi [10] showed that the optimal map preserving locality can be found by solving the following optimization problem on the manifold:

$$\min_{\|f\|_{L^2(M)}=1} \int_M \|\nabla f\|^2 \tag{3.94}$$

which is equivalent to

$$\min_{\|f\|_{L^2(M)}=1} \int_M L(f) f \tag{3.95}$$

where L is the Laplace Beltrami operator on the manifold, i.e., $L(f) = div\nabla(f)$. Thus, the optimal f has to be an eigenfunction of L. If we assume f to be linear, we have $f(x) = w^T x$. By spectral graph theory, the integral can be discretely approximated by $w^T XLX^T w$ and the L^2 norm of f can be discretely approximated by $w^T XDX^T w$, which will ultimately lead to the following eigenvalue problem:

$$XLX^T w = \lambda XDX^T w \tag{3.96}$$

The derivation reflects the intrinsic geometric structure of the manifold.

3.5.2 Statistical View of LPP

LPP can also be obtained from statistical viewpoint. Suppose the data points follow some underlying distribution. Let x and y be two random variables. We define that a linear mapping $x \to w^T x$ best preserves the local structure of the underlying distribution in the L^2 sense if it minimizes that $\|x - y\| < \varepsilon$. Namely,

$$\min_w E\left(\left|w^T x - w^T y\right|^2 \mid \|x - y\| < \varepsilon\right) \tag{3.97}$$

where ε is sufficiently small and $\varepsilon > 0$. Here, ε defines the "locality". Define $z = x - y$, then we have the following objective function:

$$\min E\left(\left|w^T z\right|^2 \mid \|z\| < \varepsilon\right) \tag{3.98}$$

It follows that,

$$E\left(\left|w^T z\right|^2 \mid \|z\| < \varepsilon\right) \tag{3.99a}$$

$$= E\left(w^T z z^T w \mid \|z\| < \varepsilon\right) \tag{3.99b}$$

$$= w^T E\left(z z^T \mid \|z\| < \varepsilon\right) w \tag{3.99c}$$

Given a set of sample points $x_1, x_2, ..., x_n$, we first define an indicator function S_{ij} as follows:

$$S_{ij} = \begin{cases} 1 & \|x_i - x_j\|^2 < \varepsilon \\ 0 & otherwise \end{cases} \quad \textbf{(3.100)}$$

Let d be the number of nonzero S_{ij}, and D be a diagonal matrix whose entries are column (or row since S is symmetric) sums of S, $D_{ii} = \sum_j S_{ji}$. By the Strong Law of Large Numbers, $E\left(zz^T \mid \|z\| < \varepsilon\right)$ can be estimated from the sample points as follows:

$$E\left(zz^T \mid \|z\| < \varepsilon\right) \quad \textbf{(3.101a)}$$

$$\approx \frac{1}{d} \sum_{\|z\|<\varepsilon} zz^T \quad \textbf{(3.101b)}$$

$$= \frac{1}{d} \sum_{\|x_i - x_j\|<\varepsilon} \left(x_i - x_j\right)\left(x_i - x_j\right)^T \quad \textbf{(3.101c)}$$

$$= \frac{1}{d} \sum_{i,j} \left(x_i - x_j\right)\left(x_i - x_j\right)^T S_{ij} \quad \textbf{(3.101d)}$$

$$= \frac{1}{d}\left(\sum_{i,j} x_i x_i^T S_{ij} + \sum_{i,j} x_j x_j^T S_{ij} - \sum_{i,j} x_i x_j^T S_{ij} - \sum_{i,j} x_j x_i^T S_{ij} \right) \quad \textbf{(3.101e)}$$

$$= \frac{2}{d}\left(\sum_i x_i x_i^T D_{ii} - \sum_{i,j} x_i x_j^T S_{ij} \right) \quad \textbf{(3.101f)}$$

$$= \frac{2}{d}\left(XDX^T - XSX^T \right) \quad \textbf{(3.101g)}$$

$$= \frac{2}{d} XLX^T \quad \textbf{(3.101h)}$$

where $L = D - S$ is the Laplacian matrix. The ith column of matrix X is x_i. By imposing the same constraint, we finally get the same minimization problem described in Section 3.5.1.

3.5.3 Learning Laplacianears for Representation

LPP is a general method for manifold learning. It is obtained by finding the optimal linear approximations to the eigenfunctions of the Laplace Beltrami operator on the manifold. Therefore, though it is still a linear technique, it seems to recover important aspects of the intrinsic nonlinear manifold structure by preserving local structure. Based on LPP, we describe our Laplacianears method for ear representation in a locality preserving subspace.

In the ear analysis and recognition problem, one is confronted with the difficulty that the matrix XDX^T is sometimes singular. This stems from the fact that sometimes the number of images in the training set (n) is much smaller than the number of pixels in each image (m). In such a case, the rank of XDX^T is at most n, while XDX^T is an $m \times m$ matrix, which implies that XDX^T is singular. To overcome the complication of a singular XDX^T, we first project the image set to a PCA subspace so that the resulting matrix XDX^T is nonsingular. Another consideration of using PCA as preprocessing is for noise reduction. This method, we call Laplacianears, can learn an optimal subspace for ear representation and recognition. The algorithmic procedure of Laplacianears is formally stated below.

1. PCA projection: The image set $\{x_i\}$ is projected into the PCA subspace by throwing away the smallest principal components. In this study, 98 percent of information is kept, in the sense of reconstruction error. For the sake of simplicity, x is still used to denote the images in the PCA subspace in the following steps. The transformation matrix of PCA is donated by W_{PCA}.
2. Constructing the nearest-neighbor graph: Let G denote a graph with n nodes. The ith node corresponds to the ear image x_i. An edge is put between nodes i and j if x_i and x_j are close, i.e., x_j is among k nearest neighbors of x_i, or x_i is among k nearest neighbors of x_j. The constructed nearest-neighbor graph is an approximation of the local manifold structure. Note that the ε-neighborhood is not used to construct the graph here. This is simply because it is often difficult to choose the optimal ε in the real-world applications, while k nearest-neighbor graph can be constructed more stably. The disadvantage is that the k nearest-neighbor search will increase the computational complexity of our algorithm. When the computational complexity is a major concern, one can switch to the ε-neighborhood.

3. Choosing the weights: In the case of unsupervised learning, if node i and j are connected, put $S_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}$ where t is a suitable constant. Otherwise, put $S_{ij} = 0$. The weight matrix S of graph G models the ear manifold structure by preserving local structure. The justification for this choice of weights can be traced back to [9]. In this study, the supervised learning was considered. If node i and j are connected, which means they are in the same class, put $S_{ij} = 1$. Otherwise, if they are not in the same class, put $S_{ij} = 0$.
4. Eigenmap: Compute the eigenvectors and eigenvalues for the generalized eigenvector problem $XLX^T w = \lambda XDX^T w$ where D is a diagonal matrix whose entries are column (or row, since S is symmetric) sums of S, $D_{ii} = \sum_j S_{ji}$ $L = D - S$ is the Laplacian matrix. The ith row of matrix X is x_i .

Let $w_0, w_1, ..., w_{k-1}$ be the solution of Eq. (3.96), ordered according to their eigenvalues, $0 \le \lambda_0 \le \lambda_1 \le ... \le \lambda_{k-1}$. These eigenvalues are equal to or greater than zero because the matrixes XLX^T and XDX^T are both symmetric and positive semidefinite. Thus, the embedding is as follows:

$$x \to y = W^T x \quad \textbf{(3.102)}$$

$$W = W_{PCA} W_{LLP} \quad \textbf{(3.103)}$$

$$W_{LLP} = [w_0, w_1, ..., w_{k-1}] \quad \textbf{(3.104)}$$

Where y is a k-dimensional vector. W is the transformation matrix. This linear mapping best preserves the manifold's estimated intrinsic geometry in a linear sense. The column vectors of W are the so-called Laplacianears.

4. EXPERIMENTS USING İTU-E DATABASE

There are many databases developed for face recognition experiments such as FERET Face Database, MIT Face Database, VCL Face Database and Georgia Tech Face Database. However, there is no adequate ear database for ear recognition experiments. That is why the author of this thesis developed an extensive ear database that consists of 50 subjects with both left and right ears. There are 12 photographs of each subject, 6 for left ear and 6 for right ear. The ear images were taken at 0 to 15 degree rotation under same day light condition. This ear database is named as İTU Ear Database (İTU-E).

The hit rates of 4 algorithms that are mentioned in previous sections were calculated by Random Subsampling and K-Fold Cross-validation. 20 random test and training set pairs were picked for random subsampling method. The hit rate of each pair was calculated. The hit rates of recognition algorithms were presented as the mean hit rate of 20 random test sets in Table 4.1, Table 4.3, Table 4.5 and Table 4.7. The dataset was divided randomly into 25 equal-sized parts for K-Fold cross-validation. One of these 25 parts was kept as validation set and remaining 24 parts were combined as training set. Each part was kept as validation set for once and hit rates were calculated. The validation set that had highest hit rate was determined and its training set was used to calculate the hit rate of the test set. This process was repeated 20 times. The means of the hit rates were presented as K-Fold cross-validation hit rates in Table 4.2, Table 4.4, Table 4.6 and Table 4.8.

4.1 Testing PCA by Random Subsampling

Table 4.1: Random Subsampling Hit Rates of PCA

	Dimension Number (Number of Feature Vectors)	Hit Rate (%)
Left Ear	4	58,6
	8	82,4
	16	91,6
	32	95,9
	64	96,8
	128	97,1
Right Ear	4	64,8
	8	87,7
	16	95,6
	32	98,9
	64	99,2
	128	99,3

4.2 Testing PCA by K-Fold Cross-validation

Table 4.2: K-Fold Cross-validation Hit Rates of PCA

	Dimension Number (Number of Feature Vectors)	Hit Rate (%)
Left Ear	4	58,6
	8	81,1
	16	89,4
	32	93,6
	64	93,5
	128	94,8
Right Ear	4	61,4
	8	84,6
	16	92,9
	32	96,8
	64	96,9
	128	97,2

Table 4.1 and Table 4.2 represent hit rates of PCA at some specific number of dimensions. Both random subsampling and K-Fold cross-validation methods were used to calculate hit rates. The purpose of this experiment is to show effects of number of selected feature vectors on hit rate.

4.3 Testing FLD by Random Subsampling

Table 4.3: Random Subsampling Hit Rates of FLD

Dimension Number (PCA)	Dimension Number (FLD)	Hit Rate (%) at Left Ear	Hit Rate (%) at Right Ear
4	2	31,2	32,9
8	4	71,1	78,5
16	4	86,1	87,5
	8	96,3	98,1
32	4	89,9	89,8
	8	98,3	98,7
	16	99,3	100
64	4	92,4	90,2
	8	98,7	99,1
	16	99,0	99,6
	32	99,6	99,9
128	4	85,7	88,0
	8	97,6	98,1
	16	98,9	99,6
	32	99,2	99,8

4.4 Testing FLD by K-Fold Cross-validation

Table 4.4: K-Fold Cross-validation Hit Rates of FLD

Dimension Number (PCA)	Dimension Number (FLD)	Hit Rate (%) at Left Ear	Hit Rate (%) at Right Ear
4	2	31,5	34,0
8	4	71,5	76,2
16	4	81,6	82,4
	8	92,3	94,7
32	4	86,4	87,3
	8	97,2	96,9
	16	98,3	99,4
64	4	88,5	87,1
	8	96,9	98,3
	16	98,3	99,6
	32	98,8	99,6
128	4	77,5	77,1
	8	92,6	94,3
	16	97,9	98,4
	32	98,3	99,1

Table 4.3 and Table 4.4 represent hit rates of FLD at some specific number of dimensions. Before processing multiple linear discriminant analysis, PCA was performed to reduce dimensions of dataset. Reducing dimensions of dataset reduces the computational cost of multiple linear discriminant analysis. Both random subsampling and K-Fold cross-validation methods were used to calculate hit rates. The purpose of this experiment is to show effects of number of selected feature vectors on hit rate.

4.5 Testing DCVA by Random Subsampling

Table 4.5: Random Subsampling Hit Rates of DCVA

	Dimension Number	Hit Rate (%)
Left Ear	4	94,0
	8	98,0
	16	99,3
Right Ear	4	91,3
	8	98,7
	16	100

4.6 Testing DCVA by K-Fold Cross-validation

Table 4.6: K-Fold Cross-validation Hit Rates of DCVA

	Dimension Number	Hit Rate (%)
Left Ear	4	91,0
	8	97,0
	16	100
Right Ear	4	96,0
	8	98,7
	16	100

Table 4.5 and Table 4.6 represent hit rates of DCVA at some specific number of dimensions. Both random subsampling and K-Fold cross-validation methods were used to calculate hit rates. The purpose of this experiment is to show effects of number of selected feature vectors on hit rate.

4.7 Testing Laplacianears by Random Subsampling

Table 4.7: Random Subsampling Hit Rates of Laplacianears

DNumber (PCA)	DNumber (LPP)	HR (%) at LE	HR (%) at RE
4	2	59,0	65,7
8	4	71,1	78,5
	8	86,2	91,5
16	4	86,1	87,5
	8	96,3	98,1
	16	93,3	97,7
32	4	89,9	89,8
	8	98,3	98,7
	16	99,3	100
	32	97,8	99,3
64	4	92,4	90,2
	8	98,7	99,1
	16	99,0	99,6
	32	99,6	100
	64	99,5	99,6
128	4	85,6	87,8
	8	97,7	98,1
	16	98,9	99,6
	32	99,1	99,8
	64	99,8	99,8
	128	99,0	99,4

4.8 Testing Laplacianears by K-Fold Cross-validation

Table 4.8: K-Fold Cross-validation Hit Rates of Laplacianears

DNumber (PCA)	DNumber (LPP)	HR (%) at LE	HR (%) at RE
4	2	58,7	61,2
8	4	71,1	77,3
	8	85,1	89,3
16	4	84,6	85,7
	8	95,2	96,8
	16	91,7	95,3
32	4	90,5	89,3
	8	98,6	98,4
	16	98,8	99,6
	32	97,6	98,8
64	4	88,5	86,3
	8	97,9	98,0
	16	98,3	99,5
	32	99,3	100
	64	99,2	99,1
128	4	81,1	83,7
	8	95,5	95,1
	16	97,3	98,3
	32	98,9	99,1
	64	99,1	99,6
	128	98,0	98,7

Table 4.7 and Table 4.8 represent hit rates of LPP at some specific number of dimensions. Before processing locality preserving projections, PCA was performed

to reduce dimensions of dataset. Reducing dimensions of dataset reduces the computational cost of locality preserving projections. Both random subsampling and K-Fold cross-validation methods were used to calculate hit rates. The purpose of this experiment is to show effects of number of selected feature vectors on hit rate.

4.9 Error Rate vs. Dimension Number

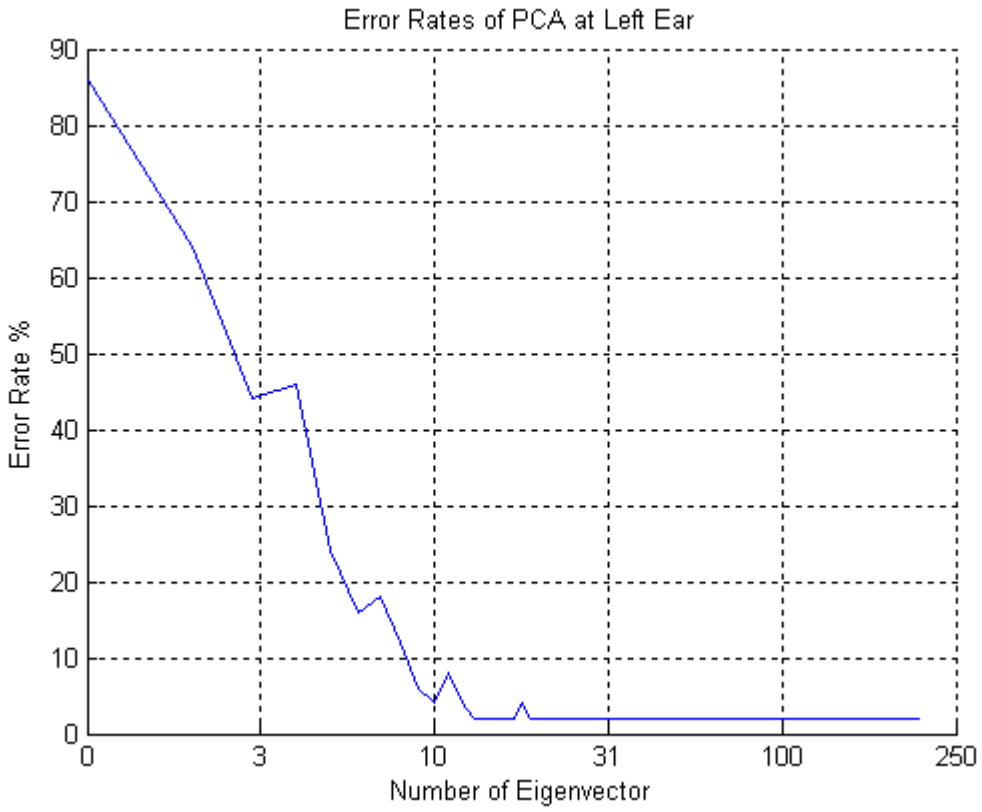


Figure 4.1: Error Rates of PCA at Left Ear

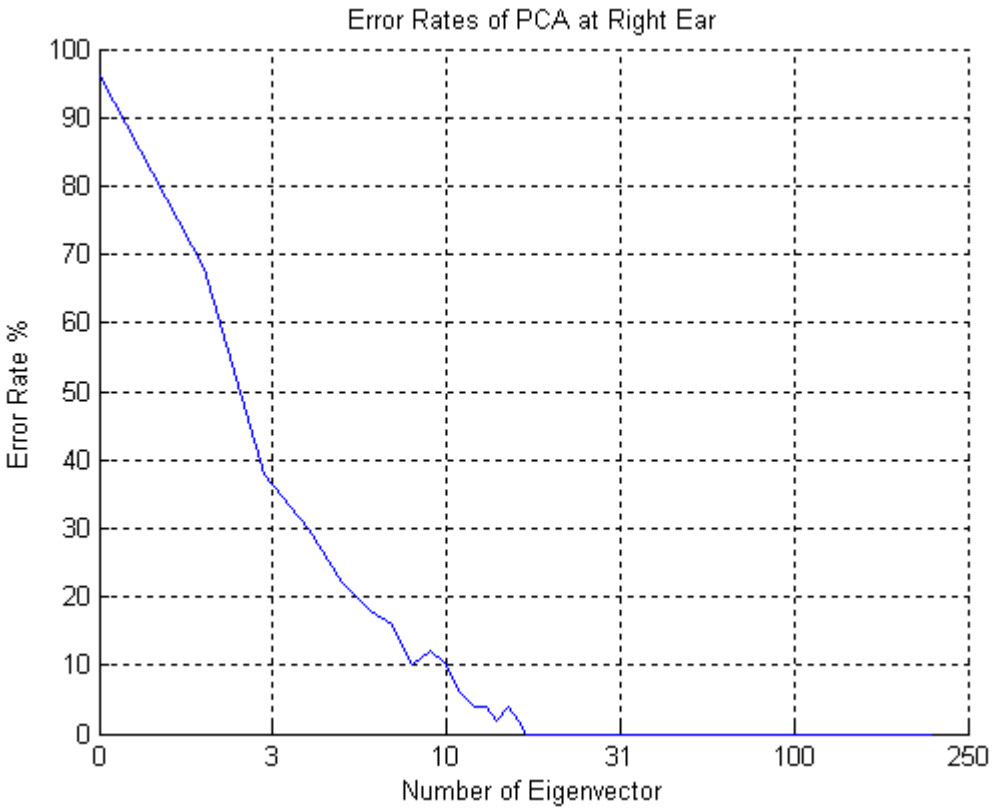


Figure 4.2: Error Rates of PCA at Right Ear

Figure 4.1 and Figure 4.2 show the relation of number of used feature vectors and hit rate of PCA.

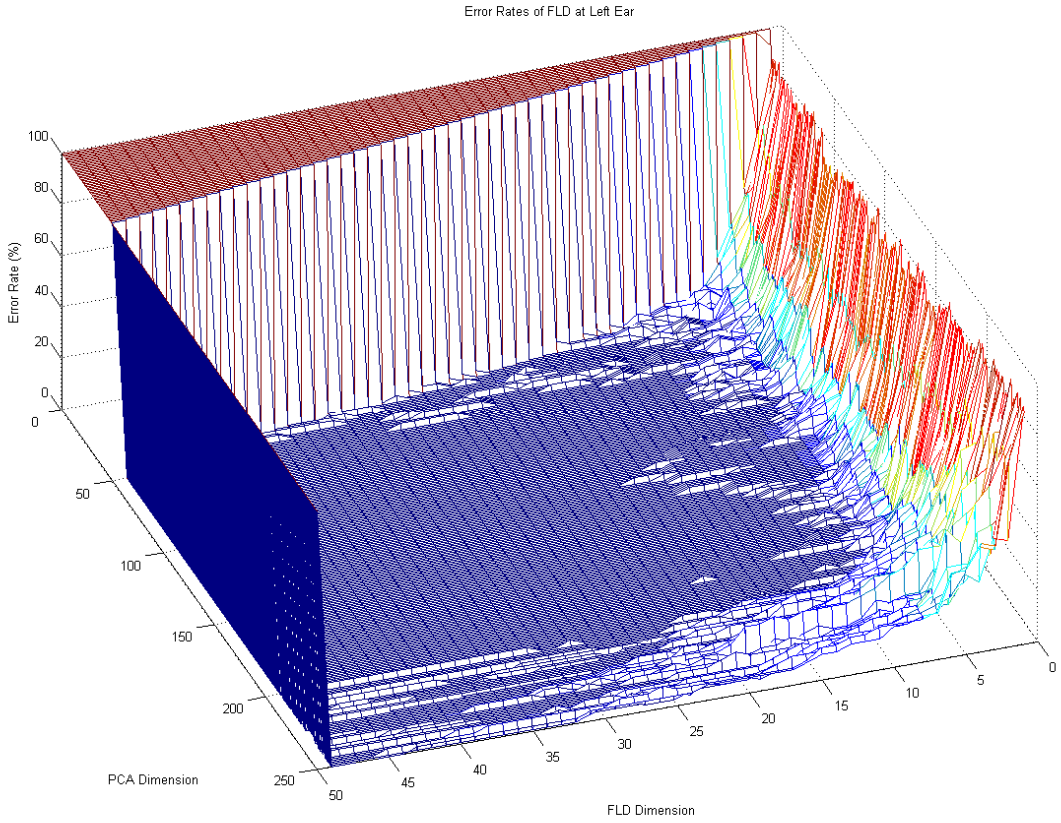


Figure 4.3: Error Rates of FLD at Left Ear

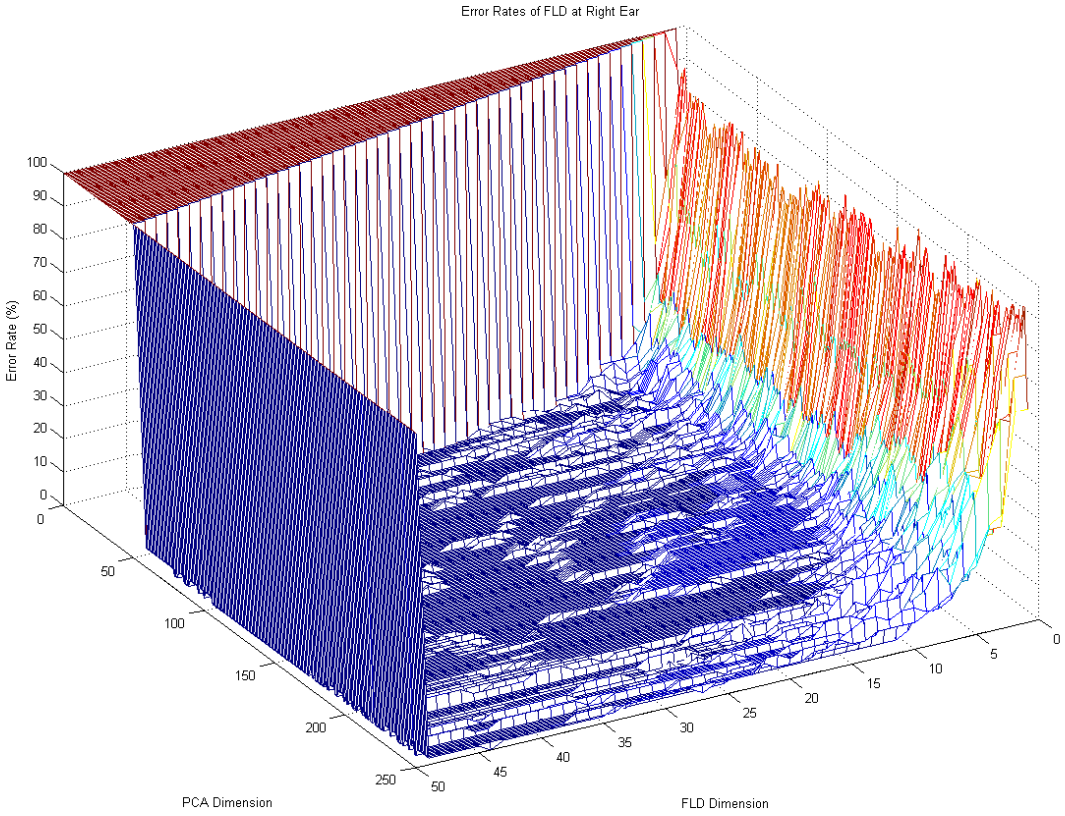


Figure 4.4: Error Rates of FLD at Right Ear

Figure 4.3 and Figure 4.4 show the relation of number of used feature vectors and hit rate of FLD. The hit rate of FLD is related to feature vectors of both LDA and PCA.

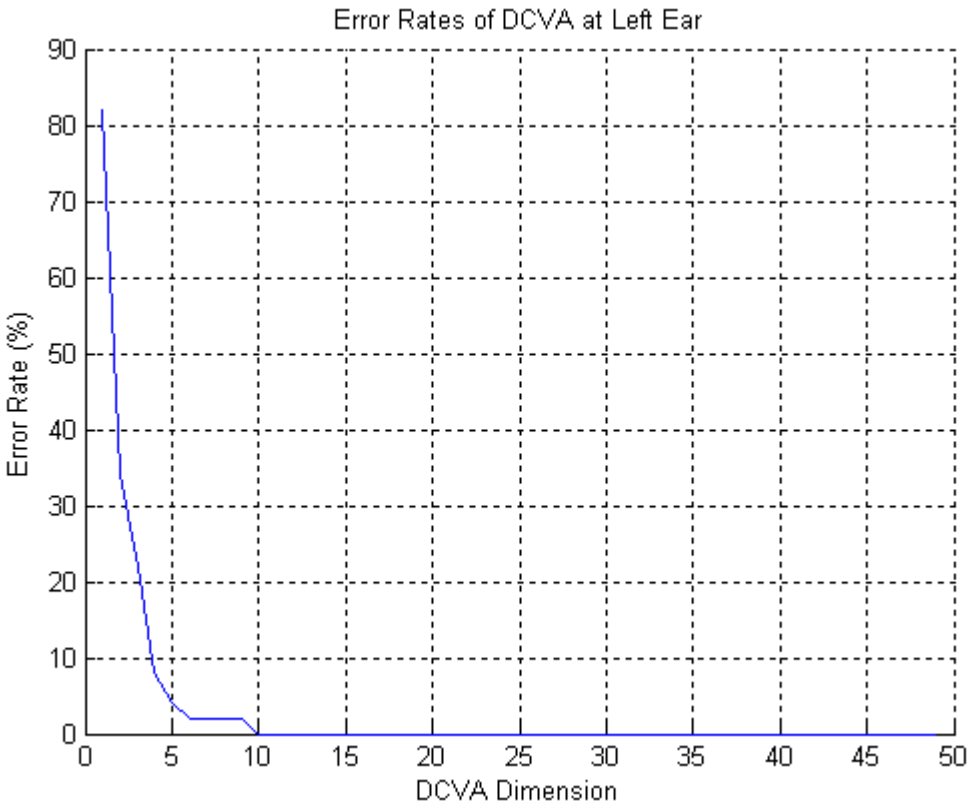


Figure 4.5: Error Rates of DCVA at Left Ear

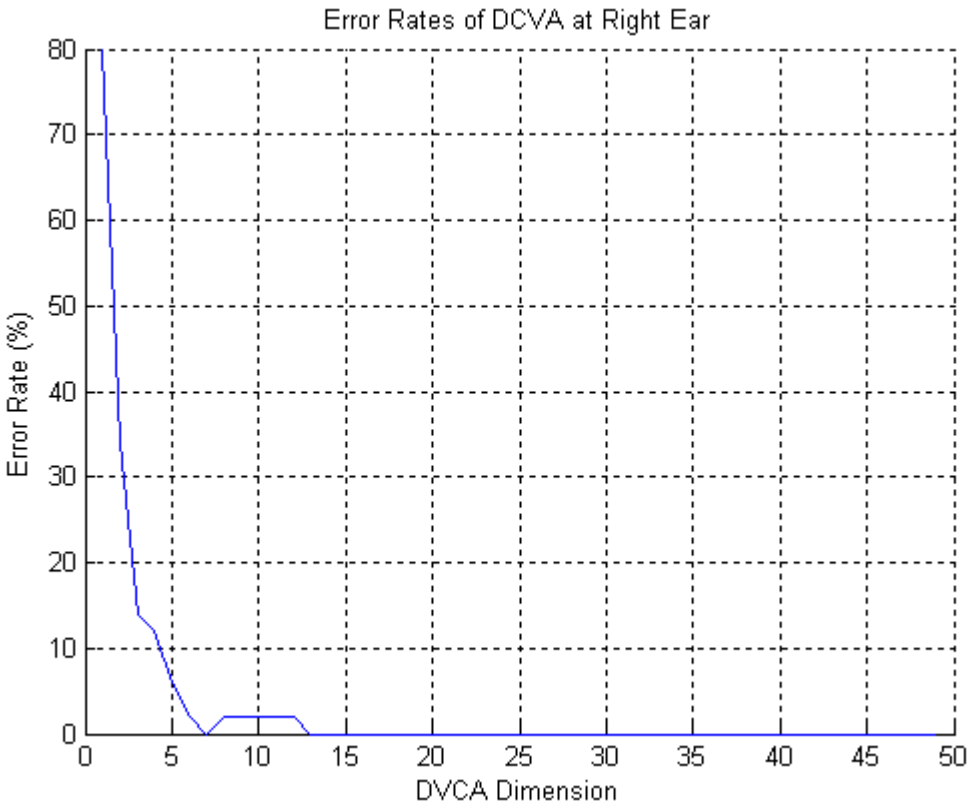


Figure 4.6: Error Rates of DCVA at Right Ear

Figure 4.5 and Figure 4.6 show the relation of number of used feature vectors and hit rate of DCVA.

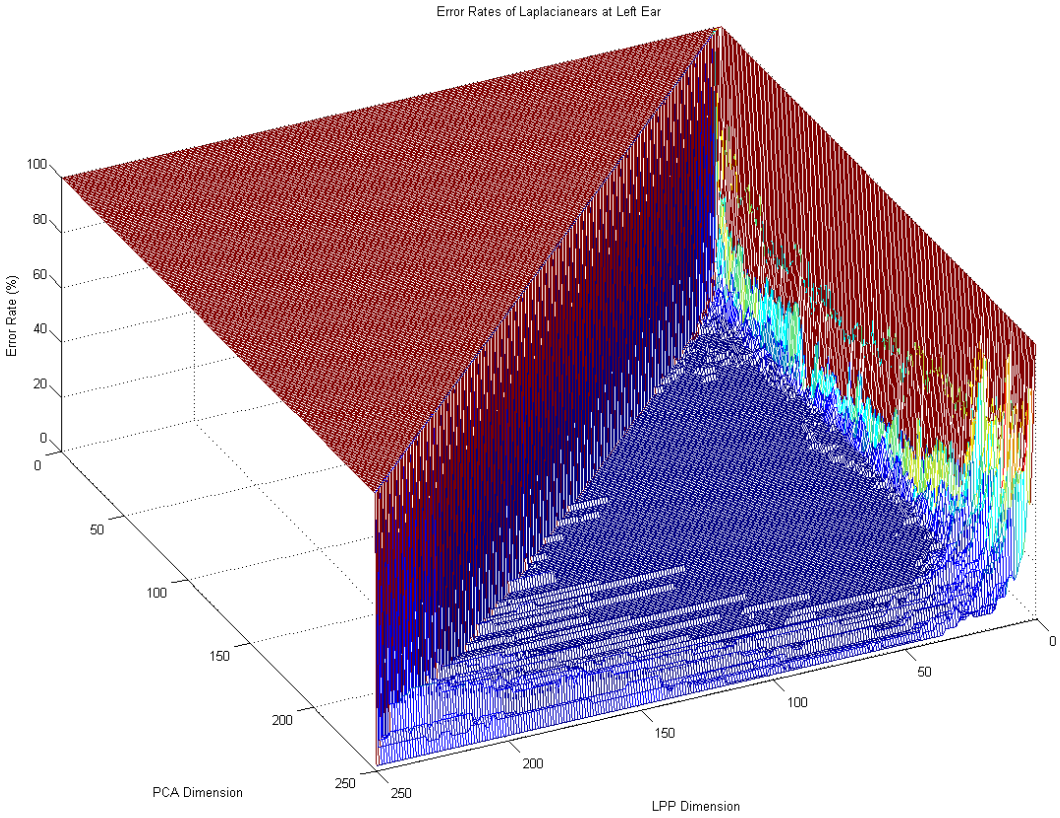


Figure 4.7: Error Rates of Laplacianears at Left Ear

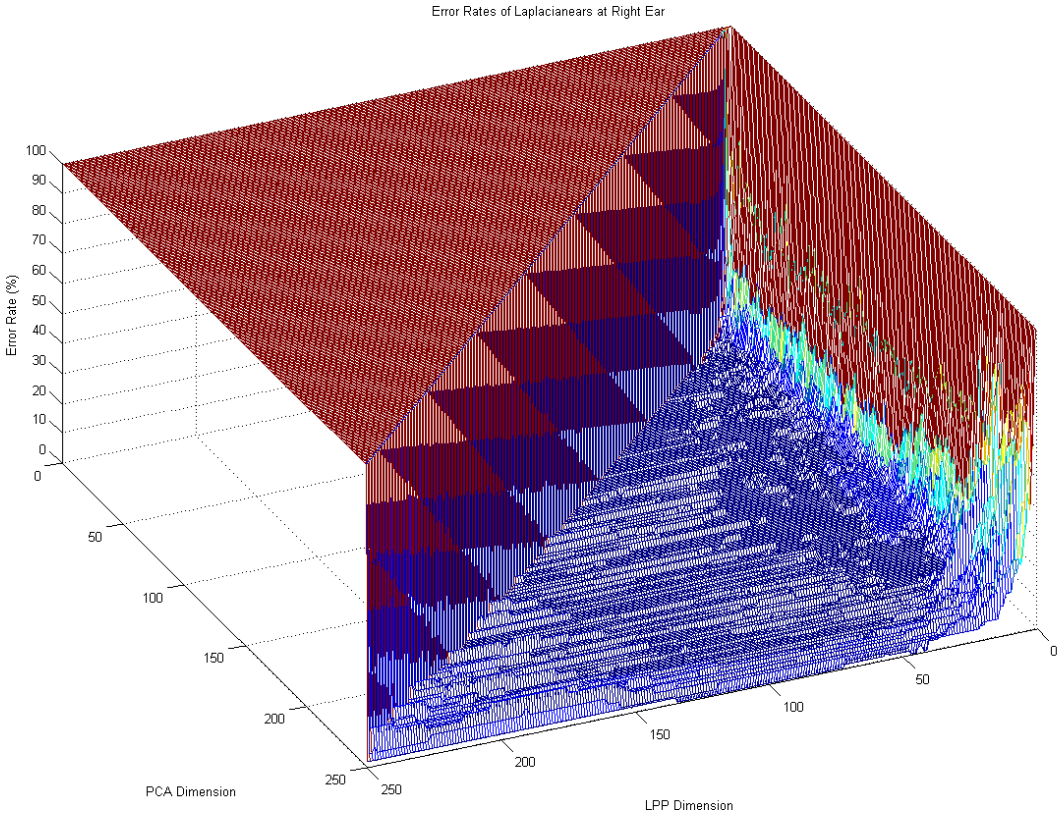
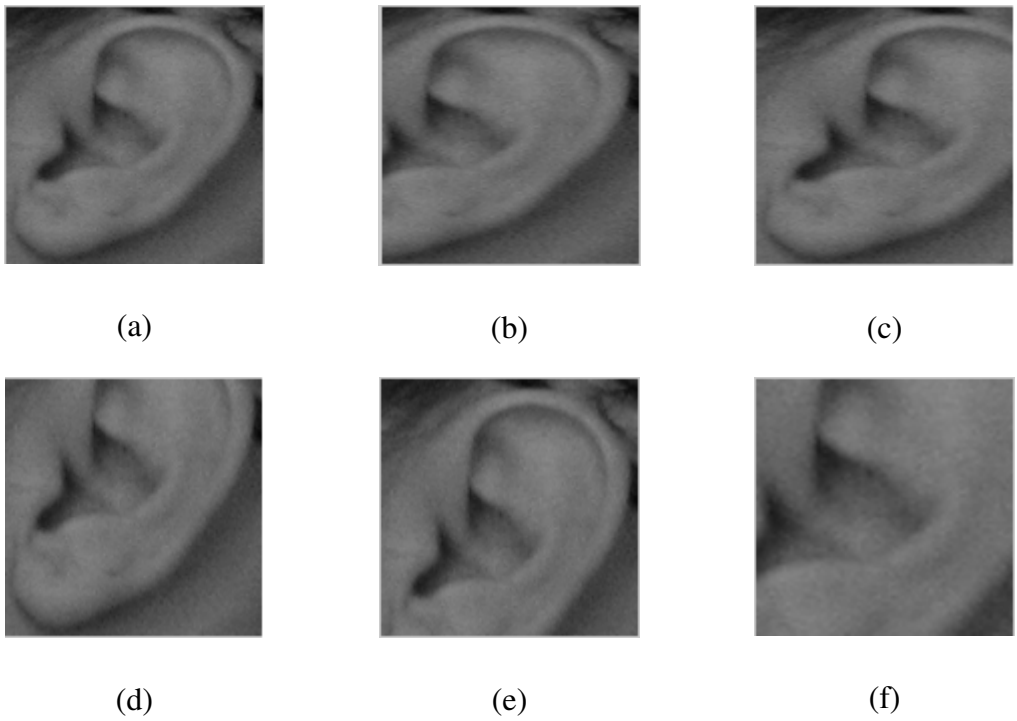


Figure 4.8: Error Rates of Laplacianears at Right Ear

Figure 4.7 and Figure 4.8 show the relation of number of used feature vectors and hit rate of Laplacianears. The hit rate of Laplacianears is related to feature vectors of both LPP and PCA.

4.10 Effects of Cropping

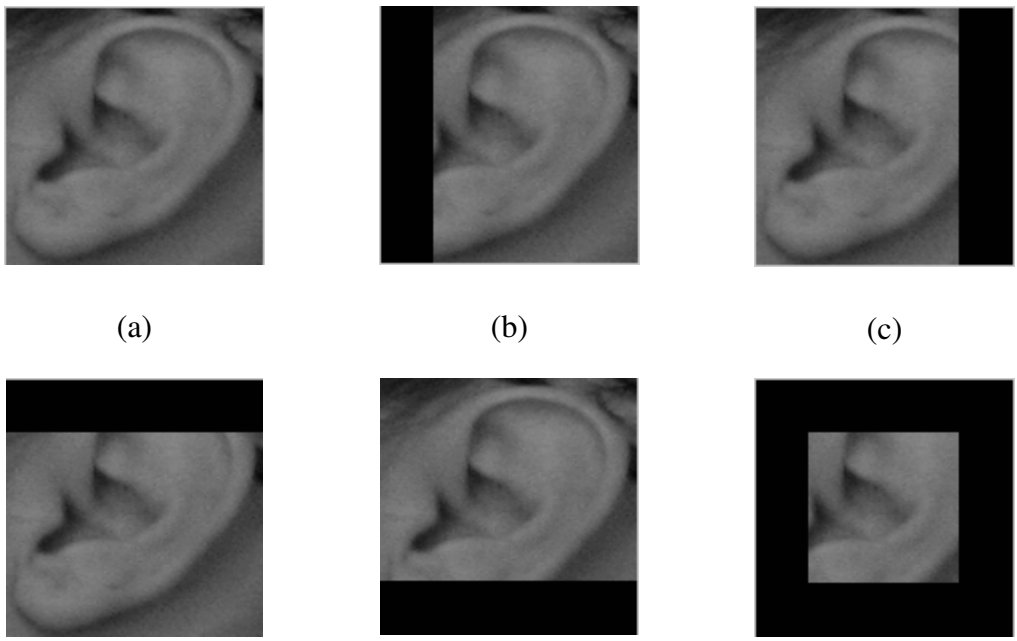
Testing images cropped in two different ways as shown in Figure 4.9 and Figure 4.10 below.



(a) (b) (c)

(d) (e) (f)

Figure 4.9: Test images are cropped and renormalized to 128*128 pixels. (a): uncropped. (b): cropped from left. (c): cropped from right. (d): cropped from top. (e): cropped from bottom. (f): cropped from all sides.



(a) (b) (c)

(d) (e) (f)

Figure 4.10: Black mask applied to test images. (a): unmasked. (b): masked from left. (c): masked from right. (d): masked from top. (e): masked from bottom. (f): masked from all sides.

Table 4.9: Hit Rates of PCA Over Cropped Left Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	92,2	82,3	63,7	48,4
Cropped from right	89,7	73,0	49,6	32,5
Cropped from top	86,4	37,8	22,7	17,5
Cropped from bottom	91,1	68,0	46,3	32,2
Cropped from all sides	66,3	18,8	9,8	6,9

Table 4.10: Hit Rates of PCA Over Masked Left Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	89,8	68,5	45,4	24,3
Masked from right	89,6	66,4	48,7	35,2
Masked from top	93,3	69,4	49,7	37,1
Masked from bottom	88,3	66,1	45,8	29,0
Masked from all sides	47,0	47,0	47,0	47,0

Table 4.11: Hit Rates of PCA Over Cropped Right Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	96,1	73,9	50,2	28,3
Cropped from right	96,5	85,8	70,6	46,4
Cropped from top	87,8	46,1	25,6	20,4
Cropped from bottom	95,5	80,4	54,8	32,5
Cropped from all sides	72,5	19,7	13,3	7,8

Table 4.12: Hit Rates of PCA Over Masked Right Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	93,8	80,0	62,8	46,9
Masked from right	93,7	72,3	42,9	22,6
Masked from top	97,9	78,5	61,4	47,2
Masked from bottom	93,9	73,7	62,2	40,4
Masked from all sides	55,3	55,3	55,3	55,3

During this experiment 32 PCA feature vector used. Hit rates of 32 dimensional uncropped/unmasked PCA process can be found in Table 4.1.

Table 4.13: Hit Rates of FLD Over Cropped Left Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	97,6	86,1	60,5	37,7
Cropped from right	97,2	80,7	60,5	37,0
Cropped from top	88,1	33,9	15,1	8,3
Cropped from bottom	94,5	60,2	39,1	25,8
Cropped from all sides	62,7	11,6	6,3	5,7

Table 4.14: Hit Rates of FLD Over Masked Left Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	88,7	50,3	23,5	13,5
Masked from right	95,2	80,0	65,5	47,1
Masked from top	93,8	40,8	24,5	20,1
Masked from bottom	87,2	62,3	36,9	19,6
Masked from all sides	54,6	22,5	11,9	6,5

Table 4.15: Hit Rates of FLD Over Cropped Right Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	99,2	87,1	66,9	46,4
Cropped from right	99,1	91,6	71,4	37,5
Cropped from top	86,9	2937	13,9	10,6
Cropped from bottom	98,2	69,1	36,6	19,1
Cropped from all sides	69,9	13,0	5,1	5,1

Table 4.16: Hit Rates of FLD Over Masked Right Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	98,9	96	85,3	57,7
Masked from right	76,6	30,5	6,7	2,2
Masked from top	94,3	53,8	29,3	25,4
Masked from bottom	89,3	70,1	59,3	32,1
Masked from all sides	46,6	6,9	3,7	3,3

During this experiment, the data set is represented in a lower dimensional space by PCA. 32 feature vectors are used for PCA process. After that, FLD applied to the dataset. 16 feature vectors are used for FLD process. Hit rates of uncropped/unmasked FLD process can be found in Table 4.2.

Table 4.17: Hit Rates of DCVA Over Ear Images

	Left		Right	
	Cropped %15	Masked %15	Cropped %15	Masked %15
Left	54,0	24,0	82,0	92,0
Right	72,0	64,0	72,0	32,0
Top	10,0	22,0	12,0	36,0
Bottom	30,0	40,0	36,0	52,0
All sides	6,0	10,0	2,0	2,0

During this experiment 16 DCVA feature vector used. Hit rates of 16 dimensional uncropped/unmasked DCVA process can be found in Table 4.3.

Table 4.18: Hit Rates of LPP Over Cropped Left Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	96,2	81,5	60,4	36,9
Cropped from right	96,7	84	60,1	35,1
Cropped from top	91,1	38,2	19,4	10,8
Cropped from bottom	93,6	61,6	37,3	21,7
Cropped from all sides	70	17,5	9,1	5,7

Table 4.19: Hit Rates of LLP Over Masked Left Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	92,2	69,4	39,4	19,1
Masked from right	95,8	88,3	74,5	60,9
Masked from top	94,3	57,7	33,8	26,7
Masked from bottom	93,9	72,8	45,5	25,9
Masked from all sides	64,5	21,9	11,9	7,2

Table 4.20: Hit Rates of LPP Over Cropped Right Ear Images

	Cropped %5	Cropped %10	Cropped %15	Cropped %20
Cropped from left	98,3	85,9	67,2	45,9
Cropped from right	97,8	85,1	61,8	32,3
Cropped from top	89,5	37,8	15,3	9,9
Cropped from bottom	96,7	76,4	46,5	26,5
Cropped from all sides	75,6	16	5,4	4,6

Table 4.21: Hit Rates of LPP Over Masked Right Ear Images

	Masked %5	Masked %10	Masked %15	Masked %20
Masked from left	97,6	93,2	85,6	63,4
Masked from right	94,2	62	29,7	9,1
Masked from top	97,3	69,2	45,5	35,3
Masked from bottom	97,2	89,5	77,5	53,6
Masked from all sides	71,2	17,5	10,8	8,7

During this experiment, the data set is represented in a lower dimensional space by PCA. 32 feature vectors are used for PCA process. After that, LPP applied to the dataset. 32 feature vectors are used for LPP process. Hit rates of uncropped/unmasked LPP process can be found in Table 4.4.

5. CONCLUSION

Several recognition algorithms were introduced in the last two decades. According to the results that are presented in this thesis, 2D image based, linear recognition algorithms have better performance at ear recognition tasks than face recognition task that are presented in the literature previously. The explanation of this situation is conditions such as make up, illumination, posing, the rotation angle of face to up/down/right/left directions and emotional expressions such as smiling and frowning brows, mustache and beard do not affect ear as much as they affect face.

PCA, FLD, DCVA and LPP are almost excellent at ear recognition tasks according to experiment have been done in this study.

The experiments of error rate versus the numbers of selected dimensions show that error rate can be minimized by choosing adequate number of dimensions for representation of ear images.

Four degrees of cropping applied to testing images to show effects of deformed test images. The experiments showed that if the cropping ratio increases, error rate increases. Hit rate and cropping are inversely proportional.

According to the experiment results, this thesis achieved its goal and showed that ear is adequate alternative to face for recognition tasks.

REFERENCES

[1] **Karimi B.,** 2006. Comparative Analysis of Face Recognition Algorithms and Investigation on the Significance of Color, *M. Sc. Thesis,* Concordia University, Montreal, Quebec.

[2] **Duda R. O., Hart P. E. and Stork D. G.,** 2001. Pattern Classification, John Wiley & Sons, Inc., New York.

[3] **Smith L. I., 2002,** A Tutorial on Principal Component Analysis, http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf (Last Access Date: 05.09.2007)

[4] **Turk M. A. and Pentland A. P.,** 1991. Face Recognition using Eigenfaces, *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591

[5] **Veksler O.,** 2004. Pattern Recognition, Lecture 8, http://www.csd.uwo.ca/~olga/Courses//CS434a_541a//Lecture8.pdf (Last Access Date: 20.09.2007)

[6] **Belhumeur P., Hespanha J. and Kriegman D.,** 1997, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 19, pp. 711-720.

[7] **Çevikalp H., Neamtu M., Wilkes M. and Barkana A,** 2005. Discriminative Common Vectors for Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no 1, pp. 4-13.

[8] **He X., Yan S., Hu Y., Niyogi P. and Zhang H. J.,** 2005. Face Recognition Using Laplacianfaces, *IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 27 no. 3, pp. 328-340.

[9] **He X. and Niyogi P.,** 2003. Locality Preserving Projections, *Proc. Conf. Advances in Neural Information Processing Systems*.

[10] **Belkin M. and Niyogi P.**, 2001, Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering, *Proc. Conf. Advances in Neural Information Processing Systems 15.*

[11] **Chang K., Bowyer K. W., Sarkar S. and Victor B.,** 2003. Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics, *IEEE Trans. On Pattern Analysis and Machine Intelligence,* vol. 25 no. 9, pp. 1160-1165.

[12] **Victor B., Bowyer K. W. and Sarkar S.,** 2002. An Evaluation of Face and Ear Biometrics, *Proc. Int'l. Conf. Pattern Recognition*, pp. 429-432.

[13] **Hurley D. J., Nixon M. S. and Carter J. N.,** 2005. Forced Field Feature Extraction for Ear Biometrics, *Computer Vision and Image Understanding,* 98, pp. 491-512.

[14] **Choras M.,** 2005. Ear Biometrics Based on Geometrical Feature Extraction, *Electronic Letters on Computer Vision and Image Analysis,* 5(3), pp. 84-95.

[15] **Bledsoe W. W.,** 1966. Man-machine Facial Recognition, Panoramic Research Inc., Palo Alto, CA

[16] **Kanade T.,** 1973. Picture Processing System by Computer Complex and Recognition of Human Faces, *Ph. D. Thesis,* Kyoto University, Kyoto, Japan.

[17] **Kanade T.,** 1977. Model Representations and Control Structures in Image Understanding, *Proc. 5th International Joint Conference on Artificial Intelligence,* pp. 1074-1082.

[18] **Niyogi P. Petejan E. and Zhong J.,** 1999. Feature based representation for audio-visual speech recognition, *Proceedings of the Audio Visual Speech Conference (AVSP'99),* Santa Cruz, CA.

[19] **Nishida S.,** 1986. Speech Recognition Enhancement by Lip-information, *ACM SIGCHI Bulletin*, 17(4), pp. 198-204.

[20] **Daugman J. G. and Downing C. J.,** 1993. Demodulation: A New Approach to Texture Vision In Harris L and Jenkin M. (Eds.), *Spatial Vision in Humans and Robots,* pp. 63-87. Cambridge University Press, Cambridge, UK.

[21] **Daugman J. G. and Downing C. J.,** 2001. Epigenetic randomness, complexity and singularity of human iris patterns, *Proceedings of the Royal Society, B,* 268, *Biological Sciences*, pp. 1737-1740.

[22] **Nixon M. S. and Carter J. N.,** 1999. Automatic Gait Recognition, In Jain A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 231-250. Kluwer Academic Publishers, Boston.

[23] **Persuad K. C., Lee D. H. and Byun H. G.,** 1999. Objective Odour Mesurements, In Jain A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 251-272. Kluwer Academic Publishers, Boston.

[24] **Sommerville B. A., McCormac J. P. and Broom D. M.,** 1994. Analysis of Human Sweat Volatiles: An Example of Pattern Recognition in the Analysis and Interpretation of Gas Chromatograms, *Pesticide Science,* 41(4), pp. 365-368.

[25] **Nalwa V. S.,** 1999. Automatic On-Line Signature Verification, In Jain, A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 143-164. Kluwer Academic Publishers, Boston.

[26] **Zunkel R. L.,** 1999. Hand Geometry Based Identification, In Jain, A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 87-101. Kluwer Academic Publishers, Boston.

[27] **Jain, A., Bolle R. and Pankanti S.,** 1999. Introduction to biometrics, In Jain, A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 1-42. Kluwer Academic Publishers, Boston.

[28] **O'Gorman, L.,** 1999. Fingerprint Verification, In Jain, A., Bolle R. and Pankanti S. (Eds.), *Biometrics: Personal Identification in Networked Society,* pp. 43-64. Kluwer Academic Publishers, Boston.

[29] **Iannarelli A.,** 1989. Ear Identification, *Forensic Identification Series,* Paramont Publishing Company, Fremont, California.

[30] **Burge M. and Burger W.,** 1999. Ear biometrics, In Jain, A., Bolle R. and Pankanti S. (Eds.) *Biometrics: Personal Identification in a Networked Society,* pp. 273-286, Kluwer Academic Publishers, Boston.

[31] **Burge M. and Burger W.,** 2000. Ear Biometrics in Computer Vision, *Pattern Recognition, 2000, Proceedings, 15th International Conference*, Barcelona 03-07.09.2000, vol. 2, pp. 822-826.

[32] **Hurley D. J., Nixon M. S. and Carter J. N.,** 2005. Forced Field Energy Functionals for Image Feature Extraction, *The Tenth British Machine Vision Conference BMVC99,* The University of Nottingham, Nottingham 13-16.09.1999, pp. 604-613.

[33] **Yan P. and Bowyer K. W.,** 2005. Empirical Evaluation of Advanced Ear Biometrics, *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference,* 20-26.06.2005, vol. 3, pp. 41-41.

[34] **Yan P. and Bowyer K. W.,** 2005. Ear Biometrics Using 2D and 3D Images, *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference,* 20-26.06.2005, vol. 3, pp. 121-121.

[35] **Middendorff C., Bowyer K. W. and Yan P.,** 2007. Multi-Modal Biometrics Involving the Human Ear, *Computer Vision and Pattern Recognition, 2007, CVPR'07 IEEE Conference,* 17-22.06.2007, pp. 1-2.

[36] **Chen L. F., Liao H. Y. M., Ko M. T., Lin J. C. and Yu G. J.** 2000. A New LDA-Based Face Recognition System Which Can Solve the Small Sample Size Problem, *Pattern Recognition,* vol. 33, pp. 1713-1726.

RESUME

Bahattin KOCAMAN was born in 1982 in Fethiye. He graduated from Eskişehir Fatih Fen Lisesi in 2000. He received his B. Sc. degree from İstanbul Technical University Electronics and Communication Engineering department in 2005. In the same year, he started M. Sc education at İstanbul Technical University Institute of Science and Technology Electronics Engineering Programme. He is working at İ.T.Ü Faculty of Electrical and Electronic Engineering Circuits and Systems division since December 2005.