ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

NEAR-INFRARED IMAGE BASED FACE RECOGNITION

M.Sc. THESIS

Nil SERİ

Department of Computer Engineering

Computer Engineering Programme

JUNE 2012

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

NEAR-INFRARED IMAGE BASED FACE RECOGNITION

M.Sc. THESIS

Nil SERİ
(504091520)

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Muhittin GÖKMEN

JUNE 2012

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

YAKIN KIZILÖTESİ GÖRÜNTÜ TABANLI YÜZ TANIMA

YÜKSEK LİSANS TEZİ

Nil SERİ
(504091520)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Muhittin GÖKMEN

HAZİRAN 2012

Nil SERİ, a M.Sc. student of ITU Graduate School of Science Engineering and Technology student ID 504091520 successfully defended the thesis entitled "Near-Infrared Image Based Face Recognition", which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Muhittin GÖKMEN** Istanbul Technical University	.....
Jury Members :	**Asst. Prof. Dr. M. Elif KARSLIGİL** Yıldız Technical University	.....
	Asst. Prof. Dr. Mustafa Kamaşak Istanbul Technical University	.....

Date of Submission : 04 May 2012
Date of Defense : 12 June 2012

FOREWORD

I would like to thank my advisor, Prof. Dr. Muhittin Gökmen, for his help and support throughout the thesis. I also wish to thank my family for their continuous support throughout my life as well as in all my studies.

May 2012

Nil SERİ
Computer Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

2.5D : Two-and-a-half-dimensional
2D : Two-dimensional
3D : Three-dimensional
ATM : Automated Teller Machine
CBSR : Center for Biometrics and Security Research
DCT : Discrete Cosine Transform
FAR : False Acceptance Rate
FIR : Far Infrared
FLD : Fisher's Linear Discriminant
FRVT : Face Recognition Vendor Tests
H-LZM^2 : Cascaded LZM transformation
IR : Infrared
IR-A DIN : Near Infrared
IR-B DIN : Short-wave-infrared
IR-C DIN : Medium/Long-wavelength infrared
LBP : Local Binary Pattern
LDA : Linear Discriminant Analysis
LED : Light Emitting Diode
LW : Short-wavelength
LWIR : Long-wavelength infrared
LZM : Local Zernike Moment
MW : Mid-wavelength
MWIR : Medium-wave-infrared
NIR : Near Infrared
OTCBVS : Object Tracking and Classification in and Beyond the Visible Spectrum
PCA : Principal Component Analysis
PMH : Phase-magnitude histogram
SWIR : Short-wave-infrared
SVM : Singular Value Decomposition

THz **:** Terahertz

USB **:** Universal Serial Bus

VIS **:** Visible Light

VL **:** Visible Light

ZM **:** Zernike Moment

LIST OF TABLES

Page

LIST OF FIGURES

Page

NEAR-INFRARED IMAGE BASED FACE RECOGNITION

SUMMARY

Humans have the ability to remember, recognize and distinguish faces and the scientists have been working on systems that can establish the same facility. The improvements in face recognition and numerous commercial face recognition systems has increased in a parallel way. Yet the need for more accurate systems still remains.

Some examples of the applications in which face recognition is being used are:

- Face-based video indexing and browsing engines
- Multimedia management
- Human-computer interaction
- Biometric identity authentication
- Surveillance systems

There are two kinds of scenarios in face recognition, namely cooperative and uncooperative. Survellience systems can be a good example for uncooperative user applications. Cooperative user applications are such as access control machine readable traveling documents, ATM, computer login, e-commerce and e-government systems. In cooperative user scenarios, a user is required provide his/her face in a proper position for the camera to have the face image captured properly, in order to be granted for the access. In fact, many face recognition systems have been developed for such applications.

The intrinsic and extrinsic factors of the face affect the performance of the face recognition. Face recognition should be performed based on intrinsic factors of the face only, like 3D shape reflectance of the facial surface. Extrinsic factors include eyeglasses, hairstyle, expression, posture, environmental lighting. They should be minimized for reliable face recognition.

A biometric system should adapt to the environment, not vice versa. Among several extrinsic factors, problems with uncontrolled environmental lighting is the topmost issue. Lighting conditions, especially the light angle, change the appearance of a face so much that the changes calculated between the images of a person under different illumination conditions are larger than those between the images of two different people under the same illumination conditions.

All of the local filters under study are insufficient by themselves to overcome variations due to changes in illumination direction. So, therefore, near infrared imaging is proposed. Studies on imaging beyond visible spectrum has been carried on recently. However, thermal imaging has many disadvantages as well as its advantages. Enviromental temperature, physical and emotional conditions, drinking

alcohol can affect the system's success drastically. Studies have shown they have not performed better than visible image based systems. 3D visible imaging had also been tried but the load created during its process and wearing sunglasses or an open mouth can fail the system's success.

There are two principles for the active lighting in near-infrared imaging:

- The lights should be strong enough to produce clear frontal-lighted face image but not cause disturbance to human eyes
- The resulting face image should be affected as little as possible after minimizing the environmental lighting.

In this work, firstly, traditional face recognition methods such as PCA, LDA and LBP have been tried on NIR images for comparison with other methods. In Eigenfaces approach, "eigenfaces" are constructed from the face images, by means of PCA. The purpose of PCA is to reduce the large dimensionality of the data space to the smaller intrinsic dimensionality of feature space. In Fisherfaces approach, where LDA is applied after PCA, the projection direction is found so that the images belonging to different class, here the different ids, are separated maximally. In other words, the projection matrix that makes the ratio of the between-class scatter matrix and within-class scatter matrix of the images maximum, is found.

Local image representations such as Gabor and LBP has arisen great interest. For robust face recognition, dealing with extrinsic properties of face is an important issue. LBP texture operator can handle the variations caused by these properties, such as illumination, so it has become a popular approach in various applications. LBP representation is used to compensate for the degree of freedom in a monotonic transform in the gray tone to achieve an illumination invariant representation of faces for indoor face recognition applications. The pixels of an image are labeled as 0 or 1, by thresholding the neighborhood of each pixel, considering the result as a binary number.

The LBP operator was extended for neighborhood of different sizes and radius by bilinearly interpolating values at non-integer pixel coordinates. Another extension is the uniform patterns. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa. Uniform LBPs that have (8,1), (8,2) and (16,2) neighborhood and radius size are computed.

LBP+LDA is also used in this work. After uniform LBP(8,1) representations of the images are obtained, they are downsampled because of the memory limitations. Then LDA is performed on the downsampled feature sets after PCA is applied to make the within-class scatter matrix nonsingular.

Zernike moments are used to further improve the face recognition performance. Global Zernike moments are modified to obtain a local representation, such as LBP, called Local Zernike moments (LZM). The moments are computed at each pixel, considering their neighborhood and moment components obtained to capture the micro structure around each pixel. A complex moment image, which has the same size of the original face image, is obtained for each moment component. Later, each moment image is divided into non-overlapping subregions and phase-magnitude histograms are extracted from each subregion. Finally, the phase-magnitude histograms are concatenated and the face representation is built.

Since the use of LDA on LBP has positive effects on the success of the recognition, LZM+LDA is implemented for this study. The process of applying LDA on LZM is the same as the process in LBP+LDA. The phase-magnitude moments are downsampled and PCA is applied before LDA operation. Afterwards, the LDA projections are calculated and cosine distance is used for the matching operation. It is found out that the success of LZM+LDA over LZM is significant.

The tests in this study are performed with the following methods:

1. PCA with Mahalanobis distance
2. LDA with cosine distance
3. LBP with chi-square distance (original uniform (8,1), (8,2) and (16,2))
4. LBP+LDA with cosine distance
5. LZM with Manhattan distance
6. LZM+LDA with cosine distance

Both identification and verification have been tested for the methods. In face identification, a system tries to figure who the person is. In face verification, the system verifies whether the identity a person claims to be is true.

CBSR NIR Face Dataset of OTCBVS Benchmark Dataset Collection is used. The database contains 3,940 NIR face images of 197 people. The images were taken by an NIR camera with active NIR lighting. 18 NIR LEDs are mounted on the camera.

It is found that LZM performs better than both the original and extended uniform LBP methods in verification and identification tests. A method's combination with LDA carries the success of face recognition to higher levels. In identification step, however, the extended LBP operators are more successful than LDA itself but in verification step, LDA is more successful than all the LBP operators. The success rate of PCA is not good enough to catch up with the other methods in face recognition.

Using NIR face images for face recognition saves the system from the load of the preprocessing steps before the recognition. With the help of LZM on NIR images, robust and highly accurate systems can be built. Yet, NIR imaging is not improved enough to handle outdoor and uncooperative user applications. Future works on this context can help the system's success carry to a higher level.

YAKIN KIZILÖTESİ GÖRÜNTÜ TABANLI YÜZ TANIMA

ÖZET

İnsanların yüzleri hatırlama, tanıma ve ayrıştırma yetenekleri doğuştandır. Yüz tanıma alanındaki gelişmeler ve çeşitli ticari yüz tanıma uygulamaları birbirlerine paralel ilerlemiştir. Yine de, daha hatasız ve doğru sistemlere olan ihtiyaç devam etmektedir.

Yüz tanımanın kullanıldığı bazı uygulama örnekleri aşağıdaki gibidir:

- Yüz tabanlı video dizinleme ve arama motorları
- Multimedya yönetimi
- İnsan-bilgisayar etkileşimi
- Biyometrik kimlik tanıma
- Takip sistemleri

Yüz tanımada işbirliği içinde ve işbirliği etmeyen olarak iki tip senaryo bulunmaktadır. Takip sistemleri, işbirliği etmeyen kullanıcı uygulamaları için iyi bir örnektir. İşbirliği içinde olan kullanıcı uygulamalarına da geçiş control makinelerinde okunabilen seyahat dökümanları, ATM, bilgisayarda oturum açma, e-ticaret ve e-devlet uygulamaları örnek verilebilir. Kullanıcının sistemle işbirliği içinde olduğu uygulamalarda, sistemin kabulu için yüzün kameraya uygun bir şekilde konumlandırıldıktan sonra yüz resminin elde edilmelidir. Aslında çoğu yüz tanıma sistemleri bu tip uygulamalar için geliştirilmiştir.

Yüze ait iç ve dış faktörler yüz tanıma işleminin performansını etkilemektedir. Yüz tanıma yüz yüzeyinin 3D şekil yansıması gibi, yalnızca yüze ait iç faktörlere dayandırılmalıdır. Dış faktörler gözlük, saç modeli, yüz ifadesi, poz ve çevresel ışıklandırma gibi özellikleri içerir. Güvenilir bir yüz tanıma için etkileri en aza indirgenmelidir.

Biyometrik bir sistem çevreye uyum sağlamalıdır, bu durumun tam tersi düşünülemez. Çeşitli dış faktörlerin arasından kontrolsüz çevresel ışıklandırma en önemli konudur. Işıklandırma koşulları, özellikle ışığın açısı, yüzün görünümünü öyle çok değiştirmektedir ki; farklı ışıklandırma altında aynı kişiye ait görüntüler ile aynı ışıklandırma altında iki ayrı kişiye ait görüntüler arasında hesaplanan farklılık daha fazladır.

Üzerinde çalışılmakta olan bölgesel filtrelerin çoğu kendi başlarına ışıklandırma yönünün sebep olduğu değişimlerin üstesinden gelmekte yetersizdir. Bu sebeple, yakın kızılötesi görüntüleme önerilmiştir. Son zamanlarda, görünür spectrum ardı görüntüleme üzerine çalışmalar yürütülmektedir. Ancak, termal görüntülemenin üstünlükleri yanısıra birçok dezavantajı vardır. Çevresel sıcaklık, fiziksel ve duygusal durum, alkol alımı sistemin başarısını çok fazla etkilemektedir. Çalışmalar, thermal görüntüleme ile yapılan tanıma işlemlerinin, görünür ışık tabanlı görüntüleme işlemlerinden daha iyi bir performans sergilemediklerini göstermiştir.

3D görüntüleme de kullanılan yöntemler arasındadır; fakat işlem yükü, görüntüleme sırasında gözlük takılması veya ağzın açık olma durumu sistemi başarısız kılabilir.

Yakın kızıl-ötesi için aktif ışıklandırmada dikkat edilmesi gereken iki önemli husus vardır.

- Işıklar net bir önden aydınlatılmış yüz resmi sağlayacak şiddette olmalı; fakat göze rahatsızlık vermemelidir.
- Elde edilen yüz resmi çevresel ışıklandırmadan minimum derecede etkilenmiş olmalıdır.

Bu çalışmada, diğer metotlarla karşılaştırma amacıyla, yakın kızılötesi (YKÖ) imajları üzerinde öncelikle PCA, LDA ve LBP gibi geleneksel yüz tanıma metotları uygulanmıştır. Eigenfaces yaklaşımında, "öz yüzler" PCA yardımıyla yüz imajlarından oluşturulmuştur. PCA'in amacı yüksek boyutlu veri uzayını, daha az boyuta sahip içsel özellik uzayına dönüştürmektir.

LDA'in PCA'den sonra uygulandığı Fisherfaces yaklaşımında, projeksiyon yönü bulunur böylece farklı id'li, farklı sınıflara ait imajlar azami ölçüde ayrıştırılacaktır. Diğer bir deyişle, sınıflar arası dağılım matrisi ve sınıf içi dağılım matrisi oranını maksimum yapan projeksiyon matrisi bulunur.

Gabor ve LBP gibi yerel görüntü temsilleri ile ilgili çalışmalar da merak uyandırmaktadır. Başarılı bir yüz tanıma için yüzün dışsal özellikleri ile uğraşmak önemli bir konudur. LBP doku operatörü, ışıklandırma gibi özellikler nedeniyle oluşan değişimlerle başa çıkabilmektedir; bu yüzden çeşitli uygulamalarda popular bir yaklaşım haline gelmiştir. Kapalı mekan için yapılan yüz tanıma uygulamalarında, ışıklandırma bağımsız yüz temsilinde, gri tonlamadaki monotonik dönüşümün serbestlik derecesini telafi etmek amacıyla LBP gösterimi kullanılmaktadır. İmaja ait pixeller, komşu piksellerin eşik değeri olarak ilgili pikselle karşılaştırılması ile 0 veya 1 olarak etiketlenir.

LBP operatörü, tamsayı olmayan piksel koordinatlarında çift doğrusal interpolasyon uygulayarak, farklı boyut ve çaplardaki komşuluklarda kullanılabilmesi için geliştirilmiştir. Başka bir değişik kullanımı ise tek biçim dokulardır. Yerel bir ikili değer dokusu, 0'dan 1'e veya tersi şeklinde en fazla iki bitsel geçiş içeriyorsa tek biçim olarak adlandırılır. Bu çalışmada, (8,1), (8,2) ve (16,2) komşu sayısı ve çap için tek biçim LBP'leri hesaplanmıştır.

LBP+LDA metotu da bu çalışmada kullanılmıştır. İmajlara ait ek biçim (8,1)'lik LBP görüntü temsilleri elde edildikten sonra, bellek kısıtlarından ötürü alt örnekleme ile boyutu düşürülür. Tekil olmayan sınıf içi dağılım matrisi için PCA işleminden sonra, alt örneklenmiş özellik sınıfları üzerinde LDA uygulanır.

Yüz tanıma performansını daha da arttırmak için Zernike momentleri kullanılmıştır. Global Zernike momentleri, LBP gibi bir yerel görüntü temsilleri eldesi için değiştirilmiştir. Komşuluklar ve her bir piksel etrafındaki mikro yapıyı yakalamak için bulunan moment bileşenleri dikkate alınarak, momentler her bir piksel için hesaplanmıştır. Asıl yüz imajı boyutlarına sahip kompleks moment imajı, her bir moment bileşeni için elde edilir. Daha sonra, her moment imajı, üst üste denk gelmeyecek şekilde alt bölgelere bölünür ve her bir alt bölgeden faz-büyüklük histogramları çıkartılır. Bu histogramlar peşi sıra birbirine eklenerek yüz temsili elde edilir.

LBP ve LDA metotlarının birlikte kullanımı yüz tanıma başarısını olumlu bir şekilde etkilemektedir. Bu yüzden LZM ile LDA de birlikte kullanılarak, başarısı test edilmiştir. LDA'in LZM üzerine uygulanma şekli LBP+LDA işlemindekinin aynısıdır. Faz-büyüklük histogramlarının alt örnekleme ile boyutu düşürülmüştür. Daha sonar, LDA projeksiyonları hesaplanmış ve cosine benzerliği formülü ile eşleşme operasyonu gerçekleştirilmiştir. Sonuçlardan anlaşıldığı üzere, LZM+LDA'in LZM üzerinde belirgin bir üstünlüğü vardır.

Bu çalışmada aşağıdaki metotlar kullanılmıştır:

1. Mahalanobis mesafesi ile PCA
2. Cosine benzerliği ile LDA
3. Ki-kare mesafesi ile tek biçim LBP (original (8,1), (8,2) ve (8,16))
4. Cosine benzerliği ile LBP+LDA
5. Manhattan mesafesi ile LZM
6. Cosine benzerliği ile LBP+LZM

Bu çalışma için oluşturulan yazılım hem kimlik tanımlama hem de kimlik doğrulama için test edilmiştir. Kimlik tanımlamada, sistem kullanıcının kim olduğunu bulmaya çalışır. Kimlik doğrulamada ise, kullanıcı belirli bir kimlik olduğunu iddia eder ve sistem bunun doğruluğunu kontrol eder.

Testler için OTCBVS kalite testi veri kümesi koleksiyonundan CBSR NIR yüz veritabanı kullanılmıştır. Veritabanında 197 farklı kişiye ait toplam 3,940 YKÖ yüz imajı bulunmaktadır. Görüntüler, aktif yakın kızıl-ötesi ışıklandırma ile yakın kızıl-ötesi kamera kullanarak çekilmiştir. Kameranın üstüne konumlandırılmış 18 adet yakın kızıl-ötesi led bulunmaktadır.

Bu çalışma için yapılan testler sonucunda, LZM'in başarısı, hem orijinal tek biçim LBP hem de farklı komşuluk sayısı ve çapta kullanım için geliştirilmiş olan tek biçim LBP metotlarından daha yüksek çıkmıştır. Metotların LDA ile birlikte kullanımı ise yüz tanıma işleminin başarısını daha üst seviyelere taşımaktadır. Kimlik doğrulama adımında, LBP operatörlerinin başarısı tek başına LDA'in başarısından daha fazladır; ancak kimlik tanımlama adımında LDA'in başarısı, LBP'nin üstünde çıkmıştır. PCA kullanımı ise hem tanımlama hem doğrulama için diğer metotların başarımlarını yakalayamamış; güvenilir bir yüz tanıma için yetersiz kalmıştır.

Bir YKÖ yüz imajı, yüz tanıma sistemleri için sorunsuz bir girdi oluşturmaktadır; çünkü tanıma aşamasından önceki ağır ön işleme adımlarını azaltmaktadır. LZM işleminin de yardımlarıyla, YKÖ görüntüleme sisteminden elde edilmiş yüz imajları ile hızlı ve yüksek başarımlı yüz tanıma sistemleri gerçekleştirilebilir. Yalnız, YKÖ görüntüleme, işbirliği etmeyen kullanıcı uygulamaları için henüz uygun değildir. Ayrıca, dış mekan kullanımı da özellikle görünür ışığın, güneşli havalar gibi baskın olacağı yerlerde başarılı olamayabilir. Gelecekte, YKÖ görüntüleme sistemlerinde yapılacak çalışmalar ile bu tür kısıtların üzerinden gelinebilir.

1. INTRODUCTION

In many applications, such as e-passport and driver's license, the enrollment of face templates is done using visible light (VIS) face images. Such images are normally acquired in controlled environment where the lighting is approximately frontal. However, authentication is done in variable lighting conditions. Matching of faces in VIS images taken in different lighting conditions is still a big challenge. The development in near infrared (NIR) image based face recognition has well overcome the difficulty arising from lighting changes [10].

1.1 Purpose of Thesis

Face recognition has received significant attention during the past years (Samal & Iyengar, 1992; Valentin et al., 1994; Chellappa et al., 1995; Zhao & Chellappa, 2002, as cited in [1]), partly due to recent technology advances initially made by work on eigenfaces (Sirovich &. Kirby, 1987; Turk & Pentland, 1991, as cited in [1]). However, the problem of face recognition remains a great challenge after several decades of research [1].

It is difficult for conventional methods to achieve high accuracy, even in cooperative-user conditions, such as access control, machine readable travelling documents, computer login and ATM. Whereas the shape and reflectance are the intrinsic properties, the appearance of a face is affected by extrinsic factors, including illumination, pose and expression [3].

To achieve reliable results, face recognition should be performed based on intrinsic factors of the face only, mainly related to 3D shape and reflectance of the facial surface. Among several extrinsic factors, problems with uncontrolled environmental lighting is the topmost issue to solve for reliable face-based biometric applications in practice. From the end-user point of view, a biometric system should adapt to the environment, not vice versa. However, most current face recognition systems, academic and commercial, are based on face images captured in the visible light (VL) spectrum; they are compromised in accuracy by changes in environmental

illumination, even for cooperative user applications indoors. Several conclusions are made [1]:

1. Lighting conditions, and especially light angle, drastically change the appearance of a face.
2. When comparing unprocessed images, the changes between the images of a person under different illumination conditions are larger than those between the images of two people under the same illumination.
3. All of the local filters under study are insufficient by themselves to overcome variations due to changes in illumination direction.

The influence of illumination is also shown in the recent Face Recognition Vendor Test ("Face Recognition Vendor Tests (FRVT)," Nat'l Inst. Of Standards and Technology, http://www.frvt.org, 2006 as cited in [1]).

For systems that have to work in the daytime and at night, infrared is a solution [11]. An illumination invariant face representation on the basis of active NIR images is derived in this study. It is shown that the active NIR imaging is mainly subject to an approximately monotonic transform in the gray tone due to variation in the distance between the face and the NIR lights and camera lens. Noting that the ordering relationship between pixels is not changed by any monotonic transform, local binary pattern (LBP) features (Ojala et al., 1996; Hadid et al. 2004, as cited in [1]) are used to compensate for the monotonic transform in the NIR images [1, 6].

Another contribution is methods for building a highly accurate face recognition engine using LBP features from NIR images. While there are a large number of LBP features, not all are useful or equally useful for face recognition. Another method that uses LDA, building a discriminative classifier from LBP features, is also presented. Both perform better than the state-of-the-art LBP method (Ojala et al., 1996; Hadid et al. 2004, as cited in [1]) [1, 6].

Whereas VL images of the same face under different lighting directions are negatively correlated, NIR imaging produces closely correlated images of faces of the same individual. However, even with the good basis offered by the NIR imaging system, a straightforward matching engine, such as correlation or PCA-based or LDA-based, is insufficient to achieve high accuracy; more advanced techniques are required to deal with the complexity in the pattern recognition problems [1].

1.2 Literature Review

Much effort has been made to model illumination on faces and correct illumination directions in order to achieve illumination invariant face recognition. Georghiades et al. (2001, as cited in [1]) proved that face images with the same pose under different illumination conditions form a convex cone, called the illumination cone. Ramamoorthi (2002, as cited in [1]) and Basri and Jacobs (2003, as cited in [1]) independently used the spherical harmonic representation to explain the low dimensionality of face images under different illumination. Nayar and Bolle (1996, as cited in [1]) and Jacobs et al. (1998, as cited in [1]) proposed algorithms for face image intrinsic property extraction by Lambertian model without shadow. Shashua and Raviv (2001, as cited in [1]) proposed a simple yet practical algorithm, called the quotient image, for extracting illumination invariant representation. Gross and Brajovic (2003, as cited in [1]) and Wang et al. (2004, as cited in [1]) developed reflectance estimation methods by using the idea of the ratio of the original image and its smooth version from Retinex (Land, 1986, as cited in [1]) and center-surround filters (Jobson et al, 1997, as cited in [1]). These works are shown to improve recognition performance, but have not led to a face recognition method that is illumination invariant [1].

Other directions have also been explored to overcome problems caused by illumination changes. Because 3D (in many case, 2.5D) data captures geometric shapes of face, such systems are less affected by environmental lighting and can cope with rotated faces because of the availability of 3D (2.5D) information for visible surfaces. The disadvantages are the increased cost and slowed speed as well as specular reflections and recognition performances achieved by using a single 2D image and by a single 3D image are similar (Chang et al., 2005, as cited in [1]). A commercial development is A4Vision ("A4Vision Technology," A4Vision, http://www.a4vision.com/, 2006, as cited in [1]). It is basically a 3D (or 2.5D) face recognition system, but it creates 3D mesh of the face by means of triangulation based on an NIR light pattern projected onto the face. While not affected by lighting conditions, background colors, facial hair, or make-up, it has problems in working under conditions when the user is wearing glasses or opening the mouth, due to limitations of its 3D reconstruction algorithm [1].

Imaging and vision beyond the visible spectrum has received much attention in the computer vision community [1]. The infrared spectrum is divided into four bandwidths (Figure 2.2): Near-IR (NIR), Short-wave-IR (SWIR), Medium-wave-IR (MWIR) and Long-wave IR (Thermal IR) [9]. Thermal or far infrared imagery has been used for face recognition (Kong et al., 2005, as cited in [1]). This class of techniques is advantageous for identifying faces under uncontrolled illumination or for detecting disguised faces [1]. Thermal IR sensors measure the emitted heat energy from the object and they don't measure the reflected energy [9]. Their disadvantages include instability due to environmental temperature, emotional and health conditions, and poor eye localization accuracy (Chen et al., 2005; Selinger & Socolinsky, 2004, as cited in [1]). A large-scale study (Chen et al., 2005, as cited in [1]) showed that they did not perform as well as visible light image-based systems, in a scenario involving time lapse between gallery and probe and with relatively controlled lighting [1]. Some researchers have suggested fusing the information from thermal and visual to improve the face recognition rates and solve the opaqueness problem (Kong et al., 2005, as cited in [9]). A simple example for data fusion is a weighted sum of pixel intensity values obtained from each sensor data. Some other researchers have proposed to get the thermal face images and extract the vascular networks (Pavlidis et al., 2005, 2007, as cited in [9]) while the others have proposed the use of blood perfusion data (Song et al., 2005, as cited in [9]) in order to overcome the ambient temperature dependency problem in thermal face recognition [9].

The use of near infrared (NIR) imaging brings a new dimension for face detection and recognition (Dowdall et al., 2003; Li & Liao, 2003; Pan et al., 2003, as cited in [1]). Wilder et al. (1996, as cited in [9]) presented one of the first works on infrared face recognition. They compared the relative performances of three face recognition algorithms for visible and IR images. They also computed the performance of their system when the visible and infrared information are fused. Cutler (1996, as cited in [9]) was one of the first researchers who used the eigenface (Turk & Pentland, 1991, as cited in [9]) method in the infrared face recognition. Three views points were used in his study (frontal, 45 degree and profile) and for each view the subject had two expressions (normal and smile). Yoshitomi et al. (1997, as cited in[9]) focused on the lighting problem in face recognition and suggested infrared face recognition as a solution [9]. Dowdall et al. (2003, as cited in [1]) presented an NIR-based face

detection method; faces are detected by analyzing horizontal projections of the face area and by using the fact that eyes and eyebrows regions have different responses in the lower and upper bands of NIR. Li and Liao (2003, as cited in [1]) presented a homomorphic-filtering preprocessing to reduce inhomogeneous NIR lighting and a facial feature detection method by analyzing the horizontal and vertical projections of the face area. Pan et al. (2003, as cited in [1]) presented an NIR-based face recognition method in which hyperspectral images are captured in 31 bands over a wavelength range of 0:7m-1:0m; multiband spectral measurements of facial skin sampled at some facial points are used for face recognition; they are shown to differ significantly from person to person. Zou et al.'s work (2003, as cited in [1]) derives their matching methods based on an LDA transform and shows that the NIR illuminated faces are better separable than faces under varying ambient illumination [1]. A new approach to cope the illumination dependency problem in face recognition was proposed in (Zou et al., 2005, as cited in [1, 9]). The difference was taken between two face images captured when the LED light is on and off. The difference image was the image of a face under just the LED illumination, and was independent of the ambient illumination [9]. Zhao and Grigat's system [11] uses DCT coefficients as features and an SVM as the classifier. In [10], face recognition is done by matching an NIR probe face against a VIS gallery face. A mechanism of correlation between NIR and VIS faces is learned from NIR→VIS face pairs, and the learned correlation is used to evaluate similarity between an NIR face and a VIS face.

Image moments are regarded as well established shape descriptors and have been frequently used for content based retrieval and pattern recognition tasks. One of the most inspiring early studies illustrating the potential of image moments is the study of Hu (1962, as cited in [8]), in which the moment invariants that enable successful recognition against scaling, translation, and rotation are presented.

In the works of Ono (2003, as cited in [8]) and Singh et al. (2011 as cited in [8]), the authors proposed methods to extract features from face images through ZMs (ZMs) and discussed the rotation invariancy of ZMs. Haddadnia et al. (2003, as cited in [8]) used Pseudo ZM invariants to extract feature vectors from faces. Another ZM based approach that considers spatial locality by partitioning input images is presented in the work of Kanan et al. (2007, as cited in [8]). This approach involves dividing the

face images into non-overlapping subregions and then extracting features from each of these subregions using pseudo ZMs.

2. NIR IMAGING

2.1 Infrared (IR)

Infrared (IR) light is electromagnetic radiation with a wavelength longer than that of visible light, measured from the nominal edge of visible redlight at 0.74 micrometers (μm), and extending conventionally to 300 μm (Figure 2.1). These wavelengths correspond to a frequency range of approximately 1 to 400 THz, and include most of the thermal radiation emitted by objects near room temperature. Microscopically, IR light is typically emitted or absorbed by molecules when they change their rotational vibrational movements.

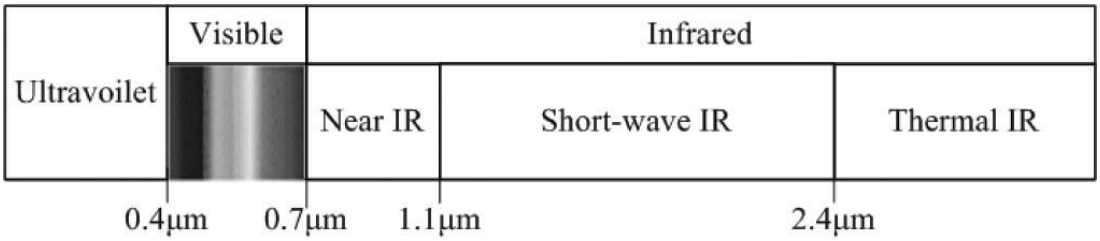


Figure 2.1 : Radiation spectrum ranges [1].

Infrared light is used in industrial, scientific, and medical applications. Night-vision devices using infrared illumination allow people or animals to be observed without the observer being detected. In astronomy, imaging at infrared wavelengths allows observation of objects obscured by interstellar dust. Infrared imaging cameras are used to detect heat loss in insulated systems, observe changing blood flow in the skin, and overheating of electrical apparatus.

Objects generally emit infrared radiation across a spectrum of wavelengths, but sometimes only a limited region of the spectrum is of interest because sensors usually collect radiation only within a specific bandwidth. Therefore, the infrared band is often subdivided into smaller sections, as can be seen in Figure 2.2 [16].

Near-infrared (NIR, IR-A DIN): (Wavelength: 0.75-1.4 μm) Defined by the water absorption, and commonly used in fiber optic telecommunication because of low attenuation losses in the SiO2 glass (silica) medium. Image intensifiers are sensitive

to this area of the spectrum. Examples include night vision devices such as night vision goggles.

Short-wavelength infrared (SWIR, IR-B DIN): (Wavelength: 1.4-3 μm) Water absorption increases significantly at 1,450 nm. The 1,530 to 1,560 nm range is the dominant spectral region for long-distance telecommunications.

Mid-wavelength infrared (MWIR, IR-C DIN, Also called intermediate infrared (IIR)): (Wavelength: 3-8 μm) In guided missile technology the 3-5 μm portion of this band is the atmospheric window in which the homing heads of passive IR 'heat seeking' missiles are designed to work, homing on to the infrared signature of the target aircraft, typically the jet engine exhaust plume.

Long-wavelength infrared (LWIR, IR-C DIN): (Wavelength: 8–15 μm) This is the "thermal imaging" region, in which sensors can obtain a completely passive picture of the outside world based on thermal emissions only and requiring no external light or thermal source such as the sun, moon or infrared illuminator. Forward-looking infrared (FLIR) systems use this area of the spectrum. This region is also called the "thermal infrared".

Far infrared (FIR): (Wavelength: 15 - 1,000 μm) Used in far-infrared lasers. It is one of the possible sources of terahertz radiation. FIR lasers have application in terahertz time-domain spectroscopy, terahertz imaging as well in fusion plasma physics diagnostics. They can be used to detect explosives and chemical warfare agents, by the means of infrared spectroscopy or to evaluate the plasma densities by the means of interferometry techniques.

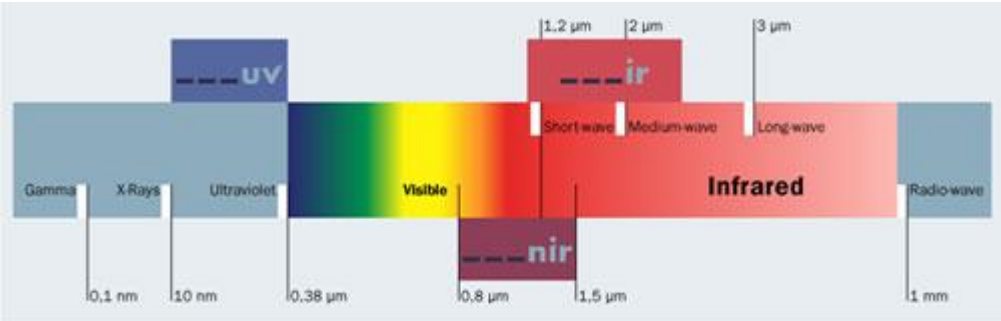


Figure 2.2 : Wavelength spectrum [17].

NIR and SWIR is sometimes called "reflected infrared" while MWIR and LWIR is sometimes referred to as "thermal infrared". Due to the nature of the blackbody

radiation curves, typical 'hot' objects, such as exhaust pipes, often appear brighter in the MW compared to the same object viewed in the LW [16].

2.2 Active NIR versus Visible Light Images

Fig. 2.3 shows example images of a face illuminated by NIR LED lights from the front, a lamp aside and environmental lights. We can see the following:

1. The lighting conditions are likely to cause problems for face recognition with the color images.
2. The NIR images, with the visible light composition cut off by the filter, are mostly frontal-lighted by the NIR lights, with minimum influence from the side lighting, and provide a good basis for face recognition [1].

Figure 2.3 : Color images (top) captured by a color camera versus NIR images (bottom) captured by an NIR imaging system [2, 7].

It is shown in [1] that the scores for intrapersonal pairs under different lighting directions are generally lower than those for extrapersonal pairs under similar lighting directions. This means that reliable recognition can not be achieved with visible light images, even using the advanced matching engine.

The impact of environmental lighting is much reduced by the present NIR imaging system. The correlation coefficients between NIR images are all positive, regardless of the visible light conditions and person identity. However, the intrapersonal correlation coefficients may not necessarily be higher than the extrapersonal ones, meaning possible recognition errors, even with the NIR images (Figure 2.4). Therefore, a better matching engine than correlation or PCA is still needed for highly accurate face recognition, even with NIR face images [1].

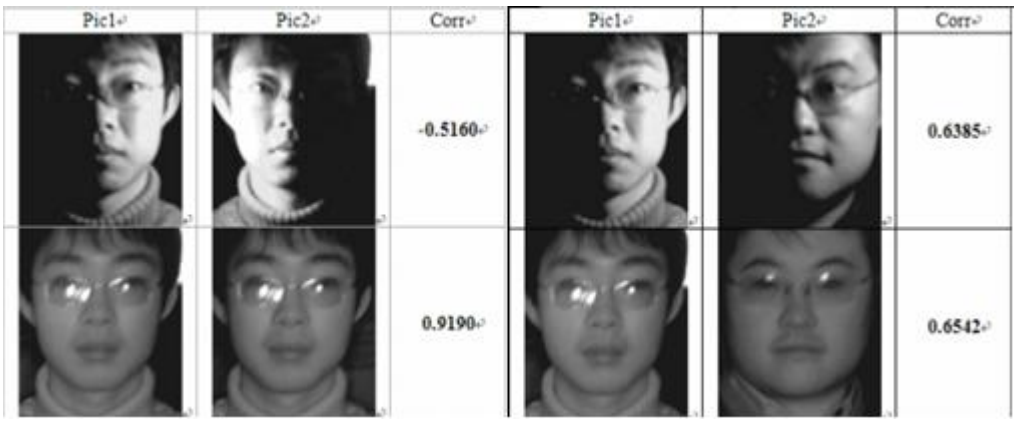


Figure 2.4 : Correlation coefficients [2].

2.3 Modeling of Active NIR Images

According to the Lambertian model, an image $I(x, y)$ under a point light source is formed according to the following:

$$\mathrm{I(x, y)} = \rho(x, y) n(x, y) s, \qquad \textbf{(2.1)}$$

where $\rho(x, y)$ is the albedo of the facial surface material at point (x, y), $n = (n_x, n_y, n_z)$ is the surface normal (a unit row vector) in the 3D space, and $s = (s_x, s_y, s_z)$ is the lighting direction (a column vector, with magnitude). Here, albedo $\rho(x, y)$ reflects the photometric properties of facial skin and hairs; $n(x, y)$ is the geometric shape of the face (Figure 2.5).

Figure 2.5 : Physical imaging model [2].

Assume $s = \kappa s^0$, where κ is a multiplying constant due to possible changes in the strength of the lighting caused by changes in the distance between the face and the LED lights (a less restrictive modeling of constant κ would be a monotonic transform instead of a constant [1]; however, such variations can be easily coped with by using the LBP representation, as will be detailed later [3]).

$s^0 = (s_x^0, s_y^0, s_z^0)$ is a unit column vector of the lighting direction. Let $\theta(x, y)$ be the incident angle between the lighting and the face surface normal at point (x, y), $\cos\theta(x, y) = n(x, y)s^0$. Equation (2.1) can be expressed as

$$\mathrm{I(x, y)} \propto \kappa\rho(x, y)\cos\theta(x, y) \qquad \textbf{(2.2)}$$

When the active NIR lighting is from the (nearly) frontal direction, i.e., $s^0 = (0,0,1)$, the image can be approximated by

$$\mathrm{I(x, y)} = \kappa\rho(x, y)n_z(x, y), \qquad \textbf{(2.3)}$$

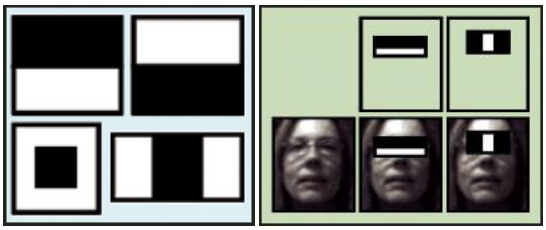
An active NIR image $\mathrm{I(x, y)}$ combines information about both surface normal component $n_z(x, y)$ and albedo map $\rho(x, y)$ and, therefore, provides the wanted intrinsic property about a face for face recognition [1].

This much simplifies the subsequent processing tasks such as face detection, facial feature detection and thereby face recognition [3].

3. PREPROCESSING BEFORE FACE RECOGNITION

3.1 Face and Eye Detection using OpenCV

OpenCV's face detector uses a method that Paul Viola and Michael Jones published in 2001. The features that Viola and Jones used are based on Haar wavelets. Haar wavelets are single wavelength square waves (one high interval and one low interval). In two dimensions, a square wave is a pair of adjacent rectangles - one light and one dark. The actual rectangle combinations used for visual object detection are not true Haar wavelets. Instead, they contain rectangle combinations better suited to visual recognition tasks. Because of that difference, these features are called Haar features, or Haar-like features, rather than Haar wavelets. Figure 3.1 shows the features that OpenCV uses [12].



(a) (b)

Figure 3.1 : (a) Examples of the Haar features used in OpenCV [12]. (b) The first two Haar features in the original Viola-Jones cascade [12].

First, a classifier (namely a cascade of boosted classifiers working with haar-like features) is trained with a few hundreds of sample views of a particular object (i.e., a face or a car), called positive examples, that are scaled to the same size (20x20, etc.), and negative examples - arbitrary images of the same size.

After a classifier is trained, it can be applied to a region of interest (of the same size as used during the training) in an input image. The classifier outputs a "1" if the region is likely to show the object (i.e., face/car), and "0" otherwise. To search for the object in the whole image, the search window is moved across the image and check every location using the classifier. The classifier is designed so that it can be

easily "resized" in order to be able to find the objects of interest at different sizes .To find an object of an unknown size in the image, the scan procedure should be done several times at different scales.

The word "cascade" in the classifier name means that the resultant classifier consists of several simpler classifiers (stages) that are applied subsequently to a region of interest until at some stage the candidate is rejected or all the stages are passed. The word "boosted" means that the classifiers at every stage of the cascade are complex themselves and they are built out of basic classifiers using one of four different boosting techniques (weighted voting). Currently Discrete Adaboost, Real Adaboost, Gentle Adaboost and Logitboost are supported. The basic classifiers are decision-tree classifiers with at least 2 leaves. Haar-like features are the input to the basic classifiers [13, 14].

In this study, face and eye detection are performed in case that there may be other images that will be input to the program but are not included in the NIR face image database. For this, OpenCV's built-in functions are used to implement the detection step before all the input images are sent to the alignment process. For the cascade file, "haarcascade_frontalface_alt" file is used to detect faces. After selecting the detected face as the region of interest, "haarcascade_eye" file is used to detect eyes. "haarcascade_eye_tree_eyeglasses" file was also tested since the face images included faces with eye-glasses, but after a number of trials on the NIR image database, it is found that "haarcascade_eye" performed a better eye detection. If a face and eyes (more than one) are detected, then the file is accepted as an input to the system. Now, the face image can be sent to face alignment transform function.

3.2 Face Alignment Transform

The main idea for aligning face is to localize the facial features in the detected face image. Given the position of two pupils, eye alignment is an affine transformation of the original detected face image (Figure 3.2). Suppose that we have acquired two pupils with coordinates (x_1, y_1), and (x_2, y_2), the expected distance between two eyes on aligned face is d, the rotation center's position is calculated as $(center_x, center_y) = (x_1 - x_2, y_{1-} y_2)$ and the rotation angle

$\theta = \arctan((y_2 - y_1)/(x_2 - x_1))$. Then the mapping matrices A_m and B_m for rotation can be created as follows [4].

$$A_m = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \tag{3.1}$$

$$B_m = \begin{bmatrix} (1-\alpha)\cdot center_x - \beta \cdot center_y \\ (1-\alpha)\cdot center_y - \beta \cdot center_x \end{bmatrix} \tag{3.2}$$

where $\alpha = d \cdot \cos\theta$ and $\beta = d \cdot \sin\theta$.

The aligned location (x', y') for the original point in (x, y) can be obtained by [4]:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A_m \cdot \begin{bmatrix} x \\ y \end{bmatrix} + B_m \tag{3.3}$$

(a) (b)

Figure 3.2 : (a) shows the found eye; image (b) is the aligned face acquired by eyes' location information [4].

3.3 Face Cropping

After aligning the faces, they should be cropped so that when images are put on top of each other, the eyes should overlap. This is also important for face recognition algorithms to produce better results.

After the face&eye detection and alignment, the images should be resized to a ratio so that the distances between the eyes are the same for each image. After resizing, since we have the eye locations from previous steps performed, we cut the image from the right and left sides, leaving the same distance between the eye and the border. The image is cropped from up and down sides, leaving upper border and lower border distances (lower distance should be nearly twice the size of upper

distance) same for each image. To find out the best distance lengths for between the eyes and for between upper/lower bounds and the eye line, a number of ratio values are tested so that the output images are of the same size (142x120 for this study). Some example output images are given in Figure 3.3.

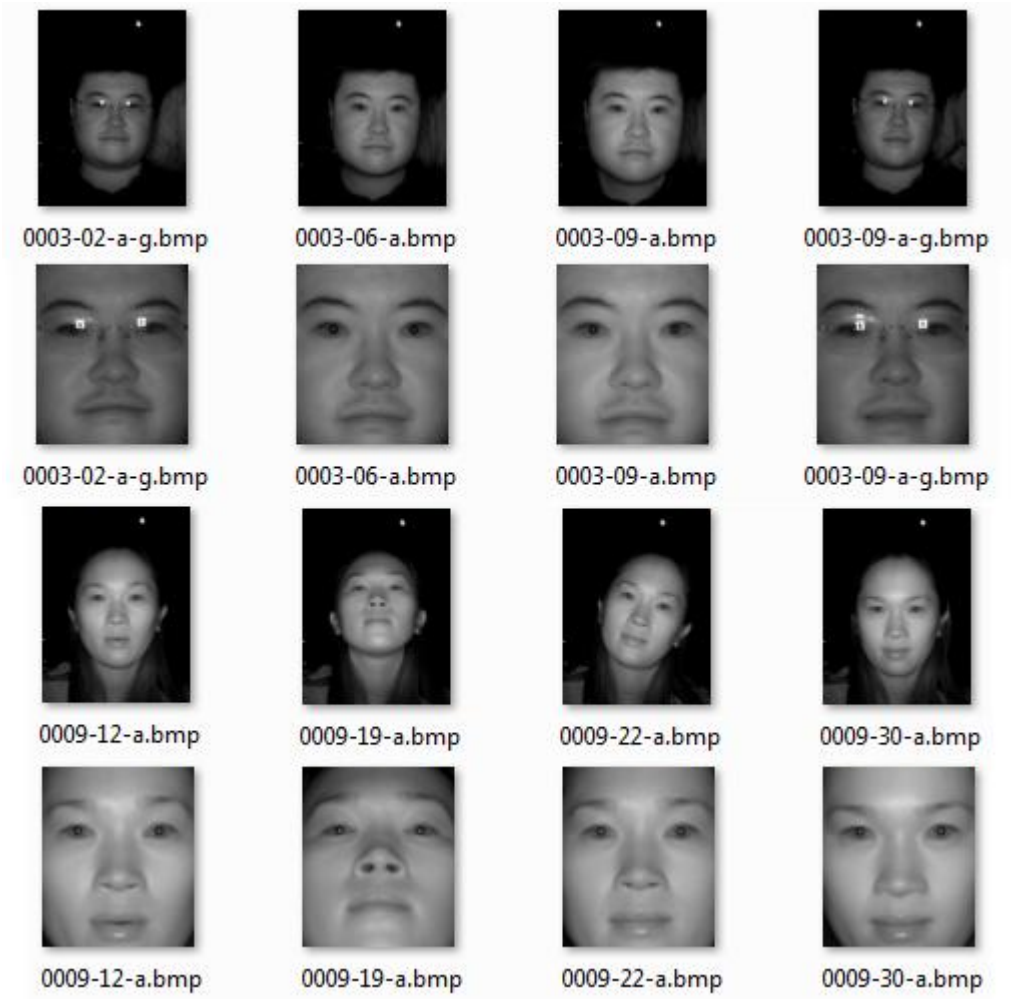


Figure 3.3 : Aligned and cropped NIR face images.

4. FACE RECOGNITION METHODS

4.1 Eigenfaces

As correlation methods are computationally expensive and require great amounts of storage, it is natural to pursue dimensionality reduction schemes.

PCA techniques, also known as Karhunen-Loeve methods, choose a dimensionality reducing linear projection that maximizes the scatter of all projected samples. More formally, let us consider a set of N sample images $\{x_1, x_2, ..., x_N\}$ taking values in an n-dimensional image space, and assume that each image belongs to one of c classes $\{X_1, X_2, ..., X_c\}$. Let us also consider a linear transformation mapping the original n-dimensional image space into an m-dimensional feature space, where $m < n$. The new feature vectors $y_k \in \Re^m$ are defined by the following linear transformation:

$$y_k = W^T x_k \qquad k = 1,2,...,N \tag{4.1}$$

where $W \in \Re^{n \times m}$ is a matrix with orthonormal columns.

If the total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{4.2}$$

where N is the number of sample images, and $\mu \in \Re^n$ is the mean image of all samples, then after applying the linear transformation W^T, the scatter of the transformed feature vectors $\{y_1, y_2, ..., y_N\}$ is $W^T S_T W$. In PCA, the projection W_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_W \left| W^T S_T W \right| \tag{4.3}$$

$$= [w_1 w_2 ... w_m] \quad \textbf{(4.4)}$$

where $\{w_i | i = 1,2,...,m\}$ is the set of n-dimensional eigenvectors of S_T corresponding to the m largest eigenvalues. Since these eigenvectors have the same dimension as the original images, they are referred to as Eigenpictures in (Sirovitch & Kirby, 1987, as cited in [5]) and Eigenfaces in (Turk & Pentland, 1991a; Turk & Pentland, 1991b, as cited in [5]).

For the comparison process, Mahalanobis distance is used. The formula is given below.

$$d(x, y) = -\sum_{i=1}^{k} \frac{1}{\sqrt{\lambda_i}} x_i y_i \quad \textbf{(4.5)}$$

Where λ_i is the i th Eigenvalue corresponding to the i th eigenvector.

A drawback of this approach is that the scatter being maximized is due not only to the between-class scatter that is useful for classification, but also to the within-class scatter that, for classification purposes, is unwanted information [5].

4.2 Fisherfaces

Fisher's Linear Discriminant (FLD) (Fisher, 1936, as cited in [5]) is an example of a class specific method, in the sense that it tries to "shape" the scatter in order to make it more reliable for classification. This method selects W in (Chellappa et al., 1995, as cited in [5]) in such a way that the ratio of the between-class scatter and the within-class scatter is maximized.

Let the between-class scatter matrix be defined as

$$S_B = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T \quad \textbf{(4.6)}$$

and the within-class scatter matrix be defined as

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \quad \textbf{(4.7)}$$

where μ_i is the mean image of class X_i, and N_i is the number of samples in class X_i. If S_w is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples, i.e.,

$$W_{opt} = \arg\max_{W} \frac{\left|W^T S_B W\right|}{\left|W^T S_W W\right|}$$

$$= \left[w_1 w_2 ... w_m\right] \qquad \textbf{(4.8)}$$

where $\{w_i | i = 1,2,...,m\}$ is the set of generalized eigen-vectors of S_B and S_W corresponding to the m largest generalized eigenvalues $\{\mu_i | i = 1,2,...,m\}$, i.e.,

$$S_B w_i = \lambda_i S_W w_i, \quad i = 1,2,...,m \qquad \textbf{(4.9)}$$

Note that there are at most $c-1$ nonzero generalized eigenvalues, and so an upper bound on m is $c-1$, where c is the number of classes.

In the face recognition problem, one is confronted with the difficulty that the within-class scatter matrix $S_W \in \Re^{n \times n}$ is always singular. This stems from the fact that the rank of S_W is at most $N-c$, and, in general, the number of images in the learning set N is much smaller than the number of pixels in each image n. This means that it is possible to choose the matrix W such that the within-class scatter of the projected samples can be made exactly zero.

In order to overcome the complication of a singular S_W, an alternative to the criterion in (4.8) is proposed. This method, which is called Fisherfaces, avoids this problem by projecting the image set to a lower dimensional space so that the resulting within-class scatter matrix S_W is nonsingular. This is achieved by using PCA to reduce the dimension of the feature space to $N-c$, and then applying the standard FLD defined by (4.8) to reduce the dimension to $c-1$.

More formally, W is given by

$$W_{opt}^T = W_{fld}^T W_{pca}^T \tag{4.10}$$

where

$$W_{pca} = \arg\max_W \left| W^T S_T W \right| \tag{4.11}$$

$$W_{fld} = \arg\max_W \frac{\left| W^T W_{pca}^T S_B W_{pca} W \right|}{\left| W^T W_{pca}^T S_W W_{pca} W \right|} \tag{4.12}$$

Note that the optimization for W_{pca} is performed over $n \times (N-c)$ matrices with orthonormal columns, while the optimization for W_{fld} is performed over $(N-c)xm$ matrices with orthonormal columns. In computing W_{pca}, we have thrown away only the smallest $c-1$ principal components [5].

For comparison, cosine distance is used. The formula is given below.

$$\cos(W_1, W_2) = \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} {p_i}^2 \sum_{i=1}^{n} {q_i}^2}} \tag{4.13}$$

4.3 Local Binary Patterns (LBP)

The original LBP operator, introduced by Ojala et al. (1996, as cited in [6]), is a powerful means of texture description. The operator labels the pixels of an image by thresholding the 3x3-neighbourhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. See Figure 4.1 for an illustration of the basic LBP operator.

Later the operator was extended to use neigbourhoods of different sizes (Ojala et al., 2002, as cited in [6]). For neighbourhoods the notation (P,R) which means P sampling points on a circle of radius of R, is used. The pixel values are bilinearly

interpolated whenever the sampling point is not in the center of a pixel. See Figure 4.2 for an example of the circular (8,2) neighbourhood.

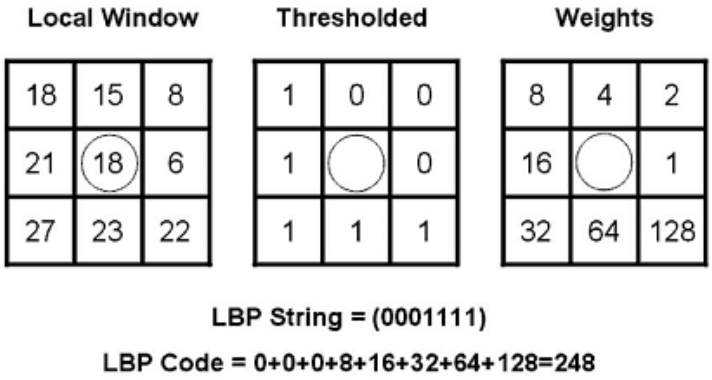


Figure 4.1 : The basic LBP operator [1, 6].

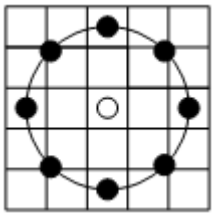


Figure 4.2 : The circular (8,2) neigbourhood [6].

Another extension to the original operator uses so called uniform patterns (Ojala et al., 2002, as cited in [6]). A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 00011110 and 10000011 are uniform patterns. Ojala et al. noticed that in their experiments with texture images, uniform patterns account for a bit less than 90 % of all patterns when using the (8,1) neighbourhood and for around 70 % in the (16,2) neighbourhood.

The following notation for the LBP operator: $LBP_{P,R}^{u2}$ is used. The subscript represents using the operator in a (P,R) neighbourhood. Superscript $u2$ stands for using only uniform patterns and labelling all remaining patterns with a single label.

This histogram contains information about the distribution of the local micropatterns, such as edges, spots and flat areas, over the whole image. For efficient face representation, one should retain also spatial information. For this purpose, the image is divided into regions $R_0, R_1, ..., R_{m-1}$ (Figure 2.7) [6]. The length of the feature vector becomes $B = mB_r$, in which m is the number of regions and B_r is the LBP

histogram length. A large number of small regions produces long feature vectors causing high memory consumption and slow classification, whereas using large regions causes more spatial information to be lost. See Figure 4.3 (a) for an example of a preprocessed facial image divided into 49 windows [6], the number of windows same as used in this study.

Figure 4.3 : (a) An example of a facial image divided into 7x7 windows. (b) The weights set for weighted $\chi 2$ dissimilarity measure. Black squares indicate weight 0.0, dark grey 1.0, light grey 2.0 and white 4.0. [6].

From the pattern classification point of view, a usual problem in face recognition is having a plethora of classes and only a few, possibly only one, training sample(s) per class. For dissimilarity measurement, Chi square statistic (4.14) is used in this study.

$$\chi^2(S,M) = \sum_i \frac{(S_i - M_i)^2}{S_i + M_i} \qquad \textbf{(4.14)}$$

When the image has been divided into regions, it can be expected that some of the regions contain more useful information than others in terms of distinguishing between people. For example, eyes seem to be an important cue in human face recognition (Zhao et al., 2002; Gong et al., 2000, as cited in [6]). To take advantage of this, a weight can be set for each region based on the importance of the information it contains (Figure 2.7 (b)) [6].

The pixel intensities in an NIR image are subject to a multiplying constant κ due to changes in the distance between the face and the LED lights. This degree of freedom can be fixed by using an LBP-based representation. Let us relax the effect of the multiplying constant to a monotonic transform, τ. Then, active NIR images can be modeled by

$$I(x,y) = \tau(p(x,y)n_z(x,y)). \quad \textbf{(4.15)}$$

Let us be given an image $I'(x,y) = p(x,y)n_z(x,y)$ and a transformed image $I''(x,y) = \tau(I'(x,y)) = \tau(p(x,y)n_z(x,y))$. The ordering relationship between pixels in an image is not changed by anymonotonic transform, namely, if $I'(x_1,y_1) > I'(x_2,y_2)$, then $I''(x_1,y_1) > I''(x_2,y_2)$. Therefore, the LBP codes generated from I'' are exactly the same as the ones generated from I'.

From the analysis, we see that the NIR imaging and LBP features together lead to an illumination invariant representation of faces for indoor face recognition applications. In other words, applying the LBP operator to an active NIR image generates illumination invariant features for faces. The illumination invariant face representation provides great convenience for face recognition [1].

4.4 LBP + LDA

The LBP features are derived from LBP histogram statistics as follows [1, 7]:

1. Computing Base LBP Features
    - Computing $LBP_{P,R}^{u2}$ codes for every pixel location in the image.
2. LBP Code Histogramming
    - A histogram of the base LBP codes is computed over a local region centered at each pixel, each histogram bin being the number of occurrences of the corresponding LBP code in the local region. There are 59 bins for $LBP_{8,1}^{u2}$.
    - An LBP histogram is considered as a set of 59 individual features.
3. Gathering LBP Histograms
    - For a face image of size $W \times H$, with the interior area of size $W' \times H'$, the total number of LBP histogram features is $D = W' \times H' \times 59$ (number of valid pixel locations times the number of LBP histogram bins).

In this study, $W \times H = 120 \times 142$ and a local region for histogramming is a rectangle of size 16×20, the interior area is of size $W' \times H' = 104 \times 122$ pixels. Then there are a total of $104 \times 122 \times 59 = 748{,}592$ elements in the LBP histogram feature pool [7].

For memory reasons, the 748,592-dimensional LBP histogram features are downsampled to 10,000 by uniformly sampling. The 10,000-dimensional data is preprocessed using the PCA transform to make the within-class scatter matrix nonsingular. The LDA projection matrix P, composed of so-called Fisherfaces (Belhumeur et al., 1997, as cited in [1]), are then computed.

Given two input vectors x_1 and x_2, their LDA projections are calculated as $v_1 = Px_1$ and $v_2 = Px_2$ and the following cosine score (or called "cosine distance" in some of the literature) is used for the matching:

$$H(v_1, v_2) = (v_1 \cdot v_2) / \|v_1\| \ \|v_2\| \qquad \textbf{(4.16)}$$

In the test phase, the projections v_1 and v_2 are computed from two input vectors x_1 and x_2, one for the input face image and one for an enrolled face image. By comparing the score $H(v_1, v_2)$ with a threshold, a decision can be made whether x_1 and x_2 belong to the same person [1, 7].

4.5 Local Zernike Moments

In Local Zernike Moments, Global Zernike Moments are modified to obtain a local representation by computing the moments at every pixel of a face image by considering its local neighborhood, thus decomposing the image into a set of images, moment components, to capture the micro structure around each pixel [8].

4.5.1 Global Zernike moments

ZMs of an image are defined as the projection of the image onto an orthogonal set of polynomials called Zernike polynomials. The Zernike polynomials are defined as

$$V_{nm}(p, \theta) = R_{nm}(p) e^{jm\theta} \qquad \textbf{(4.17)}$$

where n is the order of the polynomial and m is the number of iterations such that $|m| < n$ and $n - |m|$ is even. The radial polynomials R_{nm} are given as

$$R_{nm}(p) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s \, p^{n-2s} (n-s)!}{s!(\frac{n+|m|}{2} - s)!(\frac{n-|m|}{2} - s)!}. \quad \textbf{(4.18)}$$

The ZMs of a digital image $f(i, j)$ are calculated as

$$Z_{nm} = \frac{n+1}{\pi} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i, j) V^*(p_{ij}, \theta_{ij}) \Delta x_i \Delta y_j, \quad \textbf{(4.19)}$$

where x_i and y_i are the image coordinates mapped to the range [−1, 1], $p_{ij} = \sqrt{x_i^2 + y_i^2}$, $\theta_{ij} = \tan^{-1} \frac{y_j}{x_i}$ and $\Delta x_i = \Delta y_j = 2/N\sqrt{2}$. The terms $\frac{n+1}{\pi}$, Δx_i and Δy_j are constant and they will be ignored.

4.5.2 Local moment transformation

The key difference between the approach used in [8] and the global approaches is that moments are calculated around each pixel to obtain a new image representation through a transformation referred to as LZM transformation.

Moment-based operators V_{nm}^k, which are k × k kernels calculated using the equation $V_{nm}^k(i, j) = V_{nm}(p_{ij}, \theta_{ij})$, are derived. These operators are similar to 2D convolution kernels used for image filtering. The LZM transformation can be defined by these kernels as follows

$$Z_{nm}^k = \sum_{p,q=-\frac{k-1}{2}}^{\frac{k-1}{2}} f(i-p, j-q) V_{nm}^k(p, q). \quad \textbf{(4.20)}$$

This transformation provides a rich image representation by successfully exposing the intensity variations around each pixel. The components of the representation (i.e. complex images corresponding to different moment orders) seem to be very robust to illumination variations.

The number of components depends on the moment order n. An additional constraint is imposed on the second parameter m, such as $m \neq 0$ since the imaginary part of the kernels V_{nm}^{k} becomes zero when $m = 0$ and this is not a desirable behavior when the outcome of (4.20) is used to extract the phase-magnitude histograms as explained in Section 4.5.4. The number of effective moment components can be calculated through the following expression:

$$K(n) = \begin{cases} \dfrac{n(n+2)}{4} & \text{if } n \text{ is even} \\ \dfrac{(n+1)^2}{4} & \text{if } n \text{ is odd}. \end{cases} \qquad \textbf{(4.21)}$$

4.5.3 Face description from moment components Z_{nm}

The moment components are divided into subregions after the proposed LZM transformation is applied. A two-step partitioning is performed. First the image is divided into $N \times N$ equal-sized blocks beginning from the top-left of the image. Then, the image is divided into $N-1 \times N-1$ blocks of the same size as the previous ones with a grid shifted half a block size from top left. Hence, we have $N^2 \times (N-1)^2$ subregions for each moment component. In order to increase the robustness against illumination variations, each subregion is z-normalized.

4.5.4 Cascaded LZM transformation (H-LZM^2)

The PMH (phase-magnitude histogram) of a moment component is extracted as follows: The angle interval of $[0, 2\pi]$ is divided into b bins. Then, the magnitude value $\left(\left|Z_{nm}^{k}(i,j)\right|\right)$ at each pixel location (i, j) is added to the bin corresponding to phase value $\left(\angle Z_{nm}^{k}(i,j)\right)$ of the same pixel location.

The PMHs of moment components are extracted at each subregion as illustrated in Figure 4.4. Each local PMH is normalized to have unit norm, and all normalized local PMHs are concatenated to get the final feature vector. The weight map proposed in [6] is used by simply interpolating it to fit to the grid size.

A different, pixel-wise weighting is performed inside each subregion when computing the PMHs. A gaussian weighting kernel of same size as a subregion with

$\sigma = 8$ is used. The magnitude of each pixel is multiplied with the corresponding kernel weight before adding it to the relevant histogram bin [8].

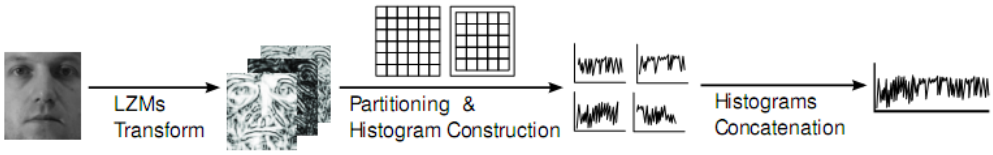


Figure 4.4 : An illustration of the proposed method [8].

The LZM transformation is applied twice: First to stimulate the local shape characteristics; second, to describe the local shape statistics of the transformed images. It is found by experiments in [8] that using the imaginary or real parts directly leads to better results.

The block diagram of the cascaded transform is shown in Figure 4.5 and an example of moment components obtained by the cascaded LZM transformation can be seen in Figure 4.6.

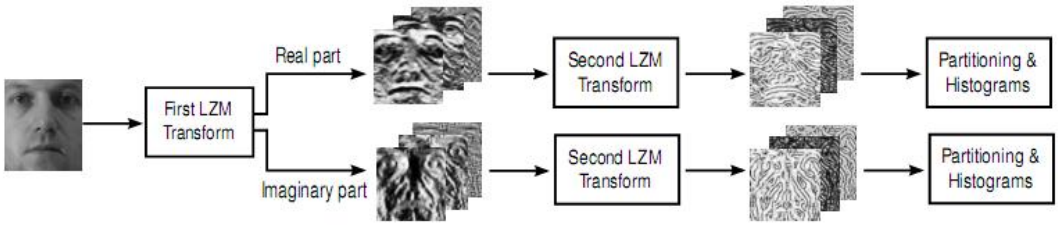


Figure 4.5 : An illustration of the cascaded approach [8].

For the first LZM transformation applied, the parameter values for the size of ZM-based operator k is 5 and n_{max} is 4. For the second application, k is 7 and n_{max} is 4. The value for N is 10, the number of face bins is 24. The final length of the feature vector can be calculated with the equation, $(N^2 + (N-1)^2) x b x K_1 x K_2 x 2$. For comparison, L1 (Manhattan) distance is used.

For comparison, L_1 (Manhattan) distance is used. The formula is given below.

$$MH(a,b) = \sum_{i=1}^{n} |x_i - y_i| \qquad \textbf{(4.22)}$$

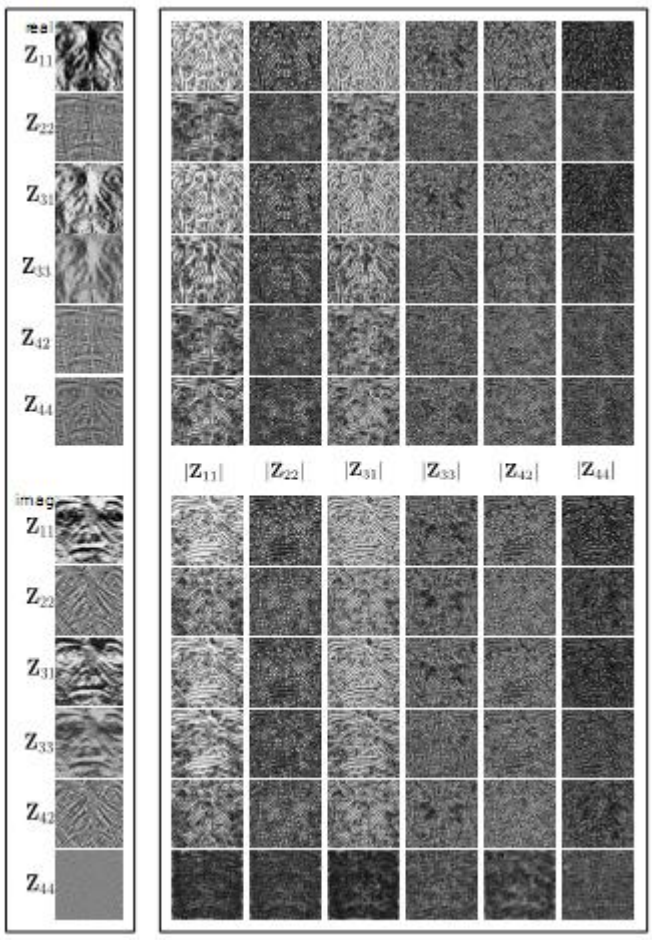


Figure 4.6 : Moment components by repeated LZM transformation. Blocks on the left and right show outputs of the first and second LZM transformations, respectively [8].

4.6 Local Zernike Moments + LDA

It is seen from experiments in [1] that the performance of LBP has increased when LDA is applied afterwards. So, LDA is applied to Zernike phase-magnitude histograms. Same steps as LBP+LDA are followed, the details are explained in Section 4.4. The size of the histograms are downsampled to 10,000 dimensional data. PCA is applied before LDA. Their LDA projections are calculated and cosine distance is used for the matching.

5. EXPERIMENTAL RESULTS

5.1 NIR Image Database

For gallery and probe NIR images, OTCBVS Benchmark Dataset Collection has been used in this project. This is a publicly available benchmark dataset for testing and evaluating novel and state-of-the-art computer vision algorithms. The benchmark contains videos and images recorded in and beyond the visible spectrum and is available for free to all researchers in the international computer vision communities. CBSR NIR Face Dataset of OTCBVS Benchmark Dataset Collection is used (Figure 5.1). Its topic of interest is NIR face detection, NIR eye detection and NIR face recognition subjects.

Figure 5.1 : CBSR NIR Face Dataset [15].

The images were taken by an NIR camera with active NIR lighting [15]. The device consists of 18 NIR LEDs, an NIR camera, a color camera, and the box (Figure 5.2). The NIR LEDs and camera are for NIR face image acquisition. The color camera capture color face images may be used for fusion with the NIR images or for other purposes. The imaging hardware works at a rate of 30 frames per second with the USB 2.0 protocol for 640 x 480 images and costs less than 20 US dollars. More information about the imaging system can be found out in [1].

The database contains 3,940 NIR face images of 197 people. The image size is 480 by 640 pixels, 8 bit, without compression. Images are divided into a gallery set and a probe set. In the gallery set, there are 8 images per person. In the probe set, 12 images per person.

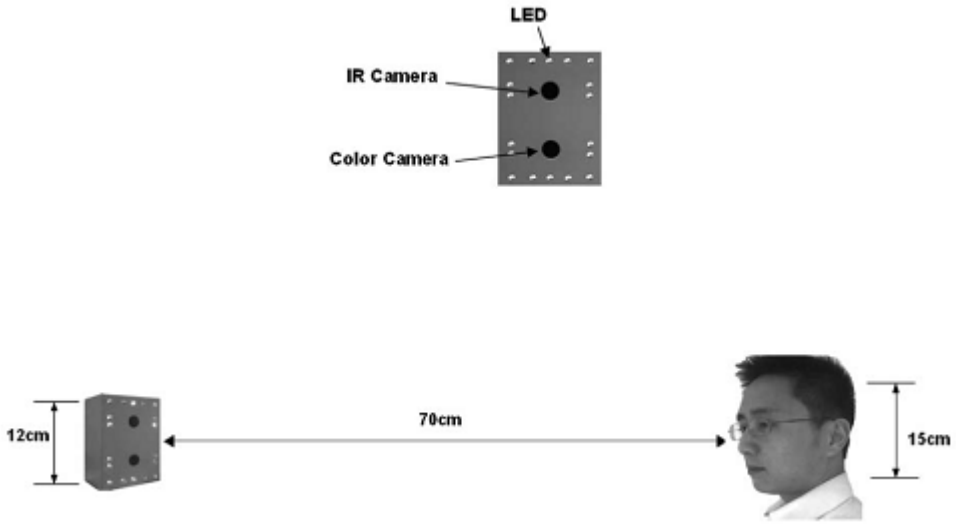


Figure 5.2 : Active NIR imaging system (upper) and its geometric relationship with the face (lower) [1].

The image information which includes the image number, person id and eye coordinates is provided. The image information is provided in gallery-groundtruth.txt and probe-groundtruth.txt, which gives the image number, person number, and eye coordinates. Every line in the ground truth text file consists of the image file name and the eye positions [15]. More specifically, every line in ground truth text is, [person id]-[image id for this person id]-[a][-g].bmp,[left eye x position],[left eye y position],[right eye x position],[right eye y position]. For example:

- "0001-00-a-g.bmp,197,345,325,349" means: Person No.0001, Image No.00, with glasses; the left eye position is (197,345), and right eye (325,349)
- "0007-22-a.bmp,175,182,285,179" means: Person No.0007, Image No.22, without glasses; the left eye position is (175,182), and right eye (285,179).

5.1.1 Results

The software for this study is written in C++. Visual Studio 2008 and OpenCV v2.1, a computer vision library written in C/C++, has been used. Both identification (1:n) and verification (1:1) has been tested for the methods explained in Section 4. In face identification, a system tries to figure who the person is. In face verification, a person claims a particular identity and the system verifies whether that is true. The details for NIR face image database are written in Section 5.1.

Both recognition and verifications tests in this study are performed with the following methods:

1. PCA with Mahalanobis distance
2. LDA with cosine distance
3. LBP with chi-square distance (original uniform (8,1), (8,2) and (16,2))
4. LBP+LDA with cosine distance
5. LZM with Manhattan distance
6. LZM+LDA with cosine distance

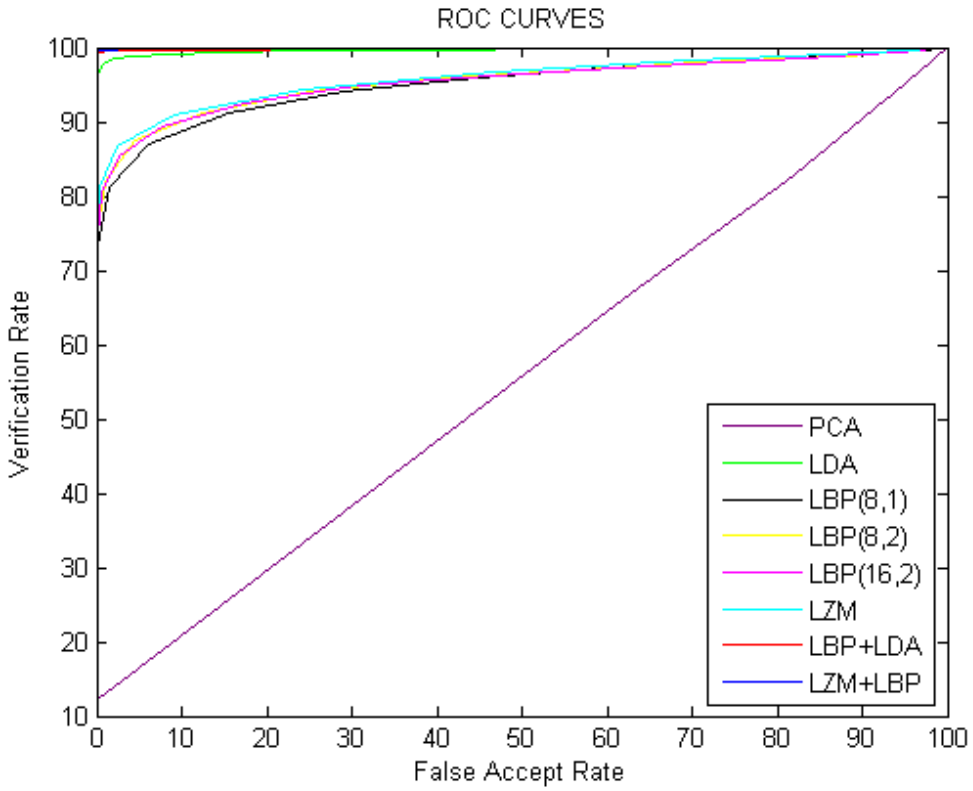
Table 5.1 shows the number of correct and incorrect results of the test performed with the probe images of the NIR face database in the recognition step, where no threshold values are applied so there are no reject values.

Table 5.1 : Comparison table for the recognition results of the methods.

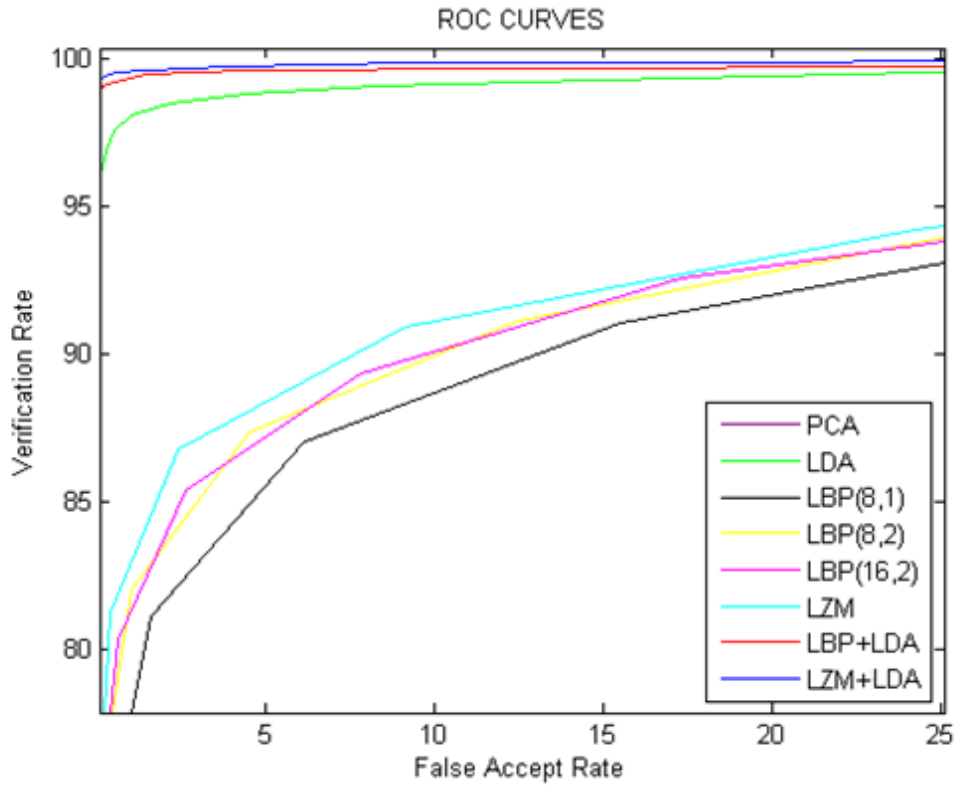
Method	# of Correct Results	# of Incorrect Results	Percentage
PCA	2127	237	% 89,97
LDA	2331	33	% 98,60
$LBP_{8,1}^{u2}$	2327	37	% 98,43
$LBP_{8,2}^{u2}$	2336	28	% 98,82
$LBP_{16,2}^{u2}$	2337	27	% 98,86
LBP+LDA	2341	23	% 99,03
LZM	2339	25	% 98,94
LZM+LDA	2343	21	% 99,11

As can be seen from Table 5.1, LZM performs well than LBP operators. When combined with LDA, the success results of both LBP and LZM increase. The extended uniform LBP operators are more successful than both the original uniform LBP and LDA itself. The original LBP method falls behind LDA when compared. The success of PCA is not enough for a robust face recognition.

For the verification graph in Figure 5.3, 20 distance values are calculated from two specific values; one is the maximum distance between two method data of an individual and the other one is the minimum distance between two method data of two individuals. False acceptance and verification rates are calculated then. A system's FAR typically is stated as the ratio of the number of false acceptances divided by the number of identification attempts. Face verification rate is defined as the rate at which legitimate end-users are correctly verified.



(a)



(b)

Figure 5.3 : (a) ROC curves for various compared methods. (b) Close caption of the graph in (a).

As can be seen in Figure 5.3, LZM outperforms uniform LBP with neighborhood and radius of (8,1), (8,2) and (16,2). When LDA is used together with a method, it carries the success of face recognition to a higher level than when only the method itself is used. The extended uniform LBP operators perform better than the original uniform LBP operator and the success of PCA operation is again not enough to catch up with the other operators mentioned.

6. CONCLUSIONS AND RECOMMENDATIONS

For visual light (VS), there are many cases where the differences between the face images of the same person's are more than the differences between two different people's face images. This means VL images of the same face are negatively correlated. With active NIR imaging, an illumination invariant representation of face images are provided. The correlation between the image faces of an individual increases. However, straightforward methods like PCA or LDA are not sufficient enough for a robust recognition system.

It is shown that the ordering of the neighboring pixels is not changed by any monotonic transform. Depending on this, local image representations are more preferred to be used for the face recognition. Local binary patterns (LBPs) are used for the representation of faces, compensating for the monotonic transform. LBP and LDA are combined and it is obviously seen the success of LBP+LDA is more than the state-of-art LBP methods.

Another method, Local Zernike Moments (LZMs), involving local feature representation is used on NIR images. Considering the neighborhood, the moments are calculated at every pixel of the image and moment images, the same size as the original one, are obtained. Later, the micro-structure information carrying moment images are divided into non-overlapping regions and this time, the phase-magnitude histograms are extracted from the complex output of each moment at each subregion. Finally, the histograms are concatenated and the face representation is obtained. Using cascaded LZM and two over-lapping grids, the face recognition's performance has increased. Moreover, since LBP+LDA together has given more successful outputs, LZM+LDA method is tried. The success of LZM+LDA over LZM is significant.

A NIR face image is a fine input for face recognition because it reduces the heavy preprocessing steps before the recognition phase. With the help of LZM on NIR images, fast and highly accurate face recognition systems can be built. Yet, NIR imaging is not suitable enough for uncooperative face recognition applications. Also,

its design is not suitable enough for outdoor usage where the effect of visual light is much more than near infrared. Future works with the NIR image capturing system will be able to overcome such limitations.

REFERENCES

[1] **Li, S.Z., Chu, R., Liao, S. and Zhang, L.** (2007) Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 29(4): 627-639.

[2] **Li, S.Z.** (2007) Face Biometric: Algorithms, Performance & Applications. [PowerPoint slides]. Retrieved from <www.comp.hkbu.edu.hk/~asi07/test/Face-ASI07.pdf>

[3] **Li, S.Z., Zhang, L.**, **Liao, S., Zhu, X., Chu, R., Ao, M. and He, R.** (2006) A Near-infrared Image Based Face Recognition System. FG 2006: 455-460

[4] **Li, H., Wang, P. and Shen, C.** (2010) A Robust Face Recognition System via Accurate Face Alignment and Sparse Representation. International Conference on Digital Image Computing: Techniques and Applications (DICTA 2010). pp. 262-269

[5] **Belhumeur, P.N., Hespanha, J.P., and Kriegman, D.J.** (1997) Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Trans. Pattern Analysis and Machine Intelligence. vol. 19, no. 7, pp. 711-720.

[6] **Ahonen, T., Hadid, A., and Pietikainen, M.** (2004) Face Recognition with Local Binary Patterns. Proc. European Conf. Computer Vision. pp. 469-481.

[7] **Li, S.Z. and Jain, A.K.** (2011) Handbook of Face Recognition. 2nd Edition. Springer. ISBN 978-0-85729-931-4.

[8] **Sarıyanidi, E., Dağlı, V., Tek, S.C., Tunç, B.¸ and Gökmen, M.** (2012) Local Zernike Moments: A New Representation For Face Recognition. IEEE 20th Conference on Signal Processing and Communications Applications (SIU). Fethiye, Muğla.

[9] **Ghiass, R.S., Bendada, A.H., and Maldague, X.** (2010) Infrared face recognition: a review of the state of the art. QIRT'10. (Quebec City, Canada). pp. 533-540.

[10] **Yi, D., Liu, R., Chu, R., Lei, Z. and Li, S.Z.** (2007) Face Matching Between Near Infrared and Visible Light Images. ICB 2007: 523-530.

[11] **Zhao, S.Y. and Grigat, R.R.** (2005) An Automatic Face Recognition System in the Near Infrared Spectrum. Proc. Int'l Conf. Machine Learning and Data Mining in Pattern Recognition. pp. 437-444.

[12] <*http://www.cognotics.com/opencv/servo_2007_series/part_2/sidebar.html*>, date retrieved 01.05.2012.

[13] <*http://opencv.willowgarage.com/wiki/FaceDetection*>, date retrieved 01.05.2012.

[14] <*http://opencv.willowgarage.com/documentation/object_detection.html*>, date retrieved 01.05.2012.

[15] <http://www.cse.ohio-state.edu/OTCBVS-BENCH/bench.html>, date retrieved 01.05.2012.

[16] <*http://en.wikipedia.org/wiki/Near_Infrared*>, date retrieved 01.05.2012.

[17] <*http://www.adphosna.com/nirtechnology.html*>, date retrieved 01.05.2012.

CURRICULUM VITAE

Name Surname: Nil SERİ

Place and Date of Birth: Istanbul / 29.10.1986

E-Mail: nilseri@gmail.com

M.Sc.: Computer Engineering, Istanbul Technical University, Istanbul

B.Sc.: Computer Engineering, Yıldız Technical University, Istanbul, 2009

High School: Kadıköy Anatolian High School, Istanbul, 2005