







**ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE**

**ROBUST FACE RECOGNITION ON NONLINEAR MANIFOLDS**

**Ph.D. THESIS**

**Birkan TUNÇ**

**Computational Science and Engineering Department**

**Computational Science and Engineering Programme**

**MARCH 2012**



**ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE**

**ROBUST FACE RECOGNITION ON NONLINEAR MANIFOLDS**

**Ph.D. THESIS**

**Birkan TUNÇ  
(702052006)**

**Computational Science and Engineering Department**

**Computational Science and Engineering Programme**

**Thesis Advisor: Prof. Dr. Muhittin GÖKMEN**

**MARCH 2012**



**DOĞRUSAL OLMAYAN MANIFOLDLAR ÜZERİNDE  
GÜRBÜZ YÜZ TANIMA**

**DOKTORA TEZİ**

**Birkan TUNÇ  
(702052006)**

**Hesaplamalı Bilim ve Mühendislik Anabilim Dalı**

**Hesaplamalı Bilim ve Mühendislik Programı**

**Tez Danışmanı: Prof. Dr. Muhittin GÖKMEN**

**MART 2012**









## **FOREWORD**

This work came into existence by a joint effort of my dear advisor, authors of hundreds of previous works, my family, lots of friends, and me. During my PhD study, I was also partially supported by TÜBİTAK-BİDEB and İTÜ-BAP.

As being close to idealism in a philosophical manner, I have been always thinking that the abstraction harms the truth; however, I could not help myself with performing yet another work concerning abstraction.

March 2012

Birkan TUNÇ



## TABLE OF CONTENTS

	<u>Page</u>
<b>FOREWORD</b> .....	<b>vii</b>
<b>TABLE OF CONTENTS</b> .....	<b>ix</b>
<b>ABBREVIATIONS</b> .....	<b>xi</b>
<b>LIST OF TABLES</b> .....	<b>xiii</b>
<b>LIST OF FIGURES</b> .....	<b>xv</b>
<b>SUMMARY</b> .....	<b>xvii</b>
<b>ÖZET</b> .....	<b>xix</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
1.1 The Problem Definition .....	3
1.2 The Classical Approaches And Their Limitations .....	4
1.3 Overview of the CDFA Framework.....	6
1.4 Connections to Previous Works.....	8
1.5 Other Related Works .....	10
<b>2. A GENERIC FRAMEWORK</b> .....	<b>15</b>
2.1 Constructing a Basis Set for a Variation Type.....	15
2.2 Proposed Generic Basis Recovery Scheme .....	16
2.3 Mathematical Background.....	17
2.3.1 Manifold Learning.....	19
2.3.2 Bootstrap: an algebraic approach .....	22
2.3.3 Bootstrap: a probabilistic approach.....	26
2.3.4 Training: recovering class factors.....	29
2.3.5 Testing: classification of novel points .....	31
2.4 Interpretation of Governing Distributions .....	32
<b>3. EXPERIMENTAL EVALUATIONS</b> .....	<b>35</b>
3.1 Tuning the Bootstrap Parameters.....	35
3.2 Classification Performance against Illumination.....	36
3.3 Classification Performance against Facial Expressions.....	38
3.4 Classification Performance against Pose .....	41
3.5 Scalability .....	43
3.6 Real Life Performance.....	47
3.7 Computational Aspects.....	47
3.8 Complexity of the Framework.....	51
3.8.1 Offline complexity .....	51
3.8.2 Online complexity .....	52

<b>4. CONCLUDING REMARKS</b> .....	<b>55</b>
4.1 Discussions on Experimental Results.....	55
4.2 Contributions .....	56
4.3 Future Works .....	57
<b>REFERENCES</b> .....	<b>59</b>
<b>APPENDICES</b> .....	<b>63</b>
APPENDIX A.1 .....	64
APPENDIX A.2 .....	66
<b>CURRICULUM VITAE</b> .....	<b>70</b>

## **ABBREVIATIONS**

<b>FA</b>	: Factor Analysis
<b>PCA</b>	: Principal Component Analysis
<b>LDA</b>	: Linear Discriminant Analysis
<b>SVD</b>	: Singular Value Decomposition
<b>LPP</b>	: Locality Preserving Projections
<b>MAP</b>	: Maximum a Posterior
<b>C DFA</b>	: Class Dependent Factor Analysis



## LIST OF TABLES

	<u>Page</u>
<b>Table 2.1</b> : Summary of the CDFA.....	19
<b>Table 2.2</b> : Detailed algorithm of the CDFA. ....	32
<b>Table 3.1</b> : Face recognition rates for Yale B Database. Performances of the other methods were taken from [16]......	38
<b>Table 3.2</b> : Recognition error rates for Yale B Database with multiple gallery images. ....	39
<b>Table 3.3</b> : Average face recognition rates on JAFFE database. 40 trials with randomly chosen gallery images were performed for each row. ....	40
<b>Table 3.4</b> : Average face recognition rates on CMU AMP database. 10 trials with randomly chosen gallery images were performed for each row. ....	41
<b>Table 3.5</b> : Initial face recognition rates (%) with changing poses. A single image is selected as a gallery image and recognition rates for $\pm 22.5^\circ$ , $\pm 67.5^\circ$ and $\pm 90^\circ$ are given. Tests are performed with 50 identities in the gallery. ....	42
<b>Table 3.6</b> : Initial face recognition rates (%) with changing poses. Multiple images are selected as gallery images and recognition rates for $\pm 22.5^\circ$ , $\pm 67.5^\circ$ and $\pm 90^\circ$ are given. Tests are performed with 50 identities in the gallery. ....	42
<b>Table 3.7</b> : New results with the proposed basis recovery scheme. Multiple images are selected as gallery images and recognition rates for $\pm 22.5^\circ$ , $\pm 67.5^\circ$ and $\pm 90^\circ$ are given. Tests are performed with 50 identities in the gallery. ....	43
<b>Table 4.1</b> : Contributions of the study and related publications.....	57





## LIST OF FIGURES

	<u>Page</u>
<b>Figure 1.1</b> : Effects of different variations: (a) illumination, (b) pose, (c) facial expression. ....	4
<b>Figure 1.2</b> : Illustration of individual manifolds of different identities. Any point on the manifold corresponds to a variation type. The intrinsic geometry is common among different manifolds. This behavior results in the same variation type for same coordinate values. ....	6
<b>Figure 1.3</b> : Demonstration of the semantic difference between (a) a common basis set generated by a classical approach (SVD was used for this example) and (b) class dependent basis sets generated by the proposed approach. Each basis set includes the class information intrinsically. For this example, images under changing illumination conditions were used. ....	8
<b>Figure 1.4</b> : Several synthesis results for a single identity with varying illumination conditions. ....	8
<b>Figure 2.1</b> : Example set of spherical harmonics for a person. This basis set can be used to synthesize images of this person under an arbitrary illumination. Images are taken from [16]. ....	16
<b>Figure 2.2</b> : Embedding results of LPP: (a) 2D embedding of the bootstrap database with changing illumination. (b) Average coordinates corresponding to different illumination conditions. These coordinates are invariant to the identity. ....	22
<b>Figure 2.3</b> : Basis sets of different identities with (a) a constraint over the combination coefficients, (b) no constraint over the combination coefficients. ....	23
<b>Figure 2.4</b> : Mean parameter, $\mu$ , is illustrated for two different variation types: (a) for illumination and (b) for expression. ....	29
<b>Figure 2.5</b> : Illustration of the governing distributions: (a) A template manifold is defined by the marginal distribution, $p(x_k)$ . (b) This template is customized by the identity drawn from the prior distribution, $p(\mathbf{w})$ . ....	34
<b>Figure 3.1</b> : Recognition rates on evaluation sets with different manifold dimensions under (a) illumination and (b) facial expression changes. Yale & Multi-PIE means that the bootstrap set is from Yale and the evaluation set is from Multi-PIE. ....	37
<b>Figure 3.2</b> : Some example images of Yale B Database ....	37
<b>Figure 3.3</b> : Several images from expression databases. ....	39
<b>Figure 3.4</b> : Example face images under different view point conditions. ....	41

<b>Figure 3.5</b> :	Several images from CAS-PEAL and CMU Multi-PIE databases....	44
<b>Figure 3.6</b> :	(a) Behavior of LDA against the illumination with increasing number of identities. Three scenarios were tried: with no bootstrap, with a bootstrap drawn from Multi-PIE, and with a bootstrap drawn from Yale. (b) Behavior of CDFA against illumination and facial expressions. Yale & Multi-PIE means that the bootstrap set is from Yale and the evaluation set is from Multi-PIE. ....	45
<b>Figure 3.7</b> :	Recognition performance of different methods on (a) Multi-PIE illumination database and (b) CAS-PEAL expression database. Values in parentheses shows the number of gallery images. ....	46
<b>Figure 3.8</b> :	Several real life recognition results with pose variations. ....	48
<b>Figure 3.9</b> :	When the system is trained only against the pose variation, it is variant to the illumination changes. Recognition is failed when the position of the light is reversed. ....	49
<b>Figure 3.10</b> :	Unlike illumination, the system is promisingly invariant to the facial expressions although no such information is introduced during the bootstrap. ....	49
<b>Figure 3.11</b> :	Another set of examples to demonstrate the invariance to the facial expressions. ....	50

# ROBUST FACE RECOGNITION ON NONLINEAR MANIFOLDS

## SUMMARY

The face recognition is one of the most studied, yet one of the most incomplete topic due to the nonlinearity and the diversity of variations which are effective during the data acquisition. Developing an algorithm that can handle illumination, pose, expression, occlusion etc. altogether still seems to be a very challenging job. There exist lots of study concerning invariant representations to handle certain variations, yet a generic approach to model different variations at once still seems to be a task to accomplish.

In this study, we define a baseline framework to handle different types of variations. The main attention is to propose a guideline that can be used for different types of variations without requiring any modifications depending on the physical or geometric characteristics of the concerned variation. In other words, the methodology can be utilized for recognition under illumination, pose changes or expression changes. The proposed method is established over the subspace analysis; therefore, the direction of the future works is also defined explicitly.

The proposed method defines the geometry of the variation space spanned by observations (images) of a class (a person) under an operative variation (illumination). This goal is achieved by constructing a coordinate system for this subspace.

Many popular face recognition algorithms use holistic approaches in conjunction with appearance-based models. Appearance-based models utilize the actual pixel intensities, and this fact alone is enough to damage the effective signal-noise ratio since individual pixels tend to change dramatically under certain variations like illumination and facial expression. A common approach to handle these variations is to define a lower dimensional subspace in which the useful statistics are more definite compared to the noise.

Under a problematic variation, individual or class statistics may be altered dramatically preventing a useful discrimination. In LDA, the idea of distinguishing the real signal source and the noise caused by the variation was exploited by controlling the inter-class and intra-class variances. To understand the face space under variations, one needs to determine its geometric structure *i.e.* to understand the distribution of images according to their illumination and pose labels. Definitely, it can not be managed in the original input space because the dimensionality is considerably large, and pixel values tend to change critically even under small environmental changes.

When the utilized appearance-based method depends on a dimensionality reduction technique as a transformation agent, factor analysis happens to be the main actor. Factor analysis is a powerful tool, especially when it is used for the dimensionality

reduction. The classification is achieved in the lower dimensional subspace instead of the noisy higher dimensional pixel space.

Regardless of the selected technique to classify the object, a numerical representation of the object is needed to perform calculations. The simplest representation is the vectorized form of the image matrix. These vectors then are assumed to span a vector space, and all calculations can be carried out in this vector space. In its initial form the vector space assumption is not able to handle real life variations effectively. This assumption is very loose and can only be useful under lots of constraints. First of all, in the real life face images do not span an Euclidean vector space in the sense of mathematical definitions. Beside the fact that the face space is not  $\mathfrak{R}^{m \times n}$  as a topological space, it is not also Euclidean in the geometrical manner since the Euclidean distance can not represent the geometric structure of face distributions. Banach and Hilbert spaces, as more generalized vector spaces, are still useless as they inherit linear scaling of the distance.

Although the face space is not Euclidean, face vectors lie on subspaces which are locally Euclidean and smooth. Differentiable manifolds are generalization of this kind of locally Euclidean and smooth subspaces. Manifold learning approaches can help with employing non-Euclidean geometries into the subspace analyses. The main idea behind manifold learning is to utilize local geodesic distances instead of global Euclidean distances.

In this study, a new subspace analysis perspective, in which a new representation is proposed implicitly, is drawn. Images of a person under a certain variation are assumed to be generated by a linear generative model. The identity of a novel observation is determined by the likelihood of being generated by this model. In other words, the generative model of each person, represents observations (images) by its model parameters. A manifold embedding technique is incorporated to handle the nonlinearity introduced by the variation; hence, a novel connection between manifold learning and generative models is proposed.

The proposed method can be summarized as a two-step probabilistic framework. The first step is a bootstrap phase in which the useful statistics are calculated. A manifold learning technique is employed at this step to define the geometry of the subspace. The second step includes regular training and testing tasks.

Numerous experiments were performed to analyze the performance of the proposed method against different variation types and with relatively large databases. In both cases, the results are very promising. Several advantages of the method can be summarized as follows: (1) different types of variation that lie on smooth manifolds can be handled by the method, (2) the scalability of the classical factor analysis is improved by a class dependent scheme, (3) the decision process is fully probabilistic, and posterior probabilities can be utilized for large scale and domain specific real life applications by incorporating priors on the identities, (4) bootstrap has less time complexity compared to 3D rendering approaches, and finally (5) a single observation for each identity is sufficient to perform reliable recognition while a way to use more images is also introduced.

## DOĞRUSAL OLMAYAN MANİFOLDLAR ÜZERİNDE GÜRBÜZ YÜZ TANIMA

### ÖZET

Günümüze dek yapılmış tüm çalışmalara rağmen, yüz tanıma konusu hala kontrollü ortamlarda gösterdiği başarının ötesinde bir ilerlemeye ihtiyaç duymaktadır. Görüntüleme sırasında etkin olan, ışıklandırma, poz, yüz ifadeleri gibi değişimler tanıma etkinliğini olumsuz yönde ve yoğun bir biçimde etkilemektedir. Belli değişimler karşısında başarı gösteren yöntemler geliştirilmiş olmasına karşın, farklı değişimleri aynı yaklaşım ile modelleyebilen bir çalışmadan bahsetmek pek mümkün olamamaktadır.

Bu çalışmanın amacı, farklı değişimleri modelleyebilecek genel bir yaklaşımın tasarlanması ve başarımının ölçülmesidir. Sunulan yaklaşımın, değişimlere özel ayarlamalara ihtiyaç duymadan, yalın hali ile kullanılabilmesi ve böylece farklı alt uzay incelemelerini aynı çatı altında toplayabilmesi hedeflenmektedir. Önerilen yöntem, genel hatları ile, alt uzay tasarımlarına dayanmaktadır ve böylece gelecekte yöntemin ne şekilde geliştirilebileceği, kapalı bir şekilde sunulmaktadır.

Çalışma içerisinde, farklı değişimlere karşılık gelen görüntülerin oluşturduğu geometrilerin incelenmesi ve bu geometrilere ait bilgilerin ışığında, kişilere ait değişim manifoldlarının oluşturulması ile tanıma işlemin gerçekleştirileceği ortamın hazırlanması söz konusu olacaktır.

Birçok tanıma yöntemi, görünüm tabanlı yöntemleri kullanmaktadır. Görünüm tabanlı yöntemler, işlenmemiş gözek (ing: pixel) parlaklık değerlerini kullanırlar ve bu durum, etkin sinyal/gürültü oranını düşürmek yönünde etkide bulunur çünkü parlaklık değerleri temel değişimler altında büyük değişiklikler gösterirler. Değişimlerin sebep olduğu sorunlar ile baş etmenin en temel yollarından birisi, parlaklık değerlerinden oluşan öznitelik vektörlerinin, daha düşük boyutlu altuzaylar içerisinde temsil edilmesi ve böylece faydalı sayımların etkin hale getirilmesidir.

Görünüm tabanlı yöntemlerler birlikte, düşük boyutlu altuzay tasarımlarından faydalanılacağı zaman, etmen çözümü baş aktör olarak karşımıza çıkar. Etmen çözümlemesinde temel mantık, öznitelik vektörlerinin daha düşük boyutlu altuzaylar içerisinde temsil edilmesi ve sınıflandırmanın bu temsiller yardımı ile gerçekleştirilmesidir. PCA ve LDA gibi boyut düşürme yöntemleri de etmen çözümlemesiyle aynı mantık ile çalışmakta ancak etkileşimin yönünü değiştirmektedir. Etmen çözümü, her gözlemi düşük boyutlu temsillerden üretirek altuzaydan gözlem uzayına doğru bir yönelim sergilerken, diğer yöntemlerde gözlem uzayından altuzaya doğru çalışan dönüşümlerden faydalanılacaktır. Bu çalışma, üretim modelleri üzerine kurulduğu için temel olarak etmen çözümlemesini alacaktır.

Etmen çözümlenmesi ve benzeri temel yöntemler, değişimlerin oluşturduğu geometri-lerin doğrusal olmayan yapıları nedeniyle yeterli etkinliğe ulaşamamaktadırlar. Bu amaçla, uygulama alanı belirlendikten sonra, bu alan için özelleşmiş yöntemlerin kullanılması genelde tercih edilen yöntem olmuştur. Yüz tanıma söz konusu olduğunda, değişimlerin etkilerinin ortadan kaldırılması ancak belli başarımlarla sağlanabilmektedir.

Etmen çözümlenmesi benzeri yöntemler ile üretilen alt uzay, tüm sınıflar (örneğin kişiler) için ortak olacak ve sınıflar arası ayırım, bu alt uzay içerisindeki yerleşim ile sağlanacaktır. Söz konusu değişimin (örneğin farklı ışıklandırmalar) öznelik vektörlerinde meydana getireceği farklılık, sınıf farklılığından daha baskın ise alt uzay içerisindeki konuşlanmalar yeterince etkili olamayacaktır.

Nesne görünümünün taradığı vektör uzayı içerisinde, söz konusu değişimlerin meydana getireceği alt uzayın geometrisi genelde doğrusal değildir. Bu durum, düşük boyutlu alt uzay konaçlarının (ing: coordinates) PCA, LDA gibi doğrusal teknikler ile anlamlı bir şekilde elde edilmesini engelleyecektir. Bu bağlamda geliştirilen ve başarı ile kullanılan Manifold öğrenimi teknikleri doğrusal olmayan geometri-lerin, genellemelere gerek kalmadan incelenebilmesine olanak tanımaktadır.

Bu çalışmada, olasılık tabanlı PCA benzeri bir çerçeve kullanılarak, doğrusallıktan belli düzeyde uzak değişimlerin modellenebilmesi ve bu değişimlerin var olduğu durumlarda sınıflandırma yapılabilmesi için genel amaçlı bir yöntem geliştirilmiştir. Yöntem iki temel aşamadan oluşmaktadır: (1) Manifold öğrenimi ve (2) olasılık temelli üretim modeli. İlk aşamada elde edilen düşük boyutlu alt uzay konaçları, ikinci aşamada sınıfa özel altuzayların belirlenmesinde kullanılmaktadır. Yöntemin en belirgin üstünlüğü, her sınıf için ayrı bir alt uzay elde edilmesi ve eğitim aşamasında her sınıfın tek bir örneğinin yeterli olmasıdır. Sınıfların bağımsız alt uzaylar içerisinde modellenmesi, yöntemin ayırım gücünü oldukça arttırmaktadır.

Çalışmanın ilk adımı, ilgilenilen değişimin meydana getirdiği altuzay geometrisinin belirlenmesi olacaktır. Bu amaçla manifold öğrenimi yöntemleri düşük boyutlu konaç değerlerinin bulunması için kullanılabilir. Bu çalışmada LPP yönteminden faydalanılmaktadır. LPP çıktısı, yeni konaçların üretimi için kullanılan bir  $\mathbf{M}$  matrisidir. Herhangi bir  $\mathbf{x}$  öznelik vektörü için yeni konaç değerleri ( $\mathbf{c} = \mathbf{M}\mathbf{x}$ ) eşitliği ile hesaplanabilir. Kullanılan  $\mathbf{M}$  matrisi tüm sınıflar için ortaktır. LPP ile modelleme sırasında, etiketleme değişim türü üzerinden yapılabilir. Böylece, iki farklı  $\mathbf{x}$  öznelik vektörü aynı değişim türüne sahipse, farklı sınıflara ait olsalar bile karşılık gelen  $\mathbf{c}$  konaç vektörleri aynı olacaktır. Örneğin, kızgınlık ifadesi taşıyan iki farklı kişinin görüntüleri aynı  $\mathbf{c}$  değerlerine sahipken, aynı kişinin kızgınlık ve üzüntü ifadelerindeki görüntüleri farklı  $\mathbf{c}$  değerlerine sahip olurlar.

Manifold öğrenimi ile ilgilenilen değişimin sebep olduğu geometri öğrenildikten sonra amaç altuzayın bir taban takımının hesaplanması olacaktır. Tüm bu amaçlar doğrultusunda bir ön inceleme veri kümesi oluşturulacak ve model öğrenimi gerçekleştirilecektir. Bu doğrultuda, ilgilenilen değişim altındaki görüntülerden oluşan herhangi bir  $\mathbf{X} = \{\mathbf{x}_{ik}\}$  veri kümesinden faydalanılabilir. Burada,  $i$  sınıfına ait ve  $k$  türünde değişikliğe sahip  $\mathbf{x}_{ik}$  öznelik vektörünün,

$$\mathbf{x}_{ik} = \mathbf{W}_i\mathbf{c}_k + \varepsilon_k$$

üretim modeli ile oluşturulduğu kabul edilecektir. Bu eşitliğin geleneksel etmen çözümlemesi yazımının geliştirilmiş bir hali olduğunu söylemek mümkündür. Görüldüğü gibi  $\mathbf{W}_i$  matrisi sınıfa özel bir etmen ağırlıkları kümesiye,  $\mathbf{c}_k$  vektörü her sınıf için ortaktır ve değişimin türünü belirtmektedir. Bu tasarım, geleneksel etmen çözümlemesinden farklı olarak her sınıf için ayrı etmen bileşen kümesi tanımlamaktadır. Eş deyişle, elde edilecek alt uzay taban takımları her sınıf için ayrı olacaktır. Böylece her sınıfa özel ayrı bir etmen çözümlemesi kurgulandığından bahsedilebilir. Bu farklı modeller arasındaki ortaklık,  $\mathbf{c}_k$  vektörleri üzerinden sağlanmaktadır. Başka bir şekilde yorumlamak istersek,  $\mathbf{c}$  vektörleri ilgilenilen değişimin oluşturduğu manifold üzerindeki yerel konaç değerlerimizdir. Benzer şekilde  $\mathbf{W}$  matrisleri de taban takımlarını ifade eder.

Hesaplamaları kolaylaştırmak amacıyla aynı modeli  $\mathbf{x}_{ik}$  vektörünün her bir ögesi için

$$x_{ik} = \mathbf{w}_i^T \mathbf{c}_k + \varepsilon_k,$$

şeklinde yeniden kurgulamak mümkün olmaktadır. Tüm hesaplamalar sırasında bu eşitlikten faydalanılacaktır. Bu eşitlikteki  $\mathbf{c}_k$  değerleri LPP sonrasında bilinir durumdadır. Ayrıca  $\mathbf{w}_i$  vektörü ve  $\varepsilon_k$  sabiti üzerinde sırasıyla

$$\begin{aligned} p(\mathbf{w}_i) &\sim \mathcal{G}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}), \\ p(\varepsilon_k) &\sim \mathcal{G}(0, \sigma_k^2), \end{aligned}$$

Gauss dağılımları kabullenmesi yapılacaktır. Böylece, amacımız bu dağılımların değiştirgelerinin belirlenmesi olacaktır. Bu noktaya kadar tüm hesaplamalar ön inceleme amacıyla gerçekleştirilmiştir. Başka bir deyişle, eğitim sırasında sisteme tanıtılacak olan sınıflara ait  $\mathbf{W}_i$  işleçleri bulunmak istendiğinde temel alınacak dağılımlarda etkin  $\boldsymbol{\Omega}$ ,  $\boldsymbol{\mu}$  ve  $\sigma_k^2$  değiştirgelerinin bulunması sağlanmıştır. Bu hesaplamalarda kullanılacak  $\mathbf{X}$  örneklem kümesinin, eğitim ve test aşamalarında kullanılacak olan kümeden farklı olması beklenmektedir. Her sınıf için birden fazla örneğin ( $\mathbf{x}_{ik}$ ) gerekeceği de unutulmamalıdır.

Ön inceleme aşamasında kullanılan  $\mathbf{X}$  örneklem kümesi, rastgele bir kümedir ve ayırım/tanımaya yapılması istenen sınıf örneklerini içermez. Eğitim aşamasında, üzerlerinde tanıma deneyleri yapılacak kişiler için  $\mathbf{W}_i$  matrislerinin bulunması amaçlanmaktadır. Bu amaçla MAP tahminlemesi,

$$\mathbf{w}_{MAP} = \operatorname{argmax}_{\mathbf{w}} p(\mathbf{w}|\mathbf{x})$$

şeklinde kullanılabilir. Bayes kuralı yardımıyla  $p(\mathbf{w}|\mathbf{x}) = p(\mathbf{x}|\mathbf{w})p(\mathbf{w})$  eşitliğinden faylanmamız mümkün olur. Böylece ön inceleme aşamasında belirlediğimiz dağılımlar yardımıyla MAP tahminlemesine bir çözüm bulabiliriz. Sonuç olarak, tanımak istediğimiz kişinin örnek görüntüsü ile kişiye ait taban takımı belirlenebilecektir.

Yeni bir  $\mathbf{x}$  test örneği için, ait olunan sınıfın belirlenmesi, her sınıf için  $p(\mathbf{x}|\mathbf{W}_i)$  olasılıklarının hesaplanarak, en büyük değer seçilmesi ile olacaktır. Diğer bir yöntem de, eğitim aşamasında hesaplanan  $\mathbf{W}_i$  matrisinin, birim boylu ve dik sütunlardan oluşacak hale getirildikten sonra,  $\mathbf{x}$  vektörünün,

$$\mathbf{x}_i = \mathbf{W}_i \mathbf{W}_i^T \mathbf{x},$$

şeklinde sentezlenmeye çalışılması olabilir. Bu durumda son karar,  $\|\mathbf{x}_i - \mathbf{x}\|$  boyu üzerinden verilecektir.

Yöntemin farklı deęişimler altında çalışabildiğini göstermek amacıyla, ışılandırma, poz ve ifade farklılıkları söz konusuken yüz tanıma deneyleri yapılmıştır. Yöntem, mevcut yazında başarılı olarak nitelendirilen yöntemlerle yarışan başarımları elde etmiş ve yüksek boyutlu veritabanları için de uygun olduğunu kanıtlamıştır.

Önerilen yöntemin bazı temel artı deęerleri şu şekilde sıralanabilir: (1) Manifoldlar üzerinde tanımlı farklı deęişimler, yöntem üzerinde yenilemeye ihtiyaç duyulmadan kontrol altına alınabilmektedir. (2) Geleneksel etmen çözümlemesi yaklaşımının etkinliği ve ölçeklenebilirliği, sınıf temelli bir yaklaşım ile artırılmıştır. (3) Karar verme süreci tamamen olasılıksaldır ve böylece yüksek boyutlu veritabanlarına yönelik olarak öncül olasılıkların devreye sokulması ve alınacak kararın alan bilgisi ile kuvvetlendirilmesi mümkündür. (4) Üç boyutlu modellemeler ile kıyaslandığında, ön inceleme aşamasının zaman karmaşıklığı daha düşüktür. (5) Her kişinin tek bir örneğinin bulunması tanıma için yeterliyken, birden çok görüntünün bulunması durumunda başarımları arttıracak eklentiler de tanımlanmıştır.



## 1. INTRODUCTION

When the subject under consideration is the object recognition and specifically the face recognition a compelling question arises: why does such an exhaustively studied subject still need further attention which concludes in a PhD thesis? Answer is easy just like the question itself: Among all computer vision studies, the problem of recognizing faces is one of the most studied, yet one of the most incomplete topic due to the nonlinearity and the diversity of variations which are effective during the data acquisition. Developing an algorithm that can handle illumination, pose, expression, occlusion etc. altogether still seems to be a very challenging job. That may be realized by using 3D scanners or equivalent technologies during the recognition; however, such an solution itself produces new constraints in addition to the already exhausting real life requirements.

Regardless of the selected technique to classify the object, a numerical representation of the object -a face in our case- is needed to perform calculations. At this point, probably the most important decision should be made which in turn determines the upper boundary of the final recognition rate. The decision concerns the selection of the base representation. The utilized classification algorithm can only push the limit implicitly defined by the representation.

The simplest representation is the vectorized form of the image matrix. These vectors then are assumed to span a vector space, and all calculations can be carried out in this vector space. This simple idea was actually a corner stone for today's recognition algorithms. When for the first time, M. Turk and A. Pentland made use of Euclidean vector spaces by employing a well known dimensionality reduction technique *Principal Component Analysis* (PCA) [1] in their remarkable work [2], they opened a gate to the diverse possibilities of the matrix algebra.

Indeed, in its initial form the vector space assumption is not able to handle real life variations effectively. This assumption is very loose and can only be useful under

lots of constraints. First of all, in the real life face images do not span an Euclidean vector space in the sense of mathematical definitions. Even when the face images are considered as  $m \times n$ -dimensional vectors, there is no meaning of multiplying a face vector with a scalar (especially with a negative one). Pixel values are bounded in some intervals like  $[0, 255]$ , and image vectors can not be generated by adding two face vectors if the resulting pixels are outside this interval. Beside the fact that the face space is not  $\mathfrak{R}^{m \times n}$  as a topological space, it is not also Euclidean in the geometrical manner since the Euclidean distance can not represent the geometric structure of face distributions. Banach and Hilbert spaces, as more generalized vector spaces, are still useless as they inherit linear scaling of the distance [3]. Due to all these negative aspects, techniques relying on linear subspaces of face images are easily affected by even simple variations.

Embeddings like PCA can solve problems caused by statistically well behaving noise. However, under a problematic variation, individual or interclass statistics may be altered dramatically preventing a useful discrimination. An elegant idea is to distinguish the real signal source (identity of the image) and the noise caused by the variation (differences imposed by illumination). In *Linear Discriminant Analysis* (LDA) [4], this idea was exploited by controlling the inter-class and intra-class variances. That was the second leap towards the world of sophisticated subspace analyses. After LDA, we now know that it is possible to embed face images in a subspace which explicitly designed to handle the variation. Using a layer of abstraction (representing faces by coordinates inside the subspace instead of original pixel values), it is possible to get a new set of vectors behaving more presumably under variations.

One way to understand the face space under variations like illumination and pose changes is to determine its geometric structure *i.e.* to understand the distribution of images according to their illumination and pose labels regardless of their identities. Definitely, it can not be managed in the original input space because the dimensionality is considerably large, and pixel values tend to change critically even under small environmental changes.

Although the face space is not Euclidean, face vectors lie on subspaces which are locally Euclidean and smooth. Differentiable manifolds are generalization of this

kind of locally Euclidean and smooth subspaces. Face images taken from different viewpoints or under changing illumination conditions can be regarded as lying on smooth manifolds [5–9]. Under uncontrolled environmental settings, the manifold assumption may not hold due to the complexity of the data. However, it is still possible to utilize this assumption by only considering a small set of significant factors.

Manifold learning approaches can help with employing non-Euclidean geometries into the subspace analyses. Manifold learning can be summarized as a nonlinear dimensionality reduction technique based on the assumption that input data lie on a differentiable manifold. The main idea behind manifold learning is to utilize local geodesic distances instead of global Euclidean distances [6, 9, 10].

In this study, a new subspace analysis framework called *Class Dependent Factor Analysis* (CDFA) is proposed. During the formulation of the framework, a new representation is suggested implicitly. Images of a person under a certain variation are assumed to be generated by a linear generative model. The identity of a novel observation is determined by the likelihood of being generated by this model. In other words, the generative model of each person, represents observations (images) by its model parameters. A manifold embedding technique is incorporated to handle the nonlinearity introduced by the variation; hence, a novel connection between manifold learning and generative models is proposed.

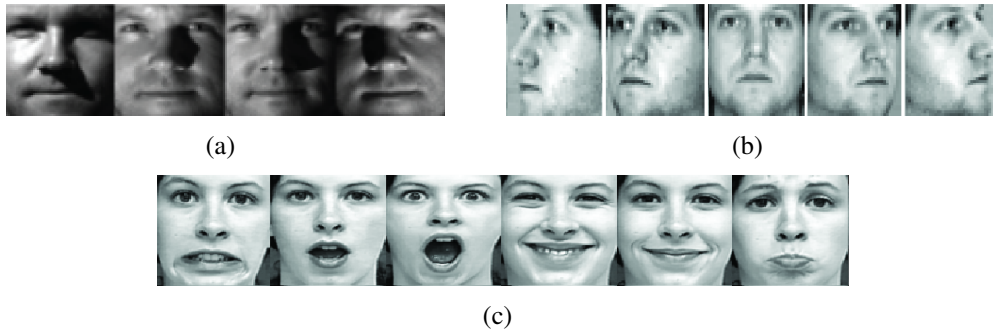
### **1.1 The Problem Definition**

The proposed framework is an alternative approach for handling different variations in the face recognition problem. The scope of the study consists of a generic way to deal with three leading factors namely, illumination, viewpoint, and facial expression. Face recognition under such variations is the main challenging task in the domain. This study addresses a common and generic solution which can be employed against such variations without any modification based on geometrical or physical aspects of the variation.

Appearance based models (*i.e.* feature vectors are constructed by raw pixel values) are utilized through the study. Input images are used in their raw gray valued form

without any preprocessing (beside z-normalization) or new representations like LBP [11]. Hence, the explicit shape information is not present in the feature vectors.

Example face images are given in Figure 1.1 to illustrate effects of illumination, facial expression, and viewpoint.



**Figure 1.1:** Effects of different variations: (a) illumination, (b) pose, (c) facial expression.

## 1.2 The Classical Approaches And Their Limitations

Many popular face recognition algorithms use holistic approaches in conjunction with appearance-based models [12]. Appearance-based models utilize the actual pixel intensities, and this fact alone is enough to damage the effective signal-noise ratio since individual pixels tend to change dramatically under certain variations like illumination and facial expression. A common approach to handle these variations is to define a lower dimensional subspace in which the useful statistics are more definite compared to the noise. As an example, PCA is used to define a subspace where the variance on principal axes is maximized.

When the utilized appearance-based method depends on a dimensionality reduction technique, factor analysis (FA) happens to be the main actor. Besides the methods which concern physical and geometric properties of the studied object, most of the modern approaches share the main ideas of this statistical tool. FA is a well known and commonly used approach in the data analysis community. Although its early development traces to the beginning of the century, it is still one of the most popular multivariate statistical analysis tools in applied science domain [13]. Its main

formulation is a linear generative model

$$\mathbf{x} = \mathbf{W}\mathbf{c} + \boldsymbol{\varepsilon}, \quad (1.1)$$

where the weighted average of lower dimensional factors,  $\mathbf{c}$ , is taken to generate a higher dimensional signal,  $\mathbf{x}$ . In this view, FA can be seen as a dimensionality reduction technique when the inverse mapping of  $\mathbf{W}$  is considered.

FA is a powerful tool, especially when it is used for the dimensionality reduction. The classification is achieved in the lower dimensional subspace instead of the noisy higher dimensional pixel space. The very same idea is exploited in PCA and LDA. They both have similar underlying generative models but different directions between the lower dimensional subspace and the higher dimensional observation space. For PCA and LDA, the direction is drawn from the higher dimensional observation space to the lower dimensional subspace as in

$$\mathbf{c} = \mathbf{W}^T \mathbf{x}, \quad (1.2)$$

when considering zero mean observations. Although the error term is omitted in this form, it is modeled implicitly by defining a distribution over observations. In PCA, the transformation matrix,  $\mathbf{W}$ , is estimated by considering the eigenvectors of the empirical covariance matrix of observations while in LDA, it is constructed by maximizing the (between variance / within variance) ratio of classes. Indeed, the most important difference between LDA, PCA, and FA is the fact that LDA is a supervised method whereas PCA and FA are unsupervised methods.

In classical approaches, the first limitation arises with the *common subspace* constraint: The mapping,  $\mathbf{W}$ , is common for all classes. The discrimination among classes is achieved by the deployment of the class centroids on the coordinate system. Such a modeling is insufficient when the effect of the variation is more dominant than the class characteristics. In such a case, the coordinates of the points are mostly determined by the variation type. A well known example is the fact that the images of different people under same illumination lie closer in such subspaces compared to the images of a single person under different illumination.

Another important drawback of a classical subspace approach is its dimensionality concerns. As new identities are introduced to the gallery, methods like FA, LDA, and

PCA require the subspace is to be re-constructed to increase the dimensionality. This is an important constraint to sustain the scalability of the method.

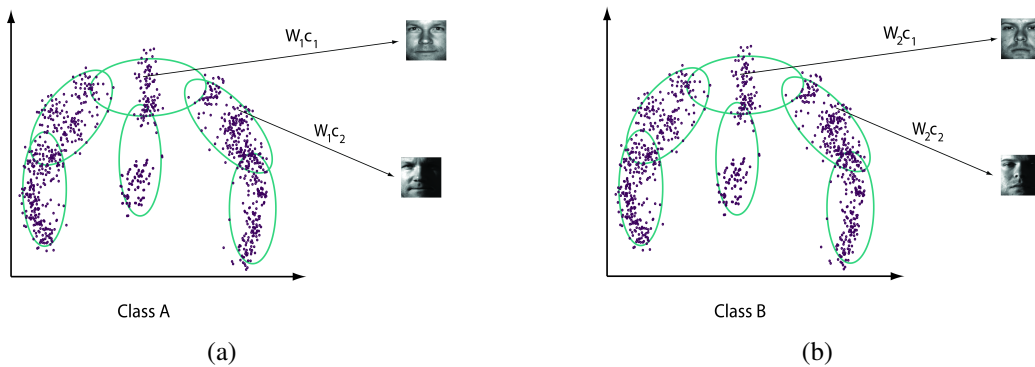
As a critical fact, classical embeddings like PCA can handle variations caused by statistically well behaving noise terms. However, variation types that are effective in the real life prevent a useful discrimination by altering individual or interclass statistics dramatically.

### 1.3 Overview of the CDFFA Framework

The design of the framework starts with the reformulation of the factor analysis model under a variation such as illumination. An observation  $\mathbf{x}_{ik}$ , which belongs to the class  $i$  and has a variation  $k$ , is generated by the model

$$\mathbf{x}_{ik} = \mathbf{W}_i \mathbf{c}_k + \boldsymbol{\varepsilon}_k. \quad (1.3)$$

With this formulation, individual factor loadings,  $\mathbf{W}_i$ , for each class  $i$ , are introduced instead of a common loading matrix for all classes. However, the factors,  $\mathbf{c}_k$  (coordinates on the lower dimensional subspace), are common for all classes and related to the variation type. The geometric interpretation yields different manifolds for different classes while all manifolds have exactly same intrinsic geometries. Inside two manifolds, points having same local coordinates correspond to the same variation type. This interpretation is illustrated in Figure 1.2.



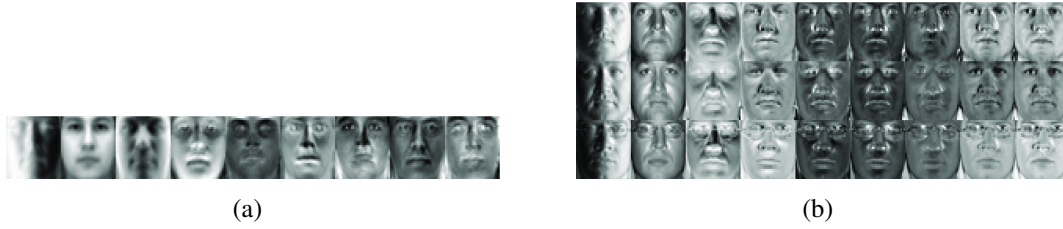
**Figure 1.2:** Illustration of individual manifolds of different identities. Any point on the manifold corresponds to a variation type. The intrinsic geometry is common among different manifolds. This behavior results in the same variation type for same coordinate values.

Several important aspects of this formulation should be mentioned:

- Each class has its own subspace/manifold. Therefore, discrimination between classes is performed by the distance to the manifold instead of the distance within the manifold. Inside each individual manifold, a mixture of Gaussians is defined to model the variation.
- Coordinate vectors,  $\mathbf{c}_k$ , represent the variation type instead of class identities. Thus, the determination of the variation value is explicitly provided. The variation of an observation,  $\mathbf{x}_{ik}$ , can be determined if the factor loadings are known.
- Class identities are stored as factor loadings in matrix  $\mathbf{W}_i$ . This property increases the scalability of the recognition as more space is left for identity. The variation does not condition the structure of the matrix since it is already modeled by the factors. Theoretically, recognition can be performed under even severe variations, as long as class dependent factor loadings are recovered successfully.
- The intrinsic dimensionality of manifolds is fixed once determined during the bootstrap. Nevertheless, the actual dimensionality in which the recognition is performed is  $n$  since the manifolds are embedded in  $\mathfrak{R}^n$ , where  $n$  is the number of pixels in images.
- A manifold learning step is employed to derive the reduced dimensional coordinates,  $\mathbf{c}_k$ . Thus, a connection between manifold learning and probabilistic generative models is proposed. This can be seen as an initial step towards nonlinear probabilistic models.

The difference between individualized and common factor loadings can be observed in Figure 1.3. The proposed method introduces basis sets which are specific to their corresponding classes.

With this setting, one can synthesize different images of a person under different conditions like changing illumination given a class dependent basis set. Figure 1.4 illustrates an example synthesis. Results can be improved by sophisticated error models; however, this work does not concern such a task as the main goal is limited with the discrimination among classes.



**Figure 1.3:** Demonstration of the semantic difference between (a) a common basis set generated by a classical approach (SVD was used for this example) and (b) class dependent basis sets generated by the proposed approach. Each basis set includes the class information intrinsically. For this example, images under changing illumination conditions were used.

A critical feature of the method is its generic structure. No physical or geometrical attributes of the concerned variation are employed during calculations. Hence, any variation lying on a smooth manifold can be modeled by the proposed method.



**Figure 1.4:** Several synthesis results for a single identity with varying illumination conditions.

#### 1.4 Connections to Previous Works

The proposed method has an analogous formulation with the probabilistic interpretation of PCA [14, 15]. Both approaches tackle with finding lower dimensional representations of observations under some prior assumptions. The main difference is that the proposed method derives class specific coordinates and accounts for the variation explicitly.

A similar framework was introduced in [16]. That work dealt with individualized subspaces. The actual improvement over [16] is that CDFA has a more generic



structure which can be used for the general classification problem whereas only illumination was considered in [16]. The authors of [16] used spherical harmonics to calculate class specific bases. The results are limited to illumination as the spherical harmonics can not be generalized to other types of variation. The relation between reflectance functions on a Lambertian surface and spherical harmonics was defined in [17] and [18].

The authors of [19] developed a cone model to solve the face recognition problem with varying illumination. They argued that the set of images of an object in a fixed pose but under all possible illumination define a convex cone. The approach requires a few images of each gallery identity to estimate its surface geometry and albedo map. After estimation is completed, synthetic images with different illumination conditions can be rendered. That model illustrates the real power of the subspace analysis; nevertheless, it is again constrained to be useful only for illumination and may not work with a single observation. The proposed method is able to work with a single observation while extra observations increase the accuracy.

Other techniques such as [20–24] suffer from being useful only for the specific variation type that they have been developed for. We try to propose a method which can be used for different variations.

A comparable work was performed in [25]. Authors defined a common subspace for class identities yet different transformation matrices (factor loadings) for different poses. Keeping the class information inside the coordinate vectors inherits an important disadvantage of classical subspace methods: as the number of classes increases, the subspace dimensions also need to be increased to sustain the scalability. The technique may work with different variations that can be discretized. The same idea was used in [26] again for pose variations.

The probabilistic approaches for the discriminative subspace analysis were proposed in [27] and [28]. Both solutions were based on LDA with different settings. In [27], authors defined a three layer decision process. At the initial layer, identity is drawn from a common Gaussian distribution. Then, at the second layer a perturbation is applied by another Gaussian. Finally, the third layer defines a projection from the latent

space to the observation space. In [28], the model introduced in [25] was improved by employing different projections from the latent space to the observation space: one for the between-individual subspace and one for the within-individual subspace. Both models still assume common subspaces for different identities.

Compressive sensing and sparse representation were utilized in [29] and [30]. The subspace analysis was performed on the basis of compressive sensing theory. Both techniques can be used for different types of variation. The technique introduced in [29] finds a discriminative sparse representation of each probe image by using the whole gallery as a dictionary *i.e.* by a linear combination of gallery images. Such a model requires each gallery identity to have a sufficiently large training set, and the space complexity is high since all training images have to be stored and accessed. The method in [30] assumes that an image of a class can be represented as a sum of a common component and an innovation component. The common component carries the main identity related information for the class while the sparse innovative component is specific to the image and includes the information related to the variation. To calculate required statistics, both techniques need several images of an identity. These methods are used in our benchmarks against facial expressions.

## 1.5 Other Related Works

The face recognition can be seen as one of the most popular and successful applications in the image processing and understanding domain [12]. However, as a challenging problem, illumination and pose invariant recognition still remains as an open study. Face images taken in an uncontrolled environment usually contain variations in viewpoint and illumination; therefore, these two factors have an important role in the robustness of the system.

It is known for a long time, the feature-based methods like *elastic bunch graph matching* [31] are promisingly successful against lots of factors including illumination and viewpoint [12]. Nevertheless, their extreme sensitivity to the feature extraction and the measurement of extracted features makes them unreliable [19]. As a result, appearance-based methods have dominated the literature.

One of the milestones for face recognition under variations can be stated as *fisherfaces* [4] technique. LDA was used in [4] to construct a subspace on which inter-person variance is optimally large while intra-person variance is efficiently small. The main drawback of the technique, same as PCA [2], is the Euclidean consideration of the data space. The method fails when data points lie on a nonlinear subspace, which is usually true with multimodally distributed face images. A promising improvement was proposed in [26] as using local linear transformations instead of one global transformation. Method finds different mapping functions for different pose classes. When a probe image is tested, its pose is determined by a soft clustering. Deciding to the number of pose clusters is a vital problem as in all clustering algorithms. Moreover, novel poses can not be handled in case of critical variations.

In [32], authors used the neighborhood structure of the input space to determine the underlying nonlinear manifold of multimodal face images. *Locality Preserving Projections* (LPP) was applied to calculate a basis set called *laplacianfaces*. Face images with different poses, facial expressions, and illumination conditions were studied and the recognition performances were shown to be higher compared to *fisherfaces* or *eigenfaces*.

Pose variation was studied in [33] by using view-based eigenfaces. For each view, eigenfaces were calculated and employed as separate transformations into a common lower dimensional subspace. Authors also introduced *eigenfeatures* by which a feature based scheme was incorporated. Their performance highly depends on the discretization as it is a fact in [22]. In [22], the eigen light fields technique was utilized to define the subspace of poses. Unfamiliar poses can be handled by the technique. Authors in [20] combined the generalized photometric stereo and eigen light field concept to design a generic method which is also insensitive to illumination changes. 3D morphable face models were used in [34], [19], and [16] to generate novel poses, and their performance values were superior to the previous research. Rendering ability for new poses and illumination conditions is exceptional with 3D morphable models [35]. However, the computational cost of generating 3D models from 2D images or using laser scanners to access 3D models decrease the feasibility of the recognition system.

Illumination variance was studied in [23]. Authors proposed quotient image as an illumination insensitive identity signature. The approach may fail when the probe image has an unpredictable shadow; however, it has the ability of recognizing probe images with illuminations different than that of gallery images. Technique requires only one gallery image per subject. The method in [24] introduced extra constraints on the albedo and the surface normal to remove the shadow constraint.

An *illumination cone* model was proposed in [19]. Authors argued that set of images of an object in a fixed pose but under all illumination conditions define a convex cone. The method requires a few images of a test identity to estimate its surface geometry and albedo map. To handle pose variations, they defined different illumination cones for each sampled viewpoint.

All sets of Lambertian reflectance functions, which can be used to generate all kinds of illumination conditions for Lambertian objects, were defined in [17] and [18]. They showed that by using only nine spherical harmonics, a wide variety of illumination can be approximated. A methodology for recognition was also proposed in [17]. In [16], spherical harmonics approach was exploited, and excellent results for recognition were represented. They implemented a 3D morphable model to achieve pose invariance, and this requires generating 3D face models from 2D images.

Authors in [36] suggested a nonlinear subspace approach using the tensor representation of faces in different conditions like facial expressions, illumination, and poses. They employed  $n$  mode tensor *Singular Value Decomposition* (SVD) to generate image basis. The method requires several images under different variations for each training identity. In [37], another nonlinear subspace analysis was proposed by the manifold assumption. For each identity, a gallery manifold is stored in the database. When a test identity with several new poses arrives its probe manifold is constructed and by help of manifold to manifold distance, its identity is determined. The requirement for multiple images of the test person is the main drawback.

A considerable idea was introduced by bilinear generative models that can be used to decompose orthogonal factors in [38]. Authors defined a separable bilinear mapping between the input space and the lower dimensional subspace. Once all parameters of

mappings are determined, one can separate identity and pose information explicitly. They analyzed the recognition and synthesizing capabilities of the technique, and the results were promising. In [39], illumination invariance was analyzed by employing a similar framework. To overcome the matrix inversion requirement in the symmetric bilinear model, authors proposed a ridge regressive technique. A modified asymmetric model was introduced in [25] to cope with pose variations. Discretization resolution for the pose space is one of the leading factors on performance. The nonlinearity for the generative models was incorporated in [40]. Authors recommended a nonlinear scheme combined with the bilinear model, and the linearity constraint of the classical generative models was tried to be removed.



## 2. A GENERIC FRAMEWORK

In this study, we define a baseline framework to handle different types of variations. The main attention is to propose a guideline that can be used for different types of variations without requiring any modifications depending on the physical or geometric characteristics of the concerned variation. In other words, the methodology can be utilized for recognition under illumination, pose changes or expression changes. The proposed method is established over the subspace analysis; therefore, the direction of the future works is also defined explicitly.

The CDFA defines the geometry of the variation space spanned by observations (images) of a class (a person) under an operative variation (such as illumination). This goal is achieved by constructing a coordinate system for this subspace.

### 2.1 Constructing a Basis Set for a Variation Type

The data geometry of subspaces spanned by the different images of a person under changing illumination has been studied by several authors [16–19]. For instance, spherical harmonics that can be employed to approximate any reflectance functions were defined as a basis set of the illumination subspace in [17] and [18]. Authors of [16] showed that this subspace can be effectively used for recognition under illumination changes. Once you have a 3D map of a person *i.e.* surface normals of the face map, spherical harmonics for this person can be defined as

$$\begin{aligned}
 w_1 &= \frac{1}{\sqrt{4\pi}}\lambda, & w_2 &= \sqrt{\frac{3}{4\pi}}\lambda n_z, & w_3 &= \sqrt{\frac{3}{4\pi}}\lambda n_x, \\
 w_4 &= \sqrt{\frac{3}{4\pi}}\lambda n_y, & w_5 &= \frac{1}{2}\sqrt{\frac{5}{4\pi}}\lambda(2n_z^2 - n_x^2 - n_y^2), \\
 w_6 &= 3\sqrt{\frac{5}{12\pi}}\lambda n_x n_z, & w_7 &= 3\sqrt{\frac{5}{12\pi}}\lambda n_y n_z, \\
 w_8 &= \frac{3}{2}\sqrt{\frac{5}{12\pi}}\lambda(n_x^2 - n_y^2), & w_9 &= 3\sqrt{\frac{5}{12\pi}}\lambda n_x n_y,
 \end{aligned}$$

where  $\lambda$  is the surface albedo and  $n_x$ ,  $n_y$ , and  $n_z$  are surface normals at  $x$ ,  $y$ , and  $z$  directions, respectively. Only 9 harmonics are sufficient to capture approximately 99.9% of the energy of any reflectance function [17]. An example set of harmonics for a person is demonstrated in Figure 2.1. Considering these harmonics as a basis set for the variation subspace yields the following interpretation: Once the coordinate system of the subspace corresponding to a person is constructed, it is possible to synthesize any image of this person under any probable illumination condition. In this setting, a given probe image can be recognized by a metric such as distance-to-manifold. Hence, the initial problem is reduced to the problem of recovering basis sets for people in the gallery.



**Figure 2.1:** Example set of spherical harmonics for a person. This basis set can be used to synthesize images of this person under an arbitrary illumination. Images are taken from [16].

Similar ideas were exploited in [19]. Again individual subspaces (illumination cones) are defined for each person in the gallery. Those approaches may only fill a limited gap for the real life recognition tasks since they are highly restricted to illumination changes. One may not define a harmonic set or a cone model analytically for facial expressions. Indeed, the main goal of this study is to eliminate this constraint. We try to find a way to define basis sets corresponding to different types of variations without using any physical or geometric properties of the concerned variation.

## 2.2 Proposed Generic Basis Recovery Scheme

The proposed scheme is an optimization procedure based on the linear generative model

$$\mathbf{x}_{ik} = \mathbf{W}_i \mathbf{c}_k. \quad (2.1)$$

An image  $\mathbf{x}_{ik}$  is assumed to be generated by the linear combination of the basis vectors (columns of the matrix  $\mathbf{W}_i$ ). The combination coefficients,  $\mathbf{c}_k$ , are the lower dimensional coordinates in the subspace defined by the range of the matrix  $\mathbf{W}_i$ . Let's



assume that we have  $K$  images of a person  $i$  with  $K$  different values of a certain variation (images with different viewpoints or illumination). The total reconstruction error inside the subspace related to that variation can be defined as

$$\begin{aligned}\mathcal{E}_i &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{W}_i \mathbf{c}_k\|^2 \\ &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{w}_{i1}c_{k1} - \mathbf{w}_{i2}c_{k2} - \dots - \mathbf{w}_{in}c_{kn}\|^2,\end{aligned}\quad (2.2)$$

in terms of bases ( $\mathbf{W}_i$ ) and coordinates ( $\mathbf{c}_k$ ) where  $\mathbf{w}_{ij}$  indicates  $j^{\text{th}}$  column of the matrix  $\mathbf{W}_i$ , and  $c_{kj}$  is  $j^{\text{th}}$  element of vector  $\mathbf{c}_k$ .

As the notation states, individual bases for different identities are defined while keeping the coordinates common to identities. This behavior is very similar to the one used in the spherical harmonics approach. The basis,  $\mathbf{W}_i$ , can be calculated by minimizing the error. This procedure is repeated for each identity.

Indeed, this method is only useful if a complete set of images for each identity is present. Unfortunately, this is not the case for real life scenarios. Therefore, a way to recover the basis matrix,  $\mathbf{W}_g$ , of a gallery identity  $g$  is required when only a few or a single observation is present,  $\mathbf{x}_{gk}$ . This may be achieved by the *Maximum a Posterior* (MAP) estimate as in

$$\begin{aligned}\mathbf{W}_{MAP} &= \arg \max_{\mathbf{W}_g} p(\mathbf{W}_g | \mathbf{x}_{gk}) \\ &= \arg \max_{\mathbf{W}_g} p(\mathbf{x}_{gk} | \mathbf{W}_g) \cdot p(\mathbf{W}) .\end{aligned}\quad (2.3)$$

Such an approach requires the prior distribution,  $p(\mathbf{W})$ , and the likelihood,  $p(\mathbf{x}_{gk} | \mathbf{W}_g)$ , is defined beforehand. Given a novel observation,  $\mathbf{x}_{pk}$ , the class label can be determined by assigning the identity  $g$  with the maximum likelihood,  $p(\mathbf{x}_{pk} | \mathbf{W}_g, \mathbf{c}_k)$ .

### 2.3 Mathematical Background

The proposed method can be summarized as a two step probabilistic framework. The first step is a bootstrap phase in which useful statistics are calculated. A manifold embedding technique is employed at this step to define the geometry of the subspace. The second step includes regular training and testing tasks. Framework starts with analyzing the underlying manifold. A bootstrap database, consisting of identities with

several observations (people with several images), is collected for this purpose. The identities of the bootstrap database are different than the ones to be recognized; any suitable database can be selected.

To simplify the calculations, the equation (1.3) is rewritten in an element-wise form as

$$x_{ik} = \mathbf{w}_i^T \mathbf{c}_k + \varepsilon_k, \quad (2.4)$$

where  $x_{ik}$  is an element of the observation vector,  $\mathbf{x}_{ik}$ . Similarly, the vector  $\mathbf{w}_i$  is the corresponding row of the matrix  $\mathbf{W}_i$ . Again,  $\varepsilon_k$  is the corresponding element of the error vector,  $\varepsilon_k$ . Such an element-wise formulation ignores the correlations among pixels while introducing new correlations among columns of  $\mathbf{W}_i$ . Unlike the classical factor analysis model, the factors are treated as deterministic variables which are calculated during the manifold learning step. Moreover, distributions

$$\begin{aligned} p(\mathbf{w}) &\sim \mathcal{G}(\boldsymbol{\mu}, \boldsymbol{\Omega}^{-1}), \\ p(\varepsilon_k) &\sim \mathcal{G}(0, \sigma_k^2), \end{aligned} \quad (2.5)$$

on the vector  $\mathbf{w}$  and the constant  $\varepsilon_k$  are defined. Along with the prior over the vector  $\mathbf{w}$ , the conditional probability  $p(x_k|\mathbf{w}, \mathbf{c}_k)$  is needed for the MAP estimate. It may be defined as another Gaussian by

$$p(x_k|\mathbf{w}, \mathbf{c}_k) \sim \mathcal{G}(\mathbf{w}^T \mathbf{c}_k, \sigma_k^2). \quad (2.6)$$

The mean and the variance of the distribution are calculated by

$$\begin{aligned} E[x_k|\mathbf{w}, \mathbf{c}_k] &= E[\mathbf{w}^T \mathbf{c}_k + \varepsilon_k] = \mathbf{w}^T \mathbf{c}_k + E[\varepsilon_k] = \mathbf{w}^T \mathbf{c}_k, \\ E[(x_k - E[x_k|\mathbf{w}, \mathbf{c}_k])^2] &= E[(x_k - \mathbf{w}^T \mathbf{c}_k)^2] = E[(\varepsilon_k - 0)^2] \\ &= E[(\varepsilon_k - E[\varepsilon_k])^2] = \sigma_k^2, \end{aligned} \quad (2.7)$$

using the generative model (2.4).

The proposed method is detailed through the following sections and summarized in Table 2.1. For all formulations, a single variation such as illumination is considered for the sake of simplicity. The bootstrap database will include multiple images of people under different conditions. However, the gallery including identities that are to be recognized may contain a single image of each identity.

**Table 2.1:** Summary of the CDFA.

---

*Bootstrap:* Given a bootstrap database,  $\mathbf{X} = \{\mathbf{x}_{ik}\}$

- Calculate lower dimensional coordinates,  $\mathbf{c}_k$ , for each observation,  $\mathbf{x}_{ik}$ , by a manifold learning technique (Section 2.3.1)
- Calculate the parameters  $\mu, \Omega^{-1}, \sigma_k^2$  (Section 2.3.2 and Section 2.3.3)

*Training:* For each identity  $g$  in the gallery,

- Recover  $\mathbf{W}_g$  specific to this identity by maximizing  $p(\mathbf{w}_g | x_{gk}, \mathbf{c}_k)$  for each element  $x_{gk}$  of the observation,  $\mathbf{x}_{gk}$  (Section 2.3.4)

*Testing:* Given a probe observation  $\mathbf{x}_{pk}$ ,

- Calculate the point to manifold distance for each identity  $g$  in the gallery, and select the one with the minimum value (Section 2.3.5)
- 

### 2.3.1 Manifold Learning

The aim of this step is to define a mapping,  $\mathbf{M}$ , from the high dimensional image space to the lower dimensional variation space as in

$$\mathbf{c}_k = \mathbf{M}^T \mathbf{x}_k . \quad (2.8)$$

The term *variation space* is chosen to emphasize that the coordinates of the subspace are related to the variation. Locality Preserving Projections (LPP) [10] is employed as a manifold embedding technique. This technique tries to preserve the intrinsic geometry and the local structure of the underlying manifold. Method starts with the one dimensional subspace assumption. In this view, the one dimensional representations of two observations  $\mathbf{x}_k$  and  $\mathbf{x}_j$  are  $c_k$  and  $c_j$ . The relation between  $\mathbf{x}_k$  and  $c_k$  is defined as  $c_k = \mathbf{m}^T \mathbf{x}_k$ , where the vector  $\mathbf{m}$  is a column of the mapping  $\mathbf{M}$ . Considering the weighted distance between data points in the one dimensional subspace as an error, the total error after the dimension reduction becomes

$$\mathcal{E} = \sum_k \sum_j (c_k - c_j)^2 S_{kj} , \quad (2.9)$$

where the coefficients  $S_{kj}$  are the similarity indices and related to the distances in the higher dimensional observation space. They may be defined by

$$S_{kj} = \begin{cases} \exp(-\|\mathbf{x}_k - \mathbf{x}_j\|^2/t), & \|\mathbf{x}_k - \mathbf{x}_j\|^2 < \delta, \\ 0, & \text{otherwise} \end{cases} \quad (2.10)$$

where the parameter  $\delta$  determines the radius of the local neighborhood. In other words, method tries to assign close coordinates to the points which are in a small neighborhood in the observation space. The cost function (2.9) can be rewritten as

$$\begin{aligned} \mathcal{E} &= \frac{1}{2} \sum_k \sum_j (c_k - c_j)^2 S_{kj} \\ &= \frac{1}{2} \sum_k \sum_j (\mathbf{m}^T \mathbf{x}_k - \mathbf{m}^T \mathbf{x}_j)^2 S_{kj} \\ &= \mathbf{m}^T \mathbf{X}(\mathbf{D} - \mathbf{S})\mathbf{X}^T \mathbf{m} \\ &= \mathbf{m}^T \mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{m}, \end{aligned} \quad (2.11)$$

where the matrix  $\mathbf{X}$  has data points,  $\mathbf{x}_i$ , as its columns.  $\mathbf{D}$  is a diagonal matrix, and its entries are column sums of  $\mathbf{S}$ .  $\mathbf{L} = \mathbf{D} - \mathbf{S}$  is the Laplacian matrix. By introducing a constraint ( $\mathbf{m}^T \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{m} = 1$ ), the minimization of (2.11) is transformed to the generalized eigenvalue problem as

$$\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{m} = \lambda \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{m}. \quad (2.12)$$

Then, the eigenvectors corresponding to minimum eigenvalues are selected to construct a linear mapping,  $\mathbf{M}$ .

The selection of the similarity indices totally determines the structure of the embedding. In the current form, LPP preserves the locality by minimizing the local variance. When  $S_{kj}$  is taken to be  $1/n^2$  for all  $k, j$  then the Laplacian matrix,  $\mathbf{L}$ , becomes the data covariance matrix. In this form, we get the solution of PCA by collecting the eigenvectors corresponding to the maximum eigenvalues. As an another choice,  $S_{kj}$  can be defined in a supervised manner by

$$S_{kj} = \begin{cases} 1/n_c, & \text{if } \mathbf{x}_k \text{ and } \mathbf{x}_j \text{ both belong to the class } c, \\ 0, & \text{otherwise} \end{cases} \quad (2.13)$$

where  $n_c$  is the number of data points in class  $c$ . By this way,  $\mathbf{X}\mathbf{L}\mathbf{X}^T$  becomes the within covariance matrix  $\mathbf{S}_W$ . Similarly the between covariance matrix  $\mathbf{S}_B$  is  $\mathbf{C} - \mathbf{X}\mathbf{L}\mathbf{X}^T$  where the matrix  $\mathbf{C}$  is the data covariance matrix. In LDA, a generalized eigenvalue problem

is solved as

$$\mathbf{S}_B \mathbf{m} = \alpha \mathbf{S}_W \mathbf{m}. \quad (2.14)$$

With the new weight configuration, this equation is equivalent to

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{m} = \lambda \mathbf{C} \mathbf{m}. \quad (2.15)$$

Finally, if the sample mean of the data set is zero,  $\mathbf{C}$  is exactly  $\mathbf{X} \mathbf{D} \mathbf{X}^T$ . Such examples show the key role of selecting  $S_{kj}$  during the embedding.

During the experiments, the following settings are used. A bootstrap database,  $\{\mathbf{x}_{ik}\}$ , is collected for the concerned variation. Each identity  $i$  has several images corresponding to different values of the variation. The distances between images are calculated in a supervised manner as in LDA. The similarity indices in (2.10) are determined based on variation labels. In other words, instead of considering local neighborhoods (the parameter  $\delta$ ), the coefficients  $S_{kj}$  becomes 0 if two data points do not have the same variation. For data points with the same variation, coefficients are calculated by the heat kernel. Details can be gathered from [10, 32].

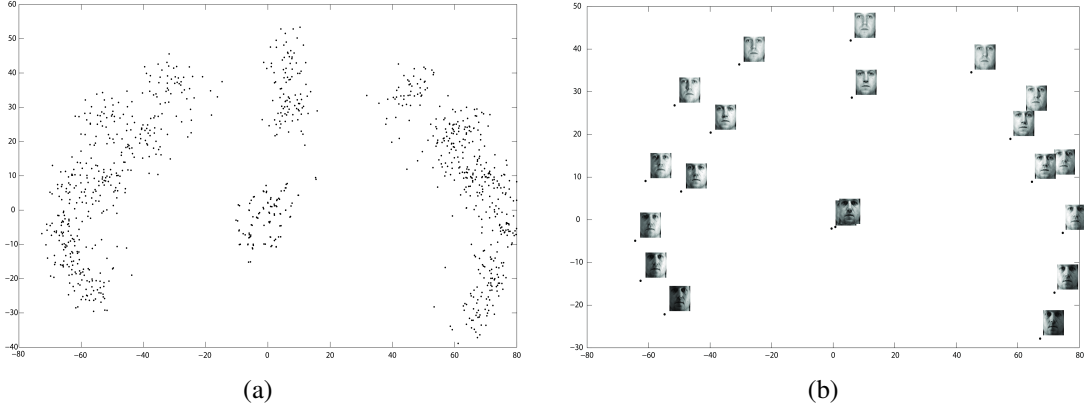
Using such a supervised approach draws an upper bound to the dimensionality of the manifold. Since the rank of the generalized eigenvalue problem in (2.12) is determined by the number of discretized variation labels (different types of illumination), the dimensionality is at most the number of different variation labels in the bootstrap database.

An example embedding of the bootstrap database into two dimensional subspace is illustrated in Figure 2.2(a). A further averaging step is performed to discard the effect of the identity completely. As shown in Figure 2.2(b), averages over identities are calculated to represent each variation type.

The averaging is applied as follows: For each observation,  $\mathbf{x}_{ik}$ , the reduced dimensional coordinates,  $\mathbf{c}_{ik}$ , are calculated by  $\mathbf{c}_{ik} = \mathbf{M}^T \mathbf{x}_{ik}$ . Then, for each variation label,  $k$ , the average over all identities is taken by

$$\mathbf{c}_k = \frac{1}{N} \sum_{i=1}^N \mathbf{c}_{ik}, \quad (2.16)$$

where  $N$  is the total number of identities in the bootstrap database.



**Figure 2.2:** Embedding results of LPP: (a) 2D embedding of the bootstrap database with changing illumination. (b) Average coordinates corresponding to different illumination conditions. These coordinates are invariant to the identity.

### 2.3.2 Bootstrap: an algebraic approach

In the bootstrap phase, the parameters  $\mu, \Omega^{-1}, \sigma_k^2$ , which define the distributions  $p(\mathbf{w})$  and  $p(x_{ik}|\mathbf{w}_i, \mathbf{c}_k)$ , are calculated. As a first attempt, the distributions are defined empirically *i.e.* the basis vectors,  $\mathbf{w}$ , are found for different identities in the bootstrap database, and then parameters over them are calculated. Such an approach seems to be non-globally-optimal; however, by performing some regularizations, it is believed to reach an appropriate solution which agrees with the assumptions on the distributions.

Here, the governing equation defined in (1.3) is taken into account. Now we consider factor loadings,  $\mathbf{W}$ , as a basis set of the variation subspace. Similarly, factors,  $\mathbf{c}_k$ , are assumed to be coordinates *i.e.* linear combination coefficients. If both parameters (basis vectors and their coefficients) are treated as unknowns to be optimized, then it is not possible to guarantee that the basis sets of different identities have similar characteristics. If the basis sets of different identities are not forced to generate a certain geometry for their own subspaces, they only adapt themselves to the observations present in the bootstrap database. This fact illustrated in Figure 2.3. When compared to the results in Figure 2.3 (a), it is easy to say that one may not define proper distributions on the basis vectors of Figure 2.3 (b) since they do not have compatible characteristics among themselves, unlike the ones in Figure 2.3 (a).

To this extend, the combination coefficients are kept fixed among different identities to have a common geometry for different manifolds. Hence, another challenge is to



(a)



(b)

**Figure 2.3:** Basis sets of different identities with (a) a constraint over the combination coefficients, (b) no constraint over the combination coefficients.

be faced: a coefficient set that represents the geometry of the underlying manifold most accurately is required. That is accomplished by the manifold learning step that is detailed in Section 2.3.1.

The problem can be described as a high dimensional reconstruction error minimization. The minimization is run for each identity in the bootstrap database separately to obtain its basis set,  $\mathbf{W}$ . Finally, measurements are performed over these basis sets to calculate the required statistics.

Let's assume that we have  $K$  images of an identity  $i$  in the bootstrap database. Then the total reconstruction error for the identity  $i$  is

$$\begin{aligned} \mathcal{E}_i &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{W}_i \mathbf{c}_k\|^2 \\ &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{w}_{i1}c_{k1} - \mathbf{w}_{i2}c_{k2} - \cdots - \mathbf{w}_{in}c_{kn}\|^2, \end{aligned} \quad (2.17)$$

where  $\mathbf{w}_{ij}$  indicates  $j^{\text{th}}$  column of the matrix  $\mathbf{W}_i$ , and  $c_{kj}$  is  $j^{\text{th}}$  element of vector  $\mathbf{c}_k$ . The index  $i$  will be omitted in the following equations for the clarity. The manifold dimension,  $n$ , is determined during the manifold embedding. Details on the dimensionality are given in Section 3.1. The combination coefficients,  $\mathbf{c}_k$ , are assumed to be calculated during the manifold embedding.

The unknown basis vectors,  $\mathbf{w}_{ij}$ , can be minimized by taking the derivatives with respect to them, and equating the derivatives to zero. To define orthogonalization constraints over the bases, one may follow an iterative approach: find one basis vector at each step. The framework starts with the 1-dimensional subspace assumption. Then, the total reconstruction error is

$$\begin{aligned}
\mathcal{E} &= \sum_{k=1}^K \|\mathbf{x}_k - c_{k1} \mathbf{w}_1\|^2 \\
&= \sum_{k=1}^K (\mathbf{x}_k - c_{k1} \mathbf{w}_1)^T (\mathbf{x}_k - c_{k1} \mathbf{w}_1) \\
&= -2 \sum_{k=1}^K c_{k1} \mathbf{x}_k^T \mathbf{w}_1 + \mathbf{w}_1^T \mathbf{w}_1 \sum_{k=1}^K c_{k1}^2 + \sum_{k=1}^K \mathbf{x}_k^T \mathbf{x}_k \\
&= -2 \mathbf{c}_1^T \mathbf{X} \mathbf{w}_1 + \mathbf{w}_1^T \mathbf{w}_1 \mathbf{c}_1^T \mathbf{c}_1 + \sum_{k=1}^K \mathbf{x}_k^T \mathbf{x}_k, \tag{2.18}
\end{aligned}$$

where the matrix  $\mathbf{X}$  includes vectors  $\mathbf{x}_k^T$  as its rows, and  $\mathbf{c}_1$  is the vector of the first coordinate terms. The last term,  $\sum_{i=k}^K \mathbf{x}_k^T \mathbf{x}_k$  can be omitted since it does not depend on the optimization variable,  $\mathbf{w}_1$ .

To reduce the condition number related to the problem, usually a normalization constraint as  $\mathbf{w}_1^T \mathbf{w}_1 = 1$  is introduced. However, since a value is already assigned to each  $c_{k1}$ , which plays a scaling role, using such a constraint on the norm of the variable may result in sticking in a local optimum. Taking derivative with respect to  $\mathbf{w}_1$ , and equating it to zero yields

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial \mathbf{w}_1} &= 0 \Rightarrow \\
0 &= -2 \mathbf{X}^T \mathbf{c}_1 + 2 \mathbf{w}_1 \mathbf{c}_1^T \mathbf{c}_1. \tag{2.19}
\end{aligned}$$

Therefore, the first basis vector is

$$\mathbf{w}_1 = \frac{\mathbf{X}^T \mathbf{c}_1}{\mathbf{c}_1^T \mathbf{c}_1}. \tag{2.20}$$

To calculate the second basis vector, a similar minimization formulation with an extra constraint ( $\mathbf{w}_1^T \mathbf{w}_2 = 0$ ) can be used. By finding the minimum point for

$$\mathcal{E} = \sum_{k=1}^K \|\mathbf{x}_k - c_{k1} \mathbf{w}_1 - c_{k2} \mathbf{w}_2\|^2 + \lambda (\mathbf{w}_2^T \mathbf{w}_1), \tag{2.21}$$

the second basis can be found. The error term is now

$$\mathcal{E} = -2 \mathbf{c}_2^T \mathbf{Y} \mathbf{w}_2 + \mathbf{w}_2^T \mathbf{w}_2 \mathbf{c}_2^T \mathbf{c}_2 + 2 \lambda \mathbf{w}_2^T \mathbf{w}_1, \tag{2.22}$$



after omitting constant terms. The matrix  $\mathbf{Y}$  is defined as  $\mathbf{Y} = \mathbf{X} - \mathbf{c}_1 \mathbf{w}_1^T$ . Again the partial derivatives are taken with respect to unknown parameters,  $\mathbf{w}_2$  and  $\lambda$  as

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \mathbf{w}_1} &= 0 \Rightarrow \\ 0 &= -\mathbf{Y}^T \mathbf{c}_2 + 2\mathbf{w}_2 \mathbf{c}_2^T \mathbf{c}_2 + \lambda \mathbf{w}_1, \end{aligned} \quad (2.23)$$

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \lambda} &= 0 \Rightarrow \\ 0 &= \mathbf{w}_2^T \mathbf{w}_1. \end{aligned} \quad (2.24)$$

If the equation (2.23) is multiplied by  $\mathbf{w}_1^T$  to use the identity in (2.24), the value  $\lambda$  is found to be

$$\lambda = \frac{\mathbf{w}_1^T \mathbf{Y}^T \mathbf{c}_2}{\mathbf{w}_1^T \mathbf{w}_1}. \quad (2.25)$$

Then the second basis is

$$\mathbf{w}_2 = \frac{\mathbf{P} \mathbf{Y}^T \mathbf{c}_2}{\mathbf{c}_2^T \mathbf{c}_2}, \quad (2.26)$$

where  $\mathbf{P} = \mathbf{I} - \frac{1}{\mathbf{w}_1^T \mathbf{w}_1} \mathbf{w}_1 \mathbf{w}_1^T$  is a projection matrix that projects into the complementary of the subspace spanned by the first basis vector,  $\mathbf{w}_1$ .

Following the same procedure, the  $n^{\text{th}}$  basis is

$$\mathbf{w}_n = \frac{\mathbf{P} \mathbf{Y}^T \mathbf{c}_n}{\mathbf{c}_n^T \mathbf{c}_n}, \quad (2.27)$$

and similarly  $\mathbf{Y} = \mathbf{X} - \sum_{i=1}^{n-1} \mathbf{c}_i \mathbf{w}_i^T$  and  $\mathbf{P} = \mathbf{I} - \sum_{i=1}^{n-1} \frac{1}{\mathbf{w}_i^T \mathbf{w}_i} \mathbf{w}_i \mathbf{w}_i^T$ .

As the complete basis set  $\mathbf{W}_i$  of each identity  $i$  in the bootstrap database is calculated, the parameters of the distribution  $p(\mathbf{w})$  can be estimated by the empirical formulas

$$\mu = \frac{1}{N} \sum_{i=1}^N \mathbf{w}_i, \quad (2.28)$$

$$\Omega = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{w}_i - \mu)(\mathbf{w}_i - \mu)^T. \quad (2.29)$$

One should be careful with this notation. Here, we turned back to the form defined in (2.4). Therefore, the vector  $\mathbf{w}_i$  is a row (not a column) of the matrix  $\mathbf{W}_i$ , and the averages are taken over identities. After calculating the matrix  $\mathbf{W}_i$  for an identity  $i$ , the parameters corresponding to different rows are determined independently.

The parameters of the error distribution,  $p(\varepsilon_k)$ , can be again estimated by an empirical approach. The error for each identity  $i$  and the variation  $k$  is defined by (2.4) as

$$\varepsilon_{ik} = x_{ik} - \mathbf{w}_i^T \mathbf{c}_k. \quad (2.30)$$

Then, the mean value and the variance is calculated by averaging over identities. Here, the mean value is most probably not zero contrary to the definition in (2.5). One may redefine the distribution  $p(\varepsilon_k)$  to have a nonzero mean value. Nevertheless, in such a case the formulation of MAP estimate should be corrected in Section 2.3.4.

To relax the minimization problem, we included orthogonalization constraints over the basis vectors. Nevertheless, it is possible to get a unique solution without using such constraints provided that there exists a linearly independent set of coordinate vectors,  $\mathbf{c}_k$ , corresponding to different values of the variation. This fact is shown in Appendix A.1.

### 2.3.3 Bootstrap: a probabilistic approach

To provide an improved framework in terms of completeness, a new probabilistic interpretation is presented. Here, the distributions  $p(\mathbf{w})$  and  $p(x_{ik}|\mathbf{w}_i, \mathbf{c}_k)$  are re-defined. In other words, another way to calculate the parameters  $\mu, \Omega^{-1}, \sigma_k^2$  is shown. A way to calculate these distributions were already given in Section 2.3.2. This time, a complete probabilistic approach is utilized.

The parameters of prior distributions defined in (2.5) are calculated again using the bootstrap database,  $\mathbf{X} = \{\mathbf{x}_{ik}\}$ . Considering the element-wise formulation (2.4) and priors, the conditional and the marginal distributions over the variable  $x_k$  are

$$\begin{aligned} p(x_k|\mathbf{w}, \mathbf{c}_k) &\sim \mathcal{G}(\mathbf{w}^T \mathbf{c}_k, \sigma_k^2), \\ p(x_k) &= \int p(x_k|\mathbf{w}, \mathbf{c}_k) p(\mathbf{w}) d\mathbf{w}. \end{aligned} \quad (2.31)$$

Both the prior and the conditional distributions are Gaussians in (2.31), and this makes the resulting marginal distribution,  $p(x_k)$ , to be another Gaussian. Indeed, this integral form does not have to be solved analytically since the mean value and the variance can be easily evaluated by employing the equation (2.4) as in

$$\begin{aligned} E[x_k] &= E[\mathbf{w}^T \mathbf{c}_k + \varepsilon_k] = E[\mathbf{w}^T] \mathbf{c}_k + E[\varepsilon_k] = \mu^T \mathbf{c}_k, \\ E[(x_k - E[x_k])^2] &= E[(\mathbf{w}^T \mathbf{c}_k + \varepsilon - \mu^T \mathbf{c}_k)^2] = E[((\mathbf{w}^T - \mu^T) \mathbf{c}_k)^2 + \varepsilon] \\ &= \mathbf{c}_k^T E[(\mathbf{w} - \mu)(\mathbf{w} - \mu)^T] \mathbf{c}_k + 2E[\varepsilon(\mathbf{w} - \mu)^T] \mathbf{c}_k + E[\varepsilon^2] \\ &= \mathbf{c}_k^T \Omega \mathbf{c}_k + \sigma_k^2, \end{aligned} \quad (2.32)$$

where we assumed zero correlation between the error term  $\varepsilon$  and the basis vector  $\mathbf{w}$ . The mean and the variance parameters are sufficient to define the marginal as

$$p(x_k) \sim \mathcal{G}(\boldsymbol{\mu}^T \mathbf{c}_k, \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2). \quad (2.33)$$

The bootstrap database can be used at this point to calculate the unknown parameters,  $\boldsymbol{\Omega}$ ,  $\boldsymbol{\mu}$ , and  $\sigma_k^2$  by maximizing the likelihoods. The Likelihood to be maximized is the empirical likelihood of the observed points,  $x_{ik}$ . Assuming i.i.d observations, the total log likelihood over observations is

$$\ln p(X|\boldsymbol{\mu}, \boldsymbol{\Omega}, \sigma_k^2) = \sum_i^N \sum_k^K \ln p(x_{ik}), \quad (2.34)$$

where the upper bounds  $N$  and  $K$  denote the number of identities and different values of the variation in the bootstrap gallery, respectively. After omitting the constant terms which are not related to the unknown parameters, the cost functional becomes

$$\mathcal{E} = - \sum_i^N \sum_k^K \ln (\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2) - \sum_i^N \sum_k^K \frac{(x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2}. \quad (2.35)$$

In order to determine the unknown parameters  $\boldsymbol{\mu}, \boldsymbol{\Omega}, \sigma_k^2$  which minimize the cost functional, the partial derivatives are to be taken with respect to those and set them equal to zero. By this way, equations

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \sigma_k^2} &= 0 \Rightarrow \\ \frac{N}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} &= \frac{\sum_i^N (x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2}{(\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2)^2} \Rightarrow \\ N (\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2) &= \sum_i^N (x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2 \Rightarrow \\ \sigma_k^2 &= \frac{1}{N} \sum_i^N (x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2 - \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k, \end{aligned} \quad (2.36)$$

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \boldsymbol{\Omega}} &= 0 \Rightarrow \\ N \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} &= \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T}{(\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2)^2} \sum_i^N (x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2, \end{aligned} \quad (2.37)$$

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial \boldsymbol{\mu}} &= 0 \Rightarrow \\
0 &= \sum_i^N \sum_k^K \frac{\mathbf{c}_k (x_{ik} - \mathbf{c}_k^T \boldsymbol{\mu})}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} \Rightarrow \\
\left( N \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T \boldsymbol{\mu}}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} \right) &= \frac{1}{N} \sum_i^N \sum_k^K \frac{x_{ik} \mathbf{c}_k}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} \Rightarrow \\
\left( \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} \right) \boldsymbol{\mu} &= \frac{1}{N} \sum_i^N \sum_k^K \frac{x_{ik} \mathbf{c}_k}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2}, \tag{2.38}
\end{aligned}$$

define a system of nonlinear equations. The solution for (2.36) is also a solution for (2.37), thus the system is rank deficient. It has infinite solutions; therefore, we can not assume any optimality. To overcome this problem, one may calculate the covariance matrix as defined in Section 2.3.2. We expect that such an empirical solution leads us to an optimal solution. The experiments on changing illumination conditions, facial expressions, and poses indicate that this assumption holds for real life scenarios. Thus, two usefull equations are

$$\sigma_k^2 = \frac{1}{N} \sum_i^N (x_{ik} - \boldsymbol{\mu}^T \mathbf{c}_k)^2 - \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k, \tag{2.39}$$

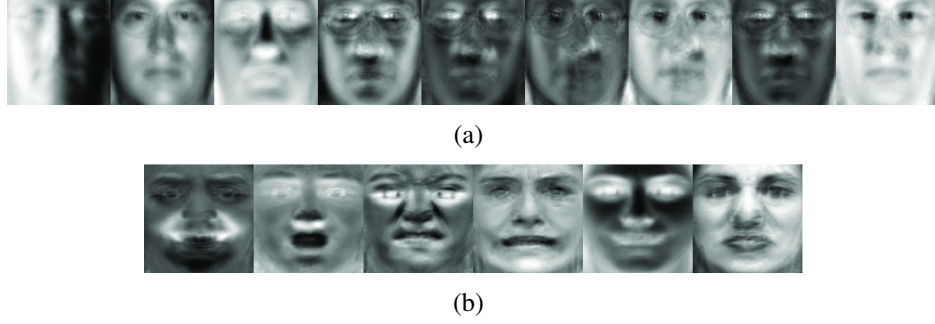
$$\left( \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2} \right) \boldsymbol{\mu} = \frac{1}{N} \sum_i^N \sum_k^K \frac{x_{ik} \mathbf{c}_k}{\mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2}. \tag{2.40}$$

Analytic solution to these nonlinear equations is not trivial. Thus, a fixed point iteration is employed to approximate the solution. Let  $\zeta_k = \mathbf{c}_k^T \boldsymbol{\Omega} \mathbf{c}_k + \sigma_k^2$  and  $a(t)$  indicates the value of the variable  $a$  at  $t^{th}$  iteration step. Then, at each iteration step, the equations

$$\left( \sum_k^K \frac{\mathbf{c}_k \mathbf{c}_k^T}{\zeta_k(t)} \right) \boldsymbol{\mu}(t) = \frac{1}{N} \sum_i^N \sum_k^K \frac{x_{ik} \mathbf{c}_k}{\zeta_k(t)}, \tag{2.41}$$

$$\zeta_k(t+1) = \frac{1}{N} \sum_i^N (x_{ik} - \boldsymbol{\mu}(t)^T \mathbf{c}_k)^2, \tag{2.42}$$

are solved. With an appropriate initial guess, this procedure converges fast. Two example solutions for  $\boldsymbol{\mu}$  corresponding to different variation types are illustrated in Figure 2.4. For all experiments, the initialization is done as  $\zeta_k(1) = 1$ , and the stopping criteria is determined as  $|\zeta_k(t+1) - \zeta_k(t)| \leq 10^{-6}$ . These calculations must be repeated for each pixel location as the element-wise formulation is employed.



**Figure 2.4:** Mean parameter,  $\mu$ , is illustrated for two different variation types: (a) for illumination and (b) for expression.

### 2.3.4 Training: recovering class factors

Having the conditional probability  $p(x_{gk}|\mathbf{w}_g, \mathbf{c}_k)$  and the prior probability  $p(\mathbf{w}_g)$  defined in the bootstrap, the MAP estimation can be applied to recover the factor loadings of a gallery identity  $g$ , given an observation  $x_{gk}$  by

$$\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}_g} p(\mathbf{w}_g|x_{gk}, \mathbf{c}_k).$$

Using Bayes' rule we get

$$\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}_g} p(x_{gk}|\mathbf{w}_g, \mathbf{c}_k) \times p(\mathbf{w}_g),$$

where the constant term  $p(x_{gk})$  is omitted. Then, MAP estimate is

$$\mathbf{w}_{MAP} = \arg \max_{\mathbf{w}_g} \mathcal{G}(\mathbf{w}_g^T \mathbf{c}_k, \sigma_k^2) \times \mathcal{G}(\mu, \Omega^{-1}). \quad (2.43)$$

If the log propabilities are considered and the constant terms are ommitted, the cost functional equals to

$$\mathcal{E} = \left( -\frac{1}{2} \left( \frac{x_{gk} - \mathbf{w}_g^T \mathbf{c}_k}{\sigma_k^2} \right)^2 - \frac{1}{2} (\mathbf{w}_g - \mu)^T \Omega^{-1} (\mathbf{w}_g - \mu) \right). \quad (2.44)$$

Then, the derivative with respect to  $\mathbf{w}_g$  is taken as

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \mathbf{w}_g} &= 0 \Rightarrow \\ \frac{2}{\sigma_k^2} (x_{gk} - \mathbf{w}_g^T \mathbf{c}_k) \mathbf{c}_k &= 2\Omega^{-1} (\mathbf{w}_g - \mu). \end{aligned} \quad (2.45)$$

The MAP estimate for  $\mathbf{w}_g$  is the solution to the set of linear equations [16]

$$\mathbf{A}\mathbf{w}_g = \mathbf{b}, \quad (2.46)$$

where

$$\mathbf{A} = \frac{1}{\sigma_k^2} \mathbf{c}_k \mathbf{c}_k^T + \Omega^{-1}, \quad \mathbf{b} = \frac{x_{gk}}{\sigma_k^2} \mathbf{c}_k + \Omega^{-1} \mu. \quad (2.47)$$

When the algebraic approach defined in Section 2.3.2 is employed to estimate bootstrap parameters, the error term has a nonzero mean value. Therefore, the conditional distribution  $p(x_{gk} | \mathbf{w}_g, \mathbf{c}_k)$  becomes an Gaussian defined by

$$\mathcal{G}(\mathbf{w}_g^T \mathbf{c}_k + \mu_{\varepsilon_k}, \sigma_k^2), \quad (2.48)$$

where  $\mu_{\varepsilon_k}$  is the mean value of the error term  $\varepsilon_k$ . Then, the vector  $\mathbf{b}$  in the final solution of the vector  $\mathbf{w}_g$  is

$$\mathbf{b} = \frac{x_{gk} - \mu_{\varepsilon_k}}{\sigma_k^2} \mathbf{c}_k + \Omega^{-1} \mu. \quad (2.49)$$

In these formulations, a single observation is enough for each class, while having more points will increase the reliability of the recovery. When multiple observations exist for an identity  $g$ , the coefficient matrix and the right-hand side vector are determined by summations over observations as in

$$\mathbf{A} = \sum_k \frac{1}{\sigma_k^2} \mathbf{c}_k \mathbf{c}_k^T + \Omega^{-1}, \quad \mathbf{b} = \sum_k \frac{x_{gk}}{\sigma_k^2} \mathbf{c}_k + \Omega^{-1} \mu. \quad (2.50)$$

Factors,  $\mathbf{c}_k$ , are assumed to be calculated by the mapping  $\mathbf{M}$  of LPP. First, the identity dependent factors  $\mathbf{c}_{gk}$  are calculated by

$$\mathbf{c}_{gk} = \mathbf{M}^T \mathbf{x}_{gk}. \quad (2.51)$$

Then, the identity invariant factors are obtained by finding the closest (in terms of Euclidean distance)  $\mathbf{c}_k$  that is calculated by (2.16) during the bootstrap. Instead, one may take the average of  $k$  nearest  $\mathbf{c}_k$  to increase the ability of handling novel values. During our tests, we took the average of 3 nearest  $\mathbf{c}_k$ .

Equations in (2.50) include weighted averages over gallery observations. Therefore, when there exist multiple observations, the MAP estimate finds a basis set that is the most representative of an average observation. Such an approach may fail in case of pose variation because images with different viewpoints are not suitable to take averages. It is possible to register feature points for images of different illumination or facial expressions (if no severe expression); however, that is not the case with the pose

variation. Therefore, the average observation can not represent the multiple samples with different poses.

Considering the fact with the pose variation, we develop another recovery approach that does not find a basis set which is a good representative for averages but can synthesize each observation individually. Let's assume we have  $K$  different images of an identity  $g$ . Then, the main formulation

$$\mathbf{w}_g = \arg \max_{\mathbf{w}} -\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1}(\mathbf{w} - \boldsymbol{\mu}) + \sum_k^K \lambda_k (x_{gk} - \mathbf{w}^T \mathbf{c}_k), \quad (2.52)$$

is an optimization scheme with several constraints where the constraints are introduced as  $x_{gk} = \mathbf{w}^T \mathbf{c}_k$ . The first term is the Mahalanobis distance related to the prior,  $p(\mathbf{w})$ . The coefficients  $\lambda_k$  are the unknown Lagrange multipliers. A solution to this new approach is derived in Appendix A.2.

### 2.3.5 Testing: classification of novel points

Given a novel observation  $\mathbf{x}_{pk}$ , the class label can be determined by assigning the class with the maximum likelihood  $p(\mathbf{x}_{pk} | \mathbf{W}_g, \mathbf{c}_k)$ .

Another approach which is used during our experiments is to minimize the distance between the novel point and its synthesized counterparts (distance to manifold) as

$$d_g = \|\mathbf{x}_{pk} - \mathbf{x}_{gk}\|, \quad (2.53)$$

where  $\mathbf{x}_{gk} = \mathbf{W}_g \mathbf{W}_g^T \mathbf{x}_{pk}$  is calculated for each gallery identity,  $g$ .

As a third choice, posterior probabilities may be used to decide the identity of the novel point. The decision is made by selecting the maximum posterior  $p(\mathbf{W}_g | \mathbf{x}_{pk}, \mathbf{c}_k)$ . Bayes' rule transforms the posterior into the multiplication of the likelihood and the prior:  $p(\mathbf{W}_g | \mathbf{x}_{pk}, \mathbf{c}_k) = p(\mathbf{x}_{pk} | \mathbf{W}_g, \mathbf{c}_k) \cdot p(\mathbf{W}_g)$  (the constant denominator  $p(\mathbf{x}_{pk})$  is omitted). This approach can be very useful in large scale real life scenarios as it lets us to employ priors over gallery identities.

The second approach is employed for all of our experiments. For this approach, the orthonormal  $\mathbf{W}_g$  matrices are considered, whereas no such constraint was employed during the recovery. In the experiments, Gram-Schmidt orthonormalization process is

applied after solving (2.46). The detailed algorithm of the CDFA is given in Table 2.2.

**Table 2.2:** Detailed algorithm of the CDFA.

---

*Bootstrap:* Given a bootstrap database,  $\mathbf{X} = \{\mathbf{x}_{ik}\}$

- Calculate the lower dimensional coordinates,  $\mathbf{c}_k$ , by (2.16)
- For each pixel location
  - \* Calculate the empirical covariance matrix,  $\Omega$ , as defined in Section (2.3.2)
  - \* Calculate  $\mu$  and  $\sigma_k^2$  using (2.41) and (2.42)

*Training:* Given gallery observations,  $\mathbf{G} = \{\mathbf{x}_{gk}\}$ , for each identity  $g$ ,

- Calculate the lower dimensional coordinates,  $\mathbf{c}_k$ , by (2.51)
- Recover  $\mathbf{w}_g$  for each pixel location by (2.46)
- Construct the matrix  $\mathbf{W}_g$  so that it has vectors  $\mathbf{w}_g$  as its rows
- Apply Gram-Schmidt orthonormalization to the columns of  $\mathbf{W}_g$

*Testing:* Given a probe observation,  $\mathbf{x}_{pk}$ ,

- Calculate  $d_g$  for each gallery identity  $g$  using (2.53)
  - Select the identity with the minimum distance
- 

## 2.4 Interpretation of Governing Distributions

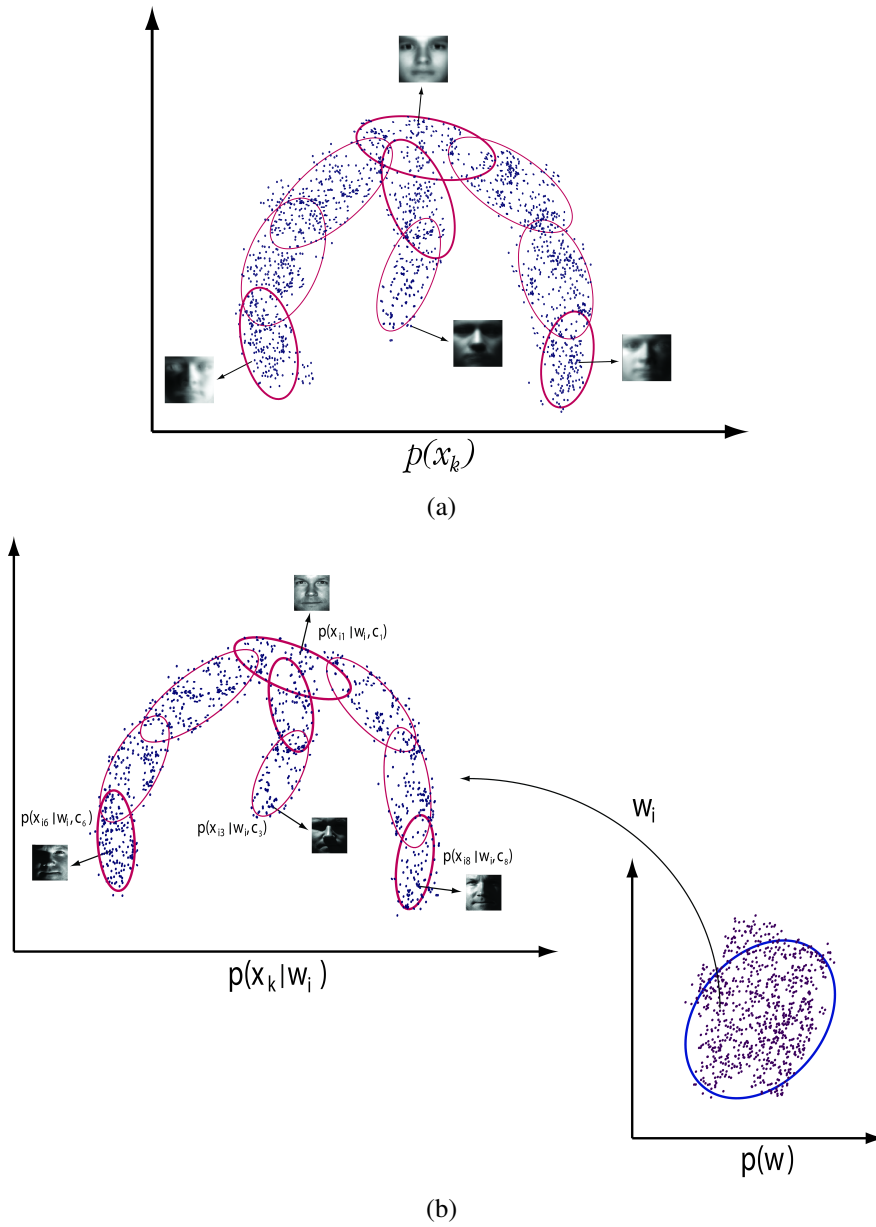
Beside the geometrical interpretation of the generative model described in Section 1.3, another probabilistic interpretation is given here, regarding the formulation of the CDFA framework. The marginal distribution  $p(x_k)$  specifies a mixture of Gaussians in which Gaussians are determined by the variation label  $k$ . Each Gaussian is characterized by parameters  $\mu^T \mathbf{c}_k$  and  $\mathbf{c}_k^T \Omega \mathbf{c}_k + \sigma_k^2$ . Hence, the variation defines the shape of each Gaussian.

Initially, the geometry of the manifold consisting of this mixture does not depend on the identities, but only on the mean identity. Thus, the manifold can be considered



as a template that will be customized after selection of an identity. When an identity is drawn from the prior distribution  $p(\mathbf{w})$ , it re-defines the mixture by the conditional distributions  $p(x_k|\mathbf{w}, \mathbf{c}_k)$ . This procedure also eliminates a considerable amount of uncertainty in each Gaussian as the variance decreases to  $\sigma^2$  from  $\mathbf{c}_k^T \mathbf{\Omega} \mathbf{c}_k + \sigma_k^2$ . Whole process is illustrated in Figure 2.5.

CDFA is defined as a two-layer decision process. At the first layer, class identities are drawn from a prior distribution. The second layer defines a mixture of Gaussians depending on a template manifold characterized by  $p(x_k)$ , and the conditional distributions  $p(x_k|\mathbf{w}, \mathbf{c}_k)$ . The assignment of observations to each Gaussian is achieved by the manifold embedding. In this view, the manifold embedding can be considered as a clustering scheme.



**Figure 2.5:** Illustration of the governing distributions: (a) A template manifold is defined by the marginal distribution,  $p(x_k)$ . (b) This template is customized by the identity drawn from the prior distribution,  $p(w)$ .

### 3. EXPERIMENTAL EVALUATIONS

Several experiments were conducted to explore two important aspects of the CDFA framework: (1) the recognition performance against extreme variations and (2) scalability in relatively large databases. For the first evaluation, we selected databases with extreme variations. Nevertheless, the sizes of such databases are usually small, including at most 30-40 identities. To analyze the real life performance of the method, a second group of experiments was performed on another set of databases with moderate variations but large number of identities.

The main characteristic of the method is its ability to be used for different types of variation. This claim was verified by different experiments under different types of variation. Three types of variation were used during tests: (1) changing illumination, (2) changing facial expressions, and (3) changing viewpoints.

#### 3.1 Tuning the Bootstrap Parameters

Each test begins with the manifold embedding on the selected bootstrap database to decide the geometrical features of the manifold. One parameter that should be determined is the dimension of the underlying manifold. The manifold learning technique LPP relies on the solution of a generalized eigenvalue problem; therefore, the spectrum of eigenvalues may help with determining the dimension. However, using an evaluation dataset is a better choice since the characteristics of the variation may prevent a meaningful spectrum analysis.

As indicated in Section 2.3.1, the intrinsic dimensionality is bounded by the number of different variation labels present in the bootstrap database. For instance, when using Multi-PIE [41] as the bootstrap database, the dimensionality is bounded by 20 since there are 20 different illumination conditions. However, this does not mean that the recognition is performed in a 20 dimensional subspace. This value represents the number of basis vectors to span the variation subspace of each identity. It is only

related to the range of the generative model, *i.e.* how the method deals with novel variations. The recognition is performed by the point-to-manifold measure which is calculated in the original observation space  $\mathfrak{R}^n$ , where  $n$  is the number of pixels of the input images.

Certain properties of the manifolds like dimensionality are totally determined by the bootstrap database. This is a clear and an understandable behavior since the bootstrap database reflects the way that the operative variation is modeled. The best practice is to use a bootstrap database that is the most compatible with the testing requirements.

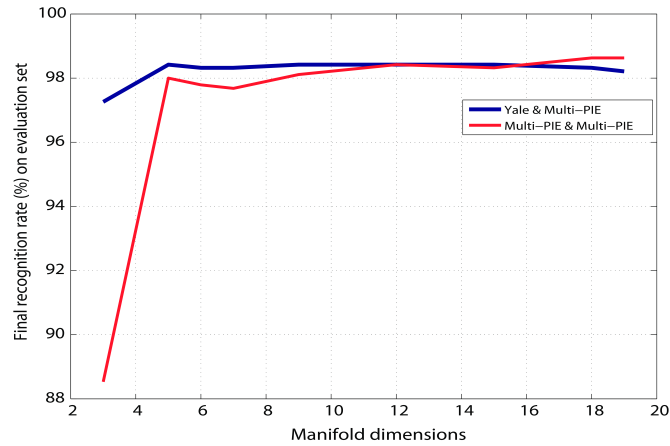
The effect of the manifold dimension is given in Figure 3.1. For two types of variation (illumination and facial expression), evaluation datasets were collected. Scenarios with different bootstrap and evaluation sets are demonstrated to grasp the characteristics completely. All tests were performed with evaluation sets containing 50 identities. A single image was selected as the gallery and all remaining images were used as probes. Those identities collected for the evaluation sets were not used during the further experiments to reflect a real life behavior.

Experiments indicate that the method behaves similarly in terms of dimensionality even if the bootstrap database is changed. The results are comparable when the dimension is fixed among different evaluation sets. Moreover, slight changes in dimension do not affect the recognition performance, considerably.

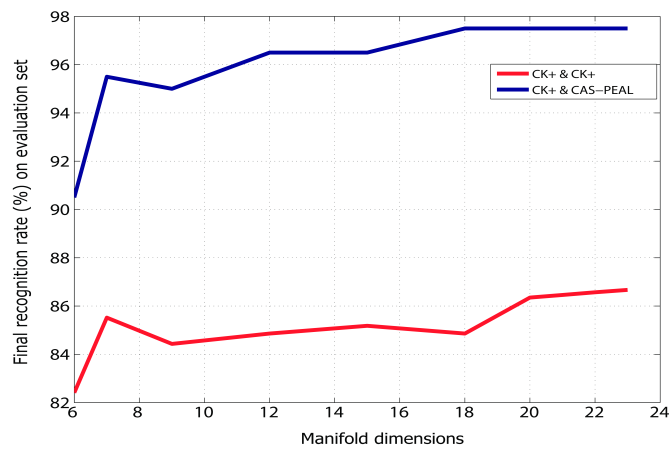
### **3.2 Classification Performance against Illumination**

Tests with changing illumination conditions were performed with Yale B Database [19]. This database includes 10 identities with 45 different illumination conditions. The database can be split into 4 subsets according to the illumination direction, which also highlights the difficulty of the recognition. Figure 3.2 shows several example images from each subset.

The Extended Yale Face Database [42] was used as the bootstrap database. This database is an extension of the original Yale B with 28 identities which are not present in the original database. At the bootstrap phase, a subset of 41 illumination types out

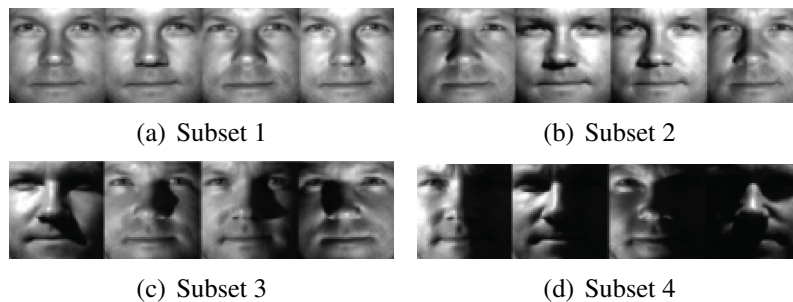


(a)



(b)

**Figure 3.1:** Recognition rates on evaluation sets with different manifold dimensions under (a) illumination and (b) facial expression changes. Yale & Multi-PIE means that the bootstrap set is from Yale and the evaluation set is from Multi-PIE.



**Figure 3.2:** Some example images of Yale B Database

of 45 was used due to several corrupted images. Hence, the gallery and probe images had novel variations which were not present in the bootstrap database.

The size of images used in the experiments was  $100 \times 90$ . As a preprocessing step, all images were normalized so that they have zero mean and unit variance. The dimension

of the manifold was fixed as 9. Totally, 19 tests were performed, and the average was taken as the final performance. For each test, a single image from subset 1 or subset 2 was selected as the gallery image, and all remaining images were used as probes. In other words, 440 recognition attempts were performed for each test, resulting in 8360 recognition attempts in total. The recognition rates with this configuration are given in Table 3.1. Recognition rates are very promising considering the moderate

**Table 3.1:** Face recognition rates for Yale B Database. Performances of the other methods were taken from [16].

Methods	Subset 1-2	Subset 3	Subset 4
Correlation	100	76.7	26.4
Eigenfaces	100	74.2	24.3
Linear Subspaces	100	100	85
Cones-attached	100	100	91.4
Cones-cast	100	100	100
9PL	100	100	97.2
Spherical Harmonics	100	99.7	96.9
C DFA	100	99.2	95

requirements for the bootstrap and the training. CDFA is trained by a single image for each identity unlike methods Cones-attached, Cones-cast, and 9PL which need number of images between 5 and 9. Compared to the spherical harmonics, CDFA is a more generic approach since it is not related to the physical aspects of the variation. The behavior of the CDFA with increasing number of gallery images is demonstrated in Table 3.2. Random images from subsets 1 and 2 are selected as gallery images for each test. The increase in the recognition performance makes the proposed method more comparable to other methods.

### 3.3 Classification Performance against Facial Expressions

As a second set of experiments, the performance of the CDFA with facial expressions was analyzed. For this purpose, three databases were selected: Cohn-Kanade AU-Coded facial expression database (CK+) [43], Japanese female facial expression database (JAFFE) [44], and CMU AMP face expression database [45]. Several example images of these databases are given in Figure 3.3.

**Table 3.2:** Recognition error rates for Yale B Database with multiple gallery images.

# images	Subset 1-2	Subset 3	Subset4
1	0.0	0.8	5.0
2	0.0	0.2	1.4
3	0.0	0.1	0.6
4	0.0	0.0	0.3
5	0.0	0.0	0.1
6	0.0	0.0	0.0



(a) CMU AMP



(b) CK+



(c) JAFFE

**Figure 3.3:** Several images from expression databases.

CK+ database is a collection of video sequences starting with a neutral pose and ending with a peak expression. 7 universal expressions [46] are included in the database: (1) surprise, (2) anger, (3) happiness, (4) sadness, (5) disgust, (6) contempt, and (7) fear. The contempt expression was discarded since only a few identities had this expression. This database is used as a common bootstrap gallery. 4 images were sampled from each sequence. Including one additional neutral image, at most 25 different images per person (24 images corresponding to 6 expression and 1 neutral image) were collected. The manifold dimension was determined to be 20.

Two groups of tests were performed using databases JAFFE and CMU AMP. JAFFE includes 213 images of 10 Japanese women with number of facial expressions varying between 20 and 23. These expressions can be different from the expressions which exist in the bootstrap database. Therefore, the ability of the method with handling novel variations was verified. CMU AMP have 13 identities with 75 different expressions.

Expressions present in this database are extremely severe as they also cause slight pose changes along with changes in face geometries.

CDFA is compared against two state-of-art techniques CS [30] and SRC [29]. To make fair comparisons, we followed the same scenarios with the compared methods, and the gallery selection procedure and the structure of random tests were kept same. Image size was set to be  $32 \times 32$  since the compared methods had selected to use such a small image size. For each identity, several gallery images were selected randomly, and the remaining images were used as probes. Images were used after *zero mean-unit variance* normalization. Results of two classical subspace techniques, PCA and LPP, are also analyzed to understand the marginal improvements. The transformation matrices for PCA and LPP are obtained using the CK+ bootstrap database. LPP is trained in kNN mode with distances being calculated by the heat kernel. Table 3.3 and Table 3.4 show test results for JAFFE and CMU AMP. Results for CS and SRC were taken from [30]. To give an impression of the significance of the presented results, the second columns list the number of actual recognition attempts for each experiment. These values are simply calculated as (the number of test images  $\times$  the number of random trials).

CDFA steadily outperforms others for both databases. However, the main intention here is to highlight that the same framework can be utilized for different types of variation without any modification in the base configuration. Indeed, these databases happen to be trivial although they include severe variations. Even a classical approach like PCA can achieve high recognition rates on them.

**Table 3.3:** Average face recognition rates on JAFFE database. 40 trials with randomly chosen gallery images were performed for each row.

# Gallery Images	Recognition Attempts	CDFA	CS	SRC	PCA	LPP
2	7720	93.04	89.94	90.1	85.84	83.84
3	7320	94.50	93.22	92.1	89.1	89.32
4	6920	96.17	95.12	95.13	91.62	91.33
5	6520	96.33	96.12	96.01	93.54	93.87



**Table 3.4:** Average face recognition rates on CMU AMP database. 10 trials with randomly chosen gallery images were performed for each row.

# Gallery Images	Recognition Attempts	CDFA	CS	SRC	PCA	LPP
4	9230	99.92	98.95	98.9	99.6	99.91
5	9100	100	99.91	99.8	99.66	99.71
6	8970	99.99	99.97	99.75	99.68	99.84
7	8840	100	100	99.74	99.71	99.75
8	8710	100	100	99.87	99.89	99.87
9	8580	100	100	100	99.94	99.97
10	8450	100	100	99.49	99.85	99.95

### 3.4 Classification Performance against Pose

Experiments are performed using Multi-PIE database [41]. Example images are illustrated in Figure 3.4. Pose variations result in highly nonlinear geometries. Without using a preprocessing or keypoint identification step, a single image can not be employed effectively during the training to perform recognition with both  $0^\circ$  and  $\pm 90^\circ$  probe images. Therefore, several gallery images of an identity are required during the basis recovery.











**Figure 3.4:** Example face images under different view point conditions.







In Table 3.5, the recognition performances with single gallery images are given. One pose is picked for training, and all remainings are used for testing. The reliability of the method fails with severe pose differences.

This behavior can be improved by using multiple gallery images. To demonstrate this behavior, several images which can model the variation in an appropriate manner were sampled. Recognition rates with different number of gallery images are given in Table 3.6. It was examined that the number of gallery images is not the main parameter determining the final rates. The ability of the gallery set to represent the whole population is a more important factor. This fact requires us to select gallery poses with a high range. The recognition rates in Table 3.6 are not reliable enough although

**Table 3.5:** Initial face recognition rates (%) with changing poses. A single image is selected as a gallery image and recognition rates for  $\pm 22.5^\circ$ ,  $\pm 67.5^\circ$  and  $\pm 90^\circ$  are given. Tests are performed with 50 identities in the gallery.

Gallery Image / Probe Images	$\pm 22.5^\circ$ 	$\pm 67.5^\circ$ 	$\pm 90^\circ$ 
	50.0	10.0	10.0
	52.0	51.0	28.0
	51.0	55.0	24.0
	23.0	43.0	54.0
	17.0	55.0	38.0

**Table 3.6:** Initial face recognition rates (%) with changing poses. Multiple images are selected as gallery images and recognition rates for  $\pm 22.5^\circ$ ,  $\pm 67.5^\circ$  and  $\pm 90^\circ$  are given. Tests are performed with 50 identities in the gallery.


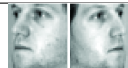




Gallery Image / Probe Images	$\pm 22.5^\circ$ 	$\pm 67.5^\circ$ 	$\pm 90^\circ$ 
	80.0	83.0	31.0
	28.0	78.0	83.0
	74.0	90.0	77.0

several images were used. The main reason is the quality of the basis recovery. Equations in (2.50) include weighted averages over gallery samples. This average observation can not represent the multiple samples since the appearance differences among different poses are critical.

Considering the fact with the pose variation, we develop another recovery approach that is an alternative to the MAP estimate. The approach is designed over the optimization problem defined in 2.52. This new scheme can be employed during basis

recovery when multiple observations per subjects are present. The final recognition rates using the new approach are given in Table 3.7. The new method has a better representation power compared to the previous MAP estimate.

**Table 3.7:** New results with the proposed basis recovery scheme. Multiple images are selected as gallery images and recognition rates for  $\pm 22.5^\circ$ ,  $\pm 67.5^\circ$  and  $\pm 90^\circ$  are given. Tests are performed with 50 identities in the gallery.

Gallery Image / Probe Images	$\pm 22.5^\circ$ 	$\pm 67.5^\circ$ 	$\pm 90^\circ$ 
	81.0	92.0	45.0
	28.0	79.0	82.0
	89.0	94.0	89.0

### 3.5 Scalability

Further experiments were performed to examine the scalability of the proposed method. Two relatively large databases were selected for the testing: CMU Multi-PIE Database [41] and CAS-PEAL Database [47]. Subsets of databases consisting of 249 identities for Multi-PIE and 267 identities for CAS-PEAL were collected. CAS-PEAL was used for the evaluation against facial expressions and Multi-PIE for the illumination. Multi-PIE includes 20 different illumination conditions, and CAS-PEAL serves 5 facial expressions for each identity. Several example images are shown in Figure 3.5.

The behavior of a classical subspace method against the increasing number of gallery identities is demonstrated in Figure 3.6 (a). LDA against illumination was used for the demonstration. All tests were performed on Multi-PIE with 2 random images of each identity being selected as the gallery and the remaining 18 images as probes.



(a) CAS-PEAL



(b) CMU Multi-PIE

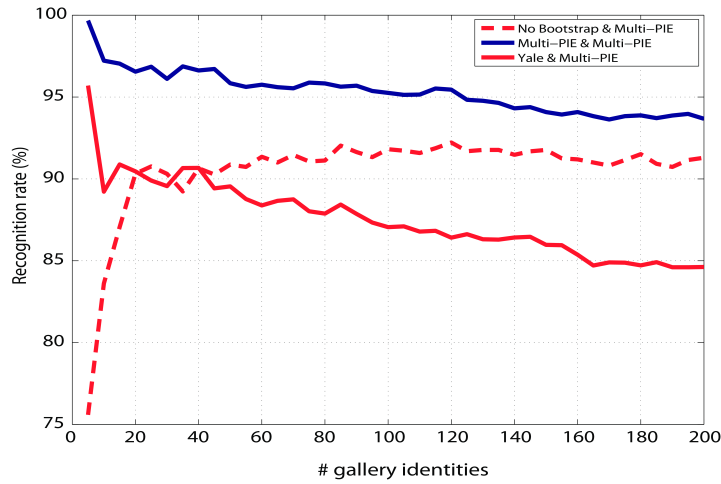
**Figure 3.5:** Several images from CAS-PEAL and CMU Multi-PIE databases.

LDA can perform steadily in terms of recognition rate with its usual configuration. The subspace is re-constructed with each new identity, and the subspace dimension becomes  $(n_i - 1)$  where  $n_i$  is the number identities. However, as new identities are introduced, LDA needs to be re-trained to get a promising recognition rate. This behavior is illustrated in Figure 3.6 (a) with "No bootstrap" label. One may eliminate such a training requirement by using a bootstrap database. In this new setting, the subspace is constructed only once by using the bootstrap database, yet the recognition rate decreases as the number of gallery identities is increased. Moreover, different bootstrap databases may result in significantly different recognition rates.

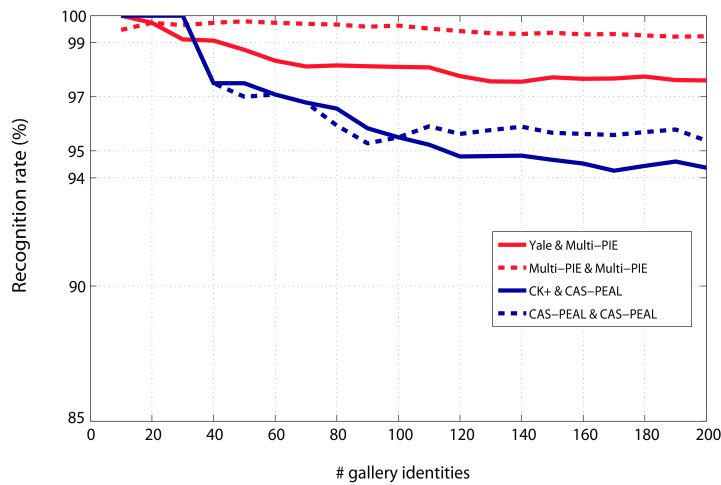
CDFA framework can improve the scalability as shown in Figure 3.6 (b). The method was tested with several scenarios both for illumination and facial expressions. When Yale<sup>1</sup> or CK+ was used as the bootstrap database, all settings like manifold dimensionality were kept same as the ones in Section 3.2 and Section 3.3. We observe that the final recognition rates are not affected significantly as bootstrap databases are switched. The largest performance difference caused by changing the bootstrap database was between 1% – 2%.

The results in Figure 3.6 (b) also suggest that it is possible to fix the template manifold for a certain type of variation since same bootstrap database can be used in different tests: CK+ was employed successfully in tests with CAS-PEAL, JAFFE, and CMU AMP while the Yale database is compatible both for Multi-PIE and Yale itself.

<sup>1</sup>There are two different Yale databases used during tests: Yale B Database [19] and Extended Yale Face Database [42]. However, when a common name 'Yale' is mentioned, it means that an augmented database which is established by concatenating two is used.



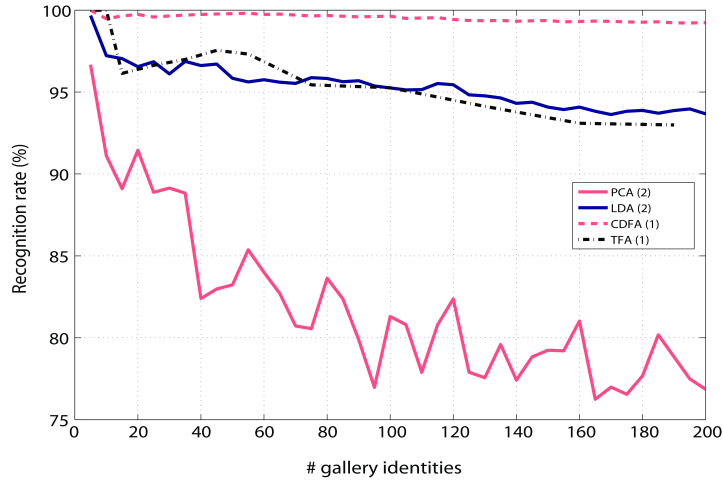
(a)



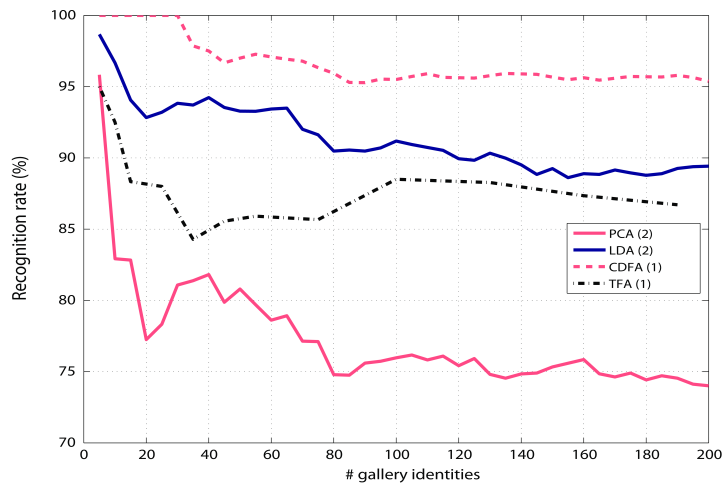
(b)

**Figure 3.6:** (a) Behavior of LDA against the illumination with increasing number of identities. Three scenarios were tried: with no bootstrap, with a bootstrap drawn from Multi-PIE, and with a bootstrap drawn from Yale. (b) Behavior of CDFA against illumination and facial expressions. Yale & Multi-PIE means that the bootstrap set is from Yale and the evaluation set is from Multi-PIE.

Figure 3.7 gives recognition rates of several methods with increasing number of identities in the gallery. CDFA was compared with PCA [2], LDA [4], and Tied Factor Analysis (TFA) [25] since they share very common aspects with CDFA, in terms of subspace analysis. The method in [25] was initially developed to handle the pose variation; however, the authors proposed the algorithm as a generic factor analysis framework just like CDFA. Multi-PIE and CAS-PEAL were used for testing against illumination and facial expression, respectively. To provide a fair comparison, a common bootstrap database with 50 identities was collected to learn the subspace



(a)



(b)

**Figure 3.7:** Recognition performance of different methods on (a) Multi-PIE illumination database and (b) CAS-PEAL expression database. Values in parentheses shows the number of gallery images.

parameters for all methods. For tests with Multi-PIE, the bootstrap includes 1000 images while this value is 250 for the tests with CAS-PEAL. Subspace dimensions were optimized individually for each method.

For both tests, bootstrap and the training/testing images were drawn from the same databases. Therefore, the manifold dimension was 4 for tests with CAS-PEAL since there are 5 different expressions in database, and the upper bound is limited by the number of expressions. In both sets of experiments, the image size was  $100 \times 90$ . Images were normalized with *zero mean-unit variance* normalization.

For CDFA and TFA, a single gallery image was selected and all remaining images were used as probes. Then, for a test having  $N$  gallery identities,  $19 \times N$  recognition attempts were performed for Multi-PIE and  $4 \times N$  recognition attempts were performed for CAS-PEAL. These attempts were repeated for each random gallery image selection, and the averages were noted.

The recognition rates tend to decrease with other methods whereas CDFA performs steadily as the number of identities increases. This fact is depicted in Figure 3.7.

### 3.6 Real Life Performance

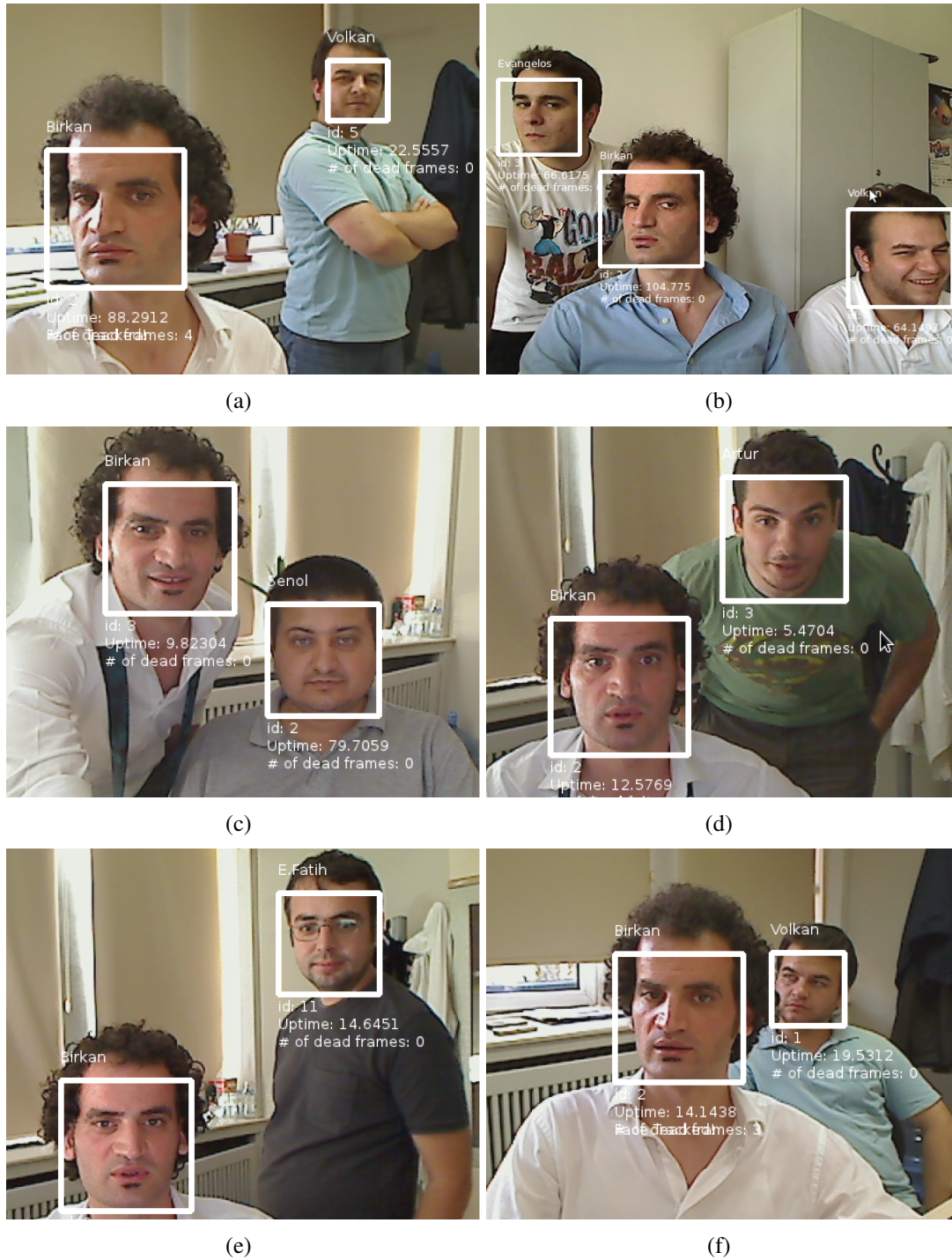
Figure 3.8 represents some real life recognition results. We employed our face detection / recognition system to get these results. The face detection system finds the surrounding rectangle and then, feature detection is run inside the rectangle. Detected features are only used for alignment and cropping. Their location information is not favored for recognition purposes. During the training, multiple images of people are used. Gallery includes 115 identities among them are 105 identities from Multi-PIE database. For each identity 10 – 12 images taken on poses between  $\pm 45^\circ$  are collected. Bootstrap tasks were performed against only pose variation. Therefore, the system is variant to illumination changes. This fact can be observed in Figure 3.9. Nevertheless, a system trained against only pose variations may handle minor facial expressions. Figures 3.10 and 3.11 show this fact promisingly. Again, no information on facial expressions was employed during the bootstrap. The real life performance against the facial expressions are much robust than that of illumination.

### 3.7 Computational Aspects

The proposed framework consists of four successive tasks: (1) manifold Learning, (2) bootstrap, (3) training, and (4) testing.

The manifold learning includes two main calculations. First, a graph is constructed and the distance of each edge is calculated (2.10). Then, a generalized eigenvalue problem is solved for this graph (2.12). Beside, there exist several column summations and one matrix-matrix subtraction to calculate matrices  $\mathbf{D}$  and  $\mathbf{L}$  in (2.12). As a final step,



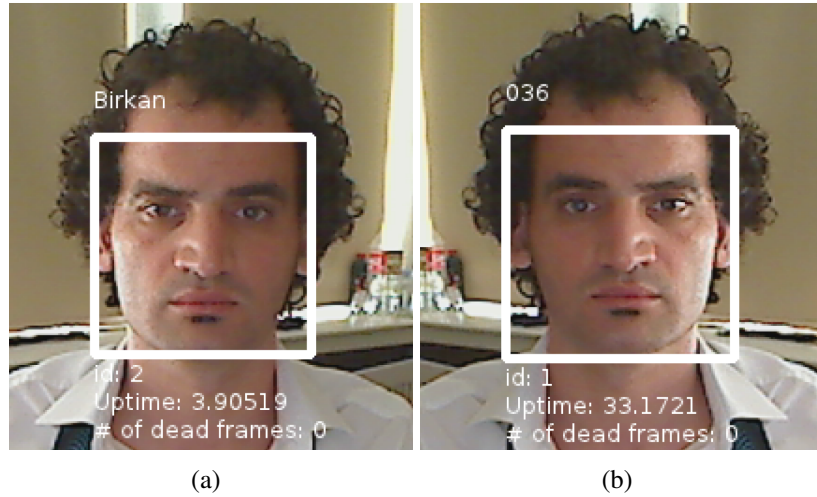


**Figure 3.8:** Several real life recognition results with pose variations.

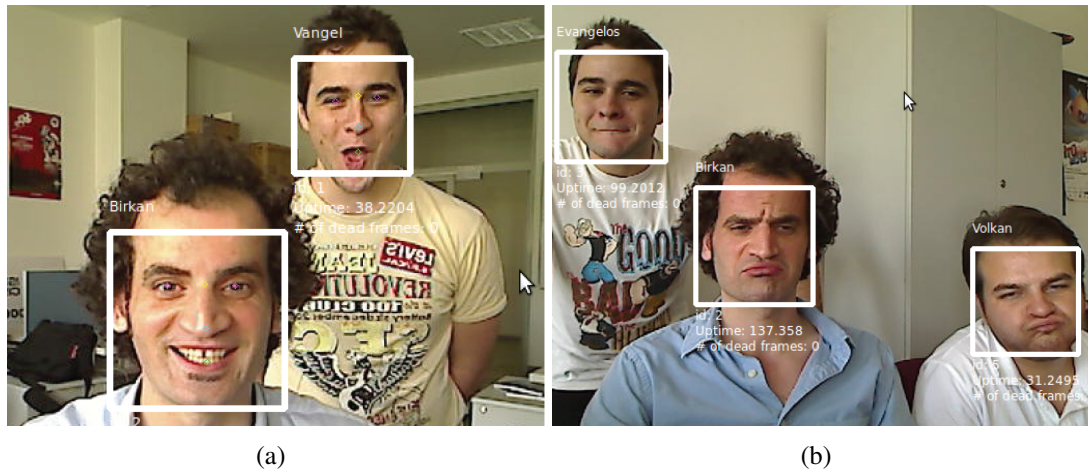
lower dimensional coordinates are calculated by (2.8) which consists of a matrix-vector multiplication.

In the bootstrap, equation (2.27) is solved for each basis vector. There are several outer products, one matrix-matrix subtraction, one matrix-vector product and other minor calculations for this equation. Then the mean vector and the covariance matrix





**Figure 3.9:** When the system is trained only against the pose variation, it is variant to the illumination changes. Recognition is failed when the position of the light is reversed.



**Figure 3.10:** Unlike illumination, the system is promisingly invariant to the facial expressions although no such information is introduced during the bootstrap.

is calculated for these basis vectors by regular calculations, (2.28) and (2.29). Statistics for the error term are obtained after a matrix-vector product and a vector-vector subtraction as in (2.30). If one wants to employ the probabilistic interpretation of Section 2.3.3, the fixed point iteration defined in (2.41) and (2.42) is required to be solved.

For the training, the main calculation is the solution of the system of linear equations defined in (2.46). To construct the system, one outer product, one matrix inversion, one matrix-vector product and summation of matrices and vectors are required. Since the



**Figure 3.11:** Another set of examples to demonstrate the invariance to the facial expressions.

size of the system relatively low depending on the dimension of the variation manifold, both the matrix inversion and the solution of the system are easy tasks.

Testing is the final step and the only part that affects the real-life performance of the framework. All previous calculations are performed offline. The main calculation is the vector norm defined in (2.53). The synthetic image,  $\mathbf{x}_{gk}$  is calculated by one

matrix-vector product and one matrix-matrix product. This is repeated for each gallery identity, then the minimum value of (2.53) is selected.

### 3.8 Complexity of the Framework

The determination of the complexity of the framework is divided into two sections: (1) offline complexity, (2) online complexity. Offline complexity includes calculations which are performed only once during the model learning or during introducing new identities. These calculations do not affect the real-life performance of the framework. Online complexity consists of calculations which are performed for each probe image, continuously. Online complexity totally determines the real-time capability of a recognition system.

#### 3.8.1 Offline complexity

Let's assume we have  $N$  bootstrap identities, each having  $K$  images corresponding  $K$  variation values. The number of features (pixels in an image) is  $n$ . The dimension of the variation manifold is fixed as  $d$ , which means that  $d$  minimum eigenvalues and their corresponding eigenvectors are calculated. During training, the complexity of introducing one gallery identity with a single gallery image is considered.

For the manifold learning (see Section 2.3.1),  $\frac{(N*K)!}{(N*K-2)!2!}$  distance calculations are performed, which are simple vector norms, resulting in  $O\left(\frac{(N*K)!}{(N*K-2)!2!}n\right)$  complexity. The final generalized eigenvalue problem has a size of  $n \times n$ . Therefore, the complexity is assumed to be between  $O(n^2)$  and  $O(n^3)$  depending on the algorithm used. The repetition factor  $d$  is omitted since it is usually too small (like 9 or 20) compared to  $n$ .

Basis vectors corresponding to each identity are calculated individually. Each basis calculation requires matrix subtractions and matrix-vector multiplications. Therefore, the complexity is  $O(n^2)$ . Again the calculations are repeated for each dimensions ( $d$ ) and for each identity ( $N$ ). Finally, obtaining statistics over vectors results in another  $O(n^2)$  complexity.

In the training, the size of the system is  $d \times d$ . In the worst case, the complexity is  $O(d^3)$ . Since the calculations are repeated for each feature (*i.e.* image pixels),

final complexity is  $O(nd^3)$ . To calculate the lower dimensional coordinates of the probe image, one matrix-vector multiplication with a matrix of size  $d \times n$  is performed. Hence, there is another  $O(nd)$  complexity.

To summarize, for the manifold learning and the bootstrap, the complexity is at most  $O(n^3)$ . This seems to be a huge amount of calculations; however, they are performed only once when the model is constructed. During the training, which is performed for each identity that is wanted to be recognized, the complexity is only  $O(nd^3)$ . Although the complexity is relatively small (depending on the dimension  $d$ ), high wall clock times are expected because to construct the system of linear equations, several readings over memory are required to get covariance matrices etc. In other words, data access may result in high waiting periods.

### 3.8.2 Online complexity

The real performance of the system is totally related to the testing phase. After the model learning and the training are completed, the testing is performed for each probe image, continuously. During testing, there exist one matrix-matrix and two matrix-vector products followed by a vector norm. Hence, the complexity is  $O(n^2)$ . Calculations are repeated for each gallery identity. This complexity does not change as different methods are employed. Both likelihood and posterior distributions (see Section 2.3.5) result in the same complexity values.

The main drawback of the proposed framework is its space complexity. For each gallery identity, the whole subspace is defined. Compared to classical methods, which store a low dimensional vector for each identity, storing a high dimensional matrix requires more space. Moreover, the testing has relatively higher time complexity since at least two matrix-vector products ( $\mathbf{M}^T \mathbf{x}_p$  and  $\mathbf{W}_g \mathbf{c}_k$ ) are performed to make decision, whereas the classical factor analysis only employs a norm calculation.

When speaking in terms of *wall clock time*, the training and the testing per image take approximately 0.3 seconds and 40 milliseconds, respectively on a regular PC (Intel Core 2 Duo 2.2 GHz and 3 GB RAM). These values are valid on a development

environment. The real life performance is better with approximately 20 milliseconds for testing on the same PC.



## **4. CONCLUDING REMARKS**

A linear generative model was developed to improve the general factor analysis framework. The main novelty is the complete probabilistic structure that individualizes manifold charts resulting in a class dependent design. Modeling nonlinear variations like illumination and facial expression is achieved by incorporating a manifold embedding technique to obtain a linear representation of the effective variation. This is not a surprising approach considering the fact that such variations can be modeled linearly on some geometries. For instance, illumination can be modeled as a linear combination of spherical harmonics on a unit sphere.

A probabilistic framework that can be employed in general classification problems when a problematic variation is exhibited on class samples is proposed. The only assumption which is used implicitly is that the variation can be modeled on a smooth manifold. If the nonlinear embedding fails, the resulting lower dimensional coordinates may disturb the final performance.

The initial results are very promising indicating the potential of the proposed framework as a replacement to regular subspace analysis methods. The proposed approach defines a novel connection between the manifold embedding and the probabilistic models.

Combining different variations is left as a future work. The first step towards this goal may be using factor tensors instead of factor matrices.

### **4.1 Discussions on Experimental Results**

Numerous experiments were performed to analyze the performance of the proposed method against different variation types and with relatively large databases. In both cases, the results are very promising.

Several advantages of the method can be summarized as follows: (1) different types of variation that lie on smooth manifolds can be handled by the method, (2) the scalability of the classical factor analysis is improved by a class dependent scheme, (3) the decision process is fully probabilistic, and posterior probabilities can be utilized for large scale and domain specific real life applications by incorporating priors on the identities, (4) bootstrap has less time complexity compared to 3D rendering approaches, and finally (5) a single observation for each identity is sufficient to perform reliable recognition while a way to use more images is also introduced.

## **4.2 Contributions**

The proposed method is a generic framework that can be used for object recognition under certain variations. In this view, a novel technique to derive a basis set of the variation subspace related to the observations of an object under an operative variation is defined.

A new framework, which is a class dependent derivative of the classical factor analysis, is proposed and two different solution schemes are provided for the optimization of the corresponding error functions. Both algebraic and probabilistic views are analyzed to offer a complete understanding of the underlying logic.

Defining separate manifolds for different classes is not a new approach; nevertheless, drawing a generic framework for such an approach is a novel and significant step. Moreover, defining a hard link between the coordinate systems of the manifolds and the variation is another improvement over the classical subspace techniques.

A novel connection between manifold learning and the probabilistic generative models is introduced in this study. Such an connection results in relaxation over the optimization problems defined by the technique while it also provides new geometric interpretations related to the algebraic and probabilistic solutions. Deriving solutions directly on the manifolds without embedding them into an Euclidean space can be a promising future work.

Handling different variations by employing a generic framework also suggests initial steps towards combining different types of variations. Utilizing factor tensors and



exploiting multilinear algebra decomposition techniques can be a first choice on this task.

These contributions are introduced and explained in the following publications.

**Table 4.1:** Contributions of the study and related publications

Construction of coordinate systems for variation manifolds	[48, 49]
Introduction of Class Dependent Factor Analysis (CDFA) model	[50, 51]
Algebraic solution to CDFA	[52]
Probabilistic solution to CDFA	[53]

### 4.3 Future Works

In this work, a generic framework that can be used against different types of variation is introduced. However, handling multiple variation types at once is left as a major future work. This may be managed by employing factor tensors instead of matrices, yet in its initial formulation such a task seems to be difficult mainly due to the fact that each possible variation has to be modeled explicitly, that with the inclusion of several variation types either through tensors or some other mathematical tool the manifold dimensionality quickly becomes unmanageable.

Another important improvement can be achieved by inclusion of a dynamic kernel that is optimized for each variation type automatically. The current formulation considers raw pixel values; however, it is a well known fact that the recognition rates can be boosted by employing different kernels or transformations for different variation types.

Partitioning the face into subregions is a must to accomplish a promising real life performance. Similarly, automatic feature detection can be very useful to make robuster and more reliable decisions. By this way, the recognition can be performed locally, and several decisions from subregions or areas around features points can be fused to obtain better results. Indeed, a feature point detection procedure seems to be an obligation to perform reliable pose invariant recognition.



## REFERENCES

- [1] **Alpaydin, E.**, (2004). *Introduction to Machine Learning*, The MIT Press.
- [2] **Turk, M. and Pentland, A.**, (1991). Face recognition using eigenfaces, *Procs. of Computer Vision and Pattern Recognition*, pp.586–591.
- [3] **Lebanon, G.**, (2005). Riemannian Geometry and Statistical Machine Learning, PhD Thesis, School of Computer Science - Carnegie Mellon University.
- [4] **Belhumeur, P.N., Hespanha, J. and Kriegman, D.J.**, (1996). Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *Procs. of European Conference on Computer Vision*, pp.45–58.
- [5] **Chang, Y., Hu, C. and Turk, M.**, (2003). Manifold of facial expression, *Procs. of International Workshop on Analysis and Modeling of Faces and Gestures*, pp.28–35.
- [6] **Roweis, S.T. and Saul, L.K.**, (2000). Nonlinear Dimensionality Reduction by Locally Linear Embedding, *Science*, **290**, 2323–2326.
- [7] **Shashua, A., Levin, A. and Avidan, S.**, (2002). Manifold Pursuit: A New Approach to Appearance Based Recognition, *Procs. of International Conference on Pattern Recognition*, volume 3.
- [8] **Seung, S.H. and Lee, D.D.**, (2000). COGNITION: The Manifold Ways of Perception, *Science*, **290**, 2268–2269.
- [9] **Tenenbaum, J.B., de Silva, V. and Langford, J.C.**, (2000). A global geometric framework for nonlinear dimensionality reduction., *Science*, **290**, 2319–2323.
- [10] **He, X. and Niyogi, P.**, (2003). Locality Preserving Projections, *Procs. of Advances in Neural Information Processing Systems, 2003*.
- [11] **Ahonen, T., Hadid, A. and Pietikäinen, M.**, (2004). Face Recognition with Local Binary Patterns, *Proc. 8th Eur. Conf. Computer Vision*, pp.469–481.
- [12] **Zhao, W., Chellappa, R., Rosenfeld, A. and Phillips, P.J.**, (2003). Face Recognition: A Literature Survey, *ACM Computing Surveys*, 399–458.
- [13] **Gnanadesikan, R. and Kettenring, J.R.**, (1984). A pragmatic review of multivariate methods in applications, **H.A. David and H.T. David**, editors, In *Statistics: An Appraisal*, Iowa State Univ. Press, pp.309–337.

- [14] **Tipping, M.E. and Bishop, C.M.**, (1999). Probabilistic Principal Component Analysis, *Journal of the Royal Statistical Society, Series B*, **61**, 611–622.
- [15] **Roweis, S.**, (1998). EM Algorithms for PCA and SPCA, *M.I. Jordan, M.J. Kearns and S.A. Solla, editors, Advances in Neural Information Processing Systems*, volume 10, The MIT Press.
- [16] **Zhang, L. and Samaras, D.**, (2006). Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, *Pattern Analysis and Machine Intelligence*, **28**(3), 351–363.
- [17] **Basri, R. and Jacobs, D.**, (2003). Lambertian reflectance and linear subspaces, *Pattern Analysis and Machine Intelligence*, **25**(2), 218–233.
- [18] **Ramamoorthi, R.**, (2002). Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object, *Pattern Analysis and Machine Intelligence*, **24**(10), 1322–1333.
- [19] **Georghiades, A., Belhumeur, P. and Kriegman, D.**, (2001). From few to many: illumination cone models for face recognition under variable lighting and pose, *Pattern Analysis and Machine Intelligence*, **23**(6), 643–660.
- [20] **Zhou, S.K. and Chellappa, R.**, (2004). Illuminating light field: image-based face recognition across illuminations and poses, *Procs. of Automatic Face and Gesture Recognition*, pp.229–234.
- [21] **Lee, K.C., Ho, J. and Kriegman, D.**, (2001). Nine Points of Light: Acquiring Subspaces for Face Recognition under Variable Lighting, *Procs. of Computer Vision and Pattern Recognition*, p.519.
- [22] **Gross, R., Matthews, I. and Baker, S.**, (2002). Eigen light-fields and face recognition across pose, *Procs. of Automatic Face and Gesture Recognition*, pp.1–7.
- [23] **Shashua, A. and Riklin-Raviv, T.**, (2001). The quotient image: class-based re-rendering and recognition with varying illuminations, *Pattern Analysis and Machine Intelligence*, **23**(2), 129–139.
- [24] **Zhou, S.K. and Chellappa, R.**, (2003). Rank constrained recognition under unknown illuminations, *Procs. of International Workshop on Analysis and Modeling of Faces and Gestures*, pp.11–18.
- [25] **Prince, S., Warrell, J., Elder, J. and Felisberti, F.**, (2008). Tied Factor Analysis for Face Recognition across Large Pose Differences, *Pattern Analysis and Machine Intelligence*, **30**(6), 970–984.
- [26] **Kim, T.K. and Kittler, J.**, (2005). Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image, *Pattern Analysis and Machine Intelligence*, **27**(3), 318–327.
- [27] **Ioffe, S.**, (2006). Probabilistic Linear Discriminant Analysis, *Procs. of European Conference on Computer Vision*, pp.531–542.

- [28] **Prince, S., Li, P., Fu, Y., Mohammed, U. and Elder, J.** Probabilistic Models for Inference about Identity, *Pattern Analysis and Machine Intelligence*, (PrePrints).
- [29] **Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S. and Ma, Y.**, (2009). Robust face recognition via sparse representation, *Pattern Analysis and Machine Intelligence*, **31**, 210–227.
- [30] **Nagesh, P. and Li, B.**, (2009). A compressive sensing approach for expression-invariant face recognition, *Procs. of Computer Vision and Pattern Recognition*, pp.1518–1525.
- [31] **Wiskott, L., Fellous, J.M., Krüger, N. and von der Malsburg, C.**, (1997). Face Recognition by Elastic Bunch Graph Matching, *Procs. of Conf. on Computer Analysis of Images and Patterns (CAIP'97)*, 1296, pp.456–463.
- [32] **He, X., Yan, S., Hu, Y., Niyogi, P. and Zhang, H.J.**, (2005). Face recognition using Laplacianfaces, *Pattern Analysis and Machine Intelligence*, **27**(3), 328–340.
- [33] **Pentland, A., Moghaddam, B. and Starner, T.**, (1994). View-based and modular eigenspaces for face recognition, *Procs. of Computer Vision and Pattern Recognition*, pp.84–91.
- [34] **Blanz, V. and Vetter, T.**, (2003). Face recognition based on fitting a 3D morphable model, *Pattern Analysis and Machine Intelligence*, **25**(9), 1063–1074.
- [35] **Blanz, V., Scherbaum, K., Vetter, T. and Seidel, H.P.**, (2004). Exchanging Faces in Images, *EUROGRAPHICS 2004*, volume 23, pp.669–676.
- [36] **Vasilescu, M. and Terzopoulos, D.**, (2003). Multilinear subspace analysis of image ensembles, *Procs. of Computer Vision and Pattern Recognition*, volume 2, pp.93–99.
- [37] **Wang, R., Shan, S., Chen, X. and Gao, W.**, (2008). Manifold-Manifold Distance with application to face recognition based on image set, *Procs. of Computer Vision and Pattern Recognition (CVPR'08)*, pp.1–8.
- [38] **Tenenbaum, J.B. and Freeman, W.T.**, (2000). Separating Style and Content with Bilinear Models, *Neural Computation*, **12**(6), 1247–1283.
- [39] **Shin, D., Lee, H.S. and Kim, D.**, (2008). Illumination-robust face recognition using ridge regressive bilinear models, *Pattern Recognition Letters*, **29**(1), 49–58.
- [40] **Elgammal, A. and Lee, C.S.**, (2004). Separating style and content on a nonlinear manifold, *Procs. of Computer Vision and Pattern Recognition*, volume 1, pp.478–485.
- [41] **Gross, R., Matthews, I., Cohn, J.F., Kanade, T. and Baker, S.**, (2008). Multi-PIE, *Procs. of International Conference on Automatic Face and Gesture Recognition*.

- [42] **Lee, K., Ho, J. and Kriegman, D.**, (2005). Acquiring Linear Subspaces for Face Recognition under Variable Lighting, *Pattern Analysis and Machine Intelligence*, **27**(5), 684–698.
- [43] **Lucey, P., Cohn, J., Kanade, T., Saragih, J., Ambadar, Z. and Matthews, I.**, (2010). The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression, *Procs. of Computer Vision and Pattern Recognition Workshops*.
- [44] **Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J.**, (1998). Coding Facial Expressions with Gabor Wavelets, *Procs. of International Conference on Face & Gesture Recognition*.
- [45] **Liu, X., Chen, T. and Kumar, B.V.K.V.**, (2003). Face authentication for multiple subjects using eigenflow, *Pattern Recognition, Special issue on Biometric*, **36**, 313–328.
- [46] **Ekman, P., Friesen, W.V. and Hager, J.C.**, (2002). *Facial Action Coding System: The Manual*, Network Information Research Corp., Salt Lake City.
- [47] **Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X. and Zhao, D.**, (2008). The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, *IEEE Transactions on Systems, Man, and Cybernetics*, **38**(1), 149–161.
- [48] **Tunç, B. and Gökmen, M.**, (2009). Illumination Invariant Face Recognition on Nonlinear Manifolds, *AAAI Fall Symposium Series*.
- [49] **Tunç, B. and Gökmen, M.**, (2010). Face Recognition with Multimodal Subspace Analysis, *Conference on Signal Processing and Communications Applications*.
- [50] **Tunç, B., Dağlı, V. and Gökmen, M.**, (2011). Robust face recognition with class dependent factor analysis, *International Joint Conference on Biometrics*, pp.1–6.
- [51] **Tunç, B., Dağlı, V. and Gökmen, M.**, (2011). Class-Specific Factor Analysis For Object Recognition, *Conference on Signal Processing and Communications Applications*.
- [52] **Tunç, B. and Gökmen, M.**, (2011). Manifold learning for face recognition under changing illumination, *Telecommunication Systems*, **47**(3-4), 185–195.
- [53] **Tunç, B., Dağlı, V. and Gökmen, M.**, (Accepted). Class Dependent Factor Analysis and Its Application to Face Recognition, *Pattern Recognition*.

## **APPENDICES**

**APPENDIX A.1 : Unconstrained Minimization of (2.17)**

**APPENDIX A.2 : An Alternative to MAP Estimate**

## APPENDIX A.1

For the unconstrained minimization problem, we again consider factor loadings,  $\mathbf{W}$ , as a basis set of the variation subspace. Therefore, factors,  $\mathbf{c}_k$ , are assumed to be coordinates *i.e.* linear combination coefficients.

Let's assume that we have  $K$  images of an identity  $i$  in the bootstrap database. Then the total reconstruction error for the identity  $i$  is

$$\begin{aligned}\mathcal{E} &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{W}_i \mathbf{c}_k\|^2 \\ &= \sum_{k=1}^K \|\mathbf{x}_{ik} - \mathbf{w}_{i1}c_{k1} - \mathbf{w}_{i2}c_{k2} - \dots - \mathbf{w}_{in}c_{kn}\|^2,\end{aligned}\quad (\text{A.1})$$

where  $\mathbf{w}_{ij}$  indicates  $j^{\text{th}}$  column of the matrix  $\mathbf{W}_i$ , and  $c_{kj}$  is  $j^{\text{th}}$  element of vector  $\mathbf{c}_k$ . The error can be minimized by equating the derivatives with respect to unknown basis vectors,  $\mathbf{w}_{ij}$ , to zero.

Normalization constraints  $\|\mathbf{w}_{ij}\| = 1$  are not introduced, since the scaling factors  $c_{kj}$  are already known and fixed. Thus, relaxations on the norms of the vectors are required to assure a global minimum. Similarly, orthogonality is not considered.

The optimization problem can be restated by a trace minimization as in

$$\begin{aligned}\mathcal{E} &= Tr[(\mathbf{X} - \mathbf{c}_1 \mathbf{w}_1^T - \dots - \mathbf{c}_n \mathbf{w}_n^T)^T (\mathbf{X} - \mathbf{c}_1 \mathbf{w}_1^T - \dots - \mathbf{c}_n \mathbf{w}_n^T)] \\ &= Tr(\mathbf{X}^T \mathbf{X}) - 2Tr(\mathbf{X}^T \mathbf{c}_1 \mathbf{w}_1^T) - \dots - 2Tr(\mathbf{X}^T \mathbf{c}_n \mathbf{w}_n^T) \\ &+ Tr(\mathbf{w}_1 \mathbf{c}_1^T \mathbf{c}_1 \mathbf{w}_1^T) + 2Tr(\mathbf{w}_1 \mathbf{c}_1^T \mathbf{c}_2 \mathbf{w}_2^T) + \dots + 2Tr(\mathbf{w}_1 \mathbf{c}_1^T \mathbf{c}_n \mathbf{w}_n^T) \\ &+ 2Tr(\mathbf{w}_2 \mathbf{c}_2^T \mathbf{c}_1 \mathbf{w}_1^T) + Tr(\mathbf{w}_2 \mathbf{c}_2^T \mathbf{c}_2 \mathbf{w}_2^T) + \dots + 2Tr(\mathbf{w}_2 \mathbf{c}_2^T \mathbf{c}_n \mathbf{w}_n^T) \\ &+ \dots \\ &+ 2Tr(\mathbf{w}_n \mathbf{c}_n^T \mathbf{c}_1 \mathbf{w}_1^T) + 2Tr(\mathbf{w}_n \mathbf{c}_n^T \mathbf{c}_2 \mathbf{w}_2^T) + \dots + Tr(\mathbf{w}_n \mathbf{c}_n^T \mathbf{c}_n \mathbf{w}_n^T).\end{aligned}\quad (\text{A.2})$$

to simplify calculations where the notation is changed slightly. The matrix  $\mathbf{X}$  has the vector  $\mathbf{x}_{ik}$  as its  $k^{\text{th}}$  row. The vector  $\mathbf{c}_j$  is the collection of constants  $c_{kj}$ . The index  $i$  of vectors  $\mathbf{w}_{ij}$  is dropped for the clarity. The objective functional is

$$\begin{aligned}\mathcal{E} &= Tr(\mathbf{X}^T \mathbf{X}) - 2\mathbf{c}_1^T \mathbf{X} \mathbf{w}_1 - \dots - 2\mathbf{c}_n^T \mathbf{X} \mathbf{w}_n \\ &+ \mathbf{c}_1^T \mathbf{c}_1 \mathbf{w}_1^T \mathbf{w}_1 + 2\mathbf{c}_1^T \mathbf{c}_2 \mathbf{w}_1^T \mathbf{w}_2 + \dots + 2\mathbf{c}_1^T \mathbf{c}_n \mathbf{w}_1^T \mathbf{w}_n \\ &+ 2\mathbf{c}_2^T \mathbf{c}_1 \mathbf{w}_2^T \mathbf{w}_1 + \mathbf{c}_2^T \mathbf{c}_2 \mathbf{w}_2^T \mathbf{w}_2 + \dots + 2\mathbf{c}_2^T \mathbf{c}_n \mathbf{w}_2^T \mathbf{w}_n \\ &+ \dots \\ &+ 2\mathbf{c}_n^T \mathbf{c}_1 \mathbf{w}_n^T \mathbf{w}_1 + \dots + \mathbf{c}_n^T \mathbf{c}_n \mathbf{w}_n^T \mathbf{w}_n.\end{aligned}\quad (\text{A.3})$$

Derivatives with respect to each basis vector yield a set of linear equations as

$$\begin{aligned}-\mathbf{X}^T \mathbf{c}_1 + \mathbf{c}_1^T \mathbf{c}_1 \mathbf{w}_1 + \dots + \mathbf{c}_1^T \mathbf{c}_n \mathbf{w}_n &= 0, \\ &\vdots \\ -\mathbf{X}^T \mathbf{c}_n + \mathbf{c}_n^T \mathbf{c}_1 \mathbf{w}_1 + \dots + \mathbf{c}_n^T \mathbf{c}_n \mathbf{w}_n &= 0.\end{aligned}\quad (\text{A.4})$$

In the matrix form,



$$\begin{bmatrix} \mathbf{c}_1^T \mathbf{c}_1 & \mathbf{c}_1^T \mathbf{c}_2 & \dots & \mathbf{c}_1^T \mathbf{c}_n \\ \mathbf{c}_2^T \mathbf{c}_1 & \mathbf{c}_2^T \mathbf{c}_2 & \dots & \mathbf{c}_2^T \mathbf{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{c}_n^T \mathbf{c}_1 & \mathbf{c}_n^T \mathbf{c}_2 & \dots & \mathbf{c}_n^T \mathbf{c}_n \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{c}_1 \\ \mathbf{X}^T \mathbf{c}_2 \\ \vdots \\ \mathbf{X}^T \mathbf{c}_n \end{bmatrix}.$$

The size of the system is relatively small depending on the dimension of the subspace. The rank of the coefficient matrix is usually  $n$  provided that a linearly independent set of vectors,  $\mathbf{c}_i$ , exists. Hence, there is a unique solution for the problem. As the complete basis set  $\mathbf{W}_i$  of each identity  $i$  in the bootstrap database is calculated, the covariance matrix for the distribution  $p(\mathbf{w})$  can be estimated by the empirical formula

$$\Omega = \frac{1}{N} \sum_{i=1}^N (\mathbf{w}_i - \bar{\mathbf{w}})(\mathbf{w}_i - \bar{\mathbf{w}})^T, \quad (\text{A.5})$$

where  $\bar{\mathbf{w}}$  is the mean value. One should be careful with this notation. Here, the form defined in (2.4) is employed. Therefore, the vector  $\mathbf{w}_i$  is a row (not a column) of the matrix  $\mathbf{W}_i$ . After calculating the matrices  $\mathbf{W}_i$  for all identities in the bootstrap gallery, the same procedure as defined in Section 2.3.2 is followed to obtain required statistics.

## APPENDIX A.2

The recognition rates in Table 3.6 are not reliable enough although several images are used. The main reason is the quality of the basis recovery. Equations in (2.50) include weighted averages over gallery samples. Such an average observation can not represent the multiple samples since pose changes cause critical appearance differences. Thus, the improvement in basis recovery is not linear with increasing number gallery images.

At this point, another recovery approach that does not try to find a basis set which is good for averages but can synthesize each observation individually is developed. Let's assume that there exist  $K$  different images of an identity  $g$ . The main formulation is then an optimization scheme with several constraints,  $x_{gk} = \mathbf{w}^T \mathbf{c}_k$ . The cost functional is

$$\mathcal{E} = -\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1}(\mathbf{w} - \boldsymbol{\mu}) + \sum_k^K \lambda_k (x_{gk} - \mathbf{w}^T \mathbf{c}_k), \quad (\text{A.6})$$

where the coefficients  $\lambda_k$  are the unknown Lagrange multipliers. A straight forward optimization is applied and derivatives with respect to unknown variables  $\mathbf{w}$  and  $\lambda_k$  are taken to get a solution by

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \mathbf{w}} &= 0 \Rightarrow \\ \boldsymbol{\Omega}^{-1} \mathbf{w} &= \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \sum_k^K \lambda_k \mathbf{c}_k \Rightarrow \\ \mathbf{w} &= \boldsymbol{\mu} + \sum_k^K \lambda_k \boldsymbol{\Omega} \mathbf{c}_k, \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} \frac{\partial \mathcal{E}}{\partial \lambda_k} &= 0 \Rightarrow \\ \mathbf{c}_k^T \mathbf{w} &= x_{gk}. \end{aligned} \quad (\text{A.8})$$

If both sides of the equation (A.7) are multiplied by  $\mathbf{c}_j^T$  and the identity defined in equation (A.8) is employed, equations

$$\begin{aligned} \mathbf{c}_j^T \mathbf{w} &= \mathbf{c}_j^T \boldsymbol{\mu} + \sum_k^K \lambda_k \mathbf{c}_j^T \boldsymbol{\Omega} \mathbf{c}_k, \\ x_{gj} - \mathbf{c}_j^T \boldsymbol{\mu} &= \sum_k^K \lambda_k \mathbf{c}_j^T \boldsymbol{\Omega} \mathbf{c}_k, \quad (j = 1, \dots, K), \end{aligned} \quad (\text{A.9})$$

emerge. The resulting set of equations can be solved by the matrix identity

$$\begin{bmatrix} \mathbf{c}_1^T \boldsymbol{\Omega} \mathbf{c}_1 & \mathbf{c}_1^T \boldsymbol{\Omega} \mathbf{c}_2 & \dots & \mathbf{c}_1^T \boldsymbol{\Omega} \mathbf{c}_K \\ \mathbf{c}_2^T \boldsymbol{\Omega} \mathbf{c}_1 & \mathbf{c}_2^T \boldsymbol{\Omega} \mathbf{c}_2 & \dots & \mathbf{c}_2^T \boldsymbol{\Omega} \mathbf{c}_K \\ \vdots & \dots & \ddots & \vdots \\ \mathbf{c}_K^T \boldsymbol{\Omega} \mathbf{c}_1 & \mathbf{c}_K^T \boldsymbol{\Omega} \mathbf{c}_2 & \dots & \mathbf{c}_K^T \boldsymbol{\Omega} \mathbf{c}_K \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_K \end{bmatrix} = \begin{bmatrix} x_{g1} - \mathbf{c}_1^T \boldsymbol{\mu} \\ x_{g2} - \mathbf{c}_2^T \boldsymbol{\mu} \\ \vdots \\ x_{gK} - \mathbf{c}_K^T \boldsymbol{\mu} \end{bmatrix}, \quad (\text{A.10})$$

to determine the Lagrange multipliers. Then, the optimum  $\mathbf{w}_g$  vector can be calculated by equation (A.7). In the case of single observation, the optimum  $\mathbf{w}_g$  is

$$\mathbf{w} = \Omega \mathbf{c}_k \frac{(x_{gk} - \mathbf{c}_k^T \boldsymbol{\mu})}{\mathbf{c}_k^T \Omega \mathbf{c}_k}. \quad (\text{A.11})$$

To analyze the result, (A.11) can be plugged into the prior  $p(\mathbf{w})$  as

$$\begin{aligned} \ln p(\mathbf{w}) &\sim -(\mathbf{w} - \boldsymbol{\mu})^T \Omega^{-1} (\mathbf{w} - \boldsymbol{\mu}) \\ &= -\frac{(x - \mathbf{c}^T \boldsymbol{\mu})}{\mathbf{c}^T \Omega \mathbf{c}} \mathbf{c}^T \Omega \Omega^{-1} \Omega \mathbf{c} \frac{(x - \mathbf{c}^T \boldsymbol{\mu})}{\mathbf{c}^T \Omega \mathbf{c}} \\ &= -\frac{(x - \mathbf{c}^T \boldsymbol{\mu})^2}{\mathbf{c}^T \Omega \mathbf{c}} \sim \ln p(x). \end{aligned}$$

In other words, the more probable the observation  $x$  is drawn from the distribution  $p(x)$ , the more appropriate the recovered basis  $\mathbf{w}$  in terms of compatibility to its prior distribution  $p(\mathbf{w})$ .



## CURRICULUM VITAE

**Name Surname:** Birkan Tunç

**Place and Date of Birth:** Diyarbakır, 07.04.1981

**E-Mail:** ahmethungari@gmail.com

**B.Sc.:** Industrial Engineering, ITU

**M.Sc.:** Computational Science and Engineering, ITU

### **Professional Experience and Rewards:**

- Siemens Corporate Research Inc., Princeton, NJ USA, February – August 2007  
Research Intern, Integrated Data Systems Department
- Adato Consulting, İstanbul, Turkey 2004 – 2005  
CRM Engineer
- İstanbul Technical University, İstanbul, Turkey 2005 – 2012  
Research Assistant

### **List of Publications and Patents:**

- E. Sarıyanidi, V. Dağlı, S. C. Tek, **B. Tunç**, M. Gökmen, A Novel Face Representation Using Local Zernike Moments, In the Procs. of Conference on Signal Processing and Communications Applications (SIU), Turkey, 2012.
- **B. Tunç**, S. K. Zhou, J. H. Park, M. Gökmen, Context Ranking Machine and Its Application to Rigid Localization of Deformable Objects, In the Procs. of International Conference on Image Processing (ICIP), Egypt, 2009.
- **B. Tunç**, H. Dağ, Generating Classification Association Rules with Modified Apriori Algorithm, In the Procs. of WSEAS AIKED, Spain, 2006.

### **PUBLICATIONS/PRESENTATIONS ON THE THESIS**

- **B. Tunç**, V. Dağlı, M. Gökmen, Class Dependent Factor Analysis and Its Application to Face Recognition, Pattern Recognition, (ACCEPTED).
- **B. Tunç**, M. Gökmen, Manifold learning for face recognition under changing illumination, Telecommunication Systems, Vol. 47 (3), pp. 185-195, 2011.
- **B. Tunç**, V. Dağlı, M. Gökmen, Robust face recognition with class dependent factor analysis, In the Procs. of International Joint Conference on Biometrics (IJCB), USA, 2011.

- **B. Tunç**, V. Dağlı, M. Gökmen, Class-Specific Factor Analysis For Object Recognition, In the Procs. of Conference on Signal Processing and Communications Applications (SIU), Turkey, 2011
- **B. Tunç**, M. Gökmen, Face Recognition with Multimodal Subspace Analysis, In the Procs. of Conference on Signal Processing and Communications Applications (SIU), Turkey, 2010.
- **B. Tunç**, M. Gökmen, Illumination Invariant Face Recognition on Nonlinear Manifolds, AAAI Fall Symposium on Manifold Learning and its Applications, USA, 2009.