

SEYRELTİK KODLAMA İLE YÜZ İFADELERİNİN SINIFLANDIRILMASI

CLASSIFICATION OF FACIAL EXPRESSIONS BY SPARSE CODING

Nesli Erdoğan, Jean-Luc Dugelay

Multimedia Communications Department
EURECOM, Sophia-Antipolis, FRANCE
{nesli.erdogmus,jean-luc.dugelay}@eurecom.fr

ÖZETÇE

Yüz imgelerindeki ifadelerin çeşitliliği, yüz tabanlı bilgisayarla görü uygulamalarında karşılaşılan en önemli ve en zor problemlerden biridir. Günümüze kadar yüz ifadelerine karşı dayanıklılık üzerine birçok sistem önerilmiş olsa da, bu sorun halen geçerliliğini korumaktadır. Yüz imgesindeki ifadenin çeşidi hakkında edinilecek bilginin çözüme yapacağı katkı göz önünde bulundurularak, bu makalede 2B frontal görüntülerde yüz ifadelerinin sınıflandırılması üzerine bir analiz sunuyoruz. Seyreltik kodlama metodunun yüz tanımda yakaladığı başarıdan kaynaklanan motivasyonla, benzer prensipler hem orijinal hem de (PCA yolu ile) boyutu indirgenen imgelere uygulanmış, elde edilen kodlar iki farklı yaklaşım kullanılarak sınıflandırılmıştır: minimum kalan (residual) hatası ve maksimum sınıf-İçi katsayı toplamı. 105 kişi için farklı yüz ifadelerinin yer aldığı Bosphorus veritabanı kullanılarak kapsamlı deneyler yapılmıştır.

ABSTRACT

Expression variations in facial images is one of the most crucial and difficult problems in face-based computer vision applications. Although numerous systems have been proposed for robustness against facial expressions, so far it still persists to be an open problem. Considering that the knowledge on the type of the expression in a facial image would greatly facilitate the solution of this issue, in this paper we present an analysis for facial expressions classification in 2D frontal views. With the motivation of the success that sparse coding achieved in face recognition, similar principals are applied for to both original and dimension-reduced (via PCA) images and the resulting codes are classified based on two different approaches: minimum residual error and maximum interclass summation of the coefficients. Extensive tests are conducted on Bosphorus database, in which different expressions are available for 105 persons.

1. GİRİŞ

Kolay erişilebilen ve yüksek derecede ayırteci olan yüz, on yıllardan beri bilgisayarla görü arařtırmacılarının gözde biyometrisi olmuřtur. Yüz teřhis ve tanıma arařtırmaları güvenli ve gözetim sistemlerine katkı saęlamanın yanısıra nörolojik ve psikolojik bilimlerde de geliřmelere yol açmıřtır. Yüz imgelerinin incelenmesinde kafa duruřu, aydınlatma řartları ve yüz ifadeleri başlıca olumsuzluklar olarak karřımıza çıkmaktadır.

Bu problemler, üzerine birçok arařtırma sunulmuř olsa dahi geçerliliklerini halen korumaktadırlar. Özellikle 3B yüz

bilgisinin eriřim ve iřlemesinin kolaylařması ile birlikte, bu verinin doęası gereęi sahip olunan poz ve aydınlatma deęiřimsizlięi sebebiyle çözüme büyük katkılar saęlanmuřsa da, yüz ifadeleri için benzer bir geliřme söz konusu deęildir.

İfadeler kiřinin duygusal durumu, biliřsel aktiviteleri ve ilgi doęrultusu hakkında ipuçları verdięinden yüzüze iletiřimde büyük rol oynarlar. Bu da ifadelere makine-insan etkileşiminde yadsınamaz bir önem kazandıır. Yüz ifadelerinin sınıflandırılması bu deęiřimlere karřı alınacak önlemler aęısından da büyük önem tařımaktadır. Örneęin, güvenli uygulamaları aęısından bakılacak olursa, 2B yüz imgelerinin barındırdıkları ifadeye göre sınıflandırılması sonucu yüz teřhis ve tanıma sistemleri daha doęru sonuçlara ulařabileceklerdir.

Bu amaçla önerilen sistemlerin bir kısmında [1, 2] başlangıçta yüz tanıma için geliřtirilen teknikler kullanılmıřtır. Yüz imgesinde ifade ile birlikte kimlik bilgisi de bulunmaktadır ve ifadelerin ayrıntılı analizi için kimlik bilgisini imgeden çıkarmak faydalıdır. Bu bakımdan, [1]'de videodan alınan imge dizileri kullanılarak fark imgeleri oluşturulmuř ve optik akıř tahmini ile ifadeler farklı uzamsal analizlerle sınıflandırılmıřlardır. Buna benzer bir yöntem ayrıca [3]'te de geliřtirilmiřtir. Burada belirtmekte fayda olan nokta řudur ki; video dizisinin getirdięi kolaylık maalesef sabit fotoęraf karelerinde bulunmaz.

Dięer yaklařımda [4] ise imgeye nokta-tabanlı yüz modelleri oturtularak yüzdeki hareket birimleri (Action Units) saptanmıř ve bu hareketler duygusal ifadelerin sınıflandırılmasında kullanılmıřtır. Yine [5]'te, yüz modellemesinde öznitelik noktalarından yararlanılmıřtır. Bu tarz yaklařımların zayıf noktası ise öznitelik noktalarının doęruluęuna olan ařırı baęımlılıktır.

Son yıllarda birçok teorik ve deneysel çalıřma, insanlardaki görme sisteminin, nesnelere tespit ve tanıma için hiyerarşik ve tutumlu bir řekilde çalıřtıęını ortaya koydu. Sinir hücrelerinin seçicilięi ve karmařık öznitelikleri kademeli olarak iřlemeleri, benzer modeller geliřtirmek için arařtırmacılara ilham kaynaęı oldu. Alıcı alanları (receptive field) modelleyen seyreltik kodlama [6, 7] da bu yaklařımla ortaya çıkan yöntemlerden biridir. [6]'da belirtildięi üzere doęal imgelerin seyreltik kodlaması, birincil görsel korteksteki (V1) basit hücrelerin alıcı alanlarına çok benzer öznitelikler saęlamaktadır.

Bu noktadan yola çıkan arařtırmacılar seyreltik kodlamanın birçok alanda güçlü bir gösterim aracı olduęunu ispatladılar. [8]'de Yang ve dięerleri, seyreltik kodlama ile öznitelik seçiminin öznitelik sayısından daha önemsiz olduęunu ve hatta yeterli sayıda olmaları durumunda rastgele seçilen özniteliklerin dahi Eigenface ya da Fisherface gibi gözde öznitelikler kadar iyi performans gösterebildiklerini

kanıtlamışlardır. [9]'da ise benzer şekilde seyreltik gösterim ve l1 enküçültme yöntemi ile gerçekleştirilen yüz tanıma sisteminin dayanıklılığı ispatlanmıştır. Bunlara benzer örnekler çoğaltılabileceği gibi seyreltik kodlama gün begün daha çok araştırmacının dikkatini çekmeye ve etki alanını genişletmeye devam etmektedir.

Bu makalede, seyreltik kodlamanın yüksek kabiliyeti noktasından hareket ederek oluşturulan bir yüz ifadesi sınıflandırma sistemi önerilmektedir. [10]'dan farklı olarak amaç, sabit 2B frontal yüz imgelerindeki ifadeyi sınıflandırmaktır. Daha önce verilen örneklere benzer şekilde [10]'da video görüntülerinden elde edilen, kişiye ait nötr yüz imgesinden istifade edilmektedir. Maalesef, bu avantaj her koşulda mevcut değildir. Bu makalede göz önüne alınan senaryoda, analiz edilecek imgeye dair kimlik bilgisi mevcut değildir. Aynı zamanda, kişinin eğitim adımında kayıtlı olması mecburiyeti de bulunmamaktadır. İfade sınıflandırılması yapılacak kişilerden bağımsız toplanan yüz ifadesi örnekleri taban alınarak seyretik kodlama yapılmış, iki farklı yaklaşım ile elde edilen kodlar sınıflandırılmıştır: minimum kalan (residual) hatası ve maksimum sınıf-içi katsayı toplamı. Geliştirilen yöntem Bosphorus veritabanı kullanılarak test edilmiştir.

Makalenin geri kalanı şu şekilde yapılandırılmıştır: İkinci kısımda seyreltik kodlamanın detaylarına yer verilmiş, ardından üçüncü kısımda önerilen yöntem anlatılmıştır. Dördüncü kısımda yapılan testler ve sonuçlar sunulmuş ve son olarak ise varılan neticeler ve yapılması planlanan işler sıralanmıştır.

2. SEYRELTİK KODLAMA

Seyreltik kodlamadaki temel fikir aslında oldukça basittir. Gözlem kümesindeki çok boyutlu v vektörleri, W özneliklerinin (taban vektörlerinin) doğrusal üstdüşümleri (superposition) olarak modellenirler. n-boyutlu v imgesi şu şekilde ifade edilir:

$$v = Wh = \sum_{i=1}^r W_i h_i \quad (1)$$

W 'nin r adet sütunu taban imgeleridir ve her biri n-boyutludur. Her taban imgesinin v 'ye olan katkısını gösteren h_i parametreleri gözlem kümesinin her elemanına göre değişkenlik gösterir ve *seyreltiklik* özelliğine sahiptir. Kapsamlı bir taban küme bulunduğu takdirde v 'yi doğru vesityrek bir şekilde temsil etmek mümkün olacaktır.

N adet yüz imgesinin K farklı kategoriye ayrılacağını varsayalım. Her bir örnek için v vektörleri imgeler satır satır taranarak oluşturulur. Bu vektörler doğrudan ya da boyutları indirgenerek ve her biri bir sütun teşkil edecek şekilde W matrisini, ya da bir diğer deyişle *sözlük* meydana getirirler. [9]'da seyrek h ağırlıklarını bulmak için l1 eniyileme yaklaşımı önerilmiştir. Bu durumda, test imgesi y için h vektörünü bulma problemi (ϵ eşik değeri olmak üzere) şu şekli alır:

$$\tilde{h} = \arg \min \| h \|_1 \Rightarrow \| Wh - y \| < \epsilon \quad (2)$$

(2)'ye göre en iyi \tilde{h} vektörü hesaplandıktan sonra y imgesinin K sınıftan birine atanması gerekmektedir. Bu görev için iki farklı yol denenmiş ve sonuçları rapor edilmiştir:

2.1. Minimum Kalan Hatası (MKH)

Test imgesi y sırasıyla her bir sınıfa göre ayrı ayrı yaklaşılan. Bunun için, h 'deki yalnızca bir sınıfa ait girişler kullanılır. Yaklaşıklama sonuçlarına göre imge, kalan hatasını (3) minimuma indiren sınıfa atanır.

$$l(y) = \arg \min_k r_k(y) = \| y - y_k \|_2 \quad (3)$$

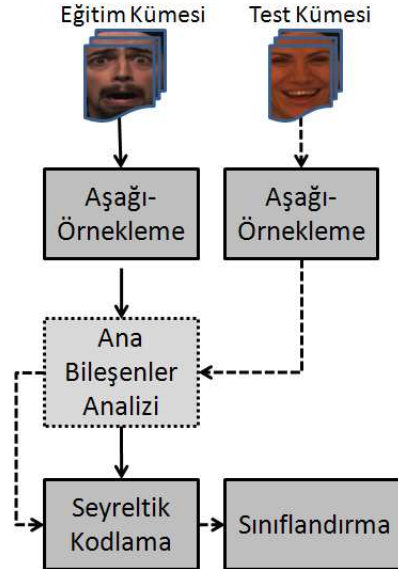
2.2. Maksimum Sınıf-içi Katsayı Toplamı (MKST)

Farklı bir yaklaşımda ise sınıflandırma en yüksek katsayı toplamına ait sınıfa göre yapılır. h vektörünün her bir sınıfa ait girdileri ayrı ayrı toplanır ve imge, en yüksek toplama sahip olan sınıfa atanır.

3. SİSTEM

[10]'da da bahsedildiği gibi yüz imgesinde ifadenin yanı sıra kimlik bilgisinin de mevcut olmasından dolayı, seyreltik kodlama sonucunda kişi aynı ifade sınıfından çok kendi kimliği ile eşleşmektedir. Öncelikle bu sorunu kısmen de olsa bertaraf etmek için eğitim kümesindeki yüzler test kümesinde olmayan kişilerden seçilmiştir. Zaten daha gerçekçi olan yaklaşım da budur.

Mutluluk, üzüntü, öfke, korku, şaşırma ve iğrenme ifadelerinin incelendiği sistemimizde, her bir sınıf için oluşturulan eğitim kümeleri taban vektörleri olarak kullanılmıştır. Aşağı-örneklenen imgeler sonrasında boyut büyüklüğünü daha da azaltmak için ana bileşenler analizine (PCA) tabi tutulmuşlardır. (Şekil 1)



Şekil 1: İşlem akış şeması.

Son olarak, test imgeleri aynı işlemlerden geçirilmiş ve herbiri için 11 eniyileme yöntemi ile h vektörleri saptanmıştır. Eniyileme probleminin çözümü için 11-magic yazılım pakedi kullanılmıştır. [11]

4. DENEYLER

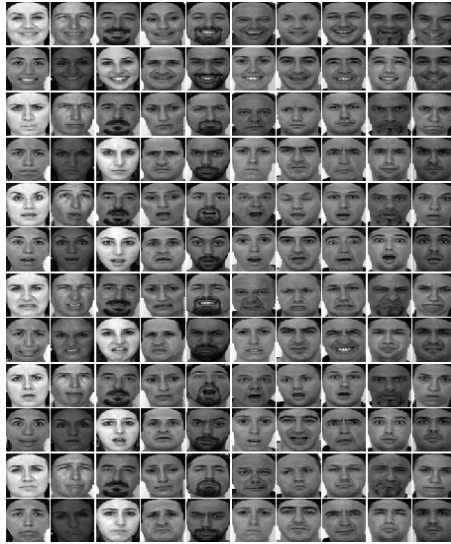
Seyreltik kodlamanın yüz ifadeleri üzerinde ayırteci gösterimler üretebilme potansiyeli Bosphorus veritabanında [12] test edilmiştir. 105 farklı kişinin çeşitli poz ve ifadelerinden oluşan bu veritabanı, aslında 3B özelliği ile öne çıksa da sağladığı ifade çeşitliliği ve örnek çokluğu açısından, 2B imgelerde yüz ifadelerini inceleme alanında yaygınca kullanılan diğer veritabanlarına (Yale yüz veri tabanı [13] – 15 kişi/4 ifade, ORL yüz veri tabanı[14] – 40 kişi/2 ifade) üstünlük sağlamıştır.

Bosphorus veritabanından alınan ifadeli yüz imgeleri ile Tablo 1’de verilen alt küme oluşturulmuştur.

Tablo 1: Yüz ifadelerinin incelendiği alt küme bileşenleri

İfade	Taban Küme Eleman Sayısı	Test Kümesi Eleman Sayısı
Mutluluk	30	76
Öfke	30	41
Şaşırma	30	41
İğrenme	30	39
Korku	30	40
Üzüntü	30	36

Her bir ifade için 30 farklı kişiden alınan imgeler *sözlük*’ü yani seyreltik kodlama için kullanılacak taban kümeyi (Şekil 2) oluşturmaktadır. Test kümesindeki örnekler taban kümede varolmayan kişilerden alınmıştır. Böylece kimlikten kaynaklanan benzerliğin, ifadenin kaynaklanan benzerliğin önüne geçmesi problemi bir ölçüde giderilmiştir.



Şekil 2: Sözlükten örnekler (30 kişiden 20si gösterilmekte)

Başlangıç için hem eğitim hem de test imgeleri ortalanarak 1000x1000 piksel büyüklüğünde olacak şekilde

kırpılmış, ardından da işlemleri hızlandırma amacıyla 64x64 piksel ölçüsüne küçültülmüşler ve gri ölçeğine dönüştürülmüşlerdir. Böylece her imge için büyüklüğü 4096 olan bir öznitelik vektörü elde edilmiştir.

Ardından, ilk olarak tüm eğitim kümesi kullanılarak bir ana bileşenler uzayı oluşturulmuş, test kümesi de dahil tüm örneklerin bu uzaya izdüşümleri hesaplanmıştır. Sonraki aşamalarda bu izdüşümlerin yalnızca ilk 45 girdisi kullanılmış, bir diğer deyişle tek bir imge için öznitelik vektörünün büyüklüğü 45’e düşürülmüştür.

Daha sonra, yapılan ilk denemede sınıflar arası etkileşimi görme ve her sınıfın ayırteci özelliklerini ölçme amacıyla tüm test imgeleri, taban kümesinin tamamı kullanarak seyreltik kodlanmış ve ifade için maksimum sınıf-ıçi katsayı toplamını veren grup seçilmiştir.

		Beklenen Sonuç						
		M	Ö	Ş	İ	K	Ü	
Alınan Sonuç	M	69 25.3%	2 0.7%	2 0.7%	9 3.3%	4 1.5%	4 1.5%	76.7% 23.3%
	Ö	0 0.0%	23 8.4%	2 0.7%	14 5.1%	2 0.7%	12 4.4%	43.4% 56.6%
	Ş	0 0.0%	1 0.4%	25 9.2%	3 1.1%	24 8.8%	3 1.1%	44.6% 55.4%
	İ	2 0.7%	4 1.5%	1 0.4%	5 1.8%	0 0.0%	2 0.7%	35.7% 64.3%
	K	3 1.1%	2 0.7%	10 3.7%	3 1.1%	7 2.6%	2 0.7%	25.9% 74.1%
	Ü	2 0.7%	9 3.3%	1 0.4%	5 1.8%	3 1.1%	13 4.8%	39.4% 60.6%
		90.8% 9.2%	55.1% 43.9%	61.0% 39.0%	12.8% 87.2%	17.5% 82.5%	36.1% 63.9%	52.0% 48.0%

Şekil 3: Tüm ifadeleri kullanarak yapılan test sonucu elde edilen hata matrisi

Şekil 3’te verilen sonuçlar oldukça ilginç veriler sunmaktadır. Bu sonuçlara göre önerilen yöntemle en fazla test imgesine sahip mutluluk ifadesi başarılı bir şekilde sınıflandırılırken, başta iğrenme ve korku olmak üzere diğer ifadeler karıştırılmışlardır. Verilen matrisde hataların öbeğlenmesi bazı ifadelerin ayırımının zorluğunu göstermektedir. Örneğin, korku beklenirken şaşırma ya da şaşırma beklenirken korku olarak sınıflanan imgeler bu iki test kümesinin yaklaşık %42’sini teşkil etmektedirler. Bu sonuçların sebebi iki şekilde yorumlanabilir:

- İfadeler için kullanılan yüz hareket birimlerinin benzerliği
- İfadelerin taklidinin güçlüğü

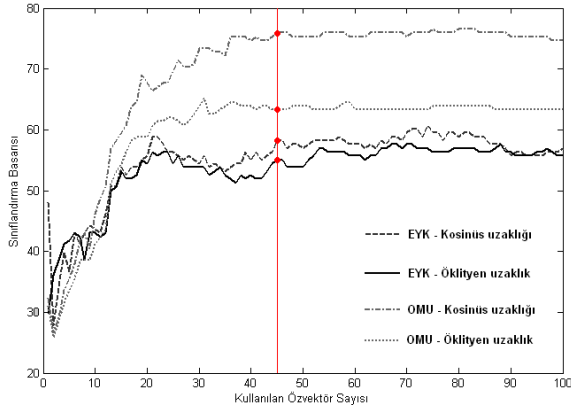


Şekil 4: Test imgelerinden bazı örnekler. İfadeleri tahmin edebilir misiniz?

Şekil 4'te verilen örneklerin tamamı korku kümesinden olup görüldüğü üzere büyük farklılıklar içermektedirler.

Varılan bu sonuçlardan yola çıkarak ek testler en iyi ayırdedilebilen 3 sınıf (mutluluk, öfke ve şaşırma) üzerinde yapılmıştır.

Öncelikle dayanak yaklaşım olarak ana bileşenler kullanılarak yapılan en yakın komşuluk ve minimum uzaklıktaki ortalama ile sınıflandırma benimsenmiştir. İlk yaklaşımda test imgesinden elde edilen öznelik vektörüne en yakın komşu taban kümesi (EYK) tespit edilmiş ve bu elemanın sınıfı seçilmiştir. İkincisinde ise her sınıf için ortalama model hesaplanmış ve test öznelik vektörüne en yakın model (OMU) bulunmuştur. Şekil 4'te verilen grafikte de gösterildiği gibi en yüksek başarı ortalama modele kosinüs uzaklığı ile elde edilmiş ve 45 özvektör kullanıldığında %75,95 başarı sağlanmıştır.



Şekil 5: Test imgelerinden bazı örnekler. İfadeleri tahmin edebilir misiniz?

Önerilen metod iki farklı sınıflandırma yöntemi ile hem boyut küçültülmüş hem de orjinal öznelik vektörleri ile test edilmiş, sonuçlar Tablo 2'de verilmiştir. Bu test sonucunda en yüksek sınıflandırma başarısı %90,5 ile boyut küçültülmüş verinin en yüksek sınıf içi toplamı yöntemi kullanılarak sınıflandırılmasıyla elde edilmiştir.

Tablo 2: Dört farklı yöntem için sınıflandırma sonuçları

Yöntem	Başarı
Boyut Küçültme + MKH	%44,9
Boyut Küçültme + MKST	%90,5
MKH	%75,9
MKST	%75,3

5. SONUÇ

Bu makalede, seyreltik kodlamadan faydalanarak yüz ifadeleri incelenmiştir. Test kümesiyle ilintisiz örnek ifade imgeleri doğrudan taban küme olarak kullanılmıştır. Birçok kaynaktan, test kümesinden bağımsız sayısız ifade örneği toplanabileceği göz önünde bulundurulduğunda bunun gerçekçi bir yaklaşım olduğu aşikardır. Bosphorus veritabanında yapılan analizler sonucu ayırdedilmesi zor ifadeler saptanmıştır. Buna ek olarak, önerilen yöntem dayanak (baseline) yaklaşımla karşılaştırılmış ve dayanak sonucundan %15 daha fazla başarı elde edildiği görülmüştür.

6. KAYNAKÇA

- [1] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, "Classifying Facial Actions", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 21, No.10, pp. 974-989, 1999.
- [2] M. Pantic, L.J.M. Rothkrantz, "Automatic Analysis of Facial Expressions: the State of the Art", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 22, No.12, pp. 1424-1445, 2000
- [3] S. Park, D. Kim, "Spontaneous facial expression classification with facial motion vectors," *Automatic Face & Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, vol., no., pp.1-6, 17-19 Sept. 2008
- [4] M. Pantic, L.J.M. Rothkrantz, "An expert system for multiple emotional classification of facial expressions," *Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on*, vol., no., pp.113-120, 1999
- [5] A. Colmenarez, B. Frey, T.S. Huang, "Embedded face and facial expression recognition," *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol.1, no., pp.633-637 vol.1, 1999
- [6] B. A. Olshausen, D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, Vol. 381, No. 6583. (13 June 1996), pp. 607-609.
- [7] P. O. Hoyer, "Modeling receptive fields with non-negative sparse coding," *Neurocomputing*, vol. 52-54, pp. 547-552, 2003.
- [8] A. Y. Yang, J. Wright, M. Yi, S.S. Sastry, "Feature selection in face recognition: A sparse representation perspective," *Tech. Report*, 2007.
- [9] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, M. Yi, "Robust Face Recognition via Sparse Representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.31, no.2, pp.210-227, Feb. 2009
- [10] S. Zafeiriou, M. Petrou, "Sparse representations for facial expressions recognition via l1 optimization," *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, vol., no., pp.32-39, 13-18 June 2010
- [11] E. Candes, J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," Available in URL: www.acm.caltech.edu/l1magic/downloads/l1magic.pdf, 2008.
- [12] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, "Bosphorus Database for 3D Face Analysis," *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID 2008)*, Roskilde University, Denmark, May 2008.
- [13] P. N. Bellhumer, J. Hespanha, D. Kriegman. "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Special Issue on Face Recognition, 17(7):711--720, 1997.
- [14] F. Samaria, A. Harter, "Parameterisation of a Stochastic Model for Human Face Identification," *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota FL, December 1994.