

Konum ve Doku Bilgisinin Ortak Dağılımını Kullanarak Yüz Özniteliklerinin İstatistiksel Çıkarımı

Statistical Facial Feature Extraction Using Joint Distribution of Location and Texture Information

Mustafa Berkay Yılmaz, Hakan Erdoğan, Mustafa Ünel

Mühendislik ve Doğa Bilimleri Fakültesi
Sabancı Üniversitesi

{berkayyilmaz}@su.sabanciuniv.edu, {haerdogan, munel}@sabanciuniv.edu

Özetçe

Bu bildiriye, konum ve doku bilgisini beraber kullanarak yüz üzerindeki kritik noktaların çıkarılmasını sağlayan bir yöntem anlatılmaktadır. Önerilen yöntem, konum ve doku bilgilerini otomatik olarak eğitim verisinden öğrenebilmektedir. Yüz özniteliklerinin konum ve doku bilgisinin ortak dağılımı maksimuma çıkarılarak, yüz öznitelikleri için en iyi konumlar bulunmaktadır. Yöntemin, 100 resimlik bir sinama verisi üzerindeki başarımı incelenmiş ve ümit verici bulunmuştur. Ayrıca bu yöntemin, aynı sinama verisi kullanıldığında, aktif görünüm modellerinden daha başarılı olduğu gözlemlenmiştir.

Abstract

A facial feature extraction method is proposed in this work, which uses location and texture information given a face image. Location and texture information can automatically be learnt by the system, from a training data. Best facial feature locations are found by maximizing the joint distribution of location and texture information of facial features. Performance of the method was found promising after it is tested using 100 test images. Also it is observed that this new method performs better than active appearance models for the same test data.

1. Giriş

Esnek cisimlerin modellenmesi önemli bir problemdir. Buna bağlı olarak, yüz özniteliklerinin çıkarılması problemi de; yüz tanıma, yüz ifadesinin analizi, dudak okuma gibi bir çok alanda kullanılabilen önemli bir başka problemdir. Bu amaçla, aktif görünüm modelleri (Active Appearance Models - AAM) ve aktif şekil modelleri (Active Shape Models - ASM) [1] gibi yaklaşımlar kullanılmıştır. Bu yöntemler popüler olmakla beraber, kişiye bağımlı eğitim durumunda iyi çalışmaktadır. Fakat kişiden bağımsız genel modeller iyi başarımlar gösterememektedir.

Bunların dışında, farklı yöntemler kullanarak yüz özniteliklerinin çıkarılmasını amaçlayan; [2], [3], [4] gibi çalışmalar bulunmaktadır. Bu yöntemler genellikle bazı önsel kurallar kullanılır ve bu durum da farklı kontrol noktalarının yöntemlere sonradan eklenmesini güçleştirir.

Biz bu bildiriye, hem doku hem de konum bilgilerini otomatik olarak öğrenebilen bir yöntem öneriyoruz. Böylece, sadece doku bilgisine bağımlı olmayan daha güçlü bir algoritma geliştirilebilir. Bu algoritmaya girdi olarak, bir yüz tanımlayıcıdan çıkmış yüz resminin verileceği varsayılır. Uygun girdiler sağlandığı zaman, istatistiksel konum ve doku bilgisi birlikte kullanılarak, AAM'ye göre daha az piksel hatası ile özniteliklerin yüz üzerindeki konumlarını bulmak mümkündür.

Bu makalenin geri kalan kısmı şöyle organize edilmiştir: Kullanılan istatistiksel model, Bölüm 2'de açıklanmıştır. Yöntemin gerçekleştirilmesi için önerilen algoritma Bölüm 2.2.2'de ve deneysel sonuçlar Bölüm 3'te açıklanmıştır. Son olarak Bölüm 4'te makale sonuçlandırılmış ve ileride uygulanabilecek bazı iyileştirmeler önerilmiştir.

2. Yüz Özniteliklerinin Modellenmesi

Yüz öznitelikleri; her insanın yüzünde bulunan, göz kenarları, burnun uç noktası ve dudak köşeleri gibi kritik noktalar. Her yüz özniteliği, konum ve doku bileşenleriyle ifade edilecektir. $f = (l, t)$, bir özniteliğin konum ve doku bileşkesi olsun. $l = (x, y)$ özniteliğin konumu ve $t = t(l)$ de bu özniteliğin doku bileşeni olarak tanımlanır. Her yüz özniteliği için konum vektörünün boyutu 2, doku vektörünün boyutu p olacaktır. $f_i = (l_i, t_i)$, i numaralı yüz özniteliğinin konum ve dokusunu belirtir. $F = (f_1, f_2, \dots, f_N)$ bileşke niteliklerin, $L = (l_1, l_2, \dots, l_N)$ konum bileşenlerinin ve $T = (t_1, t_2, \dots, t_N)$ de doku bileşenlerinin tüm yüz özniteliklerini içeren koleksiyonları olarak tanımlanabilir.

Buradaki amaç, yüz özniteliklerinin konum ve doku bilgilerinin ortak dağılımını maksimuma çıkararak, yüz

öznitelikleri için en iyi konumları bulmaktır. Tüm öznitelikler için bileşke olasılık aşağıdaki gibi tanımlanır:

$$P(F) = P(T, L). \quad (1)$$

Bu fonksiyonu hesaplayabilmek ve optimize edebilmek için, değişik varsayımlar ve sadeleştirmeler yapılmalıdır. En uygun yüz öznitelik konumları, aşağıdaki optimizasyon problemini çözerek bulunabilir:

$$\hat{L} = \operatorname{argmax}_L P(T, L). \quad (2)$$

Elbette, bazı basitleştirici varsayımlar yapmadan bu problemi çözmek kolay değildir. Dolayısıyla, takip eden bölümlerde bazı olası varsayımlardan bahsedilecektir.

2.1. Bağımsız konum-doku modeli

Her bir özneliğin, bir başkasından bağımsız olduğunu varsayarak, bu formülü sadeleştirebiliriz. Böylece, şunu elde ederiz:

$$P(T, L) \approx \prod_{i=1}^N P(t_i, l_i). \quad (3)$$

$P(t_i, l_i)$ bileşke olasılığı, doku ve konum vektörlerinin birleştirilmesiyle oluşturulmuş $p + 2$ boyutlu bir vektör üzerinde tanımlanır. Daha sonra, bu birleştirilmiş vektörün parametrik bir dağılıma sahip olduğu varsayılarak, eğitim verisi yardımıyla bu parametreler kestirilebilir. Burada, parametrik dağılım için bir seçim, kolay kestirilebilen parametrelerle çok-modlu (multi-mode) dağılım sağlayan Gauss karışım modeli (Gaussian mixture model - GMM) olabilir. Buna göre, her özneliğin konumu bağımsız olarak kestirilir, böylece paralel hesaplama yapılabilir.

$$\hat{l}_i = \operatorname{argmax}_{l_i} P(t_i, l_i), \quad (4)$$

olduğundan dolayı, her öznitelik diğerlerinden bağımsız olarak aranıp optimize edilebilir. Bu varsayım her ne kadar tutarlı öznitelik konumlarına ulaşabilse de, tipik bir insan yüzündeki yüz özniteliklerinin birbirlerine olan konumsal bağılıkları gözardı edildiği için, sonuçta elde edilen konumlar en iyi sonuçlar değildir.

2.2. Konum-bağımlı bileşik model

Yapılabilecek bir başka varsayım da, özniteliklerin konumlarının birbirlerine bağımlı olduğunu ama özniteliklerin dokularının bağımsız olduğunu varsaymaktır. Bu durumda, bileşke olasılık şöyle yazılır:

$$P(T, L) = P(L)P(T|L) \approx P(L) \prod_{i=1}^N P(t_i|l_i). \quad (5)$$

Burada gerçeğe uygun olarak, her bir yüz özneliğinin dokusunun sadece o özneliğin kendi konumuna bağlı olduğu

ve diğer konum ve dokulardan bağımsız olduğu varsayılır. Böylece:

$$P(t_i|L) = P(t_i|l_i). \quad (6)$$

Bu modele göre, yüz özniteliklerinin civarındaki dokuların, diğer doku ve konumlardan bağımsız olduğu varsayılmıştır. Konumlar birleşik olarak modellendiği için, özniteliklerin konumlarının birbirlerine bağlı olduğu varsayılmış olur. Bunun mantıklı bir varsayım olması beklenir. Örneğin bir insanın burnunun görünümü, aynı insanın göz veya dudağının görünümü kadar bilgi vermeyebilir. (Aynı insanın yüz resmi hem eğitim hem de test verisinde kullanılmadığı sürece) Elbette, eğer belirli bir kişi daha önceden eğitim verisinde görüldüyse, bu varsayım yeterince doğru olmayabilir. Bununla birlikte, gerçekçi olması için, en kötü durum senaryosuna göre, sistemin eğitim ve test aşamalarının farklı kişileri kapsadığı varsayılmaktadır. Öznitelik konumları daha baskın bir bağımlılıktır ve insanların yüz geometrileriyle ilişkilidir. Örneğin gözlerin konumu, burun ucunun konumu için iyi bir göstergedir. Bundan dolayı, konumların bağımlılıklarının modellenmesi gerekmektedir.

(2) denkleminin çözümü, yüzdeki her bir özneliğin en uygun konumunu bulacaktır.

2.2.1. Konumun ve Dokunun Modellenmesi

Doku bilgisini modellemek için Gabor ya da SIFT benzeri öznitelikler kullanılabilir. Biz bu çalışmada algoritmanın hızı açısından daha hızlı bir alternatifi tercih ettik. Doku bilgisini yüzdeki öznitelik noktalarının etrafındaki dikdörtgensel bölgelerden çıkardık. Bu doku bilgileri için, [5]'de olduğu gibi, altuzay modelleri (PCA gibi) eğittik ve doku bilgisinin ifadesi için p adet altuzay katsayısı kullandık. Konum parametreleri doğrudan x ve y koordinatları olarak ifade edilebilir. Konum parametrelerinin ($P(L)$) bileşik dağılımının modellenmesi, parametrik bir olasılık dağılım fonksiyonu kullanılarak yapılabilir.

2.2.2. Algoritma

Denklem (5)'in logaritması alınarak log-olabilirlik elde edilir:

$$\log(P(T, L)) = \log(P(L)) + \sum_{i=1}^N \log(P(t_i|l_i)). \quad (7)$$

Burada her bir konumun doku bilgisi de konuma bağımlı olduğu için,

$$\phi(L) = \log(P(T, L)), \quad (8)$$

şeklinde sadece konuma bağlı olarak tanımlanan bir fonksiyon tanımlayabiliriz. Buna göre amacımız, en büyük $\phi(L)$ değerini veren bileşke konum vektörü L 'yi bulmaktır:

$$\hat{L} = \operatorname{argmax}_L \phi(L). \quad (9)$$

Bu vektörü bulabilmek için aşağıdaki gibi bir gradyan tırmanış algoritması kullandık:

$$L^{(n)} = L^{(n-1)} + k_n \nabla \phi(L^{(n-1)}). \quad (10)$$

Burada n iterasyon numarasını göstermektedir. Konum vektörü L , N yüzdeki toplam nokta sayısı olmak üzere şu şekilde de ifade edilebilir:

$$L = [x_1, y_1, x_2, y_2, \dots, x_N, y_N]. \quad (11)$$

Bu durumda ϕ fonksiyonunun gradyanı şöyle bulunur:

$$\nabla \phi(L) = [\partial \phi / \partial x_1, \partial \phi / \partial y_1, \dots, \partial \phi / \partial y_N]^t \quad (12)$$

. Tek bir öznitelik i için:

$$\partial \phi / \partial x_i = \frac{\partial}{\partial x_i} \log P(L) + \sum_{i=1}^N \frac{\partial}{\partial x_i} \log P(t_i | l_i) \quad (13)$$

ve

$$\partial \phi / \partial y_i = \frac{\partial}{\partial y_i} \log P(L) + \sum_{i=1}^N \frac{\partial}{\partial y_i} \log (P(t_i | l_i)). \quad (14)$$

Biz bu çalışmada, $P(L)$ dağılımını, dışbükey bir hedef fonksiyonuna yol açan çok-değişkenli bir Gauss dağılımı olarak modelledik. $\log (P(t_i | l_i))$ şeklinde ifade edilen doku parametrelerinin dağılımının modellenmesi için ise bağımsız modeldeki gibi GMM modeli kullanılmıştır. Her yüz özniteliği i için, aşağıdaki gibi GMM doku skoru imgesi bulunur:

$$I_i(x, y) = \log (P(t_i(x, y) | l_i)). \quad (15)$$

Bu modelleri kullanarak hedef fonksiyonu ϕ 'yi aşağıdaki gibi yazabiliriz:

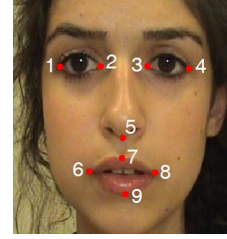
$$\phi(L) = \frac{-\beta}{2} (L - \mu)^T \Sigma^{-1} (L - \mu) + \sum_{i=1}^N I_i(x_i, y_i). \quad (16)$$

Burada μ ortalama konum vektörü ve Σ^{-1} duyarlılık matrisi (ortak değişinti matrisinin tersi), eğitim aşamasında öğrenilen konum bilgileridir. β , ayarlanması mümkün olan bir katsayıdır. Sonuç olarak; $G_{x,i}$ ve $G_{y,i}$, i numaralı özniteliğin Gauss karışım modelinden elde edilen skor alanı resminin, yani $I_i(x, y)$ 'nin x ve y yönlerindeki gradyanları ve \mathbf{G} de tüm öznitelik konumları için gradyan değerlerinin koleksiyon vektörü olmak üzere, konumlar için gradyan tırmanış adımı aşağıdaki gibi yazılır:

$$L^{(n)} = L^{(n-1)} + k_n (-\beta \Sigma^{-1} (L^{(n-1)} - \mu)) + \mathbf{G}, \quad (17)$$

$$\mathbf{G} = \begin{bmatrix} G_{x,1}(l_1^{(n-1)}) \\ G_{y,1}(l_1^{(n-1)}) \\ \dots \\ G_{x,N}(l_N^{(n-1)}) \\ G_{y,N}(l_N^{(n-1)}) \end{bmatrix}. \quad (18)$$

Burada k_n , her adım n 'de farklı değerler alacak şekilde uyarlanabilen adım boyudur. Algoritma iterasyon sonucu elde edilen konumlar fazla değişmeyinceye kadar çalıştırılır.



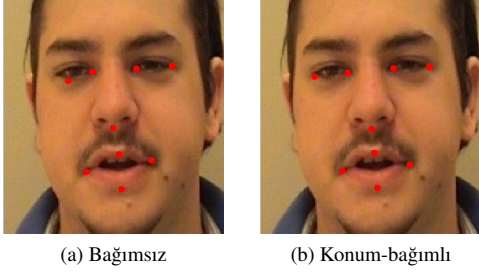
Şekil 1: Eğitim ve test aşamasında kullanılan yüz öznitelikleri

Tablo 1: Farklı öznitelikler için kullanılan eğitim parametreleri

Öznitelik	P	B	H
1	30	8x8	2
2	30	8x8	1
3	30	10x10	2
4	30	5x5	2
5	20	5x5	2
6	50	12x12	1
7	50	10x13	2
8	50	12x12	2
9	50	19x19	2

3. Deneysel Sonuçlar

Eğitim amacıyla; yüz özniteliklerinin doğru konumlarının önceden el ile işaretlenip kaydedildiği, toplam 316 adet insan yüzü resminden oluşan bir veritabanındaki resimler kullanılmıştır. Bu veritabanında, yüz öznitelik bölgelerindeki doku bilgisi, bir PCA altuzay modeli eğitmek için kullanılmıştır ([5],[6]). Kullanılan yüz öznitelikleri; gözlerin sağ ve sol köşeleri, burnun uç noktası, dudağın sağ, sol, alt ve üst köşeleri olmak üzere toplam dokuz adettir. Bu dokuz öznitelik için, dikdörtgensel pencere alanlarındaki doku verileri kullanılarak, farklı boyutlarda PCA altuzayları elde edilmiştir. Ayrıca özniteliklerin konum bilgileri kullanılarak her öznitelik için ayrı olarak ortalama konumlar ve konumların ortak değişinti matrisleri çıkarılmıştır. Resmin farklı bölgelerindeki aydınlanma farklarının etkisini azaltabilmek için, eğitim ve test aşamalarında, [7]'de önerilen 4 bölgeli uyarlanırlar histogram denkleştirme yöntemi uygulanır. Daha hızlı bir öznitelik konum araştırması yapabilmek için, 320x300 boyutlu resimler, eğitim ve test aşamalarından önce 80x75 boyutlu olacak şekilde küçültülmüştür. Eğitilen ve daha sonra test resminde aranan yüz özniteliklerinin yüz üzerindeki konumları, Şekil 1'de gösterilmiştir. Her öznitelik için kullanılan farklı eğitim parametreleri, Tablo 1'de gösterilmiştir. Bu tabloda P sütunu, öznitelik için oluşturulan PCA altuzay modelinin boyutunu; B sütunu, özniteliği eğitmek ve daha sonra aramak için kullanılan pencere boyutunu; H sütunu ise histogram denkleştirme yöntemini gösterir. H sütununun 1 olduğu özniteliklerde



Şekil 2: Bağımsız ve konum-bağımlı modelden elde edilen yüz öznitelik konumları

Tablo 2: Bağımsız ve konum-bağımlı modellerdeki piksel hatası miktarları ve AAM ile karşılaştırma.

Hata	Bağımsız	Konum-bağımlı	AAM
Ortalama	5.86	4.70	8.13
En büyük	29.55	15.78	17.84

kırmızı, yeşil ve mavi bantlar için ayrı ayrı histogram denkleştirme yapılır ve resim daha sonra gri tonuna dönüştürülür; bu değerin 2 olduğu özniteliklerde ise resim ilk önce gri tonuna dönüştürülür ve daha sonra tek bantta histogram denkleştirme yapılır. Tablo 1’de gösterilen eğitim parametreleri, deneysel olarak bulunmuştur ve yapılan deneylerde en iyi sonuçları veren değerler kullanılmıştır. Kişiden kişiye büyük farklılıklar gösterebilen dudak ve çene bölgelerinde, daha büyük boyutlu PCA altuzay modelleri eğitmek ve daha büyük bir doku arama penceresi kullanmak gerekmiştir. Sınama aşamasında, gerçek koşullara uygunluk düşünülerek, daha önceden eğitim verisinde kullanılmamış 100 adet insan yüzü resmi kullanılmıştır. Bağımsız model için bölüm 2.1’de anlatıldığı gibi PCA katsayıları ve konum vektörleri birleştirilir ve GMM modelleri kullanılarak skorlar hesaplanır. Her öznitelik için en yüksek skoru veren piksel seçilir. Bağımsız modelden gelen sonuçlar konum-bağımlı modelin çözümünü ilklendirmek için kullanılır. Bölüm 2’de açıklanan yöntem kullanılarak bir döngü içinde özniteliklerin doku ve konumları iyileştirilir. Bağımsız ve konum-bağımlı model için örnek sonuçlar Şekil 2’de görülebilir.

Bu yöntem kullanılarak 100 adet, 320x300 boyutlu sınama resminde, bağımsız ve konum-bağımlı modellerle elde edilen konumlardaki hata miktarları Tablo 2’de gösterilmiştir. Bu sınama resimleri için, daha önceden el ile işaretlenen noktalar referans olarak alınmış ve yöntemin bulduğu sonuç ile bu noktalar arasındaki Öklid mesafeleri hesaplanmıştır.

3.1. Aktif Görünüm Modeli ile Karşılaştırma

Hazır bir Aktif Görünüm Modeli (AAM) gerçekleştirimi olan AAM-API [8] kullanılarak, bu çalışmada önerilen yöntem ile karşılaştırma yapılmıştır. Eğitim ve sınama için bölüm 3’de tarif edilen veri seti kullanılmıştır. Ortalama ve en

yüksek hata değerlerinin bu çalışmada önerilen yöntem ile karşılaştırılması, Tablo 2’de gösterilmiştir.

AAM metodunun bir avantajı araştırma algoritmasının global poz değişimlerini göz önüne almasıdır. Bizim algoritmamız ise yüz resimlerinin global poz olarak çok fazla farklı olmadığı durumlarda yüz noktalarının insanlar arasında farklardan dolayı ortaya çıkan olasılıksal dağılımlarını modellemektedir. Bu yüzden bizim algoritmamıza yüz tanılama sonucunun girdi olarak verilmesi kritik öneme sahiptir. Algoritmamızın global poz değişimlerini de kapsayacak şekilde geliştirilmesi ilerdeki çalışma planlarımız arasındadır.

4. Vargılar ve Gelecek Çalışmalar

Bu çalışmada önerilen bağımsız ve konum-bağımlı istatistiksel modeller yüzdeki kritik noktaların bulunmasında ümit verici başarımlar göstermiştir. Konum-bağımlı model, bağımsız modelin üzerine dikkat çekici bir iyileştirme sağlamaktadır. Ayrıca önerilen konum-bağımlı yöntem aynı veri üzerinde AAM modelinden daha iyi başarımlar göstermiştir.

İleride, daha iyi başarımlar veren doku parametreleri geliştirmek üzerine çalışacağız. Ayrıca global poz telafisinin yöntemle eklenmesi önemli bir kazanç olacaktır.

5. Kaynakça

- [1] T.F. Cootes, C.J. Taylor, 'Statistical models of appearance for medical image analysis and computer vision', SPIE Medical Imaging, San Diego, CA, USA, 2001.
- [2] Stan. Z. Li, Juwei Lu, 'Modeling Bayesian estimation for deformable contours', in Proceedings of 7th IEEE International Conference on Computer Vision, Kerkyra, Corfu, Greece, 1999.
- [3] Fei Zuo, Peter H. N. de With, 'Fast facial feature extraction using a deformable shape model with Haar-wavelet based local texture attributes', International Conference on Image Processing, Singapore, 2004.
- [4] Cooray, Saman and O'Connor, Noel E., 'Facial feature extraction and principal component analysis for face detection in color images', International Conference on Image Analysis and Recognition, Porto, Portugal, 2004.
- [5] P.M. Hillman, J.M. Hannah, P.M. Grant, 'Global fitting of a facial model to facial features for model-based video coding', International Symposium on Image and Signal Processing and Analysis, Rome, Italy, pp. 359-364, 2003.
- [6] Erol Ozgur, Berkay Yilmaz, Harun Karabalkan, Hakan Erdogan, Mustafa Unel, 'Lip Segmentation Using Adaptive Color Space Training', International Conference on Auditory and Visual Speech Processing, Tangalooma, Australia, 2008.
- [7] Uwe Meier, Rainer Stiefelhaagen, Jie Yang, Alex Waibel, 'Towards unrestricted lip reading', International Journal of Pattern Recognition and Artificial Intelligence, Hong Kong, 1999.
- [8] The AAM-API, <http://www2.imm.dtu.dk/aam/aamapi/>