

TPS Yöntemiyle MPEG-4 Uyumlu Canlandırılabilir 3B Yüz Modelleri Elde Etme

Obtaining MPEG-4 Compliant Animatable 3D Face Models by Using TPS Method

Nesli Erdoğan¹, Rémy Ethève¹, Jean-Luc Dugelay¹

1. Multimedia Communications Department
EURECOM

{nesli.erdogmus, remy.etheve, jean-luc.dugelay}@eurecom.fr

Özetçe

Yüz tanıma, poz ve aydınlatma değişikliklerine karşı önemli katkılar sağlayan 3D yüzey verisi, yüz ifadeleri için tek başına yetersiz kalmaktadır. Bu kapsamda önerilen sistemler genel olarak, ifadelerden etkilenmeyen bölgeleri ele alma, yüz bölgelerini dayanıklılıklarına göre ağırlıklandırma gibi yüz ifadesi etkisini yok etme ya da bu etkiden kaçınma yoluna gitmektedirler. İfadelerin yüz tanımadaki olumsuz etkilerinin üstesinden gelmenin bir diğer yöntemi olarak bu ifadelerin test özneleri için elde edilmesi ve/veya öğrenilmesi önerilebilir. Bu bildiride, bu amaç üzerine TPS (Thin Plate Spline – İnce Levha Kobra Eğrisi) yöntemi kullanılarak statik yüz modellerinden, ifadelerin canlandırılabilceği dinamik modeller elde edilmesi yer almaktadır. Başlangıç için, öznelerin 3B, önden ve nötr yüz modellerinin ve bu modeller üzerinde MPEG-4’te tanımlı bazı öznelik noktalarının elde mevcut olduğu varsayılmaktadır. Daha sonra bu noktalar ve modellerin kendileri kullanılarak canlandırılabilir soysal modelin biçimi değiştirilmekte ve böylece öznelerin canlandırılabilir modelleri elde edilmektedir.

Abstract

3D surface data which yields significant improvements in face recognition against pose and illumination variations, is insufficient by itself in the presence of facial expressions. The proposed methods in this scope, generally choose to eliminate or avoid the adverse effect of the facial expressions by analyzing only the expression-proof regions or weighing facial regions according to their robustness. In order to overcome facial expressions issue for face recognition, the facial expressions for each object can be proposed to be obtained and learned. In this paper, we present an approach to obtain dynamic models on which the facial expressions can be animated from the static ones by utilizing the TPS (Thin Plate Spline) method. Initially, 3D frontal and neutral face models of the subjects and some of the MPEG-4 specified feature points on that models are assumed to be available. Afterwards, an animatable generic model is warped by using those models and points and hence animatable models for the subjects are obtained.

1. Giriş

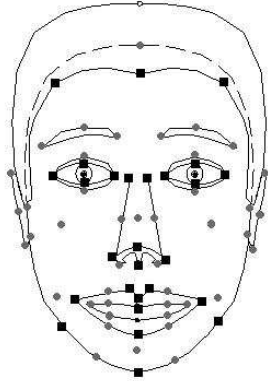
Yüz tanıma problemi çok uzun zamandır bilim camiasının ilgi odaklarından biri olmasına rağmen, yüz ifadelerinden kaynaklanan sınıf içi çeşitlilikler bir sorun olmaya devam etmektedir. Başlarda kullanılan yüz imgelerine son yıllarda yaygın olarak 3B yüz yüzey bilgisi eklenmiştir. 3B şekil verisi doğası gereği poz ve aydınlatma çeşitliliğinden etkilenmez olduğundan, bu durumlarda yüz tanımaya önemli katkılar sağlamıştır. Fakat konu yüz ifadeleri olduğunda, 3B veri bazı durumlarda imgeden de fazla kötü etkilenmektedir. Bu sorunun çözümü için yalnızca 3B ya da 2B+3B çözümler üretilmeye çalışılmıştır.

Yüz ifadesi çeşitliliğın getirdiğı olumsuz etkiyi bertaraf etme adına yüz ifadelerinden etkilenen bölgeleri dikkate almama ya da sisteme etkisini asgariye indirme çalışmaları mevcuttur. Bu yaklaşım hem 2B hem de 3B veriye uygulanabilir. 3B uygulamalarına yönelik örneğın, [1] ve [2]’de önerilen yüz tanıma yöntemlerinde bölgesel uygunluklar göz önüne alınarak yerel yüzey değişikliklerinin eşleştirmeye etkisi azaltılmıştır. [1]’de yapılan yüzey eşleştirmesinde benzerlik derecesi sabit bölgelere daha fazla önem verilerek hesaplanırken, [2]’de çakıştırma için “bölgesel” ortalama modeller kullanılmıştır. Bunların dışında [3]’te ise 3B yüz tanıma yalnızca ifadelerden etkilenmeyen bölgelerin ele alınmasının sonuçlara olumlu etkisi gösterilmiştir.

Bu bildirideki amaç verilen örneklerdeki gibi yüz ifadelerinden kaçınmaktansa bu ifadeleri kayıtlı veritabanındaki modellere uygulamaktır. Benzer fakat tersten bir yaklaşımda [4], 3B Biçimi Değıştirilebilir Model (3DMM – 3D Morphable Model), çakıştırma işleminin ön adımı olarak ifadelı yüz modeline yakınlştırılmış ve daha sonra iki model arasındaki bağıntı kestirilerek çakıştırılmış, ardından ifade düzelenmiştir. Fakat bu ve bunun gibi yaklaşımların en büyük dezavantajı istatistiksel ve üretken bir 3B yüz modeli olan “3B Biçimi Değıştirilebilir Model”i geliştirme zorunludur.

Bu bildiride ise varsayılan ön bilgi bir soysal model ve canlandırılacak yüz modeli üzerinde işaretli 29 öznelik noktadır. Aslında MPEG-4 [5] uyumlu canlandırma motorları için daha fazla nokta gereklidir fakat bu çalışmada kalan noktalar yalnızca soysal modelde işaretlenmiş ve çarpıtma sonucu özne modeli için bulunmaları gereken noktalara yakınsayacakları varsayılmıştır. Şekil 1.’de MPEG-4’te tanımlı noktalar gösterilmiştir. Bunlardan dilde ve

dişlerde olanları kullanılmamış, gözlerde 10 (5x2), burunda 6, ağızda 7 ve yüz çeperinde 6 olmak üzere toplam 29 nokta tüm modeller için, kalan noktalar ise yalnızca sosyal model için işaretlenmiştir.



Şekil 1. Canlandırılacak modellerde (■) ve yalnızca sosyal modelde (●) işaretlenecek öznelik noktaları

Bildirini kalan kısmı şu şekilde düzenlemiştir: 2. bölümde TPS temelli yöntem ile canlandırılabilir modellerin elde edilmesi ayrıntılı olarak anlatılmaktadır. Önerilen yaklaşımın uygulandığı Bosphorus veritabanı ve özelliklerine 3. bölümde yer verilmiştir. Son olarak 4. bölümde uygulama sonuçları ve 5. bölümde gelecekte yapılacak çalışmalar açıklanmıştır.

2. Önerilen Yöntem

3B canlandırılabilir yüz modelleri elde etmek için bu bildiride canlandırılabilir sosyal bir modeli TPS yöntemi ile çarpıtma (warping) yolu benimsenmiştir. TPS yöntemi 1989'da Fred L. Bookstein tarafından biyomedikal imge analizi bağlamında kullanıldı[6]. Ardından daha da yaygınlaşan bu yöntem ince metal levhaların nokta kısıtları altında sahip olduğu eğilme enerjisi analogisi üzerine kurulmuştur.

3B yüzeyler S ve T ve sırasıyla bu yüzeylerde tanımlı karşılıklı noktalar kümesi P_i ve M_i için, S'yi değiştirerek T'ye yaklaşan T' yüzeyi şu şekilde tanımlanabilir:

$$T = \{(x', y', z') \mid \forall (x, y, z) \in S, x' = x, y' = y, z' = z + f(x, y)\} \quad (1)$$

TPS algoritması buradaki $f(x, y)$ fonksiyonunu hesaplar.

$$f(x, y) = a_1 + a_2 x + a_3 y + \sum w_i U(|P_i - (x, y)|) \quad (2)$$

$$U(r) = r^2 \ln(r) \quad (3)$$

$$r = \sqrt{x^2 + y^2} \quad (4)$$

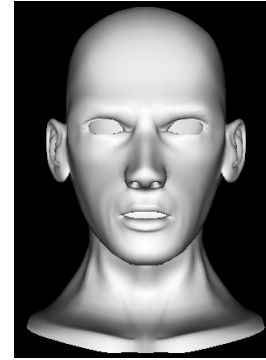
Burada w_i , $i \in \{1, 2, \dots, n\}$ ağırlıkları göstermektedir. (2)de verildiği üzere aradeğerleme fonksiyonu iki kısımdan oluşmaktadır: yüzeyin kısıt noktalarına uyumu için gerekli ilgin (affine) dönüşümü sağlayan ilgin kısım ($a_1 + a_2 x + a_3 y$) ve çarpıtma kısmı ($\sum w_i U(|P_i - (x, y)|)$).

Sosyal model olarak Şekil 2.'de gösterilen açık ağız ve göz deliklerine sahip model kullanılmıştır. Bu sayede statik yüz modellerinin tersine gözler ve ağız canlandırılabilir

durumdadır. TPS algoritmasının çokterimliliğinden dolayı kısıt noktalarından uzakta kalan bölgelerde ıraksama gözlemlendiğinden, sosyal kafa modelinin arka kısmına hem hedef hem kaynak olarak kullanılacak noktalar eklenmiştir.

2.1. Hizalama

TPS algoritması uygulanmadan önce hedef yüz (özne) ile kaynak yüzün (soyasal) doğru bir şekilde hizalandığından emin olmak gerekmektedir. Bu sebeple, ilk olarak kaynak model hedef modele göre kabaca ölçeklenir. Bu ölçekleme için katsayılar basit bir yaklaşımla x, y ve z açıklıkları karşılaştırılarak hesaplanır. Ardından, iki modelde de işaretlenmiş iki noktalar kümesi en küçük kareler yöntemi ile bir dönüşüm tanımlar ve böylece kaynak model hedefe göre hizalanır.



Şekil 2. Sosyal model

2.2. Çarpıtma

Bir sonraki adım çarpıtma adımdır. Bu adım iki kısımda gerçekleştirilir:

2.2.1. Kaba Çarpıtma

Bu kısımda iki modelde işaretlenmiş olan ortak noktalar hedef ve kaynak nirengi noktaları olarak kullanılır ve böylece doğrusal olmayan bir çarpıtma dönüşümü tanımlar. Bu dönüşüm ile kaynak modeldeki noktalar eşleniklerine yaklaşırlar ve kalan diğer noktalar için Bookstein'in TPS algoritması kullanılarak aradeğerlendirme yapılır.

2.2.2. İnce Çarpıtma

Bu evreye gelindiğinde artık iki yüzey birbirine oldukça yaklaştırıldığından, sosyal modeldeki tüm noktaların, hedef modeldeki eşlenikleri olarak o yüzey üzerindeki en yakın noktalar seçilir. Fakat bütün noktaları kullanmak yüklü ve anlamsız olacağından, hedef modelde alt örnekleme yapılır. Böylece kaynak yüzey daha fazla sayıda eşlenik nokta sayesinde hedefe en yakın duruma gelir ve düzgünlüğünü korur.

Bu kısımda ayrıca, gözler etrafında işaretlenen dört nokta kullanılarak göz boşlukları yerini dolduracak kürenin çapı (R) ve merkezi (x, y, z) bulunur.

2.3. Doku kaplama

Son olarak, hedef model dokusu kaynak modele kopyalanır.

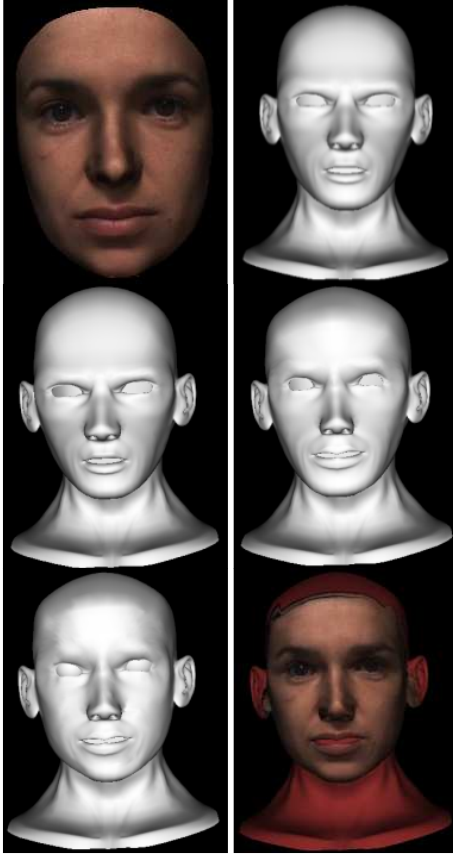
3. Bosphorus Veritabanı

Önerilen yöntemin denenmesi için İstanbul, Boğaziçi Üniversitesi'nde gerçekleştirilen Enterface'07 çalışmayı kapsamında hazırlanan Bosphorus 2B-3B yüz veritabanı kullanılmıştır[7]. Zengin bir poz ve ifade çeşitliliğine sahip bu veritabanında 105 kayıtlı özne için çeşitli poz ve ifadelerde toplam 4666 3B yüzey ve buna öakışık 2B renk verisi bulunmaktadır. Ayrıca, tüm modeller için elle işaretlenmiş 24 mirengi noktası için 2B ve 3B koordinatlar da sunulmaktadır.

4. Sonuçlar

4.1. Uygulama Sonuçları

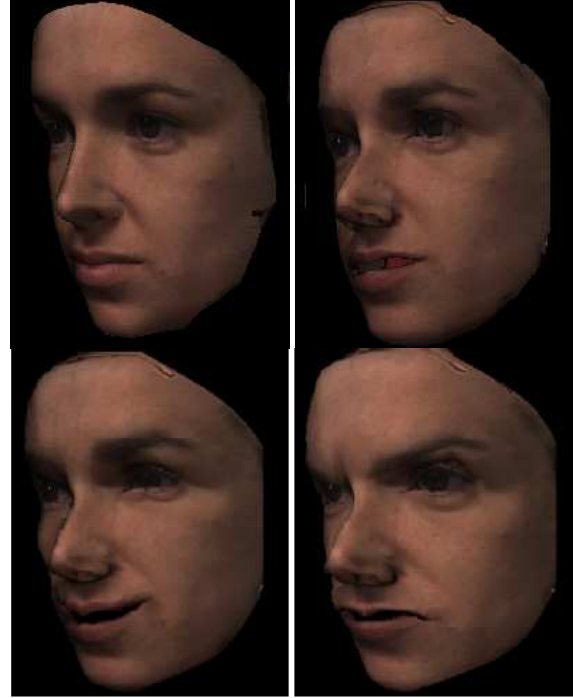
Ön-işleme safhasında gürültülü yüzey verisi iki yanlı filtreleme yöntemi ile düzgünleştirilmiştir[8]. Bu işlemden sonra 2. bölümde değinildiği üzere sırasıyla ölçekleme, hizalama, kaba ve ince çarpıtma adımları uygulanmış, son olarak doku kopyalanmıştır. Bu işlemler sonucu elde edilen soysal model Şekil 3.'te verilmektedir.



Şekil 3. Statik modelden canlandırılabilir modele doğru adımlar: a. Hedef model b. Ölçekleme c. Hizalama d. Kaba Çarpıtma e. İnce Çarpıtma f. Doku kopyalama

Canlandırılabilir model elde edildikten sonra canlandırma motoru kullanılarak yüz ifadelerinin benzetimi yapılmıştır. Bu benzetim için the visage|life™ canlandırma aracı kullanılmıştır[9]. MPEG-4 FA (Facial Animation)

standardına göre çalışan bu araçta önceden tanımlı bir çok ifade bulunabileceği gibi yeni ifadeler de tanımlamak mümkündür. Şekil 4.'te bir hedef model, elde edilen canlandırılabilir model ve iki örnek ifade gösterilmektedir. Bu benzetimde elbette canlandırma aracının kalitesi ve içerdiği ifadelerin düzgünlüğü büyük rol oynamaktadır. Fakat şekilde de görüldüğü gibi basitçe tasarlanmış ifadeler bile yüksek doğrulukta canlandırılabilirlerdir.



Şekil 4. a. Hedef model b. Elde edilen canlandırılabilir model c. Gülümserken d. Kaş çatarken

Statik bir modelin bu metot ile canlandırılabilir bir modele dönüştürülmesi "Pentium(R) Dual-Core" bir işlemci ile yaklaşık 30 saniye almaktadır.

Şekil 5.'te daha çok sayıda örnek için hedef model, elde edilen canlandırılabilir model ve gülümseyen model üçlüleri gösterilmektedir.

4.2. Genel Sonuç

Sonuçlarda görüldüğü üzere, kolay ve hızlı bir yöntemle doğal 3B verisi bulunan bir öznenin gülümser ya da somurtur halde nasıl görüleceğinin benzetimi gerçekçi bir biçimde yapılabilmektedir. TPS algoritması sayesinde elde edilen canlandırılabilir model özneye oldukça benzemekte ve yüz yüzeyi düzgün olmaktadır. Bu sayede, gülümser şekilde yakalanan 2B ve/veya 3B veri ile daha doğru bir karşılaştırma yapmak mümkün olacaktır.

5. Gelecekte Yapılacak Çalışmalar

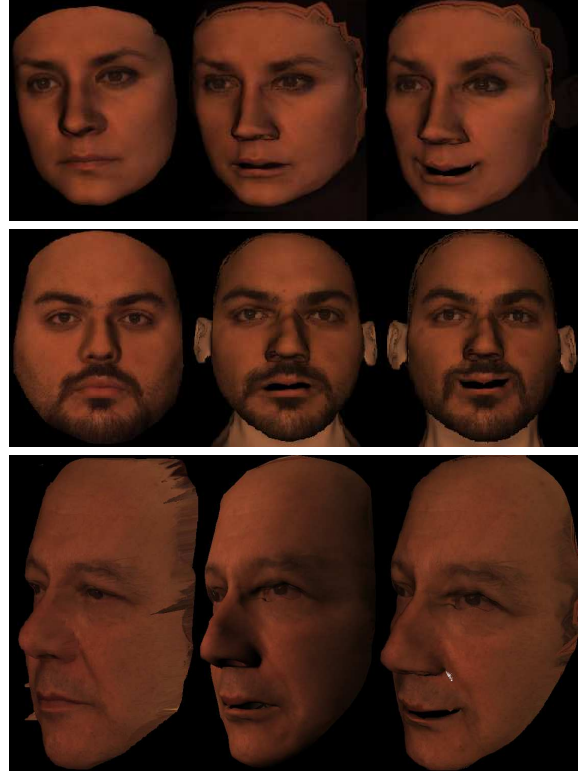
Gelinen bu noktadan ileriye yönelik birçok iyileştirme ve uygulama mümkündür. Öncelikle, iyileştirmeler ele alınırsa, özellikle yüz çeperlerinde bulunan noktaların belirgin tanımlayıcıları olmamasından kaynaklanan ve özellikle ölçekleme aşamasını etkileyen problemler bu noktaların yerini

daha iyi tanımlı noktalara bırakmasıyla çözümlenmelidir. Ayrıca tüm sistemin otomatikleştirilmesi adına her yüzde işaretlenen 29 noktanın otomatik olarak tesbiti üzerinde çalışmalarımız sürmektedir. Bu iyileştirmelere ek olarak veritabanında bulunan gülümser ifadelerle 2B ve 3B'de karşılaştırmalar yapılmalı, benzerlik dereceleri ölçülmelidir.

Uygulama aşamasında ise birçok seçenek mevcuttur. İki ana seçenek salt 3B ya da asimetric 2B-3B arasındadır. Salt 3B yaklaşımlarda, doğal gülümser ve canlandırılmış gülümser modeller için yapılan yüz tanıma testleri performansları, doğal nötr modeller kullanılarak yapılan testleri ile karşılaştırılabilir. Benzer bir uygulama asimetric için de geçerlidir. Bu iki yaklaşımda da test verisindeki ifade tespit edilip galeriye uygulanabilir veya tüm olası ifadeler tanıma sistemine öğretilir.

6. Kaynakça

- [1] Ben Amor, B.; Ardabilian, M.; Liming Chen, "Toward a region-based 3D face recognition approach," *IEEE International Conference on Multimedia and Expo, 2008*, vol., no., pp.101-104, June 23 2008-April 26 2008.
- [2] Alyuz, N.; Gokberk, B.; Akarun, L., "A 3D Face Recognition System for Expression and Occlusion Invariance," *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, 2008. BTAS 2008.*, vol., no., pp.1-7, Sept. 29 2008-Oct. 1 2008.
- [3] Chang, K.I.; Bowyer, W.; Flynn, P.J., "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , vol.28, no.10, pp.1695-1700, Oct. 2006.
- [4] Amberg, B.; Knothe, R.; Vetter, T., "Expression invariant 3D face recognition with a Morphable Model," . *8th IEEE International Conference on Automatic Face & Gesture Recognition, 2008. FG '08*, vol., no., pp.1-6, 17-19 Sept. 2008.
- [5] Moving Picture Experts Group, ISO/IEC 14496 – MPEG-4 International Standard, www.cse.it/mpeg.
- [6] Bookstein, F.L., "Principal warps: thin-plate splines and the decomposition of deformations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.11, no.6, pp.567-585, Jun 1989.
- [7] A. Savran, O. Çeliktutan, A. Akyol, J. Trojanova, H. Dibeklioglu, S. Esenlik, N. Bozkurt, C. Demirkır, E. Akagündüz, K. Çalışkan, N. Alyüz, B. Sankur, İ. Ulusoy, L. Akarun, T. M. Sezgin, "3D Face Recognition Performance Under Adversarial Conditions", in *Proc. eINTERFACE'07 Workshop on Multimodal Interfaces*, Istanbul, Turkey, July 2007.
- [8] N. Bozkurt, E. Akagündüz, İ. Ulusoy, U. Halıcı, "Yüz Tarayıcısı Verilerinin İyileştirilmesi için 3B Veri İşleme", *Sinyal İşleme ve İletişim Uygulamaları Kurultayı '09*, Antalya, Turkey, April 2009.
- [9] Visage Technologies, The Character Animation Company
www.visagetechologies.com/products_life.html



Şekil 5. Bazı örnek sonuçlar