

Yüz Özniteliklerinin Takibi ve İşaret Dili için İfade Tanıma

Facial Feature Tracking and Expression Recognition for Sign Language

İsmail Arı, Lale Akarun

Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
{ismailar,akarun}@boun.edu.tr

Özetçe

Yüz ifadesi, işaret dilinde olumlu/olumsuz vurgusu yapan, anlamı kuvvetlendiren ya da belirsizleştiren kritik bir öneme sahiptir. Bu çalışmada, işaret dilinde yaygın kullanılan yüz ifadelerinin tanınması için, Aktif Şekil Modeli tabanlı bir yüz öznitelik izleyicisi ve Saklı Markov Modeli tabanlı bir ifade sınıflandırıcı içeren bir sistem geliştirdik. Yüz izleyici, değişik pozlardaki yüzleri hızlı ve etkin bir şekilde tanıyabilmek için çok-çözünürlüklü ve çok-pozlu olarak çalışmaktadır. İzleyici ve sınıflandırıcı başarımının ölçümü için ayrıntılı testler yapılmış ve sistemin yüksek başarımla güvenilir olarak çalıştığı görülmüştür.

Abstract

Facial expressions play a vital role in sign language by changing the meaning of the performed sign. In this work, we propose a system composed of a facial feature tracker based on Active Shape Models and a classifier based on Hidden Markov Models to recognize common facial expressions used in sign language. Face tracker works in multi-resolution and multi-view to track faces in different poses fast and effectively. Detailed tests are prepared to report the accuracy of the tracker and classifier, and it is seen that the system works in high accuracy and robustly.

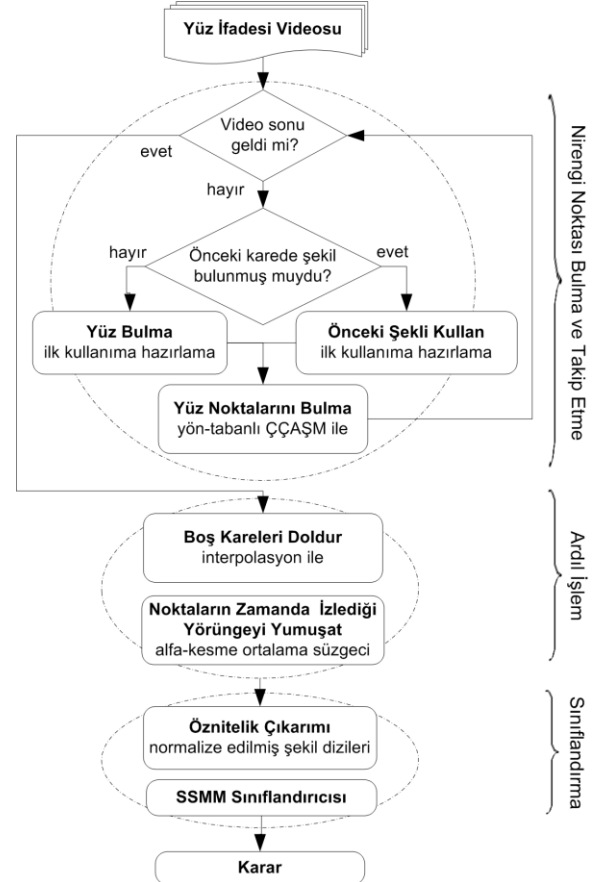
1. Giriş

Bir imge dizisinde, yüz özniteliklerinin otomatik takibi, ifade tanımayı da kapsayan birçok uygulamanın ilk adımıdır. İşaret dili özelinde bakarsak, kelimeler/tümceler yalnızca elle yapılan jestlerden oluşmaz, eş zamanlı yapılan yüz ifadeleri ve baş hareketleri ile bir bütündür. Bu sebeple, işaret dili tanıma amacı güden bir sistem, yüz ifadelerini ve baş hareketlerini doğru tanıyabilmelidir.

Yüz ifadesi takibi ve ifade sınıflama konusunda birçok çalışma yapılmıştır [1,2]. Yapılan çalışmalar temelde üç adımdan oluşur: Yüz bulma ve takibi, yüz özniteliği çıkarımı ve takibi, yüz ifadesi tanıma. Nirengi noktalarından oluşan şeklin yüze uydurulmasını amaçlayan Aktif Şekil Modelleri (AŞM), başarılı sonuçlar verdiği bilinen bir yaklaşımdır [3] ve öznitelik çıkarımında kullanılmaktadır. Özniteliklerin zaman içinde değişiminin modellenmesi ve ardarda gelen özniteliklerden sınıflandırma yapmak için ise Saklı Markov Modelleri (SMM) kullanmak yaygın bir yöntemdir [4]. Yüz ifadesi videolarında, zaman içinde yüz şeklinin nasıl

değiştirdiği bilgisini içeren özniteliklerin sınıflandırılması için de SMM uygun bir yaklaşımdır. Özniteliklerin sürekli değerlerden oluştuğu durumda Sürekli Saklı Markov Modelleri (SSMM) kullanılır.

Bu çalışmanın konusu, işaret dillerindeki yüz ifadelerini ve kafa hareketlerini bir video içinde otomatik olarak tanımadır. Önerdiğimiz sistem iki aşamadan oluşmaktadır: İlkinde, verilen bir videoda, yön-tabanlı Çok-çözünürlüklü AŞM (ÇÇAŞM) ile yüzdeki nirengi noktaları otomatik olarak saptanır ve takip edilir. İkinci aşamada, takip edilen noktalardan oluşan şekil dizisi normalize edilir ve SSMM temelli sınıflandırıcı ile ifade tanınması yapılır. Önerilen sistemin akış şeması Şekil 1'de verilmiştir.



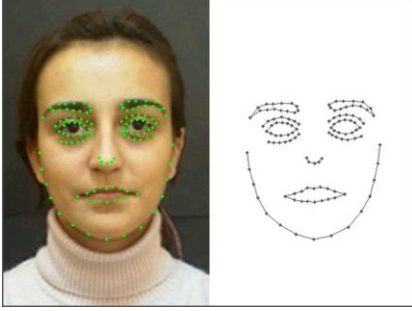
Şekil 1: Önerilen Sistemin Akış Şeması

2. Yüz Özneliklerinin Takibi

2.1. Şekillerin İstatistiksel Analizi

Şekillerin istatistiksel analizi üç temel adımda incelenebilir; şekil bilgisinin toplanması, normalizasyon ve Temel Bileşenler Analizi (TBA). İlk adımda, veritabanında bulunan videolardan N adet kare rasgele alınır ve her bir yüzdeki L adet nirengi noktası elle işaretlenir. Şekil 2'deki gibi işaretlenmiş tüm s_i şekilleri, eğitim kümesi Φ_s 'i oluşturur.

$$s_i = (x_1, y_1, x_2, y_2, \dots, x_L, y_L), \quad i = 1, \dots, N \quad (1)$$



Şekil 2: Seçilen Nirengi Noktaları

Eğitim kümesindeki şekiller Procrustes Analizi ile normalize edilir ve son olarak TBA ile şekil değişimi modellenir. Model; özvektörleri ($\mathbf{E}_{2L} = \mathbf{e}_1 \dots \mathbf{e}_{2L}$), özdeğerleri ($\lambda_i, i = 1, \dots, 2L$) ve ortalama şekli ($\bar{\mathbf{s}} = \frac{1}{N} \sum_{i=1}^N s_i$) içerir. Böylelikle kartezyen düzlemde tanımlanan herhangi bir şekil vektörü, denklem (2) kullanılarak şekil parametrelerine, $\mathbf{b} = (b_1 \dots b_{2L})$, dönüştürülebilir ve tüm özvektörler kullanılırsa denklem (3) ile kayıpsız olarak geri elde edilebilir.

$$\mathbf{b} = \mathbf{E}_{2L}^T (\mathbf{s} - \bar{\mathbf{s}}) \quad (2)$$

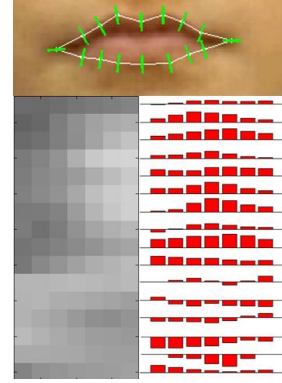
$$\mathbf{s} = \bar{\mathbf{s}} + \mathbf{E}_{2L} \mathbf{b} \quad (3)$$

TBA kullanmamızdaki ilk motivasyon, eldeki veriyi daha az boyutta ifade edebilmektir. Toplam varyansın (özdeğerler toplamının) %95'ini ifade edebilecek sayıda özvektör seçilerek (2) ve (3) işlemleri, bu az sayıdaki özvektör ile yapılır. Boyut azaltma, kartezyen düzlem ile şekil parametreleri arasındaki dönüşümde kayba neden olur fakat bu kayıp gözardı edilebilir, hatta gürültü temizleme olarak yorumlanabilir.

2.2. Aktif Şekil Modelinin Oluşturulması

i 'ninci şeklin j 'ninci nirengi noktasına \mathbf{p}_{ij} diyelim ve bu noktadan geçen profilin gradyan vektörü \mathbf{g}_{ij} olsun. Şekil 3'te, üstte, dudaklardaki noktalarda hesaplanan profiller yeşil ile gösterilmiştir. Her bir profildeki kırmızı kanal değerleri soldaki resmin bir satırında gösterilmiş, sağda ise o satıra karşılık gelen gradyanlar gösterilmiştir. Her bir nirengi noktası için, tüm şekiller üstünden hesaplanan

ortalama gradyan $\bar{\mathbf{g}}_j$ ve eşdeğişinti matrisi \mathbf{C}_j elde edilerek profiller boyunca gözlemlenen aydınlık değeri değişiklikleri modellenmiş olur. Dolayısıyla AŞM, TBA modeli ve profil gradyanları modelinden oluşur.

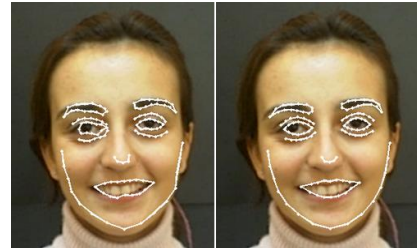


Şekil 3: Dudaklardaki profiller (üstte), her profildeki kırmızı aydınlık değeri (solda) ve profil boyunca gradyan değerleri (sağda)

2.3. Şekli, Test Yüzüne Uydurma

AŞM ile şekli yüze uydurmak için önce ortalama şekilden başlanır. Her bir tekrarlama, en iyi uyan şekil bulunur ve kısıtlanarak uydurulmaya çalışılan şekilde iyileştirme sağlanır ve yakınsadığı duruma kadar döngüye devam edilir. En iyi uyanı bulma işlemi şöyle yapılır: Her bir nokta için, profil boyunca birer kaydırma yapılarak gradyanlar örneklerin (\mathbf{g}_j) ve öğrenilen gradyan modeline en uygun örneğe ait kaydırma miktarı saptanır. Bu miktar, bir sonraki adıma geçmeden önce, o noktanın profil boyunca ne kadar kayması gerektiğini verir. Uzaklık bilgisi için Mahalanobis uzaklığı, $(\mathbf{g}_j - \bar{\mathbf{g}}_j)^T \mathbf{C}_j^{-1} (\mathbf{g}_j - \bar{\mathbf{g}}_j)$, kullanılır.

En iyi olarak bulunan nirengi noktaları, öğrenim kümesindeki şekiller ile uyumsuz (yüze benzemeyen) bir şekli oluşturabilir. TBA kullanılması başat motivasyon, bulunan bu şekli yüze benzeyecek biçimde kısıtlamaktır. Özvektörlerin dik (bağımsız) olduğunu ve özdeğerlerin bu uzaydaki her bir boyuta oturtulan Gauss dağılımının varyansı olduğunu hatırlayalım. Şekli, yüze benzeyecek biçimde kısıtlamak için TBA modeli (2) ile şekil parametrelerine dönüştürürüz ve her bir b_i parametresini $\pm 3\sqrt{\lambda_i}$ arasına düşecek şekilde güncelleriz.



Şekil 4: En iyi uyan şekil ve kısıtlanmış hali

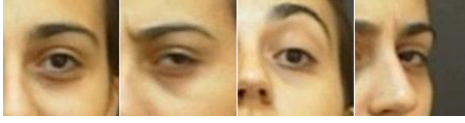
Ardından (3) ile şekli geri dönüştürürsek, Şekil 4'teki gibi eğitim kümesindeki şekillere benzediğini gözlemleriz.

2.4. Çok-çözünürlüklü Yaklaşım

Uydurulmaya çalışılan yüz şeklinin lokal en iyilerde takılmasını engellemek için çok-çözünürlüklü yaklaşım uygulanır: İmge piramidindeki her bir basamak için ayrı bir AŞM eğitilir ve aramaya en az çözünürlüklü olandan başlanır. Bu basamakta bulunan en iyi şekil, bir sonraki basamağın başlangıcı olur ve bu biçimde devam edilerek her basamağın sonunda nihai şekle daha çok yaklaşılır.

2.5. Yön-tabanlı AŞM

İşaret dilinde yüz hareketlerinin yanında kafa hareketleri de bulunduğundan tüm açıları aynı modelde toplamak istenilen varyasyon değişikliğini iyi modelleyemeye sebep olur. Şekil 5'te görüldüğü üzere aynı bölgeye ait renk değerleri her yön için büyük değişiklik gösterebilmektedir. Eğitim kümesi yöne göre altkümelere ayrılır ve her bir altkümeden ayrı bir AŞM eğitilir. Test imgesi geldiğinde modellerle ayrı ayrı aramalar yapılır ve en iyi sonuç seçilir.



Şekil 5: Aynı bölgenin dört farklı yöndeki görünümü

2.6. Videolarda Şekil Takibi

Videolarda yüze ait nirengi noktalarını takip edebilmek için ilk karede döngüye ortalama şekil ile başlanarak en iyi sonuç veren yön bulunur. Peşisıra gelen karelerde bir önceki karenin verdiği sonuç, o karedeki şeklin başlangıç konumu olur ve aramaya oradan başlanır [5]. En iyi uyan yön bulunurken, eşik değerini geçen birden fazla aday çıkarsa önceki karenin sonucuna en çok benzeyeni seçilir. Eşik değerini geçen aday yön çıkmaması durumunda ise o kare boş bırakılır ve bir sonraki karede aramaya sıfırdan başlanır. Bu sayede yanlış bulunmalarla şeklin bozunması engellenmiş olur.

Ardıl işlem olarak, şekil dizisindeki boş bırakılmış kareler interpolasyon ile doldurulur ve bölgesel hataları ortadan kaldırmak için alpha-kesme ortalama süzgeci uygulanır.

3. İfade Tanıma

Her bir şekil dizisindeki tüm şekiller, dizideki ilk şeklin merkezini orijine taşıyan ve birim çembere ölçekleyen dönüşüm matrisi ile çarpılır. Böylelikle normalize edilen şekil dizileri, ifade tanıma için kullanılan öznitelikleri oluşturur.

Özniteliklerin sürekli değerlerden oluşması sebebiyle SSMM kullanılır. İfade tanıma şöyle yapılır: Eğitim aşamasında her sınıf için ayrı bir SSMM eğitilir. Sınıfı bilinmeyen bir test videosu geldiğinde, önce takip ile öznitelikler elde edilir ve ardından en yüksek olasılığı veren modelin sınıfı seçilir.

4. Veritabanı

Önerdiğimiz yöntemi sınavabilmek için BUHMAP veritabanını kullandık [6]. Bu veritabanı hem baş hareketleri

hem de yüz ifadelerinden oluşmakta ve 8 farklı sınıfa ait (1. Yalın, 2: Yukarı doğru başı bir kez kaldırma, 3: Sağa sola baş sallama, 4: Öne doğru başı uzatma, 5: Büzülüş dudaklar, 6: Yukarı aşağı baş sallama, 7: Gülümseme, 8: Yukarı aşağı baş sallama ve gülümseme) toplam 440 video içermektedir. 6 kadın, 5 erkek öznenen 5'er tekrar ile elde edilmiştir. Bu videolardan 4 öznenen alınan 4 sınıfa ait (3, 4, 6 ve 7 numaralı sınıflar) 3'er tekrardan oluşan toplam 48 tanesi elle işaretlenmiştir. İşaretlemede yüz başına 60 nirengi noktası kullanılmıştır. Aynı 4 özne için rasgele alınmış yaklaşık 240 imge, elle işaretlenmiş ve yön tabanlı AŞM eğitimi için kullanılmıştır (30 ön, 10 sağ, 10 sol, 10 yukarı yön; kişi başı 60). AŞM eğitiminde her yüz için 116 nokta kullanılmıştır.

5. Deneyler ve Sonuçlar

5.1. Takip için Deney Kurulumu

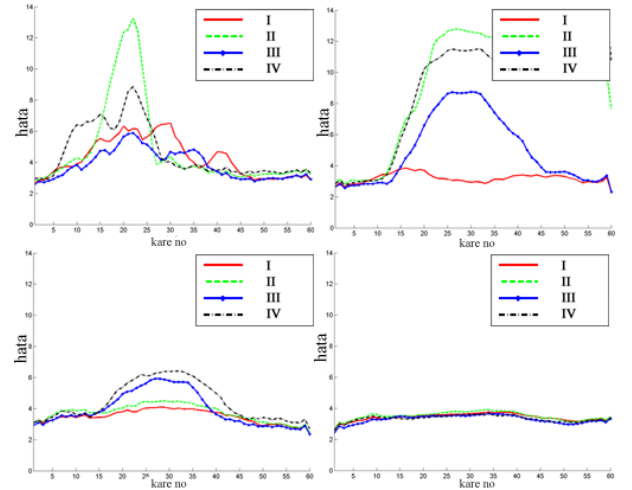
Yön-tabanlı/tek-yönlü ve kişiye özel/genel ÇÇAŞM takipçilerini karşılaştırabilmek için bu dört tipi Tablo 1'deki gibi numaralandırdık. Tek-yönlü yaklaşımda tüm örneklerden tek bir model oluştururken, yön-tabanlı yaklaşımda her bir yön için ayrı modeller oluşturduk. Yine benzer biçimde, genel yöntemde tüm özneleri kullanırken kişiye özel yöntemde her özne için ayrı bir model eğittik.

Tablo 1: ÇÇAŞM Tipleri

	Kişiye özel	Genel
Yön-tabanlı	I	II
Tek-yönlü	III	IV

5.2. Takip için Deney Sonuçları

Şekil 6'da görüldüğü üzere kişiye özel yaklaşımın genel yaklaşımdan daha iyi olduğunu gözlemledik. Yön-tabanlı yöntem, kişiye özel yaklaşımda başarıyı artırırken, genel yaklaşımda başarıyı düşürdü. Figürlerde yatay eksen zamanı gösterirken, dikey eksen yapılan ortalama piksel hatasını göstermektedir.



Şekil 6: Her sınıf için dört tip takipçinin verdiği ortalama sonuçlar

Takip edilen noktaları (116 adet) videolarda işaretlenen noktalar (60 adet) ile karşılaştırabilmek için ortak 52 nokta seçilmiştir. Videoları teker teker inceleyince Şekil 6'da alt sınır olarak gözüken 3-4 piksellik hatanın, işaretli videolar ile ÇÇAŞM'nin eğitildiği işaretli imgeler arasındaki işaretleme farkından kaynaklandığını gözlemledik ve bu değeri hata için eşik olarak aldık.

5.3. İfade Tanıma için Deney Kurulumu

Sınıflandırma deneylerini çok-boyutlu SSMM kullanarak yaptık [7]. Kullanılan Gauss dağılımı ve saklı durum sayısını sırasıyla 2 ve 6 olarak seçtik. Altta deney kümelerini hazırlayıp onları kullanarak Tablo 2'deki testleri oluşturduk:

$\Phi_{1:4}$ kümesi: Yalın ifade hariç 7 sınıftan, 4 özneye ait 5'er tekrar içerir (140 örnek).

Φ_{gt} kümesi: Veritabanı bölümünde verilen 4 sınıftan, 4 özneye ait 3'er tekrar içerir (48 örnek).

Tablo 2: Tanıma başarımı için düzenlenen testler

Test no	Eğitim kümesi	Test kümesi	Eğitim örneği sayısı	Test örneği sayısı
I	Φ_{gt} 'deki 2 tekrar	Φ_{gt} 'deki kalan tekrar	32	16
II	$\Phi_{1:4}$ 'teki 3 tekrar	$\Phi_{1:4}$ 'teki 2 tekrar	84	56
III	$\Phi_{1:4}$ 'teki 3 özne	$\Phi_{1:4}$ 'teki kalan özne	105	35
IV	Φ_{gt} 'deki 3 tekrar	$\Phi_{1:4}$ 'teki kalan tekrarlar	48	32

I., II. ve III. testlerde çapraz doğrulama kullanılmıştır.

5.4. İfade Tanıma için Deney Sonuçları

Tablo 3: İfade tanıma deney sonuçları

Test no	Sınıf sayısı	Eğitim örneği sayısı	Test örneği sayısı	Takipçi tipi	Doğruluk (%)
I	4	32	16	Gerçeklik değerleri	100.00
II	7	84	56	I	84.82
				II	83.57
				III	81.79
				IV	73.93
III	7	105	35	I	72.86
				II	68.57
				III	64.29
				IV	53.57
IV	4	48	32	I	100.00
				II	93.75
				III	93.75
				IV	87.50

Tablo 3'de görüldüğü üzere, SSMM sınıflandırıcısı gerçeklik değerlerini kullanırsa tam başarı sağlıyor (Test I). Takipçinin bulacağı öznitelikleri kullanırsa en iyi başarıyı kişiye özel yön-tabanlı takipçi veriyor. Gerçeklik değerleri ile eğitilen

sistem tam başarıyı yakalayabilirken (Test IV), sınıf sayısı artırılıp yalnızca takip edilen şekil bilgileri kullanılıncaya genel başarı düşüyor (Test II, Test III). Sonuçlar karşılaştırıldığında SSMM, [5]'te Destek Vektör Makineleri ile elde edilen sonuçlardan daha başarılı gözüküyor.

6. Sonuç

Bu çalışmada, videolardan yüz nirengi noktalarının otomatik takibini ve ifade tanımayı inceledik. Çok-çözünürlüklü Aktif Şekil Modelini yön-tabanlı olarak genişlettik ve ardışık imgelerde takip yapabilmek için geliştirdik. Takip edilen nirengi noktalarını normalize ederek Sürekli Saklı Markov Modeli tabanlı sınıflandırıcıya vererek ifade tanımayı hedefledik. Çalışmanın temel katkısı, işaret dilinde yapılan yüz ifadesi ve baş hareketlerini içeren videolarda nirengi noktası takibi için ÇÇAŞM'nin geliştirilmesi ve SSMM tabanlı sınıflandırıcı ile birlikte tamamlanmış bir ifade tanıma sistemini sunmasıdır.

Önerilen yöntemi yüz ifadesi ve baş hareketi içeren videolarda sıradık [6] ve başarılı sonuçlar elde ettik. Takip aşamasında kişiye özel/genel ve yön-tabanlı/tek-yönlü yaklaşımları karşılaştırdık ve kişiye özel yön-tabanlı yöntemin en başarılı sonuçları verdiğini gözlemledik. Bunun yanında, takipçiyi çok-çözünürlüklü yapmanın sistemin başarısında başat rol oynadığını gözlemledik.

Geliştirilmeyi bekleyen adımlar; eğitim kümesinin otomatik olarak bölütlenmesi, sistemin gerçek zamanda çalışması ve araçutusu haline getirilerek başka sistemlerle kolayca entegre edilebilmesi olarak gözükmektedir.

7. Teşekkür

Bu çalışma, TÜBİTAK/107E021 projesince desteklenmiştir.

8. Kaynakça

- [1] Fasel, B. ve Luetin, J., "Automatic Facial Expression Analysis: A Survey", *Pattern Recognition*, Vol. 36, No. 1, pp. 259-275, 2003.
- [2] Pantic, M. ve Rothkrantz, L., "Automatic Analysis of Facial Expressions: the State of the Art", *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, Vol. 22, No. 12, pp. 1424-1445, 2000.
- [3] Cootes, T. ve Taylor, C., "Statistical models of appearance for computer vision", *World Wide Web Publication*, Şubat, 2001.
- [4] Rabiner, L., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257-286, 1989.
- [5] Ari, I., Uyar, A. ve Akarun, L., "Facial Feature Tracking and Expression Recognition for Sign Language", *International Symposium on Computer and Information Sciences*, 2008.
- [6] Aran, O., Ari, I., Guvensan, A., Haberdar, H., Kurt, Z., Turkmen, I., Uyar, A. ve Akarun, L., "Türk İşaret Dili Yüz İfadesi ve Baş Hareketi Veritabanı", *Sinyal İşleme ve Uygulamaları Konferansı*, Eskişehir, 2007.
- [7] Murphy, K., "HMM Toolbox for Matlab", <http://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html>, 03 Ocak 2009.