

Türk İşaret Dili Yüz İfadesi ve Baş Hareketi Veritabanı

A Database of Non-Manual Signs in Turkish Sign Language

Oya Aran¹, İsmail Arı¹, Amaç Güvensan², Hakan Haberdar², Zeyneb Kurt², İrem Türkmen², Aslı Uyar¹, Lale Akarun¹

¹ Bilgisayar Mühendisliği Bölümü, Boğaziçi Üniversitesi, İstanbul
² Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi, İstanbul
¹{aranoya, ismailar, asli.uyar, akarun}@boun.edu.tr
²{amac, hakan, zeyneb, irem}@ce.yildiz.edu.tr

Özetçe

İşaret dilleri görsel dillerdir. Verilmek istenen mesaj sadece el hareketleri ile değil, aynı zamanda baş/vücut hareketi ve yüz ifadesi ile de verilir. Bu makalede Türk İşaret Dili (TİD)'nde kullanılan baş hareketlerini ve yüz ifadelerini içeren bir veritabanı tanıtıyoruz. Veritabanı, TİD'de sıklıkla kullanılan sekiz işaretin, yarı kontrollü bir ortamda toplanmış videolarından ve bu videoların bir kısmı için yüzdeki 60 adet yer gerçekliği noktasının elle işaretlenmiş koordinatlarından oluşuyor.

Abstract

Sign languages are visual languages. The message is not only transferred via hand gestures (manual signs) but also head/body motion and facial expressions (non-manual signs). In this article, we present a database of non-manual signs in Turkish Sign Language (TSL). There are eight non-manual signs in the database, which are frequently used in TSL. The database contains the videos of these signs as well as a ground truth data of 60 manually landmarked points of the face.

1. Giriş

Türk İşaret Dili (TİD) ifadelerinde el, kol ve vücut hareketlerini, birlikte yapılan baş hareketleri tamamlar. İşaret dillerinde yüz ifadeleri, baş ve vücut hareketleri bazen kelimenin anlamını kuvvetlendirmek ya da zayıflatmak için, bazen tamamen değiştirmek için, bazen de tek başına kullanılır [1]. Bir işaret, el hareketi ve şekli korunarak, yüz ifadesinin ve/veya kafa hareketinin eklenmesi ile olumlu/olumsuz/soru anlamları alabildiği gibi, tamamen farklı bir anlam da alabilir. Bu işaretlerin tümü *ele ait olmayan (non-manual) işaretler* olarak adlandırılır.

Dilde kullanılan tüm işaretleri tanıyabilmek için, geliştirilen işaret dili tanıma sistemleri el hareketlerinin yanında, ele ait olmayan işaretleri de çözümlenmeli ve tanıma sırasında değerlendirilmelidir. Fakat işaret tanıma literatüründe ele ait işaretler ve ele ait olmayan işaretleri beraber kullanarak işaret analizi yapan çalışma [2], [3] sayısı az olduğu gibi işaret dillerinde kullanılan ele ait olmayan işaretler üzerine yapılmış çalışma sayısı da çok değildir. Bazı çalışmalarda, sadece yüz ifadesi kullanılır [4], bazılarında ise sadece baş hareketi [5].

İşaret dilinde kullanılan yüz ifadeleri, ifade tanıma problemlerinde sıklıkla kullanılan yedi yüz ifadesiyle (sevinç, kızgınlık, şaşırma...) benzerlikler içerse de tam olarak örtüşmez. İşaret dilinde yüz ifadeleri tek başına kullanılabildiği gibi, bir baş hareketi ile birlikte de

kullanılabilir (Şekil 1). Dolayısıyla, işaret dilinde kullanılan ele ait olmayan işaretleri çözümleyebilmek için, hareketli görüntülerde yüz bölgelerini takip edebilmek ve poz değişikliklerinden bağımsız bir şekilde ifadeyi çözümleyebilmek gerekir.

İşaret dili tanıma probleminin en büyük zorluklarından biri de yöntemleri deneyecek veritabanı bulma ya da oluşturma aşamasıdır. Bu çalışmada, işaret dili tanıma problemlerinde ele ait olmayan işaretlerin daha iyi ve ayrıntılı incelenmesi amacıyla, TİD'de sıklıkla kullanılan baş hareketi ve yüz ifadesi işaretlerinin oluşturduğu bir veri tabanı hazırladık. Bu veritabanının amacı, ele ait ve ele ait olmayan işaretler birlikte kullanılırken oluşan eşzamanlama ve ellerin yüzü kapatması problemlerinden bağımsız bir şekilde, işaret diline özel yüz ifadeleri ve baş hareketleri için tanıma yöntemleri geliştirilmesine olanak vermektir. Veritabanındaki görüntülerde yer gerçekliği noktalarını elle işaretleyerek, işaretlenmiş veri oluşturduk. Noktaları seçerken, bu noktaların kaş, dudak, göz kapağı gibi yüz bölümlerinin takibine izin verecek şekilde olmasına dikkat ettik. Ayrıca, daha sonraki çalışmalarda referans oluşturması için çeşitli tanıma yöntemlerinin veritabanı üzerinde denedik.

Veritabanının özellikleri, içeriği, elle işaretlenmiş noktaların detayları 2. bölümde anlatılmıştır. 3. bölümde aktif görünüm modelleri ile veritabanının istatistiksel çözümlemesi yapılmış ve 4. bölümde de çeşitli tanıma yöntemlerinin veritabanı üzerindeki sonuçları verilmiştir.

2. TİD Baş Hareketleri Veritabanı

2.1. Veritabanında kullanılan baş hareketleri

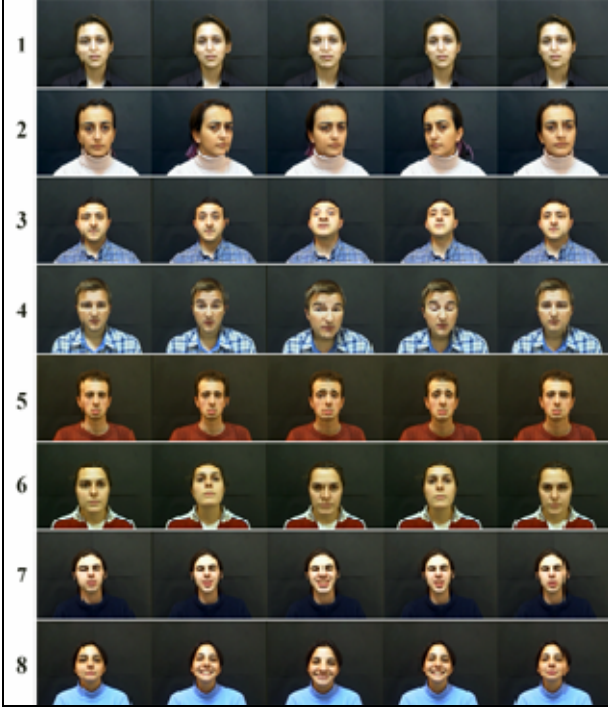
Veritabanındaki baş hareketlerini ve yüz ifadelerini TİD'de sıklıkla kullanılan ve anlamı belirgin ölçüde değiştiren hareketlerden seçtik. Seçilen bazı hareketler sadece baş hareketi ya da yüz ifadesi içerirken, bazı hareketler de hem baş hareketi hem de yüz ifadesi içermektedir.

Veritabanı aşağıdaki sekiz işaretten oluşmaktadır:

1. *Yalın*: Yüzün yalın ifadesidir.
2. *Yukarı doğru başı bir kez kaldırma*: Günlük yaşamda 'değil' demek yerine sıkça kullandığımız hareketler.
3. *Sağa sola baş sallama*: Yine günlük yaşamda 'hayır, olmaz' demek yerine sıkça kullandığımız hareketler.
4. *Öne doğru başı uzatma*: Kafanın öne doğru uzatılması ve kaşların kaldırılması ile yüzün, karşdakinden cevap bekler gibi soru sorar konuma gelmesiyle yapılır.
5. *Büzülmüş Dudaklar*: Özür dilerken, dudakları büzerek ve üzgün yüz ifadesi oluşturularak yapılır.

6. *Yukarı aşağı baş sallama*: Onaylamak için kullanılır. Yüz ifadesi içermez, sadece baş hareketinden oluşur.
7. *Gülümseme*: Baş hareketi içermez, yüz ifadesi gülümseme şeklini alır.
8. *Yukarı aşağı baş sallama ve gülümseme*: Önceki iki hareketin birleşiminden oluşur.

Şekil 1’de bu sekiz işaret için veritabanından alınan örnek videoların kareleri gösterilmiştir.



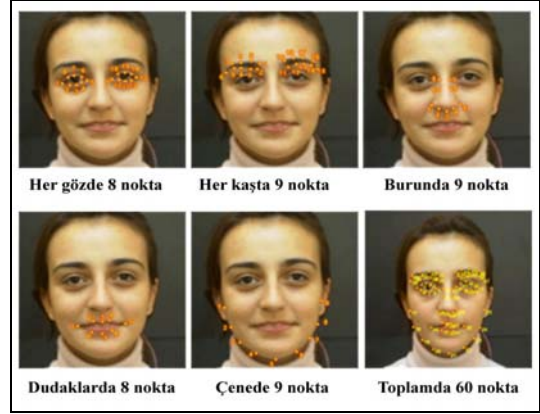
Şekil 1: Veritabanındaki işaretler.

2.2. Veritabanının Özellikleri

Veritabanı beş erkek, altı kadın olmak üzere 11 kişiden toplanmıştır. Her kişiden yukarıda belirtilen işaretler için 5 tekrar alınmıştır. TİD kafa hareketi veritabanı, uzunluğu 1-2 saniye arasında değişen toplam 440 adet hareketli görüntü içerir. Hareketler, Philips SPC900NC kamerası kullanılarak, 640x480 çözünürlükte, saniyede 30 kare hız ile kaydedilmiştir. Çekim ortamı olarak güneş ışığı almayan, günışığı lambalı halojen ve florasan ile aydınlatılmış kapalı bir oda kullandık. Görüntüler “Indeo video 5.10” kullanılarak sıkıştırılmıştır. Tüm görüntüler yalın ifade ile başlar, işaret yapılır ve yine yalın ifade ile biter. Gözlüklü, sakallı veya bıyıklı özneler içermez.

2.3. Seçilen Öznitelik Noktaları

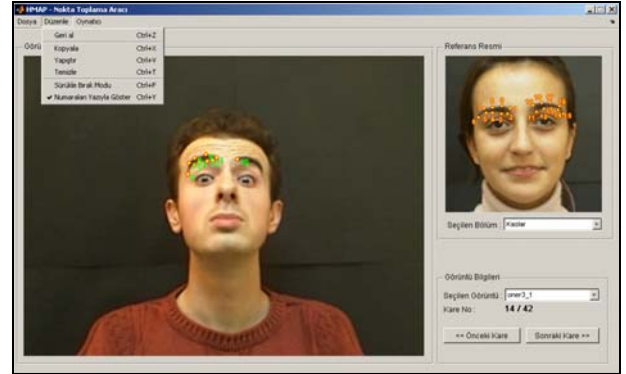
Değişik araştırmalarda ihtiyaç duyulacak noktaları barındırması için öznitelik noktalarını mümkün olduğunca kapsamlı seçtik. Yüzü 5 bileşene ayırdık; gözler, kaşlar, burun, dudaklar ve çene. Şekil 2’de bu noktalar resmedilmiştir.



Şekil 2: Seçilen Öznitelik Noktaları.

2.4. Nokta Toplama Aracı

Öznitelik noktalarını toplama işlemini kolaylaştırması için MATLAB arayüz ortamında bir araç oluşturduk. Şekil 3’de görünen araç ile veritabanındaki istenilen görüntü yüklenip karelerinde dolaşmak mümkündür. Nokta işaretleme işlemi, değişik kişilere paylaştırılabilmesi için bileşenler üzerinden yapılır. Sağ bölümdeki kaynak resim üstünde bileşenler sıralı olarak gözükmemektedir. İşaretleyicinin kaynaktaki sıraya uyması gerekmektedir. İşaretlediği noktaları kopyalayıp başka karelere yapıştırabilir, sürükle/bırak moduna geçip yerlerini değiştirebilir veya bu moddan çıkıp yeni noktalar işaretleyebilir. Son işaretlenen noktayı geri silebilir. İşaretlenen noktaları kaydettiğinde daha sonra kaldığı yerden devam edebilir.



Şekil 3: Nokta Toplama Aracı.

2.5. Nokta Toplama Aşaması

Nokta toplama işleminde tüm görüntüleri kapsamak uzun zaman gerektirdiğinden hem ifade hem de hareket içeren bir altküme için bu işlemi gerçekleştirdik. Bu altküme 2, 3, 4 ve 7 numaralı işaretleri içerir. İşbölümünü yüzü oluşturan bileşenler üzerinden yaptık ve son olarak bu bileşenlere ait noktaları toplam 60 noktayı ifade edecek biçimde birleştirdik.

2.6. Dosya Protokolü

Veritabanındaki dosyaların adlandırılması, kişi adı, yapılan işaretin numarası ve yapılan tekrar numarası kullanılarak yapılmıştır. Görüntü dosyalarının adları şu formattadır:

[kişiyadı]_[#işaret]_[#tekrar].avi

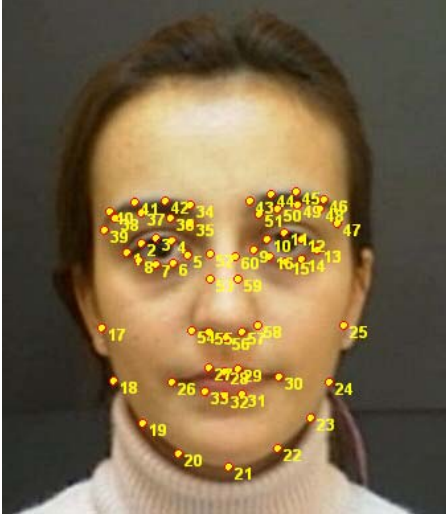
Elle işaretlenmiş yer gerçekliği noktaları metin dosyası olarak x ve y koordinatları ayrı dosyalarda olacak şekilde saklanmıştır. Noktaların dosya adları şu formattadır:

[kişiyadı]_[#işaret]_[#tekrar]_X.txt

[kişiyadı]_[#işaret]_[#tekrar]_Y.txt

Her bir metin dosyasının içinde 60 adet yer gerçekliği noktasının x ya da y koordinatları ilk görüntü karesi ilk satırda, ikinci kare ikinci satırda olmak üzere yazılmıştır. Bu 60 noktanın hangi yüz noktasına ait olduğu aşağıda listelenmiş ve Şekil 4'de gösterilmiştir:

1-8	Sol göz - saat yönünde
9-16	Sağ göz - saat yönünde
17-25	Çene - soldan sağa
26-33	Dudaklar - saat yönünde
34-42	Sol kaş - saat yönünde
43-51	Sağ kaş - saat yönünde
52-60	Burun - soldan sağa



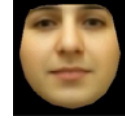
Şekil 4: Öznitelik Noktaları ve yerleri.

3. Veritabanının İstatistiksel Çözümlemesi

3.1. İstatistiksel Şekil Modelleri

Elimizde belli bir şekle ait örnekler bulunursa, bu verilerden istatistiksel bir şekil modeli oluşturabiliriz [6]. Sunduğumuz veritabanında işaretlenen öznitelik noktaları da yüze ait aynı bileşenlerin değişik öznelerde bozunmuş şekillerini oluşturmaktadır. Veritabanındaki görüntülerden, her hareketi kapsayan rastgele 128 imge ile eğitim kümesi oluşturduk. Bu imgeleri döndürme, ölçekleme ve öteleme (Procrustes Çözümlemesi) kullanarak ortalama bir şekil elde ettik.

Eğitim kümesindeki örnekleri, ortalaması Şekil 5'deki yüz olan bir olasılık yoğunluk fonksiyonu ile ifade edebiliriz. Ana Bileşenler Çözümlemesi kullanarak örnek uzaydaki şekillerin değişiminin en yüksek olduğu eksenleri bulduk. Bunları [7]'daki araç kutusunu kullanarak elde ettik.



Şekil 5: Ortalama Yüz.

Sonuçlara göre değişiminin en yüksek olduğu eksenler sırasıyla şöyle:

1. eksen: Kafanın yukarı/aşağı hareket yönü.
2. eksen: Kafanın sağa/sola hareket yönü
3. eksen: Somurtma/gülme yüz ifadeleri
- Diğer eksenler: Özneler arasındaki değişiklikler



Şekil 6: Değişiminin en yüksek olduğu ilk üç eksen. Diğer değerler sabit tutulup yalnızca ilgili eksenin değeri değiştirilerek oluşturulmuştur.

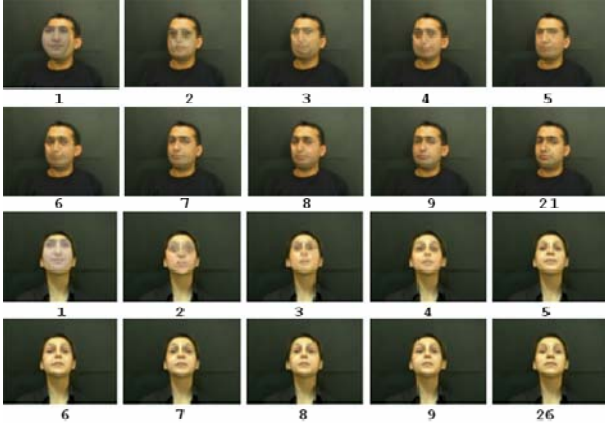
Herhangi bir yüz ifadesinin eksen değerlerine bakılarak hangi yöne baktığı ve mutluluk ifadesi hakkında yorum yapılabilir (Şekil 6). Dolayısıyla bir görüntü için ana üç bileşenin değerlerini dizi halinde tutarsak kişinin ne hareket yaptığını kestirebiliriz.

3.2. Aktif Görünüm Modelleri ile Takip

Aktif Şekil Modellerinin bir genelmesi olan Aktif Görünüm Modelleri, imgede yerel tarama yerine bütünsel tarama yapar [6]. Yeni bir yüz imgesi için önceki bölümde bulmuş olduğumuz ortalama yüzü, her yinelemede yeni imge ile en iyi örtüşecek biçimde bozunuma uğrattık (Şekil 7).

Sistemin başlangıçta, görüntünün ilk imgesiyle örtüşmek için çok sayıda yineleme yaptığını, fakat daha sonra gelen kareler için (bulunan yeni imgeyi temel alırsak) birkaç yinelemede yakınsadığını gözlemledik. Bunun sebebi, veritabanının saniyedeki kare sayısının yüksek olması ve

ardışık iki kare arasındaki görüntü değişiminin çok farklı olmamasıdır. İstatistiksel yaklaştığımız için, eğitim kümesinin kapsamadığı hareketlerde -örneğin yüz sağa/sola fazla döndüğünde takip ettiğimiz bileşenlerin gözükmemesi durumunda- noktaları birebir takip edemedik, fakat yine de son takip edilebilen noktalarda kalarak sıçramalar yaşamadık. Bu çalışmada gerçek zamanlı takip yapamadık, fakat bu sonuç ileriki çalışmalara yönelik açılım sağlıyor.



Şekil 7: Ortalama yüzün her yinelemedeki değişimi.

4. Sınıflandırma

Veritabanı üzerinde bundan sonra yapılacak olan çalışmalara referans oluşturması amacıyla, değişik sınıflandırma yöntemlerinin veritabanı üzerindeki sınıflandırma başarımlarını ölçtük. Sınıflandırma, veritabanında elle işaretlenmiş olan örnekler kullanılarak yapıldı.

Veritabanındaki işaretlenmiş kısmında, dört hareket sınıfı için 11 kişiden alınan ve her sınıf için üçer tekrardan oluşan toplam 132 adet video bulunmaktadır. Bu videolardan, her kişinin birinci tekrarlarını eğitim, ikinci ve üçüncü tekrarlarını test için kullandık. Sonuç olarak eğitim kümesi 44 videodan, test kümesi 88 videodan oluşmaktadır.

4.1. Destek Vektör Makineleri

DVM ile sınıflandırma gerçekleştirmek için öncelikle değişik uzunluktaki videolardan sabit uzunlukta öznelik vektörlerinin çıkarılması gerekmektedir. Bu çalışmada kullanılan öznelik vektörleri, yüz ve kafa bölgesine ait toplam 60 noktanın bütün hareket boyunca en büyük yer değiştirmesi hesaplanarak oluşturulmaktadır. Her bir video için 60 noktanın dört farklı yöndeki en büyük yer değiştirmelerinden oluşan toplam 240 boyutlu bir öznelik vektörü hesaplanmaktadır. Deneyler sonucunda elde edilen ortalama doğruluk oranı % 87,5 olarak hesaplanmıştır. Hata matrisi Tablo 1’de görülebilir.

Tablo 1: Hata Matrisi (Doğruluk Oranları %)

	Onaylamama	Hayır	Soru	Gülümseme
Onaylamama	95.4	4.5	0	0
Hayır	18.2	72.7	0	18.2
Soru	4.5	4.5	90.9	0
Gülümseme	0	9.1	0	90.9

4.2. Bellek Model Yöntemi

Bu yöntemde, yüz ifadesi ve kafa hareketi, kural tabanlı bir yöntem olan bellek model yöntemi ile sınıflandırılmaktadır. Eğitim kümesindeki videoların yer değiştirme hızları incelenerek, sistem tarafından daha önceden görülmemiş olan videoları tanımayı amaçlayan kurallar belirlenmiştir. Yöntemin, dört işaret üzerindeki genel tanıma başarısı %93 olarak hesaplanmıştır. Hata matrisi Tablo 2’de görülebilir.

Tablo 2: Hata Matrisi

	Onaylamama	Hayır	Soru	Gülümseme
Onaylamama	100	0	0	0
Hayır	0	86.4	9.1	4.5
Soru	0	13.6	86.4	0
Gülümseme	0	0	0	100

5. Sonuç

Bu çalışmada, TİD’de sıklıkla kullanılan, ele ait olmayan sekiz hareketten oluşan bir veritabanı hazırladık. Bu veritabanının amacı işaret diline özel yüz ifadeleri ve baş hareketleri için çözümlene ve tanıma yöntemleri geliştirilmesine olanak vermektir. İleriki bir çalışmada hem el hareketleri hem de ele ait olmayan hareketler içeren bir veritabanı hazırlanarak, bu tekniklerin el hareketi tanıma yöntemleriyle birleştirilmesi ve kapsamlı bir işaret dili tanıma sistemi geliştirilmesi amaçlanmaktadır.

6. Kaynakça

- [1] Zeshan, U., “Aspects of Türk İşaret Dili (Turkish Sign Language)”, *Sign Language and Linguistics*, vol. 6:1, 43-75, 2003.
- [2] Aran, O., Ari, İ., Benoit, A., Carrillo, A.H., Fanard, F.X., Campr, P., Akarun, L., Caplier, A., Rombaut, M. and Sankur, B., “SignTutor: An Interactive Sign Language Tutoring Tool”, *Proceedings of eNTERFACE’06*, Dubrovnik, Croatia, 2006
- [3] Ong, S.C.W. and Ranganath, S., “Automatic Sign Language Analysis: A survey and the Future beyond Lexical Meaning”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.27, no.6, pp.873-891, 2005.
- [4] Ming, K.W. and Ranganath, S., “Representations for Facial Expressions,” *Proceedings of International Conference on Control Automation, Robotics and Vision*, vol. 2, pp. 716-721, Dec. 2002.
- [5] Erdem, U.M. and Sclaroff, S., “Automatic Detection of Relevant Head Gestures in American Sign Language Communication,” *Proceedings of Int. Conference on Pattern Recognition*, vol. 1, pp. 460-463, 2002.
- [6] Cootes, T. F. ve Taylor, C. J., “Statistical models of appearance for computer vision”, *Technical report*, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, Manchester M13 9PT, United Kingdom, Eylül 1999.
- [7] Stegmann, M. B., Ersbøll, B. K. ve Larsen, R., "FAME -- A Flexible Appearance Modelling Environment", *IEEE Transactions on Medical Imaging*, IEEE, 2003
- [8] Dai, Y., Shibata, Y. “Recognition of Facial Expressions on the Hopfield Memory Model”, *Proceedings of IEEE. International Conference on Multimedia Computing*, 1999