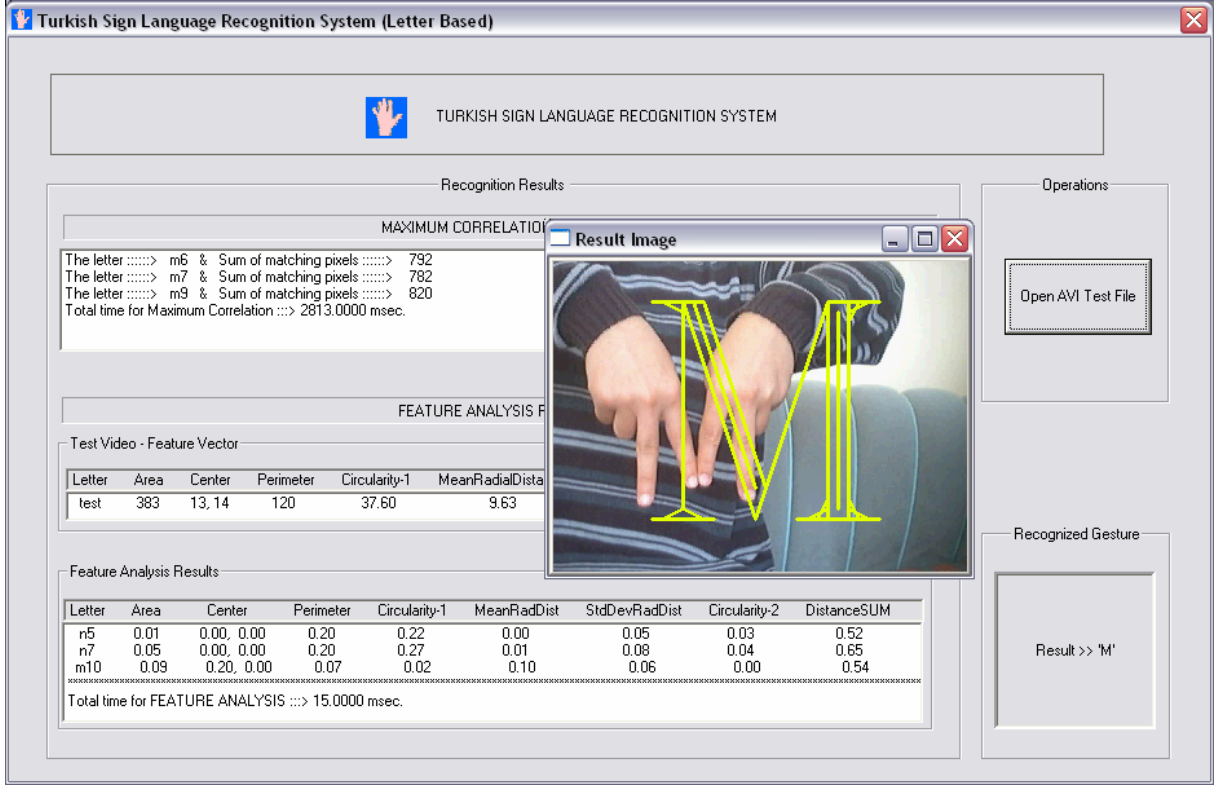


TÜRK İŞARET DİLİ TANIMA SİSTEMİ (TURKISH SIGN LANGUAGE RECOGNITION SYSTEM)



Şekil 1 İşaret dili tanma örnek ekran görüntüsü

Türk İşaret Dili Tanıma projesi 2005–2006 yılının 2. döneminde Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği 4. sınıf öğrencileri Behzat Bükün ve Ali Ekinci tarafından bitirme projesi olarak gerçekleştirilmiştir.

Giriş

Son yıllarda işaret dili tanıma ile ilgili projeler araştırma grupları tarafından büyük bir ilgi görmüştür. Bu durum insan-bilgisayar etkileşimini ön plana çıkaran alanlardaki potansiyel çalışmalara da hız vermiş olup işaret dili tanıma gibi sistemlerin hayata geçirilmesini sağlamıştır.

İşaret dili, el hareketlerinin ve yüz mimiklerinin kullanılmasıyla oluşan görsel bir dildir. Ancak sık kullanılan kelimeler için hareketler var iken, özel isimler ile bir hareket karşılığı olmayan kelimeler için **harf harf heceleme** yapılır. Bu çalışmada amaç harf harf hecelemeden oluşan işaretlerin bilgisayar tarafından tanınması hedeflenmiştir.

Bu proje kapsamında yapılan çalışmalar iki aşamadan oluşmaktadır. Birincisi, her ülkenin işaret dili farklı olduğu için işaret dili tanıma sistemlerinin geliştirilmesi her ülkenin kendi öz kaynaklarına bağlı olması ve bu kapsamda Türk İşaret Dili ile ilgili bir hareket veritabanı olmadığından, öncelikle bu aşama gerçekleştirilmiştir. Böylece ellerin alacağı değişik pozisyonlara göre Türk Dili alfabesinde bulunan 29 harften oluşacak bir işaret dili alfabesi veritabanı oluşturulmuştur. İkinci aşama ise birinci aşamayı takip ederek Türk alfabesindeki harflere karşılık gelen işaret dili hareketlerini tanıyabilen, yani tercüme edebilen bir sistem tasarlanmıştır. Tanıma işleminde *maksimum korelasyon* ve *öznitelik vektörü* kullanılarak elde edilen sonuçlar gruplandırılarak *k en yakın komşu* algoritmasıyla en iyi sonuç elde edilmiştir. Kullanılan bu tekniklerle birlikte 203 test videosundan sadece 5 yanlış sonuç çıktığından, sistemin başarısı %97.54 olarak ölçülmüştür.

Görüntü tabanlı işaret dili tanıma sistemlerinin bir çoğunda eldiven giyilerek işlemler yapılmaktadır. Ancak bu tür sistemlerin ticari veya pratiklik boyutu yok gibidir. Bu proje kapsamında yapılan yoğun çalışmalar sonucu eldiven giyme olayı elimine edilmiştir ve bazı koşullarla birlikte pratiklik boyutu yüksek olan bir sistem geliştirilmiştir.

Sistemin kısaca işleyişinden bahsetmek gerekirse şu şekilde özetlenebilir: Sistem bir bilgisayar yazılımı ve bilgisayara bağlı bir web kamerasından oluşmaktadır. Kamera önünde hareketi yapan kişinin herhangi bir eldiven veya bir başka donanım kullanmasına gerek kalmadan kamera karşısında duran kişinin yaptığı hareket, video şeklinde kaydedilmekte ve belirli bir zaman aralığında kaydedilen hareket videosundan kişinin eli ten renginden ayırt edilmektedir. Bu şekilde eli ten renginden ayırt edilen kişinin yaptığı hareket, kişiden bağımsız olarak oluşturulmuş veritabanındaki hareketlerden hangisine en yakın olduğu bulunarak yapılan işaretin yazı diline dökülmesi sağlanmaktadır.

Geliştirilen sistem banka, postane, hastane gibi ortamlarda işitme engelli insanların çevre ile iletişimini kolaylaştırmayı hedeflemektedir.

Sistemin tasarlanmasında, Microsoft Visual Studio platformu ve C++ programlama dili kullanılmıştır. Görüntü işleme için ise Intel firmasının geliştirdiği OpenCV (Open Source Computer Vision) kütüphanesi kullanılmıştır.

Kullanılan Sistemin Yapısı

Sistemde görüntü yakalama aygıtı olarak Philips® PCVC 840K model kamera kullanılmıştır. 1/3" VGA CCD alıcısı olan kamera saniyede 60 çerçeve oranına sahiptir ve 1 lüksten daha düşük ortamlara kadar görüntü alımı sağlanabilmektedir.

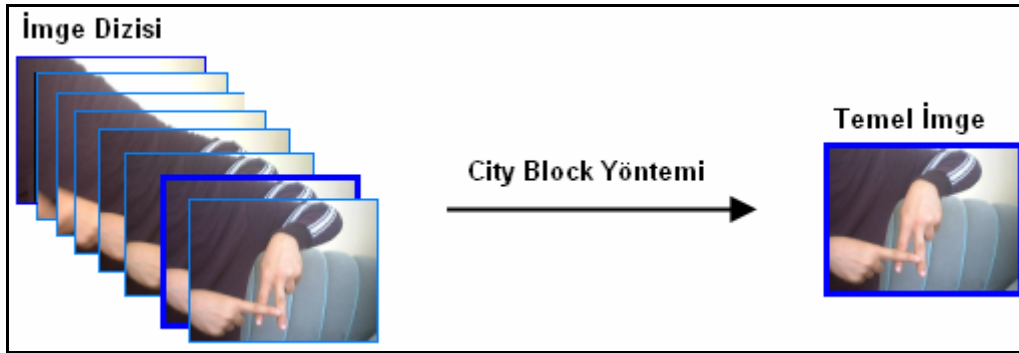
Proje kapsamında görüntü almak için kameranın özellikleri sistemin işleyişini iyileştirmek için değiştirilmiştir. İlgili kameradan saniyede 15 çerçeve oranı ile 3 saniyelik görüntüler alınmıştır. Yapılan hareketlerin çözünürlüğü ise 320x240'tır. Ancak ortamın ışık miktarını düzenlemek için de donanım ile birlikte verilen yazılımda çeşitli ayar değişiklikleri yapılmıştır.

Durağan Hareketin Bulunması

Kaydedilmiş hareket videosundan işareti anlatacak en uygun imgenin bulunması sistemde tanıma işleminin ilk adımıdır. Hareket videosu içerisinde yer alan imgeler arasındaki mesafeler art arda hesaplanarak bu değerlerin en az değiştiği aralık temel imgeyi vermektedir. Temel imge hareketin durağan olduğu durumdur. İmgeler arasındaki mesafe ölçümü ise Manhattan (City Block) uzaklığı yöntemi ile gerçekleştirilmiştir. Video içerisindeki her imge üç renk bilgisine (R: Kırmızı, G: Yeşil, B: Mavi) sahip olduğu için her bir renk bilgisine ait mesafe ölçümü ayrı ayrı yapılarak toplanır.

$$\begin{aligned}\Delta R &= R_{t+1} - R_t \\ \Delta G &= G_{t+1} - G_t \\ \Delta B &= B_{t+1} - B_t \\ D_{City} &= \sum (|\Delta R| + |\Delta G| + |\Delta B|)\end{aligned}$$

Şekil 2'de temel çerçevenin bulunması işlemi şekillendirilmiştir.



Şekil 2 Temel çerçevenin seçilmesi

Ten Rengini Tanıma

Sistemi tasarlanmanın önemli adımlarından biri de ortam özelliklerinin iyi belirlenmiş olmasına bağlıdır. Bu aşama elde edilen verinin sağlıklı bir şekilde sayısal ortama aktarılması için kritiktir. Örneğin ışık kaynağının farklı olduğu durumlar ile görüntü almak için kullanılan kameranın özellikleri de sistemin işleyişini değiştirebilme önemine sahiptir.

Ten rengi tonları lineer olmayan bir şekilde parlaklığa bağlı olduğundan, ten rengi tanımda $Y C_b C_r$ renk düzlemi kullanılmıştır. Y değeri imgenin parlaklık değerini, C_b imgenin mavi renklilik değerini, C_r ise imgenin kırmızı renklilik değerini belirlemektedir. $Y C_b C_r$ renk düzleminde renk bileşenleri ten rengi ile örtüşebilmektedir. Bu bağlamda belirli değerler arasında bulunan pikseller ten rengi olarak belirlenmiştir. Hareket videosundan alınan temel

imge normalde üç renk bilgisine sahiptir. Bunlar RGB renk düzlemi olarak R: Kırmızı, G: Yeşil ve B: Mavi renk bilgileridir. Dolayısıyla imgenin $Y C_b C_r$ renk düzlemine dönüştürülmesi gerekir.

$$Y = 0.299R + 0.587G + 0.114B$$

$$C_r = R - Y$$

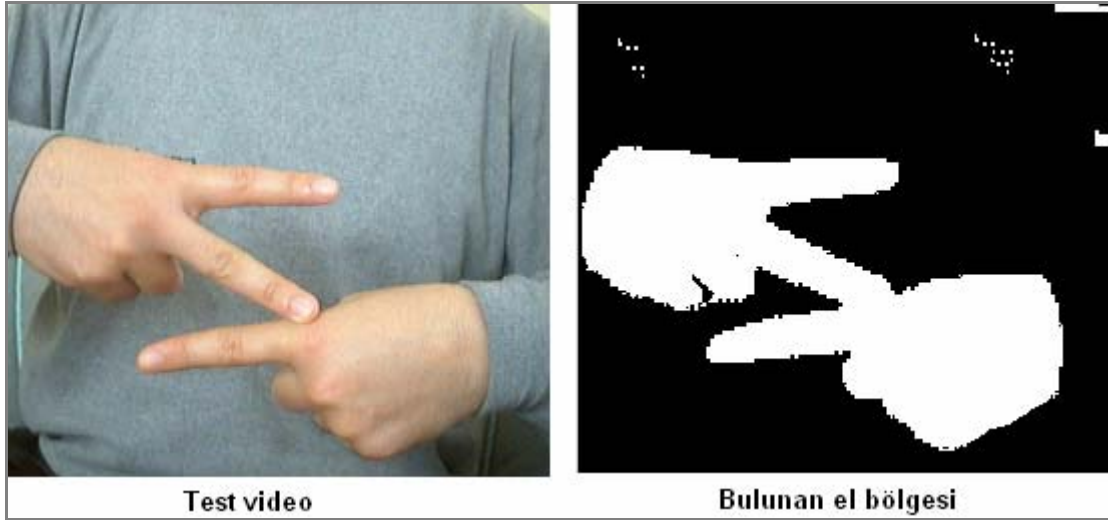
$$C_b = B - Y$$

Ten rengi olabilecek piksel değer aralıkları şu şekildedir:

$$135 < C_r < 180$$

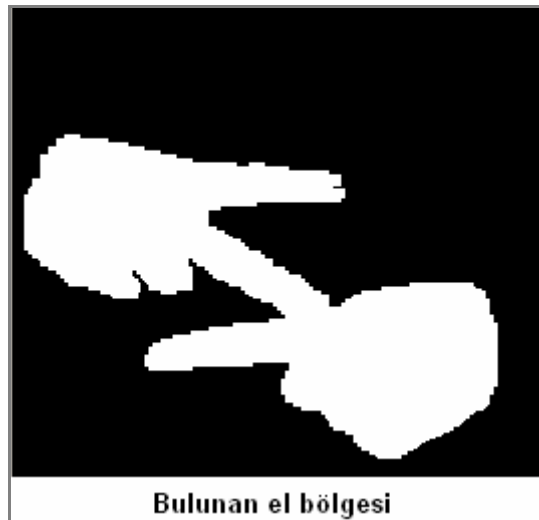
$$85 < C_b < 135$$

$$Y > 80$$



Şekil 3 El bölgesinin bulunması (Z harfi) (parazitli)

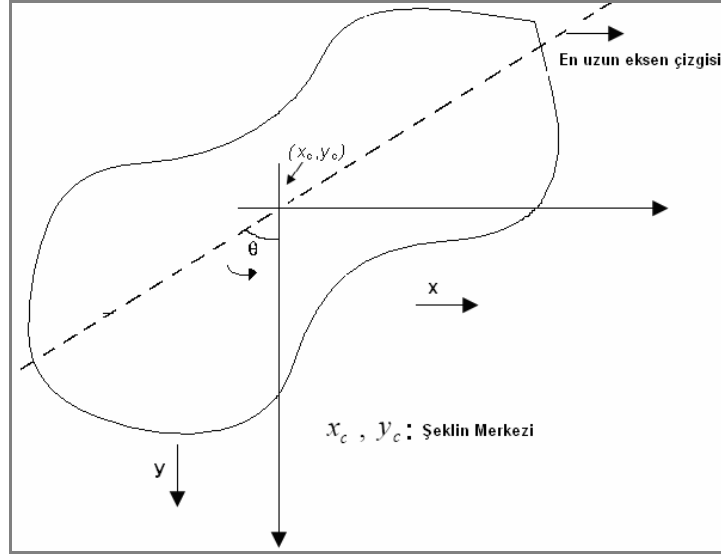
Şekil 3'te örnek bir hareket videosu üzerinden seçilen imgedeki el bölgesi gösterilmektedir. Ancak Şekil 3'te el bölgesinin bulunduğu imgede sağ üst köşede bazı parazitlerin olduğu görülebilmektedir. Bu parazitlerin temizlenmesi için Gauss filtresinden faydalanılarak bu parazitlerin ten olma olasılıkları tekrar değerlendirilmektedir. Şekil 4'te yapılan bu işlemlerin sonucu gösterilmektedir.



Şekil 4 El bölgesinin bulunması (parazitsiz)

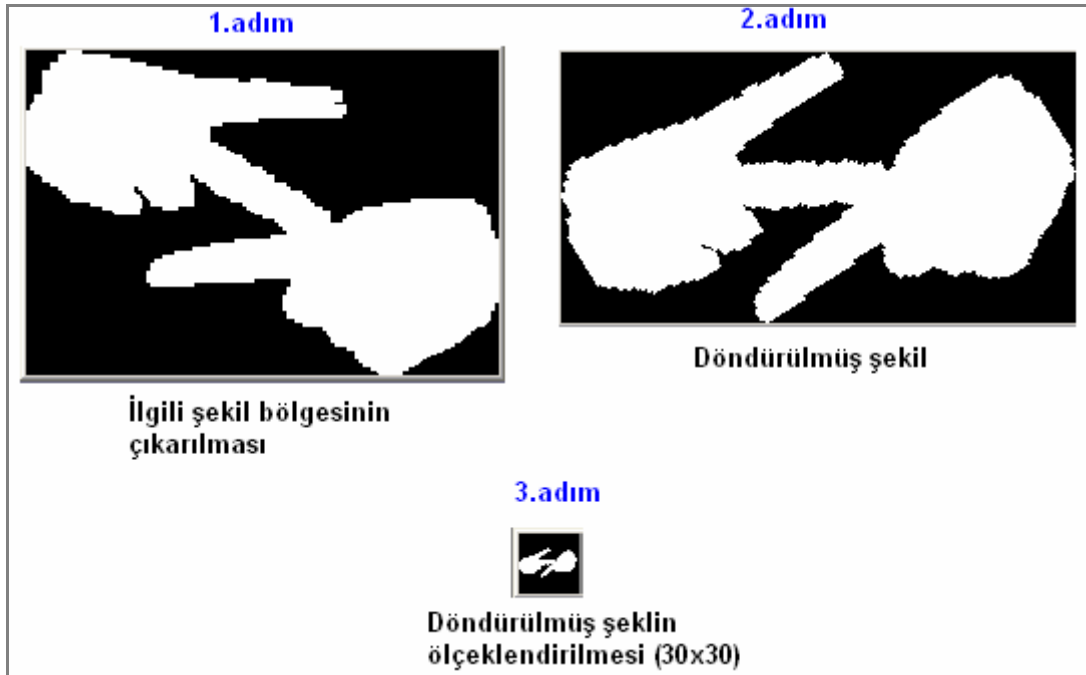
Geometrik Özelliklerin Belirlenmesi

Kamera karşısında hareketi yapan kişi bazen aynı hareket için biçim değiştirebilmektedir. Bu durumda yapılan hareketlerin yönden bağımsız olması gerekir. Bunun için şekle ait bazı geometrik özelliklerin belirlenmesi gerekir. Şöyle ki yapılan her hareket bir şekil olarak değerlendirilirse o şekilden geçen en uzun eksenin koordinat sisteminde dikey eksen veya yatay eksen ile yaptığı açı bulunursa ve bu eksenlerden herhangi birisine bulunan açı kadar döndürülürse yapılan hareketler yönden bağımsız olmuş olur. Bu durum Şekil 5'te ifade edilmiştir.



Şekil 5 Şekilden geçen en uzun eksen çizgisi

Ayrıca yapılan hareketlerin kameraya olan uzaklık mesafeleri de önemli olduğundan uzaklıktan da bağımsız olması gerekir. Bunun için döndürülmüş şekil için bir ölçeklendirme yapılmalıdır. Normalde 320x240 çözünürlükte alınan görüntüler ölçeklendirme ile 30x30 boyutlarına indirgenmiştir. Şekil 6'da 'Z' harfi için yapılmış bir harekete ait el bölgesinin bulunmasından sonra yapılan işlemler gösterilmiştir.



Şekil 6 El bölgesi için yapılan işlemler

Veritabanı Tasarımı

Hareket videolarına ait temel imgelerin en son ölçeklendirilmesinden sonra elde edilen 30x30 boyutundaki imgelerden veritabanı oluşturulmuştur. Proje kapsamında 4 ayrı kişiden Türk İşaret Dili'ne ait 29 harf için kayıtlar alınmıştır. Alınan bu görüntülerin bir kısmı eğitim seti bir kısmı ise test seti olarak değerlendirilmiştir. Eğitim setine ait görüntüler veritabanı olarak değerlendirilmiştir. Alınan bu görüntülere ait sayısal değerler Tablo 1'de görülebilir.

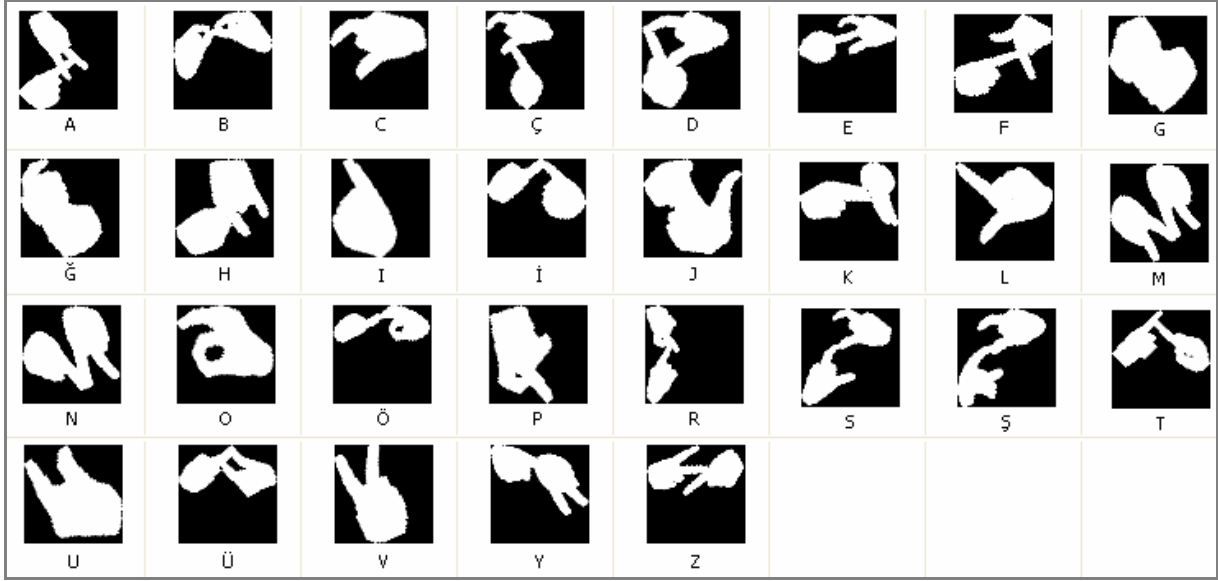
Tablo 1 Hareket videolarına ait sayısal değerler

	1. Kişi		2. Kişi		3. Kişi		4. Kişi	Toplam Eğitim	Toplam Test	Genel Toplam
	Eğitim	Test	Eğitim	Test	Eğitim	Test	Test			
Harf A	4	2	4	2	2	2	1	10	7	17
Harf B	4	2	4	2	2	2	1	10	7	17
Harf C	4	2	4	2	2	2	1	10	7	17
...	...									
...	...									
Harf Y	4	2	4	2	2	2	1	10	7	17
Harf Z	4	2	4	2	2	2	1	10	7	17
							Toplam	290	203	493

Sonuç olarak 3 kişiden alınan aynı işarete ait 4'er hareket veritabanını oluşturmuştur. 4.kişiden alınan hareketler eğitim setine dahil edilmeyerek sistemin kişiden bağımsız olup olmadığını ölçmek için test seti olarak değerlendirilmiştir. Bu ölçüm sonucunda hareketleri eğitim setinde olmayan bir kişinin yaptığı 29 hareketin 27'si sistem tarafından tanınmıştır. Dolayısıyla sistemin kişiden bağımsız olarak başarısı %93,10 olarak ölçülmüştür. Şekil 7'de veritabanında yer alan eğitim setine ait her işareten bir hareket bulunmaktadır ve bu hareketlere ait yapılan işlemlerden sonra elde edilen ölçeklendirilmiş imgeler de Şekil 8'de görülebilir.



Şekil 7 29 harfe ait hareket görüntüleri



Şekil 8 Hareketlerin ön işlemlerden sonraki son durumları

Tanıma İşlemleri

Tanıma işlemleri için iki yöntem izlenmiştir ve bu yöntemler birbirinden bağımsız olarak değerlendirilmiştir. Bu yöntemlerden ilki *Maksimum Korelasyon* tekniğidir. Maksimum korelasyon tekniğinde karşılaştırılacak iki imge üst üste oturtularak birbirine denk düşen piksellerin korelasyon değeri olarak alınır. Test imgesi, eğitim setindeki tüm imgelerle karşılaştırılarak korelasyon değeri en yüksek olan 3 imge bulunur.

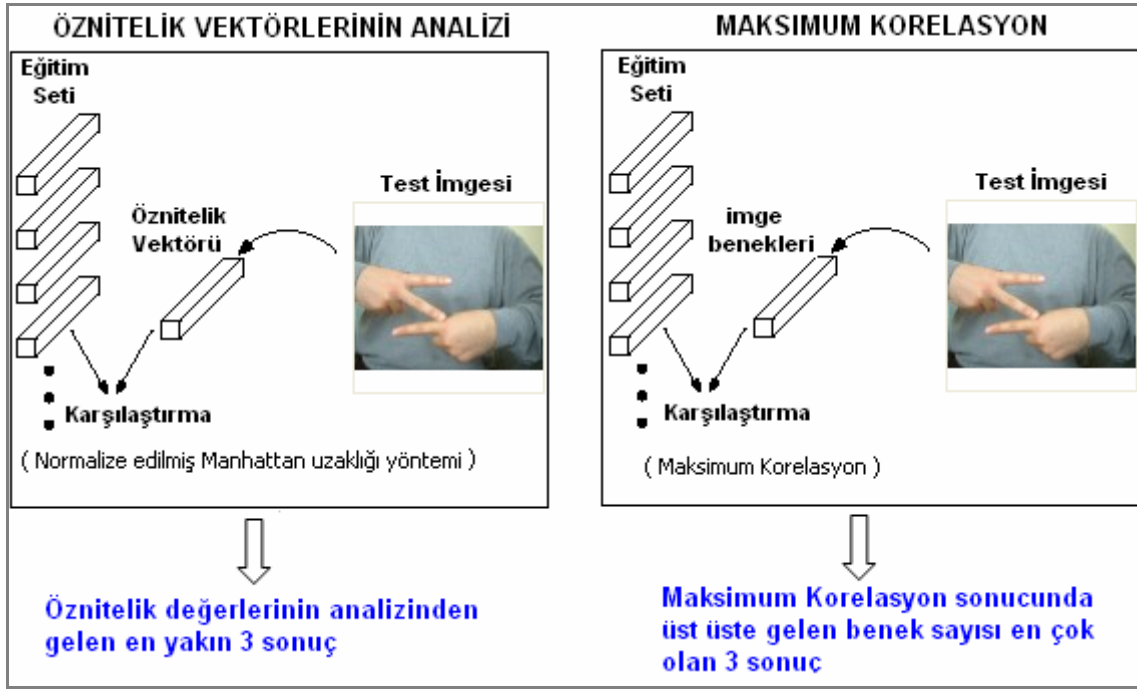
Tanıma işleminde kullanılan bir diğer yöntem ise imgelerin *Öznitelik* değerlerinin analiz edilmesidir. Veritabanı oluşturma aşamasında her bir imgeye ait öznitelik vektörü çıkarılır ve bu öz nitelik vektörü tanıma işleminde kullanılır. Bir imgeye ait öznitelik vektörü şu bilgilerden oluşmaktadır:

- Alan (*area*),
- Merkez noktası (*center point*),
- Çevre uzunluğu (*perimeter*),
- Yuvarlaklık bilgisi (*circularity*),
- Ortalama radyal uzaklık (*mean radial distance*)
- Radyal uzaklığın standart sapma miktarı (*standard deviation of radial distance*)

Alan, merkez noktası ve çevre uzunluğu imgedeki nesnenin şekil içerisindeki konumu ile ilgili bilgileri vermektedir. Diğer özellikler ise nesneye ait şekil bilgisi hakkında fikir vermektedir.

Eğitim setindeki imgelere ait öznitelik vektörleri bulunarak normalize edilir. Test imgesine ait öznitelik değerlerinin eğitim setindeki imgelere ait öznitelik vektörlerine olan uzaklık mesafeleri Manhattan uzaklığı yöntemi ile bulunur. Bu değerler her bir öznitelik değerinin sahip olduğu minimum ve maksimum değerler ile normalize edildikten sonra bulunan sonuçlardan değeri minimum olan üç imge tanıma işlemi için değerlendirilir.

Sonuç olarak, tanıma işleminde *maksimum korelasyon* ve *öznitelik vektörü* kullanılarak elde edilen sonuçlar gruplandırılarak *k en yakın komşu* algoritmasıyla en iyi sonuç elde edilmiştir. Şekil 9'da bu işlemlerin nasıl yapıldığı görülmektedir.



Şekil 9 Tanıma işlemi

Deneysel Tanıma Sonuçları

Tasarlanan sistem Türk Dili İşaretlerinden 29 harfi kişiden bağımsız, yönden bağımsız ve mesafeden bağımsız olarak tanımayı hedeflemiştir. Tablo 1'de verilen 203 test videosuna ait başarı oranı %97,54 olarak ölçülmüştür (198 test videosu DOĞRU, 5 test videosu YANLIŞ). Sistemdeki yüksek başarı, yapılan ön işlemlerin iyi yapılmış olmasına bağlıdır. Bu kapsamda sistemde yapılan döndürme ve ölçeklendirme işlemleri de sistemin başarısını arttırmıştır. Şöyleki 60x60 ölçekleme ile 174 test videosundan sadece 5 tanesi yanlış tanındığından sistemin başarısı %97,13, 30x30 ölçekleme ile 174 test videosundan sadece 3 tanesi yanlış tanındığından sistemin başarısı %98,28 olarak ölçülmüştür. 30x30'luk ölçekleme yapılmadan önce projede şöyle bir iyileştirme daha yapılmıştır. Ölçeklendirme aşamasında ilgi alanı içerisinde olmayan siyah bölgeler temizlenerek karşılaştırma dışı bırakılmıştır.

Sistem tanıma zamanı da maksimum korelasyon ve öznitelik vektörü için ayrı ayrı uygulanmıştır. Maksimum korelasyon için geçen toplam zaman yaklaşık olarak ~125ms ile ~160ms arasında değişirken, öznitelik değerlerinin analizi için geçen zaman çok daha az ve yaklaşık olarak ~0ms ile ~20ms arasında değişiklik göstermektedir. Bazı test videoları için geçen toplam zaman yani hem maksimum korelasyon hem de öznitelik vektör analizi için geçen toplam süre Tablo 2'de verilmektedir.

Tablo 2 Bazı test videoları için tanıma zamanları

Test Videosu	Toplam Tanıma Zamanı
a2	~266ms
b2	~250ms
f4	~172ms
k3	~203ms
z4	~125ms

Sistemin başarısını düşüren en büyük neden S ve Ş harflerinin birbirine şekil olarak çok yakın olmasıdır. Bu harfler için maksimum korelasyon iyi sonuç verirken öznel vektörlerinin analizi iyi sonuç verememektedir. Çünkü bu harflerin şekil bilgileri birbirine çok benzemektedir. Şekil 9’da bu harflere ait örnek gösterimler yer almaktadır.

**Şekil 10** Örnek 'S' ve 'Ş' harflerinin gösterimi

Sonuç

Bu proje ile Türk İşaret Dili için kullanılan işaretler harf tabanlı olarak bilgisayar ortamına aktarılmış ve gerekli işlemlerden geçirilerek tanıma işlemi gerçekleştirilmiştir. Bununla birlikte harfleri temsil eden hareketler için yeni bir veri tabanı oluşturulmuştur ve bu veritabanının gelecekte yapılması muhtemel projelerde de kullanılması planlanmaktadır. Sistemin başarısı gözönüne alındığında % 98.28 lik başarı ile geçmişte yapılmış uluslararası işaret dili tanıma sistemlerine oranla daha başarılı olduğu dikkat çekmektedir. Ayrıca hareketlerin mesafeden, yönden ve kişilerin ten renginden bağımsız olması ise sistemin bir başka önemli özelliğidir.

Bu projenin çok sayıda Türk İşaret Dili hareketi içermesi ve aynı işaretin farklı kişilerce tekrarlanması sonucu oluşturulan veritabanı hem araştırmacıların işlemlerini kolaylaştıracağı hem de bu tür sistemlere olan ilginin daha da artacağı ümit edilmektedir.

Sağır ve dilsiz alfabesini kullanan insanların aralarındaki iletişimin daha güçlü olması ve bu kişilerin topluma kazandırılması bu projenin getirebileceği yararların başında gelmektedir. Ayrıca işaret dili tanıma sistemlerinin popüler olması ve bu projenin insanları gelecekte bu konu ile ilgili daha güçlü sistemler geliştirilmesine teşvik edeceği düşünülmektedir.

Son olarak bu tür projeler ile insan-bilgisayar etkileşiminin güçlendirilmesi ile işitme engelli bireylerin sosyal olarak toplumda kendi yerinin olduğunu ve iletişimin sağlanamaması için herhangi bir engel bulunmadığının altı çizilmelidir.

Kaynaklar

- [1] Sazonov, V., Vezhnevets, V., and Andreeva, A. A survey on pixel based skin color detection techniques. In Proceedings of Graphicon (2003), pp. 85–92.
- [2] Chai, D., and Bouzerdom, A. A bayesian approach to skin colour classification. In IEEE Region Ten Conference (2000).
- [3] Goh, P., Recognition of Auslan Finger-spelling using Hidden Markov Models, (2005), pp. 18-19
- [4] Computer Vision and Image Processing: A Practical Approach Using CVIPtools - S. E Umbaugh, Prentice Hall PTR, Upper Saddle, NJ, 1998, pp. 135-136
- [5] Di Gesu V. and Starovoitov V. "Distance-based functions for image comparison," Pattern Recognition Letters, Vol. 20, No. 2, pages 207-214, 1999.
- [6] Umbaugh, Ph.D. S.E., (1999), "Computer Vision and Image Processing", Prentice-Hall
- [7] <http://astronomy.swin.edu.au/~pbourke/colour/colourspace>
- [8] <http://homepages.inf.ed.ac.uk/rbf/HIPR2/gsmooth.htm>
- [9] <http://www.mathworks.com/access/helpdesk/help/toolbox/images/f18-35678.html>
- [10] <http://zone.ni.com/devzone/conceptd.nsf/webmain/8CA8DE2E8881C1AB8625682E0079CE74>
- [11] <http://www.cs.bilkent.edu.tr/~duygulu/Courses/CS554/Spring2006/Notes/BinaryImageAnalysis.pdf>