Üç Boyutlu El Hareketlerinin Tanınması için İki Boyutlu Özniteliklerin Birleştirilmesi
Combination Strategies for 2D Features to Recognize 3D Gestures

Oya Aran, Lale Akarun

Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi, Istanbul
aranoya@boun.edu.tr, akarun@boun.edu.tr

Özetçe

Bu çalışmada, iki kameradan oluşan bir kurulumda üç boyutlu el hareketlerini tanıyan bir sistem tasarladık. Kameraların kalibrasyon matrislerinin hesaplanamadığı ve üç boyutta geri çatma işleminin yapılamadığı durumlarda, iki boyutlu el gezingesinin her iki kamera için ayrı ayrı çıkartılarak öznitelik ya da karar aşamasında birleştirilmesi başarımın büyük ölçüde artmasını sağlamaktadır. İki boyutlu el gezingeleri, elin orta noktası ve eli çevreleyen elipsin uzunluk, genişlik ve açısı izlenerek çıkarılmış ve bu gezingeler Kalman Süzgeci kullanılarak yumuşatılmıştır. Boyutları ve başlangıç noktaları düzgelenen gezingeler Saklı Markov Modelleriyle (HMM) veya sabit uzunlukta yeniden örneklenerek Destek Vektör Makinalarıyla (SVM) sınıflandırılmışlardır. Farklı kameralardan çıkarılan gezingelerin farklı aşamalarda birleştirilmesinin performans üzerindeki etkisi incelenmiştir. En iyi sonuç, farklı kameralardan çıkarılmış gezingelerin HMM kullanılarak modellenmesi ve karar aşamasında birleştirilmesiyle elde edilmiş, bu yöntemle toplam 210 test örneğinde sadece 1 ya da 2 örnekte hata yapılmıştır.

Abstract

In this study, using a two camera setup, we designed a system that recognizes 3D gestures. When 3D reconstruction is not possible or infeasible, combining 2D hand trajectories at feature or decision level increases the system performance drastically. The trajectories are extracted by tracking the center-of-mass of the hand and the width, height and orientation of the enclosing ellipse. Trajectories are then smoothed using a Kalman Filter. Following the translation and scale normalization, the trajectories are modelled using Hidden Markov Models (HMM) and using Support Vector Machines (SVM) by converting the trajectories to fixed length using re-sampling. Trajectories extracted from different cameras are combined at different levels and the effect to the system performance is observed. The best resut is obtained by modelling the trajectories using HMMs and combining at decision level, with %1 error in 210 test examples.

1. Giriş

El hareketleri insanlar arası iletişimde kimi zaman sadece konuşmayı tamamlamak için kimi zaman ise tek başına bir iletişim aracı olarak kullanılmaktadır. Son dönemde, bilgisayar ve kamera teknolojisindeki yenilikler el hareketlerinin aynı zamanda insan bilgisayar iletişiminde de diğer araçlarla beraber ya da tek başına kullanılabileceğini göstermiştir [1]. El hareketlerinin tanınması aynı zamanda işaret dilinin tanınması açısından da önemlidir. El hareketi, şekli ve konumunun analizine dayanan işaret dilinde yüz ifadesi, baş-vücut hareketleri gibi bileşenler el bilgisini destekleyerek işaret dilini zenginleştirirler [2].

El hareketi tanıma sistemleri elin uzay-zamansal bileşenlerini modellerler. Uzaysal bileşenin karmaşıklığı uygulamaya göre değişiklik gösterir. En karmaşık halinde elin ve parmakların yapılanışı detaylı şekilde incelenmelidir. Daha basit olarak genel el şekli veya sadece elin pozisyonu kullanılabilir. Zamansal bileşen ise iki ya da üç boyutta el gezingesinin çıkarılmasıyla elde edilir. Gezingenin çıkarılmasında çeşitli el izleme teknikleri kullanılabileceği gibi (Kalman Suzgeci [3], Parçacık Suzgeci [4], vs.) zamansal şablonlara [5] dayalı teknikler de kullanılabilir. Çıkarılan gezinge Sonlu Durumlu Makineler, Zaman Gecikmeli Sinir Ağları, Saklı Markov Modelleri ya da şablon eşleştirme teknikleriyle modellenebilir [6]. Bu teknikler içinde, Saklı Markov Modelleri, bir çok sistemde başarı sağlamışlardır.

Saklı Markov Modelleri, aynı sınıfa ait aynı ya da farklı uzunluktaki örnekleri kullanarak o sınıf için bir model oluştururlar. Her sınıf için farklı bir Saklı Markov Modeli oluşturulur ve bir test örneği geldiğinde olabilirliği en yüksek olan model / sınıf seçilir. Farklı uzunluklardaki örnekleri kullanma ve sıralı veriyi modellemedeki başarısı Saklı Markov Modellerini el hareketi tanıma problemi için uygun bir teknik haline getirmektedir. Fakat sınıflandırma problemlerinde, sınıf içi benzerlikleri olduğu kadar sınıflar arasındaki ayrımları da gözönüne alan tekniklerin daha başarılı olduğu bilinmektedir. Bu tür teknikler arasında Destek Vektör Makineleri, Yapay Sinir Ağları sayılabilir.

Bu çalışmada, farklı kameralardan çıkarılan iki boyutlu el gezingeleri ve genel el şekil bilgisi öznitelik aşamasında ya da karar aşamasında birleştirildi ve iki farklı sınıflandırıcı (HMM, SVM) kullanılarak sınıflandırıldı. İki boyutlu el gezingelerini öznitelik aşamasında ya da karar aşamasında birleştirmenin üç boyutlu el hareketlerinin tanınmasına katkısı incelendi.

2. İki Boyutlu El Gezingesinin Çıkarılması

Bu çalışmada yedi el hareketinden oluşan bir veri tabanı kullandık [7]. Veri tabanındaki el hareketleri iki el birden kullanılarak yapılan ve üç boyutlu nesneleri ittirmek ve döndürmek için kullanılabilecek hareketleri içermektedir. Kullanıcılar her iki ellerine farklı renklerde (sol elde mavi ve sağ elde sarı) eldivenler giymişlerdir. Kullanıcının sağ ve soluna iki kamera yerleştirilmiştir. Eğitim kümesinde, her kişiden 10 adet olmak üzere, dört kişiden alınmış toplam 280 örnek bulunmaktadır. Test kümesinde ise farklı üç kişiden alınmış toplam 210 örnek bulunmaktadır.

El gezingesinin çıkarılması için aşağıdaki adımlar sırasıyla uygulanır:

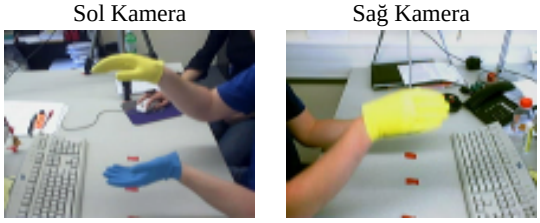
1. Elin bulunması (Giyilen eldivenin rengine göre eşikleme ve bağlantılı bileşen algoritması kullanılarak)
2. Bulunan iki boyutlu koordinatların Kalman Süzgeci kullanılarak yumuşatılması
3. Ölçek farklılıklarının düzgelenmesi (x,y koordinatlarının iki boyutlu düzgelenmesiyle)
4. Konum farklılıklarının düzgelenmesi (Gezingenin başlangıç noktasını (0,0) koordinatına çekerek)

2.1. Elin Bulunması ve İzlenmesi

Kullanıcının her iki elinin bulunması için ilk olarak giyilen eldivenlerin renklerine göre eşikleme uygulandı (Şekil 1b). Eşiklenen imgeler bağlantılı bileşen algoritmasıyla bölütlendi. Bölütleme sonucunda oluşan imgede yer alan en fazla alana sahip olan bileşenin el olduğu kabul edildi (Şekil 1c).

El gezingesi, her çerçevede bulunan elin ağırlık merkezi kullanılarak çıkarılmış ve çıkarılan gezinge Kalman Süzgeci kullanılarak yumuşatılmıştır. Gezingede, ağırlık merkezine ek olarak eli çevreleyen elipsin uzunluğu, genişliği ve açısı öznitelik olarak kullanılmıştır. Bu sayede genel el şekil bilgisi el hareket bilgisiyle birleştirilmiştir.

Bu kurulumda kameraların kullanıcıya yandan bakıyor olması sebebiyle, eller bazı hareket sınıflarında birbirlerini kapatabilirler. Bu kapatma ancak kameraların birinde olabilir. Aynı anda iki kamerada birden ellerin birbirini kapatması sözkonusu değildir. Şekil 2 sağ kamerada sağ elin sol eli kapattığı bir anı göstermektedir. Aynı anda sol kamera görüntüsünde ise herhangi bir kapatma yoktur. El bulma algoritmasının varsayımı, eşiklenmiş imgede bulunan en fazla alana sahip bağlantılı bileşenin el olduğu varsayımıdır. Bu varsayım doğru olduğu sürece, eller birbirini kısmen kapatsa da el bulma işlemi, Kalman Süzgeci tarafından yumuşatılabilen ufak bir hata ile gerçekleştirilebilir. Bu varsayım doğru değilse el olduğu varsayılan bileşenin el ile hiç bir ilgisi olmayabilir. Bu bilgi Kalman Süzgecinin parametrelerini güncellemek için kullanılırsa, Kalman Süzgecinin güvenilirliği azalabilir. Bu problemi çözmek için bulunan bileşenin alanı bir eşik değeri ile karşılaştırılır. Eğer alan bu değerin altındaysa bir elin diğerini kapattığı varsayılır ve Kalman Süzgecini parametreleri güncellenmez. Kapatmanın gerçekleştiği çerçeve sayısı bir-iki çerçeve ile sınırlı kaldığı durumda, Kalman Süzgecinin yaptığı tahminlerin kullanılması problem yaratmaz.



Şekil 2: Kameraların birinde eller birbirini kapatabilir ama aynı anda diğer kamerada iki el birden görülür.

2.2. Düzgeleme

Uygulamaya göre değişebilmekle birlikte, aynı el hareketi farklı kişiler tarafından, hatta aynı kişi tarafından, farklı ölçeklerde ve farklı yerlerde yapılabilir. Bu durumun sınıflandırma başarısına etki etmemesi için öteleme ve ölçek farklılıklarının düzgelenmesi gerekir.

$((x_1, y_1), \ldots, (x_t, y_t), \ldots, (x_N, y_N))$ iki boyutlu el gezingesi ve N gezinge uzunluğu olsun. Öteleme düzgelemesinde kullanmak için, x_m ve y_m şu şekilde tanımlanır:

$$x_m = \frac{x_{max} + x_{min}}{2}, \quad y_m = \frac{y_{max} + y_{min}}{2} \tag{1}$$

ve gezingenin x ve y koordınatlarındaki orta noktasını gösterirler. Ölçek düzgelemesinde kullanmak için, δ_x ve δ_y şu şekilde tanımlanır:

$$\delta_x = \frac{x_{max} - x_{min}}{2}, \quad \delta_y = \frac{y_{max} - y_{min}}{2} \tag{2}$$

ve gezingenin x ve y koordinatlarındaki saçılımını gösterirler. Gezingenin genel şeklini bozmamak için, ölçekleme faktörü x ve y koordinatlarındaki saçılımın maksimumu olarak seçilir:

$$\delta = max(\delta_x, \delta_y) \tag{3}$$

Bu bilgiler kullanılarak, düzgelenmiş gezinge, $((x'_1, y'_1), \ldots, (x'_t, y'_t), \ldots, (x'_N, y'_N))$, $0 \leq x'_t, y'_t \leq 1$ olmak üzere, şu şekilde hesaplanır:

$$x'_t = 0.5 + 0.5\frac{x_t - x_m}{\delta} \tag{4}$$

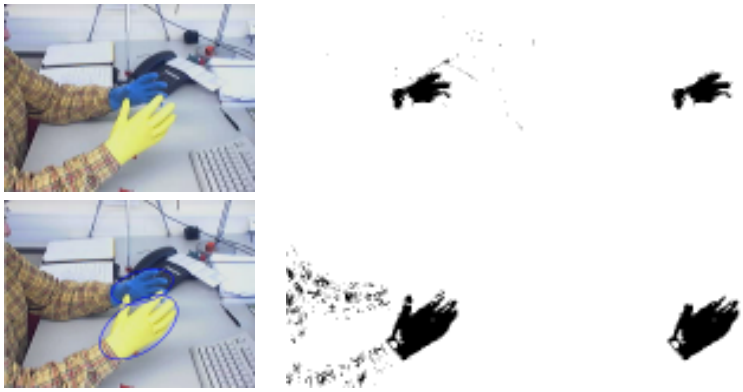
$$y'_t = 0.5 + 0.5\frac{y_t - y_m}{\delta} \tag{5}$$

Gezingedeki el şekline ait diğer bilgiler de benzer şekilde düzgelenmelidir. Elipsin genişlik ve uzunluğu aynı x ve y koordinatlarındaki düzgelemede yapıldığı gibi saçılımın maximumu kullanılarak yapılmalıdır. Aksi halde şekil bilgisi bozulabilir. Elips açısı ise bağımsız olarak düzgelenmelidir.

En son olarak, düzgelenmiş gezingelerin başlangıç noktası (0,0) koordinatına çekilir.

3. Deneyler

Her el hareketi için dört değişik gezinge çıkarılabilir: birinci kameradan sol ve sağ el gezingeleri (L1 ve R1) ve ikinci kameradan sol ve sağ el gezingeleri (L2 ve R2). Ellerin birbirini kapatma ihtimali olduğundan tek bir kameradan



(a) Orjinal çerçeve ve bulunan eller

(b) Iki el için eşikleme sonucu

(c) En yüksek alana sahip bağlantılı bileşen

Şekil 1: Çerçevede elin bulunması

elde edilen gezinge hatalar içerebilir. Ayrıca, tek bir kamera bilgisini kullanmak sınıflandırıcıyı iki boyutlu hareket bilgisiyle sınırlayacaktır. Kalibrasyon matrisinin olmaması sebebiyle üç boyuta geri çatma işlemi yapılamasa da iki kameradan gelen bilgi birleştirilerek üç boyutlu hareket bilgisi bir ölçüde elde edilebilir. İki boyutlu gezingelerin değişik aşamalarda birleştirilmesinin sistem başarımına etkisini görmek amacıyla değişik deneyler yaptık:

1. L1R1: Birinci kameradan sol ve sağ el gezingeleri (Çerçeve başına öznitelik vektörü boyutu: 10)
2. L2R2: İkinci kameradan sol ve sağ el gezingeleri (Çerçeve başına öznitelik vektörü boyutu: 10)
3. L1L2R1R2: Birinci ve ikinci kameradan sol ve sağ el gezingelerinin öznitelik aşamasında birleştirilmesi (Çerçeve başına öznitelik vektörü boyutu: 20)
4. L1R1 $\oplus$ L2R2: Birinci ve ikinci kameradan sol ve sağ el gezingelerinin karar aşamasında birleştirilmesi (Çerçeve başına öznitelik vektörü boyutu: 10+10)

Her deney için iki sınıflandırıcı eğitildi: HMM ve SVM. Değişken uzunluktaki gezingeleri SVM kullanılarak eğitebilmek için yeniden örnekleme yoluyla sabit uzunluğa çevrirdik. Gezingenin x ve y noktaları, uzamsal yeniden örnekleme yoluyla ve doğrusal aradeğerleme kullanarak 12 noktaya örneklendiler. Şekil 3'de değişik nokta sayılarıyla yeniden örnekleme sonuçları görülebilir. Örneklenmiş noktaların, genel el şeklini yansıtan öznitelikleri ise kendilerine zamansal olarak en yakın noktadan kopyalandı. SVM sonuçları LibSVM kullanılarak alındı [8]. Eğitilen SVM'lerde çekirdek fonksiyonu olarak Radyal Taban Fonksiyonları kullanıldı. Çok sınıflı SVMler bire-bir yöntemiyle eğitildi. Eğitimin öncesinde tüm öznitelikler z-normalizasyon kullanılarak düzgelendi ve ortalamadan iki standart sapma dışında kalan değerler iki standart sapmaya çekilerek aykırı değerlerin etkisi azaltıldı. HMMlerde, durum atlamasız sol-sağ modeli kullanıldı. El hareketlerinin uzunluğuna bakılarak (en uzunu 40 çerçeve) dört durumlu bir model ve her durumda bir Gauss bileşeni kullanıldı. Durum sayısını ya da bileşen sayısını arttırmanın başarıma iyi yönde farkedilir bir etkisi olmadığı görüldü.

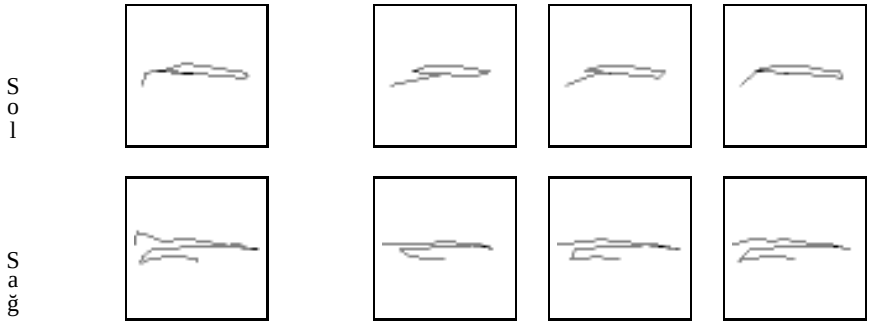
Tek kamera bilgisi kullanılarak elde edilen başarım, öznitelik ve karar aşamasında birleştirme sonuçları Tablo 1'de görülebilir. Karar aşamasında birleştirme için, HMMlerin olabilirlik sonuçları ve SVMlerin ardıl olasılıkları kullanıldı ve toplam kuralıyla birleştirildi. Tek kamera bilgisi kullanılarak elde edilen %5 civarındaki hata, iki kameradan alınan bilginin öznitelik aşamasında birleştirilerek beraber kullanılmasıyla %50'den fazla azalarak %2 civarına düşürüldü. Karar aşamasında HMM kullanarak birleştirildiğinde ise hata daha da azalarak %1'e düştü. SVM'ler kullanılarak yapılan birleştirme ise öznitelik aşamasındaki birleştirmeden daha farklı bir sonuç vermedi.

SVM sınıflandırıcısı, L1R1 kümesi hariç, HMM sınıflandırıcısından daha fazla hata yaptı. HMM'lerin sıralı bilgiyi daha iyi modellediği ve yeniden örnekleme sırasında önemli bazı bilgilerin kaybolmuş olabileceği gözönüne alınırsa bu beklenen bir sonuçtu. HMM'den daha fazla hata yapmış olmasına karşın, SVM sonuçlarının HMM sonuçlarına yakın çıkmış olması, değişik uzunluktaki gezingelerin sabit uzunluğa çevrilmesinde daha iyi bir yöntem kullanıldığı takdirde hatanın azaltılabileceğini gösteriyor.

Tablo 1: Test hataları ve standart sapmalar

Dataset	**SVM**	**HMM**
L1R1 *(cam1)*	0.057 ± 0.000	0.060 ± 0.011
L2R2 *(cam2)*	0.062 ± 0.000	0.054 ± 0.012
L1R1L2R2	0.024 ± 0.000	0.022 ± 0.007
L1R1 $\oplus$ L2R2	0.024 ± 0.000	$\mathbf{0.009 \pm 0.004}$

Karar aşamasında birleştirme uygulandığında hatanın azalması, birleştirilen sınıflandırıcıların birbirinin hatalarını yüksek oranda düzeltmesi sayesinde gerçekleşir. Kullandığımız veri kümesinde, farklı kameralardan çıkarılan gezingelerle eğitilen HMM'ler birleştirildiğinde hata,%1'e kadar azaltılabilmektedir. Tablo 2 ve 3'de birinci ve ikinci kamera bilgisiyle eğitilen HMM'lerin hata matrisleri görülebilir. Birinci kameradan çıkarılan gezingelerde daha çok ikinci ve beşinci sınıflara ait örnekler hatalı sınıflandırılmış. İkinci kamerada ise ik-



(a) Normalize edilmiş gezinge (b) Yeniden örneklenmiş gezinge

Şekil 3: Değişik nokta sayılarıyla yeniden örnekleme

inci, dördüncü ve yedinci sınıfların örneklerinde hata yapıldığı görülmeketedir. Birleştirme sonucunda, sınıflandırıcı sadece ikinci ve dördüncü sınıflarda birer hata yapmaktadır.

Tablo 2: L1R1 HMM hata matrisi

30	0	0	0	0	0	0
0	27	0	**3**	0	0	0
0	0	29	0	**1**	0	0
0	0	**1**	29	0	0	0
0	0	**2**	0	25	**3**	0
0	0	0	0	0	30	0
0	0	**1**	0	0	0	29

Tablo 3: L2R2 HMM hata matrisi

30	0	0	0	0	0	0
0	27	0	**2**	**1**	0	0
0	0	29	0	0	0	**1**
1	0	0	28	0	**1**	0
0	0	0	0	30	0	0
0	0	0	0	**1**	29	0
0	0	0	**1**	0	**1**	28

4. Sonuçlar

Bu çalışmada üç boyutta yapılan el hareketlerini iki boyut için çıkarılmış özniteliklerle tanıyan bir sistem oluşturduk. Üç boyutta geri çatma işlemi yapılmadan, iki kameradan çıkarılan iki boyutlu el gezingelerinin öznitelik ya da karar aşamasında birleştirilmesinin sistemin başarımını arttırdığı görüldü. Sistem, gezingeler öznitelik aşamasında birleştirildiğinde, toplam 210 test örneği için %2.5 hata, karar aşamasında birleştirildiğinde ise %1 hata yapmaktadır.

Bu çalışma DPT/03K120250 "Algısal İnsan Bilgisayar Etkileşimi" ve AB 6. çerçeve programı SIMILAR projeleri tarafından desteklenmektedir.

Tablo 4: L1R1⊕L2R2 HMM hata matrisi

30	0	0	0	0	0	0
0	29	0	**1**	0	0	0
0	0	30	0	0	0	0
0	0	**1**	29	0	0	0
0	0	0	0	30	0	0
0	0	0	0	0	30	0
0	0	0	0	0	0	30

5. Kaynakça

[1] Vladimir Pavlovic, Rajeev Sharma, and Thomas S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, 1997.

[2] Sylvie C. W. Ong and Surendra Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873–891, 2005.

[3] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.

[4] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 26, no. 1, pp. 5–28, 1998.

[5] A. Bobick and J. Davis, "Real-time recognition of activity using temporal templates," in *Proceedings of the Workshop on Applications of Computer Vision*, 1996.

[6] Ying Wu and Thomas S. Huang, "Hand modeling, analysis, and recognition for vision based human computer interaction," *IEEE Signal Processing Magazine*, vol. 21, pp. 51–60, 2001.

[7] Sebastian Marcel and Agnes Just, "IDIAP Two handed gesture dataset," Available at http://www.idiap.ch/∼marcel/.

[8] Chih-Chung Chang and Chih-Jen Lin, *LIBSVM: a library for support vector machines*, 2001, Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.