

# Derinlik İmgesi Temelli 3B El Pozu Bulma Sistemi

## Depth Image Based 3D Hand Pose Estimation Framework

Furkan Kırac, Yunus Emre Kara, Cem Keskin, Lale Akarun

Bilgisayar Mühendisliği Bölümü  
Boğaziçi Üniversitesi, İstanbul

kiracmus@boun.edu.tr, yunus.kara@boun.edu.tr, keskinc@cmpe.boun.edu.tr, akarun@boun.edu.tr

### ÖZETÇE

Gerçek zamanlı 3B el hareketi yakalama insan-bilgisayar etkileşiminde yeni yollar açacak bir gelişmedir. Bu çalışma, derinlik imgesi kullanarak insan eline 3B iskelet oturtmak üzere geliştirdiğimiz sistemi anlatmaktadır. İnsan elini 15 eklemden oluşan bir 3B el iskeleti ile temsil ediyoruz. Bu modeli kullanarak çeşitli sentetik el derinlik imgeleri oluşturuyoruz. Rassal karar ormanları (RDF), piksel başına el parçası sınıflandırma yapmak için eğitilir ve kullanılır. Ortalama kaydırma yöntemi, piksel sınıflandırma sonuçlarını kullanarak iskelet eklemlerinin 3B koordinatlarını bulmak üzere kullanılmaktadır. Sistemimiz, Kinect imgelerini kullanarak gerçek zamanlı olarak, 30 fps hızında çalışabilmektedir.

### ABSTRACT

Real-time 3D motion capture for the human hand opens many avenues for HCI. This work describes our framework for fitting a 3D skeleton to the human hand using depth images. We represent a human hand by a 3D skeleton with 15 joints. Using this model, various synthetic depth images are generated. Random Decision Forests (RDF) are trained and used to assign each pixel to a hand part. A mean-shift method is used for estimating joint locations using pixel classification results. Our system runs in real time at 30 fps on Kinect depth images.

## 1. GİRİŞ

Çoklu dokunmatik destekli akıllı telefonlar ve işletim sistemlerinin çıkmasından sonra başta el hareketleri olmak üzere doğal arayüzlerin kullanımı yeniden popüler oldu. Oyunlar, internet gezginleri, insan-bilgisayar etkileşimi gibi alanlarda el hareketleri kullanılmaya başlandı. Bilgisayarla görme temelli el hareketi tanıma, özellikle işaret dili tanıma, 20 yıldan uzun süredir araştırmacıların ilgisini çeken bir problemdir. Buna rağmen çıplak eli saptayıp gürbüz bir şekilde el pozunu ve hareketlerini tanıyan kapsamlı bir sistem hala geliştiremedi. Bu problemin en büyük nedenleri: i) farklı ışık koşullarında elde edilen renklerin yüksek değişintili olması; ii) ten renginin insandan insana çok değişmesi; iii) eli diğer vücut parçalarından bölütlemenin zor olmasıdır. Derinlik algılayıcı içeren bir sistem olan Kinect'in [1] çıkmasıyla, karanlıkta bile el saptama ve bölütleme

işlemleri oldukça kolaylaştı. Bunun üzerine, basit el hareketlerini tanıyan kütüphaneler geliştirildi. Fakat, bunlar sadece elin hareketini takip eden, ancak pozunu dikkate almayan sistemlerdir. 3B el iskeletini kestirimi hala tam olarak çözülememiş bir problemdir.

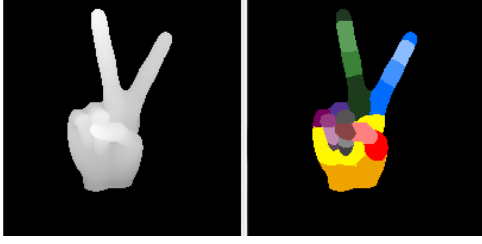
Shotton ve diğerlerinin [2], Kinect derinlik imgeleri üzerinden gerçek zamanlı insan vücudu iskeleti çıkarması doğal arayüz temelli uygulamalar için çığır açıcı olmuştur. Bu fikir el pozu kestiriminde de kullanılabilir, fakat bu problemde bazı temel farklılıklar ve zorluklar mevcuttur: i) Elin derinlik imgesi, vücudunkine göre oldukça küçüktür; ii) Vücudun baş-yukarı olduğu varsayılabilir, fakat el herhangi bir oryantasyonda olabilir; iii) Elde, konfigürasyon sayısı daha fazladır ve örtmeye daha fazla rastlanır. Öte yandan, insanlar arası el şekli ve boyutu değişimi, vücut ve kıyafet değişimine göre daha azdır.

Erol ve diğerleri [3], el pozu kestirimi problemine yaklaşımlarına bir özet vermiştir. Çoğu çalışma RGB imgeleri ile çalışırken, derinlik imgeleri ile çalışan birkaç örnek vardır [4, 5, 6, 7, 8]. Pek çok çalışmada, 3B bir model üzerinden sentetik imgeler yaratılmış ve bunlar sınıflandırıcıların eğitimi için kullanılmıştır [2, 5].

Sistemimizin amacı gerçek zamanlı olarak derinlik imgelerinden insan elinin iskeletini bulmaktır. Sistemimiz, Keskin ve diğerlerinin yöntemini kullanmaktadır [9].

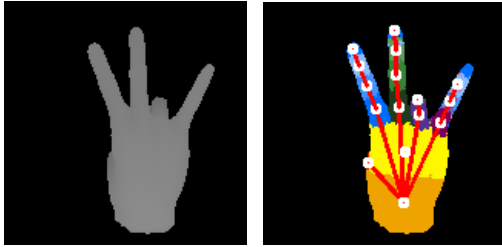
## 2. 3B EL POZU BULMA SİSTEMİ

Demomuzda Kinect kullanarak yakalanan insan eli derinlik imgelerinden gerçek zamanlı olarak iskelet bilgisini kestirmekteyiz. Gerçek zamanlı olarak çalışmaya en uygun yaklaşımlardan olan Rassal Karar Ormanı'nı kullanarak eğitilmiş karar ağaçları ile derinlik imgelerindeki el kısımları etiketlenmektedir. Karar ağaçlarımızı eğitirken sentetik olarak üretilmiş derinlik ve etiket imgeleri kullanılmaktadır. Şekil 1'de örnek imge seti görülebilir. Etiketlerin eğitimi sırasında kullanılan kısımların saha gerçeği değerleri, orta noktaları el eklemlerine denk gelecek şekilde konumlandırılmaktadır. Yeni bir derinlik resmi geldiğinde el eklemlerinin pozisyonlarını sınıflandırabilmek için karar ağaçlarımıza danışırız. Karar ağaçlarının oluşturduğu sınıflandırılmış piksel resimlerdeki kısımların orta noktalarının eklemlerin orta noktalarını verdiği varsayılır. Doğrudan karar ağaçlarımızca sınıflandırılmış resimlerdeki piksellerin orta noktalarını bulmak algoritmamız tarafından yanlış bulunacak pikseller yüzünden doğru çalışmayabilir. Uç noktalarda yanlış bulunmuş pikseller eklemlerimizin orta nokta pozisyon-



Şekil 1: Eğitim imgeleri

larını büyük miktarda bozacaktır. Bu yanlışlığı en aza indirmek için tüm piksellerin orta noktalarını bulmaktansa, ortalama kaydırma yöntemi kullanılarak eklem merkezleri bulunmaktadır. Bulunabilmiş eklem merkezleri kendilerinin hangi ekleme ait olduğunu bilmektedirler. El iskeletinin hangi eklemlerin ne şekilde birleşmesinden oluştuğu zaten önsel olarak bilinmektedir. Dolayısıyla el iskeletini elde etmiş oluruz. Şekil 2’te örnek bir el iskeleti çizimi görülebilir.



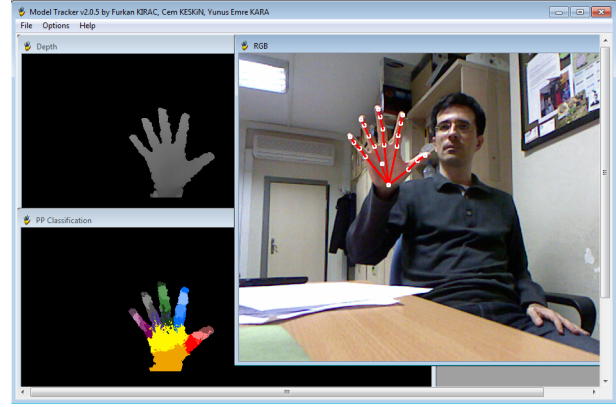
Şekil 2: El iskeleti kestirimi

### 3. DEMO TANIMI

Sistemimiz 4 çekirdekli Core i7 işlemcili bir bilgisayarda gerçek zamanlı olarak el iskeletini çıkarabilmektedir. Bu işlem sırasında herhangi bir zamansal boyut süzgeci kullanılmamaktadır. Zamansal süzgeçler kullanarak sistemin bulduğu sonuçların daha da iyileştirilmesi mümkündür.

Sistemimiz sentezleyici, el bölütleyici, piksel sınıflandırıcı ve iskelet oturtucu gibi alt modüllerden oluşmaktadır. Derinlik kamerasından gelen imge el bölütleyici modülümüz sayesinde sadece insan eline ait piksellerin derinlikleri kalacak şekilde bölütlenir. Ardından rassal karar ormanı koşan bir piksel sınıflandırıcı ile tüm derinlik pikselleri bir ekleme atanır. İskelet oturtucu modülümüz ise piksellerin eklemlere aidiyet bilgisini kullanarak eklem merkezlerini ve bütün iskeleti oluşturur. Yüksek hız gereksinimimizi karşılayabilmek için tüm algoritmalar OpenMP kullanarak işlemcinin çekirdeklerine dağıtılarak programlanmıştır. 8 sanal çekirdekli bir işlemcide 7 kat hız artışı görülmektedir. Rassal karar ormanlarının bir avantajı olarak sınıflandırma işinin büyük kısmı eğitim aşamasında çözüldüğünden karar aşamasında paralel olarak çalışmaya çok uygun algoritmalar oluşturulabilmektedir. Günümüzde popüler olarak kullanılan grafik işlemci uygulamaları ile de ekstra hız kazanmak mümkündür. Fakat sadece

işlemci çekirdeklerini dağıtık olarak kullanarak istediğimiz hızı karşılayabilmiş durumdayız. Şekil 3 sistemin çalışma esnasındaki bir görüntüsünü gösterir.



Şekil 3: Sistem çalışma görüntüsü

### 4. KAYNAKÇA

- [1] Microsoft Corp. RedmondWA. Kinect for Xbox 360.
- [2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR*, 2011.
- [3] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73, Oct. 2007.
- [4] M. de La Gorce, D. J. Fleet, and N. Paragios. Model-Based 3D Hand Pose Estimation from Monocular Video. *IEEE transactions on pattern analysis and machine intelligence*, pages 1–14, Feb. 2011.
- [5] S. Malassiotis and M. Srinivasan. Real-time hand posture recognition using range data. *Image and Vision Computing*, 26(7):1027–1037, July 2008.
- [6] P. Suryanarayan, A. Subramanian, and D. Mandalapu. Dynamic Hand Pose Recognition Using Depth Data. *2010 20th International Conference on Pattern Recognition*, pages 3105–3108, Aug. 2010.
- [7] Z. Mo and U. Neumann. Real-time Hand Pose Recognition Using Low-Resolution Depth Images. *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR'06)*, pages 1499–1505, 2006.
- [8] I. Oikonomidis, N. Kyriazis, and A. Argyros. Markerless and efficient 26-DOF hand pose recovery. In *Proceedings of the 10th Asian conference on Computer vision-Volume Part III*, pages 744–757. Springer, 2011.
- [9] C. Keskin, F. Kirac, Y.E. Kara, and L. Akarun. Real-time hand pose estimation using depth sensors. In *Proceedings Thirteenth IEEE International Conference on Computer Vision Workshops. ICCV 2011*, pages 1228–1234.