

ÖLÇEKTEN BAĞIMSIZ ÖZNETELİK DÖNÜŞÜMÜ KULLANARAK STEREO KAMERA İLE ÜÇ BOYUTLU KAFA TAKİBİ

STEREO BASED 3D HEAD POSE TRACKING USING THE SCALE INVARIANT FEATURE TRANSFORM

Batu Akan, Müjdat Çetin, Aytül Erçil

Mühendislik ve Doğa Bilimleri Fakültesi
Sabancı Üniversitesi, Tuzla, ISTANBUL
batuakan@su.sabanciuniv.edu,
{mccetin,aytulercil}@sabanciuniv.edu

Özetçe

Bu makalede üç boyutlu kafa takibi için ölçekten bağımsız öznetelik dönüşümüne (SIFT) dayalı bir yöntem önerilmektedir. 3B kafa takibi bir çok bilgisayarla görme uygulaması için önemli bir önlemdir. Önerilen yöntemin, düzlem dışı öteleme ve dönmelere karşı gürbüz olduğu belirlenmiş aynı zamanda görüntüdeki ani değişen aydınlanma farklarından da etkilenmediği gözlenmiştir. Kafa takibi ile elde edilebilecek dönmeye göre düzeltilmiş bir imge ile yüz tanınması, ifade analizi, dudak okuması gibi problemleri çözmek daha kolay olacaktır. Önerdiğimiz SIFT tabanlı yöntemin başarısını sentetik ve stereo kamera ile çekilmiş gerçek görüntüler üzerinde deneyip var olan diğer yöntemlerle karşılaştırmasını yaptık.

Abstract

In this paper a new stereo-based 3D head tracking technique, based on scale-invariant feature transform (SIFT) features is proposed. A 3D head tracker is very important preprocessing for many vision applications. The proposed method is robust to out of plane rotations and translations and also invariant to sudden changes in time varying illumination. We present experiments to test the accuracy of our SIFT based 3D tracker on sequences of synthetic and real stereo images.

1. Giriş

Kafa takibi, bir çok bilgisayarla görme uygulaması için önemli bir süreçtir. Eğer kafanın üç boyutlu uzaydaki yeri ve duruşu bilinirse, yüz tanınması, ifade analizi, dudak okuması gibi problemleri, 3B kafa izleyicisi tarafından oluşturulan dengelenmiş imgeleri kullanarak çözmek

daha muhtemeldir. Kafa takibi uygulamalarından bir tanesi, araçlarda sürücü ile etkileşime geçebilen akıllı sistemlerinin üretilmesidir. Sürücünün kafasını takip eden özışler bir sistem yardımıyla sürücünün dikkatini yoldan başka bir yere verdiği durumlar ya da sürücünün dalgın, veya yorgun olduğu durumlar anlaşılabilir. Kafa takibi aynı zamanda güçlü bir yer belirteci olarak da kullanılabilir. Kullanıcının el ve ayaklarının başka işler ile meşgul olduğu durumlarda kafa pozisyonunun bilinmesi sisteme farklı bir girdi olarak kullanılabilir. Kafa takibi, bilgi büfelerinde ve uçak kokpitlerinde insan bilgisayar arayüzünün bir parçası olarak kullanılabilir. Aynı zamanda eller serbest (handsfree) bir arayüz olarak bilgisayar oyunlarında ve engelli bilgisayar kullanıcıları için fare vazifesi görebilir.

Günümüzde derinlik tarayıcılarının ya da stereo kameraların yaygınlaşmasıyla, gerçek zamanlı 3B verinin elde edilmesi mümkün olmaya başlamıştır. Nesne takibi için stereo kamera kullanmak, normal kameralara göre oldukça avantajlıdır. Nesne bölütleme gibi problemler kolaylıkla aşılabılır. Stereo kameradan elde edilen derinlik görüntüsü aynı zamanda aydınlanma değişikliklerinden etkilenmediği için nesne takibi yöntemlerinin başarımını da arttırmaktadır.

3-B kafa hareketi, katı devinim hareketi olarak düşünülebilir ve bu hareketi video süresince takip etme problemi ise çakıştırma problemi olarak görülebilir. Derinlik imgesi kullanılarak elde edilen kafa geometrisi bir önceki video karesinden elde edilen yüzey ile çakıştırılarak, kafanın iki kare arasında ne kadar ötelendiği ve döndürüldüğü hesaplanabilir. Döngülü en yakın nokta algoritması (ICP) Besl ve McKay [1] ve Chen ve Medioni [2] tarafından önerilmiştir ve 3B nesne çakıştırması için yaygın olarak kullanılmaktadır. Morency [3] ICP algoritmasını Normal Flow Constraint (NFC) [4] yöntemi ile melezleyerek kafa duruşu takibi uygulamasında kullanmıştır.

Bu çalışma DPT'nin "Güvenli Sürüş" başlıklı ve Avrupa Komisyonunun MIRC-CT-2006-041919 ve FP6-2004-ACC-SSA-2 sayılı projeleri kapsamında desteklenmiştir.

Bu bildirinin özgün teknik katkısı ölçekten bağımsız öznetelik dönüşümünün (SIFT) 3B kafa takibi probleminde ilk defa kullanılmasını sağlayan bir yöntemin önerilmesidir.

2. Önerilen Yöntem

Bu makalede 3-B kafa takibi için döngülü en yakın nokta (ICP) algoritmasının üstüne iyileştirme olarak, ölçekten bağımsız öznetelik dönüşümüne dayalı ilişkilendirme fonksiyonu içeren bir yöntem önerilmektedir. ICP algoritması 3B uzaydaki iki nokta kümesi P ve X 'in birbirine çakıştırılmasında kullanılmaktadır. Kafa takibi uygulaması için P ve X kümeleri, yüz ve kafa üzerinde bulunan tüm noktaların 3B uzaydaki yerlerini içermektedir. ICP algoritması döngülü olarak çalışan iki işlem grubundan oluşur. Birinci aşamada iki nokta kümesi arasındaki noktalar birbirleri ile ilişkilendirilir. İkinci aşamada ise ilişkiler aşağıdaki maliyet fonksiyonu kullanılarak öteleme \mathbf{t} ve dönme \mathbf{R} üzerinden en küçüklenir.

$$e(\mathbf{R}, \mathbf{t}) = \frac{1}{N} \sum_i \|\mathbf{R}p_i + \mathbf{t} - c(p_i)\|^2, p_i \in P \quad (1)$$

Bu denklemdeki N , P kümesindeki nokta sayısını belirtmektedir. c fonksiyonu \mathbf{P} kümesindeki her noktayı \mathbf{X} kümesindeki bir nokta ile ilişkilendirir ve şu şekilde tanımlanabilir.

$$c : P \rightarrow X | \forall p_i \in P, y_i = c(p_i) \in X \quad (2)$$

Ancak farklı zamanlarda elde edilmiş derinlik imgeleri kullanılarak elde edilen 3B iki model arasındaki nokta ilişkileri bilinemediği için c fonksiyonu belirsizdir. Bu sebeple bilinmeyen c fonksiyonu yerine yaklaşık \hat{c} fonksiyonu atanır ve denklem 1'deki maliyet fonksiyonu parçalara ayrılarak döngülü olarak çözülür. Atanan \hat{c} fonksiyonu gerçek ilişkileri ne kadar iyi tanımlarsa, algoritma o kadar çabuk yakınsar. ICP algoritması \hat{c} fonksiyonunu, \mathbf{P} kümesindeki her nokta için \mathbf{X} kümesindeki en düşük Öklid uzaklığına sahip nokta nokta olacak seçer. Ancak böyle bir eşleme her zaman iyi sonuç vermediği gibi, algoritmanın döngüsel yapısı ağır bir hesapsal yük getirmektedir. Bu sebeple bu makalede c fonksiyonu yerine SIFT [5] özneteliklerini kullanarak ilişkilendirme yapan bir yaklaşık \hat{c} fonksiyonu kullanılmıştır.

2.1. Ölçekten Bağımsız Öznetelik Dönüşümü

Ölçekten bağımsız öznetelik dönüşümü (SIFT) [5] bilgisayarla görme uygulamalarında sıklıkla kullanılan özişler bir öznetelik çıkartma yöntemidir. Elde edilen öznetelikler dönme, öteleme ve ölçeğe göre değişimsizdirler aynı zamanda ilgin dönüşüme

ve aydınlatmadaki doğrusal ve doğrusal olmayan değişimlere kısmi olarak değişimsizdirler. Her öznetelik için atanan 128 boyutlu betimleyici vektör sayesinde bu öznetelik vektörleri birbirleri ile kolaylıkla eşlenebilir.

Özneteliklerin çıkartılması işlemi 4 ana başlık altında toplanabilir. İlk aşamada sistem, ölçek uzayı uç noktalarını arayarak olası öznetelik noktalarını çıkartır. Ölçek uzayı, $L(x, y, \sigma)$, sisteme girilen imgenin $I(x, y)$, çoklu-Gauss $G(x, y, \sigma)$ çekirdeği ile evriştirilmesiyle elde edilir.

$$L(x, y) = G(x, y, \sigma) * I(x, y) \quad (3)$$

Kararlı öznetelik noktası çıkarımı için aşağıdaki gibi tanımlanan Gausslar farkı (DoG) fonksiyonu kullanılır.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (4)$$

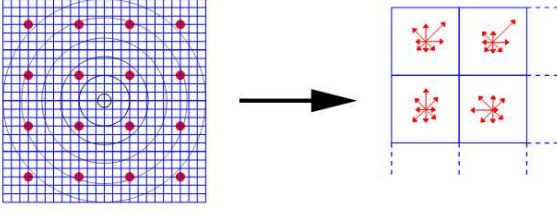
DoG imgeleri hesaplandıktan sonra, $D(x, y, \sigma)$ da tanımlanmış bir piksel üst ve alt ölçekteki 18 ve bulunduğu ölçekteki 8 komşusuyla karşılaştırılır. Eğer ki karşılaştırılan noktaların hepsinden büyük ya da küçük değere sahipse bu nokta aday öznetelik noktası olarak seçilir.

İkinci aşamada, düşük karşıtlığa sahip ya da kenarlar üzerinde seçilen aday öznetelik noktaları, iyi betimleyici özelliklere sahip olmadıkları için bu adayların elenmesi gerekmektedir. Nokta etrafındaki küçük bir alandan alınan uzamsal türevler kullanılarak hazırlanan 2×2 'lik bir Hessian matrisinin özdeğerlerinin birbirlerine oranına bakılarak kenar benzeri yerler üstünde bulunan aday öznetelik noktaları elenir.

Üçüncü aşamada belirlenen öznetelik noktaları için, buldukları ölçekten alınan uzamsal türevler kullanılarak her bir öznetelik noktası için yönelim belirlenir. Hesaplanan türevlerden bir yönelim histogramı oluşturulur ve en yüksek bir veya bir kaç tepelik öznetelik noktası için yönelim vektörü olarak atanır. Böylelikle seçilen öznetelik noktası dönmelere karşı değişimsiz hale getirilmiş olur.

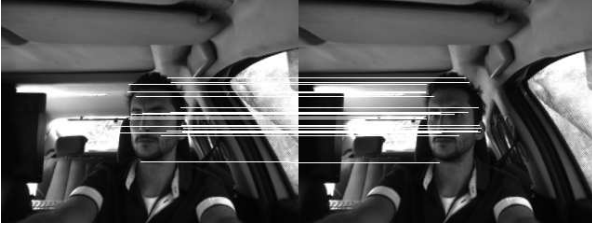
Son aşamada ise öznetelik noktaları için birer betimleyici vektör hesaplanır. Bu işlem sırasında yine öznetelik noktası etrafındaki 16×16 lık bir bölgeden alınan uzamsal türevler kullanılır. Elde edilen türevler 3. aşamada elde edilen yönelim vektörüne göre düzeltilirler ve böylelikle dönmelere karşı değişimsizlik korunmuş olur. Şekil 1'deki gibi uzamsal türevlerden 4×4 'lük öbekler için, 8 bölmeli histogramlar oluşturulur. Histogramdaki her bir bölmedeki değer 128 boyutlu betimleyici vektörün bir elemanıdır. 128 boyutlu vektör birim uzunluğa getirilerek aydınlanmadaki değişimin etkileri ortadan kaldırılır.

Öznetelik vektörleri çıkartılıp betimleyici vektörler tanımlandıktan sonra farklı iki imge arasındaki eşleştirmeler betimleyici vektörler arasındaki Öklid



Şekil 1. SIFT betimleyici vektörlerinin oluşturulması.

uzaklığına bakılarak yapılır. Eğer ki Öklid uzaklığı belirli bir eşik değerinin altında ise bu iki nokta birbiri ile eşlenir. Şekil 2 de yüz için hesaplanıp eşleştirilmiş noktalar görülmektedir.



Şekil 2. SIFT ile eşleştirilen noktalar

2.2. Hizalanmanın hesaplanması

Video süresince t ve $t - 1$ anlarında elde edilen imgeler için SIFT öznitelik vektörleri çıkartılıp birbirleri ile eşlendikten sonra 3B ilişki kümesi $C(p_i, y_i)$, derinlik imgesi kullanılarak tanımlayıcı piksellerin ters izdüşüm matrisi ile geriye çatılması ile oluşturulur. Böylece piksellerin dünya koordinatlarındaki (x, y, z) yerleri belirlenmiş olur. t ve $t - 1$ anları arasındaki dönme ve ötelemeyi hesaplamak için birim kuaterniyon [6] yöntemi kullanılır.

P ve Y nokta kümeleri 0 ortalamaya sahip olacak şekilde orijine çekilirler.

$$\mu_p = \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i, \mu_y = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i \quad (5)$$

$$\mathbf{p}'_i = \mathbf{p}_i - \mu_p, \mathbf{y}'_i = \mathbf{y}_i - \mu_y,$$

Böylece maliyet fonksiyonundaki t düşmüş olur ve denklem sadece R üzerinden en küçüklenir.

$$e(\mathbf{R}) = \sum_{i=1}^N \|\mathbf{R}\mathbf{p}'_i - \mathbf{y}'_i\| \quad (6)$$

Orijine çekilmiş nokta kümelerinden 3×3 'lük bir çapraz ilinti matrisi elde edilir.

$$\zeta_{py} = \sum_{i=1}^N \mathbf{p}'_i \cdot \mathbf{y}'_i \quad (7)$$

3×3 'lük bu matristen denklem 8 de belirtildiği gibi 4×4 'lük yeni bir A matrisi elde edilir.

$$\mathbf{A} = \begin{bmatrix} \text{trace}(\zeta_{py}) & & & \\ & \Delta & & \\ & & \zeta_{py} + \zeta_{py}^T - \text{trace}(\zeta_{py})\mathbf{I}_3 & \\ & & & \Delta^T \end{bmatrix} \quad (8)$$

\mathbf{I}_3 3×3 'lük birim matristir ve $\Delta, \mathbf{C}_{ij} = \zeta_{py} - \zeta_{py}^T$ olarak belirlenen anti simetrik matrisin $\Delta = [\mathbf{C}_{23} \ \mathbf{C}_{31} \ \mathbf{C}_{12}]^T$ elemanlarından oluşmuş bir kolon vektördür. [6] da ispatlandığı gibi A matrisinin özdeğerlerinin en küçüğüne denk düşen özvektör, P kümesini X 'e döndüren R matrisinin kuaterniyon cinsinden denkleğini verir. Basit trigonometrik fonksiyonlar kullanılarak dönmeye ait R matrisi ya da Euler açıları elde edilebilir.

3. Deneysel Sonuçlar

Önerdiğimiz SIFT öznitelik noktalarına dayalı 3B kafa takibi yöntemini hem sentetik olarak üretilmiş, hem de gerçek ortamlarda, aydınlanma değişimli ve değişimsiz olarak çekilmiş videolar üzerinde denedik ve sonuçlarını literatürde daha önceden bu iş için önerilmiş iki diğer yöntem olan NFC [4] ve ICP [3] ile karşılaştırdık. Yöntemleri birbirleri ile karşılaştırmak ve nicel sonuçlara varabilmek için kendi belirlediğimiz yer doğruluk verileri kullanarak kamera gürültüsü içermeyen sentetik veriler yarattık. Sentetik videolarda, kafa x, y ve z eksenlerinde sırasıyla 10cm ötelendi ve x, y ve z eksenleri etrafında düzlem içi ve dışı olarak 40° ye kadar döndürüldü. Tablo 1'de yer doğruluk değerlerine göre hesaplanan mutlak ortalama hata ve hataların değişimleri gösterilmektedir.

Tablo 1. Kafa takibi sistemlerinin sentetik görüntüler üzerindeki ortalama hata ve değişimleri. İlk 3 satırda x, y, z eksenlerinde cm cinsinden öteleme, 2. 3 satırda ise x, y, z eksenleri etrafında derece cinsinden dönme gösterilmektedir.

	NFC		ICP		SIFT	
	Ort.	Değ.	Ort.	Değ.	Ort.	Değ.
X	2.21	1.27	0.84	0.43	1.57	0.59
Y	5.20	11.20	5.31	7.09	1.11	0.27
Z	0.85	0.17	1.71	1.86	0.67	0.08
	NFC		ICP		SIFT	
	Ort.	Değ.	Ort.	Değ.	Ort.	Değ.
X	4.42	5.93	5.36	5.36	2.84	3.57
Y	2.52	4.78	28.25	151.93	3.78	4.88
Z	2.56	3.20	17.69	74.10	2.81	4.85

Sentetik videoların yanı sıra sistemi gerçek videolar üzerinde de denedik. Polhemus [7] adlı manyetik, 6 serbestlik derecesine sahip bir takip sistemi kullanarak yer doğruluk değerleri kaydettik. Tablo 2'de gerçek video verileri üzerinde yapılan denemelerden elde

edilen mutlak ortalama hata ve deęişintileri görülebilir. Tablo 2'deki verilere göre SIFT yöntemi dönme ve ötelemelerde dięer yöntemlere göre daha iyi sonuç vermiştir. Şekil 3'te ise zamana baęlı aydınlanma deęişiklikleri altında yöntemlerin sonuçları, girdi imgeleri üstüne bindirilerek gösterilmektedir.



Şekil 3. 90, 100, 180, 240, 360. karelerde NFC, ICP ve SIFT algoritmalarından elde edilen sonuçlar. 90 ve 100. karelerde aydınlanmanın yarattığı deęişimler görülebilir.

Tablo 2. Kafa takibi sistemlerinin zamana baęlı aydınlanma deęişimi altındaki ortalama hata ve deęişintileri. İlk 3 satırda x, y, z eksenlerinde cm cinsinden öteleme, 2. 3 satırda ise x, y, z eksenleri etrafında derece cinsinden dönme gösterilmektedir.

	NFC		ICP		SIFT	
	Ort.	Deę.	Ort.	Deę.	Ort.	Deę.
X	1.76	2.15	1.45	1.51	1.65	1.65
Y	0.69	0.55	1.16	1.43	0.98	1.23
Z	1.21	1.61	2.62	4.47	1.16	1.44
	NFC		SIFT		ICP	
	Ort.	Deę.	Ort.	Deę.	Ort.	Deę.
X	3.82	13.06	5.39	28.59	3.14	8.15
Y	6.03	28.84	10.13	107.33	4.63	19.87
Z	4.23	18.75	4.66	28.98	3.04	6.95

4. Sonuç ve İleriki Çalışmalar

Bu makalede döngülü en yakın nokta yöntemine dayalı, ilişkilendirme fonksiyonu olarak ise ölçekten bağımsız öznelik dönüşümünü kullanan bir yöntem önerilmiştir. Önerilen bu yöntem ile stereo kamera ile çekilen videolarda insan kafası 6 serbestlik derecesi ile takip edilebilmektedir. Yöntemin düzlem dışı öteleme ve dönmelere gürbüz olduğu gözlenmiş, aynı zamanda görüntüdeki ani deęişen aydınlanma farklarından da etkilenmedięi gözlenmiştir.

5. Kaynakça

- [1] P.J. Besl ve N.D. McKay, "A method for registration of 3D shapes," IEEE Trans. Pattern Anal. Mach. Intell., Cilt 14, no. 2, sayfa 239–256, 1992.
- [2] Y. Chen ve G. Medioni, "Object modeling by registration of multiple range images," cilt 3, sayfa 2724–2729, 1991
- [3] L.P. Morency, "Stereo-Based Head Pose Tracking Using Iterative Closest Point and Normal Flow Constraint," Haziran 2002
- [4] M. Harville, A. Rahimi, T. Darrell, G. Gordon ve J. Woodfill, "3D Pose Tracking with Linear Depth and Brightness Constraints," ICCV, 1999, sayfa 206–213
- [5] D. Lowe "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, cilt 20, sayfa 91–110, 2003
- [6] B.K.P. Horn, "Closed-form solution of absolute orientation using unit quaternions," Journal of the Optical Society of America, cilt 4, sayfa 629, 1987
- [7] Polhemus Inc. FasTRAK, www.polhemus.com