

Kamera görüntülerinden gidilen yolun kestirimi

M. Fatih AMASYALI*, Aykut MÜNÜK, Münir SALİ

Yıldız Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü, İstanbul, Türkiye

Özet

Çalışmada hareket eden bir nesne üzerine yerleştirilen bir kameradan alınan görüntülerle nesnenin 2 boyutlu ortamda gittiği yolun bulunması amaçlanmıştır. 2 boyutlu gidilen yolun bulunmasında temel olarak izlenen yöntemlerden biri ve kamera görüntülerinden gidilen yolun kestiriminde kullanılan yöntem olan görüntü karelerinden ortak noktaların çıkarımı ve bu ortak noktalarındaki yer değiştirmesinin bulunması ile hareketli nesnenin konumunun bulunması olmuştur. Metot olarak öncelikle kameradan alınan ardışık görüntülerdeki eşleşen SIFT (Scale-invariant feature transform- Ölçekten Bağımsız Öznitelik Dönüşümü) algoritmasının çıktıları olan kilit noktaları elde edilmektedir. Daha sonra bulunan kilit noktalardan eşleşen noktaların koordinat farklarının histogramları kullanılarak bu iki görüntü arasında nesnenin gittiği yoldaki mesafenin büyüklüğü ve gittiği yolun yönü belirlenmektedir ve gidilen yol 2 boyutlu uzayda yönlü doğru parçaları dizisiyle temsil edilmektedir. Bir cep telefonu kamerasıyla gerçek ortamlarda yapılan deneyler sonucunda gidilen yolun, hareketsiz ortamlarda başarılı, hareketli (kameranın bağlı olduğu nesne haricinde başka hareketli nesnelere olduğu) ortamlarda ise gürültülü olarak belirlenebildiği görülmüştür. Ayrıca, gürültü oluşumunun nedeni ise hareketli olan nesnelere üzerinde kilit noktalarının bulunması ve bir sonraki görüntüde olan aynı hareketli nesne üzerindeki kilit noktalarının bulunup eşleştirilmesidir. Bu yöntemin doğruluğu ise görsel odometri ile elde edilen sonuçlar ve tekerlek odometrim ile elde edilen sonuçların kümülatif hata toplamalarının karşılaştırılmasıyla ölçülebilmektedir.

Anahtar Kelimeler: *SIFT, Ölçekten Bağımsız Öznitelik Dönüşümü, Görsel Odometri, Kilit nokta, Odometre*

*Yazışmaların yapılacağı yazar: M. Fatih AMASYALI. mfatih@ce.yildiz.edu.tr

Moving path estimation from video sequences

Extended abstract

The aim of the study is to determine the 2D trajectory of a moving object by utilizing images obtained by a camera which is inserted on the object. Initially, consecutively generated images from the camera were subjected to analysis using the SIFT algorithm – a methodological step which allowed for the identification and computational description of the key points (interested point) that were to be used for the generation of the 2D trajectory. The ability to detect and match features across multiple views of a scene is a crucial first step in many computer vision algorithms for dynamic scene analysis. State-of-the-art methods such as SIFT perform successfully when applied to typical images taken by a digital camera or camcorder. However, these methods often fail to generate an acceptable number of features when applied to medical images, because such images usually contain large homogeneous regions with little color and intensity variation. Feature detection and description are possibly the most important steps in many computer vision algorithms. Distinctive image features can be used to establish matches across multiple images in a video sequence. Next, the length and direction of the object trajectory were determined by using histograms which facilitated the superimposition of the SIFT obtained key points and their corresponding points in real-time. More and more, the experiments conducted in real-time environment with a mobile phone camera, showed that the trajectory could successfully be determined if there are static objects surrounding the moving object and the camera; while the environment of the motile object is consisted of non-static objects, the determined trajectory is followed by a noisy background. The noise occurs as a result of the mapping of the key points of a motile object in the environment to the key points (interested point) corresponding to the same motile object but only few seconds later as the object moves to another position. Also, the accuracy of the method can be evaluated once the visual odometry results are compared for net cumulative error to the results obtained with wheel odometry. In other words, visual odometry makes use of an image sequence to estimate the motion of a robot and optionally the structure of the world. For example, the low-cost and small-size of cameras, combined

with the high-information content of the images they capture make them ideal for robot platforms. In the meantime we compare different approaches within this framework and show that relative orientation is superior to using absolute orientation to estimate pose. We test our algorithm on outdoor and indoor environments as well as present results showing its effectiveness. SIFT method was used to detect the identical points of the successive frames. In addition to this, the histograms of the coordinate differences of the matched points are used for estimating the path. The respective pathway was represented in two dimensional spaces by directional combining of the matched (identical) points. The SIFT method has been utilized effectively on 2D grayscale images to identify and match invariant features. Furthermore, the SIFT method works efficiently for object recognition problems where a training image of the object of interest is given as well as the number of features extracted, or number of matches between consecutive views is a good indication for the success of the algorithm and If the number of extracted features or matches between consecutive frames drops suddenly to a very low value, then it is very likely the image is blurry. Image blur is most of the time caused by fast robot rotation. Once the rotation is over, image come back to a good quality that allows extracting of a large number of features and matches. Blurry images causes by rotations are typically introduced for two or three frames. Thus, once a low number of features is detected we can stop visual odometry and wait for a good image. Once a good image is obtained, there is a very high probability.

The goal of the experiment is to be able to achieve a comparison of the experimental path and theoretical path of a certain object that holding a camera. As a result the experimental results show, the path estimation by using successive images is applicable in static environments, but in the dynamic environments, the generated paths are very noisily. The merit of the proposed approach is that it retains the benefit of fast feature extraction methods and performs the more expensive robust image matching only when needed.

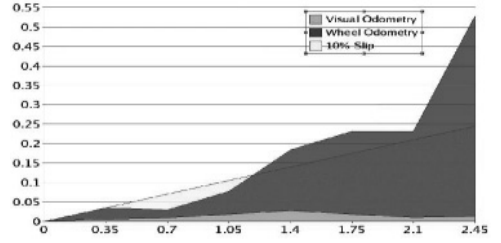
Keywords: SIFT, Visual Odometry, Keypoint, Odometry

Giriş

Otonom robotların gerek gündelik hayatta (rehber robotlar, hastabakıcı robotlar vb.) gerekse insanların gitmelerinin zor, tehlikeli (uzay, afet sonrası yıkıntılar, bomba, mayın imha vb.) olduğu alanlarda kullanımı giderek artmaktadır. Otonom bir robotun bir ortamda iş görebilmesi için başlıca koşullardan biri nerede olduğunu bilmesidir. Başlangıçta bulunduğu yeri bildiği bir durumda bir hareket işaretine göre ilerleyen robotun yeni konumunu hesaplaması ideal ortamlarda (hatasız ölçümleme ve hatasız kontrol) sadece hareket işaretine göre mümkünken gerçek uygulamalarda giderek artan bir hataya sebep olmaktadır. Bu nedenle literatürde hareket işaretinin yanında ortamdan alınan ölçümlerin de (çevredeki engel, nesnelere uzaklık vb.) kullanıldığı istatistikî algoritmalar geliştirilmiştir (Haehnel vd., 2002). Bu algoritmalar ailesine Eşzamanlı Konumlandırma ve Harita Çıkarma (SLAM) adı verilmektedir. Bu algoritmalarda robotun bir sonraki konumu ve ortam algısı, önceki hareket işaretlerine, önceki konumuna ve önceki ortam algısına bağlı olasılıklar olarak hesaplanmaktadır.

Robotun hareketinin dolayısıyla yeni konumunun hesaplanması için ortamdan alınabilen ölçümlere örnek olarak teker / paletin dönme sayısı, açısı, robotun etrafındaki cisimlere uzaklıkları, robotun üzerindeki bir kamera ile alınan görüntü verilebilir. Şekil 1’de Mars’ta kullanılan otonom robotla yapılan bir deneyin sonuçları verilmiştir. Robot hareketinin tekerlerin dönme sayısıyla hesaplandığı durumla (Wheel odometry), kamera görüntüleriyle hesaplandığı durum (Visual odometry) yapılan hatalar cinsinden karşılaştırılmıştır. Ölçekler metre cinsindedir.

Şekil 1’de görüldüğü gibi düzgün olmayan (kayalık) ortamlarda tekerlerle yapılan tahminlerin hatası, görsel odometriye göre çok daha yüksektir. Bunun sebebi tekerin dönmesine rağmen, robotun bu dönüş kadar yol alamamasıdır.



Şekil 1. Kaya tırmanışı yaparken tekerlek ve görsel odometrinin hataları (yatay düzlem gidilen yol, düşey düzlem yapılan hata) (cumulative failure of wheel and visual odometry for rock clinging)

Literatürdeki Görsel odometri (robotla bağlı bir kameradan alınan görüntüleri kullanarak robotun gittiği yolun kestirimi) sistemlerinin başarılarını, PARS grubu olarak geliştirdiğimiz robotların konumlandırma sistemlerine eklenmesi için kullandığımız metotlar ve yaptığımız deneyler bu çalışmanın içeriğini oluşturmaktadır.

Çalışmanın 2. bölümünde Görsel odometri sistemini nasıl uyguladığımız, 3. bölümde görsel odometride kullanılan SIFT algoritması, 4. bölümde ise deneysel sonuçlarımız sunulmuştur. 5. bölümde bu süreçten öğrendiklerimiz ve gelecekte yapmayı planladıklarımız yer almaktadır.

Çalışmanın önemi

Bu çalışmada görsel odometrinin tekerlek odometriye göre kümülatif hatanın daha düşük çıktığı görülmektedir. Çalışma statik ve dinamik ortamda denenmiş ve sonuçları bize büyük ölçekte hareket yönümüzü kaybetmediğimizi ve görsel odometri ile yeni konumumuzun daha az hata ile belirlendiği görülmüştür. Haritalamada gidilen yolun tekerlek odometriye göre daha az hata ile kestirimi sağlanmıştır.

Görsel odometri

Bir kamera ile çekilen bir videodan kameranın hareketinin kestirimi işlemi görsel odometri olarak tanımlanmaktadır. Bu çalışmamızda bir cep telefonu kamerasını sabit bir açıyla tutarak yürüyen bir insanın gittiği yol kestirilmiştir. Yapılan işlemin sözde kodu aşağıda verilmiştir. Kodda t zamanı, N videodaki imge sayısını göstermektedir.

for $t = 1 : N-1$

Eşleşen aday noktalar = SIFT(imge(t),
imge(t+1))

Eşleşen noktalar = Filtrele(Eşleşen aday
noktalar)

Hareket vektörü = Hareket Bul(Eşleşen
noktalar)

Yeni konum vektörü = Eski konum vektörü +
Hareket vektörü

end

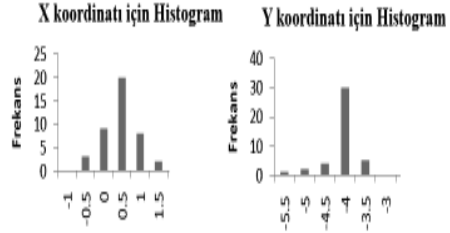
Algoritmanın her imge ikilisi için çalıştırılan ilk adımında 3. bölümde anlatılan SIFT algoritması ile iki imgedeki eşleşen noktalar bulunmuştur. Şekil 2’de örnek ardışık 2 imge üzerinde SIFT ile bulunan 34 nokta eşleşmesi verilmiştir.



Şekil 2. Ardışık 2 imgedeki eşleşen noktalar

Algoritmanın 2 adımında yer alan filtreleme işleminde eşleşen tüm noktalardan yanlış eşleşmeler çıkarılmıştır. Bunun için eşleşen noktaların iki imgedeki X ve Y boyutlarındaki koordinatlarının farkları önceden belirlenen bir eşik değerinden büyükse o nokta ikilisi eşleşen noktalar kümesinden çıkarılmıştır. Şekil 2’de

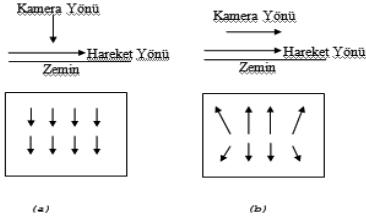
böyle filtrelenen 1 adet nokta ikilisi vardır ve daire ile işaretlenmişlerdir. Algoritmanın 3. adımında eşleşen noktaların iki imgedeki X ve Y boyutlarındaki koordinat farklarının ortalaması alınarak X ve Y boyutlarındaki hareketin büyüklüğü ve yönü belirlenmektedir. Şekil 3’te iki boyutta bulunan histogramlar kullanılarak kameranın X boyutunda 0.5 birim, Y boyutunda ise -4 birim hareket ettiği bulunmaktadır.



Şekil 3. Eşleşen noktaların X ve Y boyutlarındaki koordinat farklarının histogramları (histogram of matched keypoints for x and y coordinates)

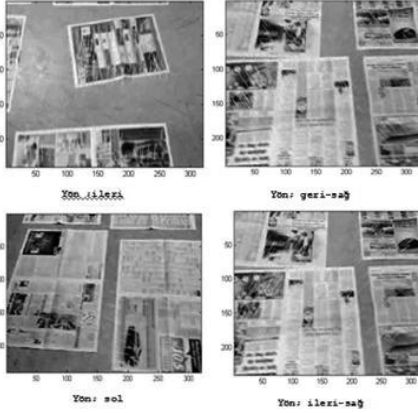
Algoritmanın son adımında ise 3. adımda bulunan hareket vektörü eski konum vektörüyle toplanarak yeni konum vektörü bulunmaktadır. Algoritmada gri seviyedeki 250*200 piksellik imgelerle çalışılmıştır. Algoritmanın çalışma hızı Pentium 4, 512 MB RAM’lik bir makinede ortalama 1 imge/sn’dir. Bu hız gerçek zamanlı bir çalışma için yeterlidir.

Çalışmanın geliştirilme aşamasında karar verilmesi gereken konulardan birisi de kameranın çekim açısıydı. Kameranın çekim açısı belirlenirken, kameranın hareketiyle, ardışık imgeler arasında eşleşen noktaların koordinat farkı hareketlerinin birbirlerine en benzer oldukları açı aranmış ve en iyi açının yerin çekildiği (hareket düzlemine dik) açı olduğu bulunmuştur. Şekil 4’te kameranın yere paralel ve dik konumlarda olduğunda ardışık imgelerde eşleşen noktaların örnek hareketleri verilmiştir.



Şekil 4. Kamera açısının eşleşen noktaların hareketleri üzerindeki etkisi, kamera açısı (a) hareket yönüne dik, (b) hareket yönüne paralel (according to camera's aspect, behaviors and sliding of keypoints)

Hareket yönüne dik açıyla çekilen videolarda 4 farklı hareket için elde edilen hareket vektörleri Şekil 5'te gösterilmiştir. SIFT metodu tamamen homojen yüzeylerde belirgin noktalar bulamadığından dolayı Şekil 5'teki denemelerde zemine gazeteler serilmiştir. Robotların gerçek ortamlarında da tamamen homojen ortamlar olmadığından yapılan bu düzenleme yöntemin kullanılabilirliğini azaltmamaktadır.



Şekil 5. 4 farklı hareket için ardışık imge ikililerinden elde edilen hareket vektörleri

Şekil 5 incelendiğinde kameranın gerçek hareketiyle, eşlenen noktalarla elde edilen hareket vektörünün uyumlu olduğu görülmektedir.

Ölçekten bağımsız öznelik dönüşümü

İki imgedeki aynı bölgelerin bulunması için literatürde birçok yöntem mevcuttur. Bu çalışmada bu işlem için Lowe (Lowe, 2004) tarafından önerilen Ölçekten Bağımsız Öznelik Dönüşümü – SIFT seçilmiştir. Bunun sebebi yöntemin imgenin boyutundan, imgenin alındığı kameranın bakış açısından, imgenin alındığı ortamın ışık koşullarından, imgedeki nesnelerin açısından bağımsız olarak eşleme işlemini başarabilmesidir. SIFT algoritmasında temel olarak izlenen 4 adım vardır (Serce vd., 2008):

- 1-Ölçeksel uzaydaki ekstrem(uç değer) noktaların tespiti
 - 2-Kilit noktalarının belirlenmesi
 - 3-Yönelim tespiti
 - 4-Düğüm noktalarının niteliklendirilmesi
- Uç değer, ölçeksel uzayda sabit olan noktalar. Ölçeksel uzay oluşturularak kilit nokta olabilecek noktalar tespit edilir ve bu noktaların Gauss filtresinden geçirilir, kenar tespit yöntemleri(LOG,DOG,.. vb.) kullanılarak iki resim arasındaki fark elde edilir. Daha sonra düşük kontrasta sahip olan noktalar elenerek daha kararlı olan noktalar elde edilir. Yönelim tespiti ise düğüm noktalarının piksel-altı hassasiyetiyle konumlandırıldıktan sonra bu noktaların yerel türevlerin yönleri kullanılarak yapılır.

Ekstremlerin elde edilmesi

Aynı nesnenin farklı pozisyonlarda tanınmasını sağlayacak olan kısımdır. Görüntü üzerinde birçok nokta tespiti yapılır fakat bu noktaların bazıları nesne ile ilgili olmadığından dolayı Gauss filtreleme fonksiyonu (Eşitlik 1) kullanılarak ayklanır.

$$G(x, y, \sigma) = [1/(2\pi\sigma^2)] * e^{-(x^2+y^2)/2\sigma^2} \quad (1)$$

Eşitlik 1'de I: orijinal resim, G:değişken orantıya sahip Gauss fonksiyonu göstermek üzere Ölçeksel uzay fonksiyonu Eşitlik 2'de verilmiştir.

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (2)$$

Kilit noktaların belirlenmesi

Görüntüde tespit edilen kilit noktalar kontrast değerleri baz alınarak elenir(düşük kontrasta sahip noktalar elenir). Bu işlemi yapmak için DOG metodu kullanılır.

Yönelim tespiti

Bu adım, iki boyutlu düzlem üzerinde nesnedeki açısal değişiklikler için nesneyi kaybetmemizi sağlayacaktır. Bu evrede büyüklük ve yön hesaplanır.

Düğüm(kilit) noktalarının niteliklendirilmesi

Bu evrede ise üç boyutlu eksende görüntü içerisindeki nesnenin kaybı engellenir.

DeneySEL Sonuçlar

Bu bölümde 2. Bölümde anlatılan adımlar uygulanarak sabit ve dinamik ortamlarda uygulanan metodun sonuçları verilmiştir. Tablo 1’de sabit, Tablo 2’de dinamik (görüntüye hareketli başka nesnelere girildiği) ortamlarda yapılan deneyler gösterilmiştir.

Tablo 2. Dinamik ortamlarda yapılan çeşitli deney sonuçları

Görüntü çöklüklerken izlenen rota	SIFT ile bulunan rota
	<p>Sadece bağlantış yönü hareketli ortam içinde gerçekleşmiştir.</p>
	<p>Çekimin tamamı hareketli ortamda yapılmıştır.</p>

Çalışmanın değerlendirilmesinde kullanılacak başarı ölçütü bulunan yolun, gerçek yola benzerliğidir. Tablo 1 ve 2 incelendiğinde sabit ortamlarda gerçek rota ile bulunan rotanın birbirine oldukça benzer olduğu, dinamik ortamlarda ise ana rotanın bir ölçüde korunduğu ancak oldukça gürültülü rotalar bulunduğu görülmektedir.

Tablo 1. Sabit ortamlarda yapılan çeşitli deney sonuçları

Görüntü çöklüklerken izlenen rota	SIFT ile bulunan rota

Bununla birlikte kamera görüntülerinin bir insanın yürürken elinde tuttuğu bir kameradan alındığı düşünülürse, kameranın ilk sütunlarda verilen gerçek rotadaki kadar düzgün bir şekilde hareket ettirilmediği anlaşılacaktır. Tablo 1’deki rotalardaki küçük oynamaların, ellerin küçük hareketlerinden ve insan yürüyüşünün düzensiz salınımından ileri geldiği düşünülmektedir.

Sonuç ve Gelecek Çalışmalar

Bu çalışmamızda bir cep telefonu kamerasını hareket düzlemine dik tutarak yürüyen bir insanın gittiği yol kestirilmiştir. Bu işlem robotların üzerlerine bağlanan kameralardan alınan görüntülerle robotların gittikleri yolun hesaplanmasının bir benzeridir. Bunun için kameradan alınan ardışık imgelerdeki eşleşen noktalar SIFT algoritmasıyla bulunmuş ve eşleşen noktaların koordinatları arasındaki farklar kullanılarak gidilen yol hesaplanmıştır.

Deneyler birbirinden 2 farklı (sabit ve dinamik) ortamda gerçekleştirilmiştir. Sabit ortamda gidilen yol başarılı bir şekilde tahmin edilirken, dinamik ortamda oldukça gürültülü olarak tahmin edilebilmiştir.

Gelecek bir çalışma olarak, dinamik ortamlardaki problemin çözümü için eşleşen noktaların hareket vektörlerinin kümelenerek kameranın ve ortamdaki diğer nesnelerin hareket vektörlerinin birbirlerinden ayırt edilmesi düşünülmektedir.

Son yıllarda literatürde robotun konumunu ve çevresini daha iyi doğru hesaplayabilmesi için algılayıcı verilerini birleştirme yoluna gidilmektedir (sensor fusion). Bu yaklaşıma paralel olarak gelecek çalışmamızda gerçek robotumuz üzerinde lokalizasyon yaparken teker odometrisine bu çalışmada yapılmış olan görsel odometrinin eklenmesi planlanmaktadır.

Bununla birlikte SIFT algoritması noktasal eşleştirmeler yaparak benzer noktaları bulmaktadır, bu işlem dokusal yüzeylerde işe yaramamaktadır. Bu olumsuz etkiden kurtulmak için bir imge segmentasyonu algoritması olan JSEG ile SIFT algoritmasının birbirleri ile senkronize çalışması veya yeni bir sentez sağlanarak çıkarılan özelliklerin benzerliklerine bağlı kilit noktaların eşleşmesi yanında dokusal yüzeylerinde eşleşmesi sağlanabilir.

Deney ortamının büyütülüp kameranın bir otoyolda hareket eden bir arabanın üzerine

(dinamik ortam- yüksek hız) takılıp arabanın gittiği yolun bulunması bir başka gelecek araştırma alanı olarak düşünülmektedir.

Kaynaklar

- Haehnel, D., Schulz, D., and Burgard, W., 2002, Map building with mobile robots in populated environments, *International Conference on Intelligent Robots and Systems (IROS)*, 496-501.
- Mark Maimone, Yang Cheng, Larry Matthies, 2007, Two Years of Visual Odometry on the Mars Exploration Rovers, *Journal of Field Robotics*, **24(3)**, 169-186.
- David G. Lowe, 2004, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision*, **60(2)**, 91-110.
- Serçe, H., Bastanlar, Y., Temizel, A., Yardımcı, Y., 2008, On Detection of Edges and Interest Points for Omnidirectional Images in Spheria I Domain, *SIU 2008, 20-22 April, Didim, Turkey*.