

DİNAMİK SAHNELER İÇİN ÖNCELİKLENDİRİLMİŞ SIRALI 3B GERİ ÇATIM[°]

PRIORITIZED SEQUENTIAL 3D RECONSTRUCTION IN VIDEO SEQUENCES OF DYNAMIC SCENES[°]

Evren İmre, Aydın Alatan, Sebastian Knorr[†], Thomas Sikora[†]

Elektrik-Elektronik Mühendisliği Bölümü
O.D.T.Ü.
Balgat, 06531 Ankara, Türkiye
E-mail: {eimre@, alatan@eee.}metu.edu.tr

[†] Communication Systems Group,
Technische Universität Berlin
Einsteinufer 17, Berlin, Germany
E-mail: {knorr, sikora}@nue.tu-berlin.de

Özetçe

Bu makalede, tek kamera tarafından çekilmiş dinamik sahnelerin çok-görüntülü hareketten yapı (ÇGHY) problemi ele alınmaktadır. Önerilen algoritma video karelerinden elde edilen öznitelikleri, eşkutupsal kısıtı kullanarak, her biri farklı bir harekete karşılık gelen kümelere ayırmakta, daha sonra bu kümeleri ayrı ayrı, sıralı geri çatım tekniğini kullanarak 3 boyutlu olarak geri çatmaktadır. Algoritma geri çatımda kullanılan görüntü çiftlerini güvenilirlikleri ve bilgi içeriklerini göz önünde bulundurarak önceliklendirebilmektedir. Mevcut bilginin verimli kullanımı, doğru bir geri çatıma hızlı bir biçimde ulaşılmasını mümkün kılmaktadır. Eşkutupsal geometrinin hesaplanması ve geri çatımın güvenilir olması için son derece önemli olan uzun taban çizgisi koşulu bir öznitelik izleyici yardımıyla sağlanmaktadır. Deneysel sonuçlar, önerilen algoritmanın tipik bir video görüntüsünde çok-nesneli ÇGHY probleminin çözümü konusunda, önemli bir potansiyel barındırdığını göstermektedir.

Abstract

In this study, an algorithm is proposed to solve the multi-frame structure from motion (MFSfM) problem for monocular video sequences in dynamic scenes. The algorithm uses the epipolar criterion to segment the features belonging to the independently moving objects. Once the features are segmented, the corresponding objects are reconstructed individually by using a sequential algorithm, which is also capable of prioritizing the frame pairs with respect to their reliability and information content, thus achieving a fast and accurate reconstruction through efficient processing of the available data. A tracker is utilized to increase the baseline distance between views and to improve the F-matrix estimation, which is beneficial to both the segmentation and the 3D structure estimation processes. The experimental results demonstrate that our approach has the potential to effectively deal with the multi-body MFSfM problem in a generic video sequence.

1. Giriş

Tipik bir videodaki bilgi artıklığının gürbüz bir 3 boyutlu (3B) geri çatım elde etmek amacıyla kullanılabilmesi için, çok-görüntülü hareketten yapı (ÇGHY) probleminin çözümü esastır. Problem sadece durağan bir sahne için tanımlanmış olduğundan, dinamik öğelerin varlığı, tanımın söz konusu öğelerin yapılarının geri çatımını ve hareketlerinin kestirimini içerecek biçimde genişletilmesini gerektirir.

Dinamik sahnelerdeki ÇGHY problemine dair literatürü en çok etkileyen nokta, dinamik sahnenin, arkaplan ve bağımsız hareket eden nesnelere (BHN) bölünebilmesidir. Bu durumda problem, her öğe için bağımsız bir statik ÇGHY problemine indirgenebilir. Bu perspektiften bakıldığında, gruplama ve geri çatım art arda çözülecek iki alt-problemdir.

Gruplama problemini çözmek için kullanılan teknikler üç ana başlıkta toplanabilir: Optik akış temelli metotlar, sahnenin farklı derinliklerdeki düzlemlerden oluştuğu varsayımıyla, optik akış kestirimlerinin kümelenmesine dayanmaktadır[5]. *İlgilik matrisi* metotları, gruplama için öznitelikler arasındaki benzerlik bilgisini içeren bir matrisin özvektörlerinden yararlanmaktadır[11]. Son olarak, geometrik metotlar eşkutupsal geometri ve katı devinim kısıtlarının kullanımı esasına dayanmaktadır[2]. En yaygın olarak kullanılan kısıt, “*temel matris*”tir.

Geri çatım problemi için başlıca yaklaşımlar, toplu ve sıralı metotlardır. Toplu metotların en iyi bilinen örneği *çarpanlara ayırma* tekniğidir[10][3]. Sıralı metotlarda ise problem ya dinamik sistemlerde durum kestirimi, ya da bilinmeyen sabit bir vektörün ters-ortalama-karesel-hata süzgeciyle kestirimi çerçevesinde ele alınmaktadır[8].

Çok sık rastlanmasa da, problemin tüm öğeler için eşzamanlı olarak çözülmesini öneren yöntemler de mevcuttur. Bunların iyi bilinen örnekleri çarpanlara ayırma metodunun çok-nesne için geliştirilmiş türü[1] ve parçacık süzgeçleridir[7].

Bu makalede, çok-nesneli ÇGHY probleminin her iki alt-problemi de ele alınacaktır. Makale, aşağıdaki gibi düzenlenmiştir: İzleyen bölüm, önerilen çözümün bir özetini içermektedir. 3. bölümde gruplama, 4. bölümde ise geri çatım algoritmaları anlatılacaktır. 5. bölümde sunulan deney

[°] Bu çalışma AT BTT 6. Çerçeve Program 3B Televizyon Mükemmeliyet Ağı tarafından desteklenmektedir

[°] This work is funded by EC IST 6th Framework 3D NoE

sonuçların ardından, makale, 6. bölümdeki tartışma ve bu çalışmanın devamı üzerine düşüncelerle noktalanacaktır.

2. Çok-Nesneli ÇGHY Probleminde Önceliklendirilmiş Geri Çatım

Önerilen algoritma, yukarıda sözü edilen iki alt-problemi çözebilecek şekilde tasarlanmıştır: Bir dinamik sahnenin onu oluşturan öğelerine indirgenmesi ve videonun içerdiği büyük miktardaki verinin verimli bir biçimde işlenerek iyi bir geri çatım elde edilmesi.

Gruplama için geometrik metot tercih edilmiştir. Bunun nedeni, geometrik metodun literatürdeki rakiplerine göre daha çok bilgiyi, daha az varsayım kullanabilmesidir. Gruplama, alışılmışın aksine öznitelikler değil, gezinimler üzerinde yapılmaktadır. Video görüntülerinin kullanılması, Kanade-Lukas izleyicisinin yardımıyla, doğru gezinimler oluşturulmasına izin vermektedir. Her grubun 3B yapısı önerilen geri çatım algoritması yardımıyla oluşturulabilir.

Gerçek çatım konusunda, toplu yaklaşımın videoda bulunan büyük miktarda veriyi işlemeye elverişli olduğu bilinmektedir. Ancak sıralı teknikler, görüntü çiftlerinin bir kısmının işlenmesiyle elde edilen ara sonuçların kalan çiftlerin işlenmesinde kullanılmasını mümkün kıldığı için, toplu tekniklere göre önemli bir avantaja sahiptir. Açık ki, bu avantajın kullanılması geri çatımı görüntü çiftlerinin işleme sırasına bağımlı hale getirmektedir. Bu çalışmada iyi bir sıralamanın nasıl yapılabileceğine dair yeni bir çözüm sunulmaktadır.

Sıralama problemini önemli kılan bir başka neden de, art arda gelen video kareleri arasındaki taban uzaklığının çok az olmasının karelerin zamansal sıralamayla işlenmesine izin vermemesidir. Bu durumda sıkça başvurulan bir yöntem, kare atlamaştır. Ancak, hareket önceden bilinmediği sürece, bu yöntemin yeterli bir taban çizgisi uzaklığını garantileyemeyeceği açıktır.

İyi bir önceliklendirme ölçeğinin tasarımında aşağıdaki ölçütler göz önünde bulundurulmalıdır:

- **Güvenilir bir kestirime hızlı yakınsama:** Sıralı yaklaşımda, bir kare çiftiyle yapılan kestirimdeki başarı o andaki ara kestirime bağlı olduğu için, güvenilir bir ara kestirime yakınsamaktaki gecikme ciddi başarıyı kayıplarına, hatta algoritmanın ıraksamasına yol açabilir.
- **Sahne yapısının hızlı bir biçimde ortaya çıkartılması:** Olabildiğince az sayıda kare çifti işleyerek sahneye ait olabildiğince çok sayıda 3B noktanın yeri hesaplanmalıdır.

Temel çizgi uzunluğunu ve eşleşmiş özniteliklerin sayısını dikkate alan bir ölçek, her iki ölçütün gereklerini de karşılayacaktır. Bu yüzden, kare çiftlerini önceliklendirmek için bu iki terimin ağırlıklı ortalaması ölçek olarak seçilmiştir.

3. Öznitelik Takibi ve Gruplama

Daha önce de belirtilmiş olduğu gibi, problemin çözümündeki ilk aşama sahnenin arkaplan ve BHNlere ayrılmasıdır. Harris metoduyla bulunan köşelerin piramitsel ayrılmasıdır. Harris metoduyla bulunan köşelerin piramitsel Kanade-Lukas izleyicisiyle tüm video boyunca takibi, gruplama ve geri çatım süreçlerinde gereksinim duyulan uzun taban çizgisini sağlamaktadır.

İzleyicide yapılan ilk değişiklik, kaybedilen izlerin yerine Harris köşe bulucusu kullanılarak yenilerinin konmasıdır. İkinci ise, yeterince uzun bir taban çizgisi elde edebilmek için gruplama probleminin çözümünde anahtar-karelerin kullanılmasıdır. Anahtar-kare seçilmesinde uygulanan yöntem, işlenen kareyle en son anahtar-kare arasındaki eşleşmiş noktaların çiftin bir üyesinden diğerine bir düzlemsel-homografi aracılığıyla taşınması ve taşınan noktalarla eşleri arasındaki ortaça uzaklığın değerlendirilmesidir[6]. Bu uzaklığın düşük olması, iki kare arasında güvenilir bir eşkutupsal geometri hesaplanamayacağını gösterir. Bu yüzden anahtar-kare çiftleri, söz konusu uzaklığın yüksek olduğu çiftler arasından seçilmektedir.

Gezinimler oluşturulduktan ve anahtar-kareler seçildikten sonra, gezinimler eşkutupsal kısıtın yardımıyla gruplanır. Bunu sağlayan, videodaki her bağımsız harekete (gruba) karşılık gelen bir temel matris (F-matrisi) olmasıdır. Bu durumda, \mathbf{x}_1 ve \mathbf{x}_2 "i"nci harekete ait öznitelikler iseler,

$$\mathbf{x}_1^T \mathbf{F}_i \mathbf{x}_2 = 0$$

denklemini sağlamak zorundadırlar. RANSAC algoritması temelli F-matrisi bulma metodu, önce hakim harekete ait öznitelikleri belirler, kalanları ise aykırı gözlem olarak etiketler. Bu aykırı gözlemler üzerinde aynı metot tekrar uygulanırsa, bir başka F-matrisini sağlayan öznitelikler belirlenir ve kalanlar reddedilir. Bu sürecin bulunan F-matrisleri güvenilir hale gelene kadar tekrarı, arkaplana ve BHNlere ait tüm gezinimlerin belirlenmesini sağlayacaktır. Algoritmanın tamamı aşağıdaki gibidir:

Gezine Gruplama Algoritması

1. ve 2. anahtar-görüntüler için RANSAC temelli F-matris kestirimi algoritmasını çalıştır. Denklemi sağlayan gezinimleri arkaplan olarak işaretleyin.
2. Kalan gezinimler üzerinde algoritmayı tekrar çalıştır ve denklemi sağlayan gezinimleri ilk BHNye ait olarak işaretleyin.
3. 2. adımı bulunan F-matrisleri güvenilir ve kalan gezinimler birbirine yakın olduğu sürece tekrarla.
4. Bir sonraki anahtar-görüntüye geç. Her bir grup için, sınıflanmış gezinimleri kullanarak bu görüntü ve bir önceki anahtar-görüntü arasındaki F-matrisini hesapla. Yeni gezinimleri 1.-3. basamakları kullanarak grupla.
5. 4. adımı tüm anahtar-görüntüler için tekrarla.

4. Önceliklendirilmiş Sıralı Geri Çatım

Algoritmanın anlatımından önce, iki kavramı tanımlamak yararlı olacaktır.

Tanım 1: Alt-kestirim, tek bir kare çiftinden üçgenleme yoluyla elde edilen sahne geri çatımıdır.

Tanım 2: Alt-geri çatım, alt-kestirimlerin birleştirilmesiyle elde edilen sahne geri çatımıdır. Bir alt-geri çatımı oluşturan kareler, bir başka alt-geri çatımın oluşturulmasında kullanılamaz. Yani, her alt-geri çatım, video karelerinin diğerleriyle kesilmeyen bir alt kümesinden oluşmuştur. Nihai geri çatım, alt-geri çatımların birleştirilmesiyle elde edilir.

Gerçek çatım algoritmasının temeli [6]da anlatılmış, uygulama detayları ise [9]da açıklanmıştır. Özetle, ilk iki kare kullanılarak bir geri çatım elde edilip, üçüncü kareyle bu geri

çatım arasındaki 3B-2B öznitelik eşleşmelerinin yardımıyla, üçüncü karede kameranın yeri bulunmaktadır. Daha sonra, ikinci ve üçüncü kareler kullanılarak gerçekleştirilen bir uçgenleme sonucu elde edilen alt-kestirim, geri çatıma yeni noktalar eklemeye, mevcutlar için yapılan kestirimi de iyileştirmeye yaramaktadır.

Anlatılan algoritma video karelerini belirli bir sırayla (örn. , $F_1-F_2, F_2-F_3, F_3-F_4$) işlemek üzere tasarlanmıştır. Farklı kare çiftlerini kullanmak, algoritmada da birtakım değişiklikleri gerektirir.

F_m-F_n ve F_p-F_q kare çiftlerini ele alalım. Eğer bu çiftlerin arasında tek bir ortak kare varsa (örn. $n=q$ ise F_m-F_n, F_n-F_p), söz konusu çiftler algoritmanın orijinal hali kullanılarak işlenebilir. Eğer aralarında hiçbir ortak kare yoksa, o halde her biri için (tek bir alt-kestirim kullanılarak oluşturulmuş) yeni bir alt-geri çatım başlatılır. İkinci durumun gerçekleştiğini varsayalım ve bu geri çatımlara T_1 ve T_2 diyelim. Şimdi de, F_r-F_s çiftini ele alalım. Eğer bu çiftin daha öncekilerle ortak bir karesi varsa, ya da hiçbir ortaklığı yoksa, yukarıda anlatılan işlem tekrarlanabilir. Ancak, eğer karelerin biri T_1 , diğeri ise T_2 'de ise (örn. $r=m$ ve $s=q$), alt-geriçatımların birleştirilmesi gerekir.

Her alt-geriçatımın koordinat sistemi, onu oluşturan ilk alt-kestirim tarafından belirlenmiştir. Bu yüzden, birleştirmenin yapılabilmesi için, alt-geriçatımların birindeki noktaları diğerine taşıyan, 4×4 lük izdüşümsel bir dönüşümün hesaplanması gerekir. Koordinat sistemleri arasındaki ilişki dönme, öteleme ve ölçeklendirme işlemleriyle tanımlanmaktadır. Bu dönüşümün kestirimi için [4]te anlatılan gürbüz 2B izdüşümsel dönüşüm kestirimi algoritması üzerinden türetilen bir algoritma kullanılmıştır. Özetle, 3B-3B eşleşmelerin belirlenmesinden sonra, RANSAC yardımıyla en çok sayıda noktayı bir alt-geriçatımdan diğerine taşıyan dönüşüm bulunmakta, daha sonra bu kestirim tüm aykırı olmayan noktaların kullanılmasıyla ve lineer olmayan optimizasyonla geliştirilmektedir.

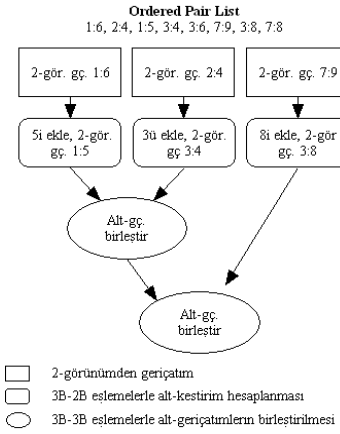
Olması muhtemel bir başka durum da, çiftteki iki karenin aynı alt-geriçatıma ait olmasıdır. Bu durumda çift yeni noktalar elde etmek amacıyla işlenebilir.

Tipik geri çatım süreci Şekil 4.1'de gösterilmektedir. Geri çatım algoritması aşağıda özetlenmiştir:

Önceliklendirilmiş Geri Çatım Algoritması

İç kalibrasyon parametrelerinin ve tüm karelerdeki eşleşmelerin bilindiği varsayımıyla,

1. Başlangıç geri çatımını hesapla.
2. 3B-2B eşleşmeleri kullanarak, her karedeki kameranın ilk kareye göre konumunu hesapla.
3. Her çift için önceliklendirme ölçevini hesapla ve kare çiftlerini sırala.
4. Tüm çiftler işlenmemişse ya da ölçev eşik değerinin üzerindeyse.
 - a. Eğer çiftteki karelerin hiçbiri mevcut alt geriçatımlar tarafından içerilmiyorsa, yeni bir alt-geri çatım başlat.
 - b. Eğer çiftteki karelerden biri mevcut alt-geriçatımların biri tarafından içeriliyorsa, diğer kareyi bu alt-geriçatıma ekle ([6]da anlatılan algoritma).
 - c. Eğer iki kare de aynı alt-geri çatım tarafından içeriliyorsa, [6]daki algoritmayla işle.
 - d. Eğer çiftin kareleri farklı alt-geriçatımlar tarafından içeriliyorsa, alt-geriçatımları birleştir.



Şekil 4.1: Önceliklendirilmiş sıralı geriçatım

5. Eğer birden fazla alt-geri çatım varsa, alt-geriçatımları birleştir.

Bu bölümü kapatmadan önce söz edilmesi gereken son konu, 1. adımdaki ilk kare çiftinin nasıl seçileceğidir. Bu amaçla, güvenilir bir çift gelene kadar, gelişigüzel kare çiftleri seçilebilir. İşe yarayacağı düşünülen güvenilirlik ölçütleri öz matrisin özdeğerleri, ya da geri-izdüşüm hatasıdır. Göz önüne alınması gereken başka bir ölçüt ise, bu çiftteki karelerin videonun kalan kareleriyle yeterli sayıda ortak özniteliği olmasıdır. Aksi halde kamera pozisyonu kestirimi güvenilir olmayabilir.

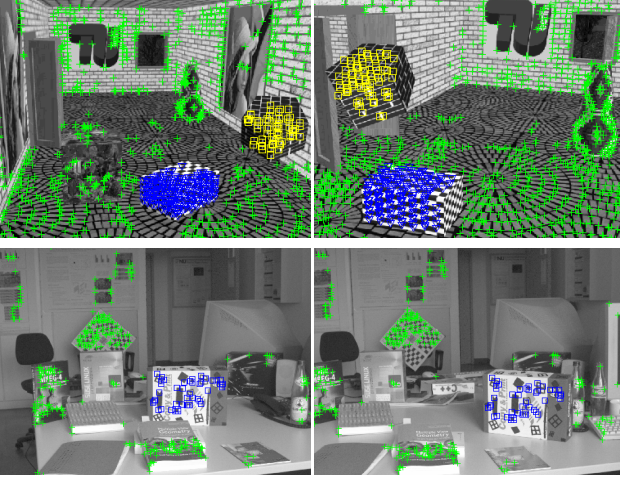
5. Deneysel

Gruplama algoritması hem gerçek, hem de yapay veriler üzerinde test edilmiştir. Gerçek veriler için kameralar manuel olarak kalibre edilmiştir. Şekil 5.1'de gruplama sonuçları görülmektedir. Üst sıradaki resimler "TUB-Room" adlı yapay videonun 1. ve 170. kareleridir. Arkaplan öznitelikleri yeşil çarpımlarla, BHNlere ait olanlar ise mavi üçgenler ve sarı karelerle işaretlenmiştir. Alt sırada ise, "Desk" adlı videonun 1. ve 13. kareleri görülmektedir. Arkaplan yeşil çarpımlarla, BHN ise mavi karelerle işaretlenmiştir. Sonuçlar, algoritmanın yüksek bir başarıyı olduğunu göstermektedir.

Geri çatım algoritması 170 kare uzunluğundaki "TUB-Room" adlı yapay videoda denenmiştir. Sonuçlar Şekil 5.2'de görülmektedir. Üst sıra BHNleri, orta sıra yalnızca temel çizgi uzunluğu göz önüne alınarak yapılan sıralama sonuçlarını, alt sıra ise hem temel çizgi uzunluğunu, hem de eşleşme sayısını göz önüne alarak yapılan sıralama sonuçlarını göstermektedir. Şekil 5.3 her iki ölçev için yakınsama hızı grafiğidir. BHNler, az sayıda noktaya sahip oldukları için, ölçevden bağımsız olarak hızla yakınsamışlardır. Bu yüzden, grafik sadece arkaplan için çizilmiştir. Beklendiği gibi, eşleşmelerin sayısının da göz önüne alınması yakınsama hızını arttırmıştır. Ayrıca, videodaki toplam 14365 kare çiftinden yalnızca 50sinin kullanıldığına da dikkat edilmelidir.

6. Sonuç

Bu makalede, dinamik sahnelerin video görüntülerinden geri çatılması için bir algoritma önerilmiştir. Algoritma, öznitelik kümesini eşkutupsal kısıtı kullanarak gruplara ayırmaktadır. Daha sonra her grup, videodaki bulunan büyük miktardaki veriyi verimli işleyecek biçimde tasarlanmış bir algoritmayla

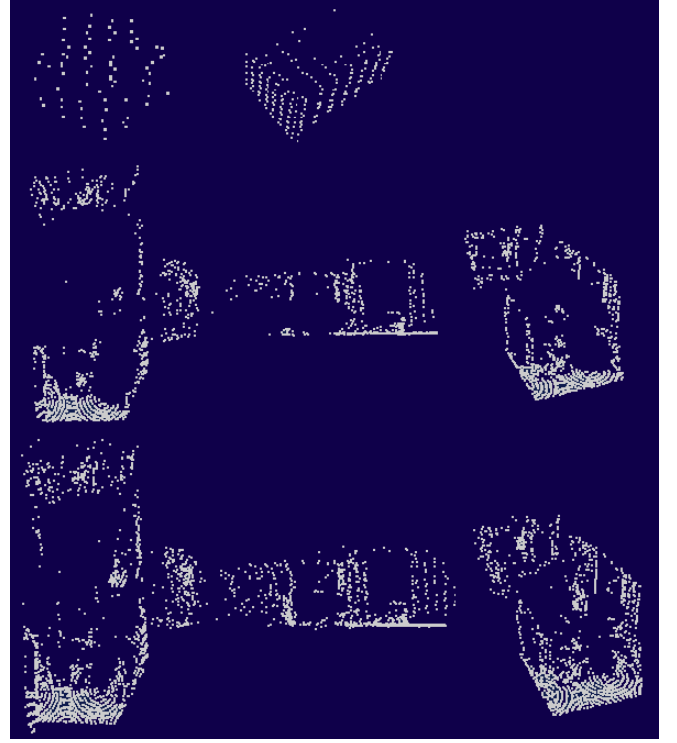


Şekil 5.1: Gruplama sonuçları. TUB-Room (üst) ve Desk-sequence (alt)

geri çatılmaktadır. Kare çiftlerinin önceliklendirilmesi, videonun içerdiği bilginin büyük bir kısmının az sayıda kare çifti kullanılarak elde edilebilmesini sağlamaktadır. Deneyler, algoritmanın, yeterli sayıda öznelik olduğu sürece, başarılı olduğunu ortaya koymuştur. Ancak, gerçek video verilerinde bu koşulun, özellikle arkaplana göre küçük BHNler tarafından sağlanmasının her zaman mümkün olmadığı unutulmamalıdır. Önerilen metot, rasgele bir 2B videodan gürbüz bir biçimde 3B bilgisi çıkartılması konusunda önemli bir adımdır.

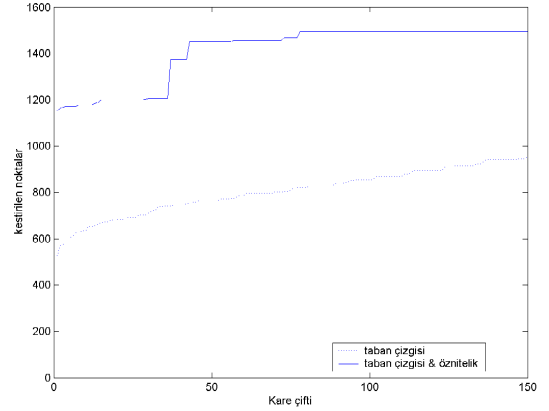
7. Kaynakça

- [1] J. P. Costeira, T. Kanade, "A Multibody Factorization Method for Independently Moving-Objects", *IJCV*(29), No. 3, p. 159-179, September 1998
- [2] W. Fitzgibbon, A. Zisserman, "Multibody Structure and Motion: 3D Reconstruction of Independently Moving Objects", *ECCV 2000*, 2000
- [3] M. Han, T. Kanade, "Perspective Factorization Methods for Euclidean Reconstruction", *CMU-RI-TR-99-22*, 1999
- [4] R. Hartley, A. Zisserman, "Multiple view geometry", Cambridge University Press, UK, 2003
- [5] M. Irani, P. Anandan, "A Unified Approach to Moving Object Detection in 2D and 3D Scenes", *IEEE Trans. on PAMI* Vol. 20, Issue 6, pp. 577-589, June 1998
- [6] M. Pollefeys, "Tutorial on 3D Modeling from Images", *ECCV 2000*, 2000.
- [7] G. Qian, R. Chellappa, Q. Zheng, "Bayesian Algorithms for Simultaneous Structure from Motion Estimation of Multiple Independently Moving Objects", *IEEE Trans. on Image Processing*, Vol. 14, No.1, January 2005
- [8] S. Soatto, P. Perona, "Reducing 'Structure from Motion': a General Framework for Dynamic Vision Part 1: Modeling", *Pattern Analysis and Machine Intelligence*, 20(9), September 1998
- [9] E. Tola, "Multiview 3D Reconstruction of a scene containing independently moving objects", Yüksek Lisans Tezi, ODTÜ Kütüphanes, 2005
- [10] C. Tomasi, T. Kanade, "Shape and Motion from Image Streams: A Factorization Method", *Journal of Computer Vision* 9(2), p.137-154, 1992



Şekil 5.2: "TUB-Room"un geriçatımı :

Üst sıra: BHNler/Orta sıra: Arkaplan, üst, sol, ön-üst-sol görüşümler, temel çizgi ölçeği/Alt sıra: Arkaplan, üst, sol, ön-üst-sol görüşümler, temel çizgi-eşleşme sayısı ölçeği



Şekil 5.3: Yakınsama grafiği

- [11] Y. Weiss, "Segmentation Using Eigenvectors: A Unifying View", *Proceedings of ICCV99*, pp. 975-982, 1999