

TÜBİTAK

2008-344
✓

TÜRKİYE BİLİMSEL VE TEKNOLOJİK ARAŞTIRMA KURUMU
THE SCIENTIFIC AND TECHNOLOGICAL RESEARCH COUNCIL OF TURKEY

Elektrik, Elektronik ve Enformatik Araştırma Grubu
Electrical, Electronical and Informatics Research Group

96091

**Ege Bölgesi'nin Kültürel Zenginlikleri için
3 Boyutlu Geriçatım ve Web-tabanlı Sanal Gezi**

Proje No: 105E187

**Prof.Dr. Yasemin YARDIMCI
Yalın BAŞTANLAR**

**ARALIK 2008
ANKARA**

Tablo Listesi

Tablo 5.1: F matrisinin farklı normalizasyon durumları ve farklı sayıda nokta ile hesaplandığında mesafe hatası ortanca değerleri.	18
Tablo 5.2: Değişen maliyet fonksiyonu ($p2l + p2c$, $p2l$ ya da $p2c$) için elde edilen sonuç hatalar.	18

Şekil Listesi

Şekil 2.1: 3B yapı çıkarımı için iki imge ile başlayıp her defasında bir imge ekleme yaklaşımı.....	7
Şekil 2.2: 35 imge için tespit edilmiş kamera konum ve yönelimleri.....	8
Şekil 3.1: Kabadan inceye geriçatım yaklaşımı ile ilk üç iterasyonda elde edilen geriçatımlar.....	9
Şekil 3.2: Dion antik kenti için 35 imge ile elde edilen iyileştirilmiş yoğun nokta kümesi tepeden ve yandan görünüşü.....	9
Şekil 3.3: Dion antik kentinden 35 imge kullanılarak yapılmış geriçatımdan elde edilen VRML'den bir görüntü.....	10
Şekil 4.1: Knossos için oluşturulan 3B modelin Google Earth™ ile görüntülenmesi.....	11
Şekil 4.2: Dion için bir hava fotoğrafını altlık olarak kullanılması, üzerine 3B geriçatımı yapılmış yüzeyin oturtulması ve bu öğelerin Google Earth™ ile görüntülenmesi.....	12
Şekil 4.4: Örnek tümyönlü görüntü ve ondan elde edilen panoramik görüntü.....	12
Şekil 4.3: Tümyönlü görüntüleme aparatı.....	12
Şekil 4.5: Panoramik imge tabanlı sanal geziden bir ekran görüntüsü.....	13
Şekil 5.1: 3B geriçatım için genel basamaklar.....	14
Şekil 5.2: Karma imgelerde nokta eşleme için deney kurulumu.....	15
Şekil 5.3: Örnek bir para-katadioptrik imge, ondan türetilmiş bir VCP imge ve aynı bölgeyi görüntüleyen bir perspektif imge.....	16
Şekil 5.4: Omni1 (a) and Omni2 (b) tümyönlü imgeleri için farklı perspektif kameralar ile doğru/toplam eşleme oranları (yüzde olarak).....	16
Şekil 5.5: Örnek Tümyönlü-Perspektif (üstte) ve VCP-Perspektif (altta) eşlemeler.....	17
Şekil 5.6: Örnek karma epipolar geometri, epipolar doğrular ve konikler.....	17
Şekil 5.7: Pers3 ve Omni1 (Şekil 5.2) kameralarının imgeleri ve üzerlerinde tespit edilmiş eşlenik noktalar.....	19
Şekil 5.8: Kestirilen noktalar ve kameraların pozisyonları.....	19

Özet

Projenin amacı arkeolojik alanların sunumu odaklı olmak üzere Web tabanlı bir sanal gezi sistemi geliştirilmesidir. Çok sayıda imgeden 3 boyutlu geriçatım ve panoramik görüntüleme gibi etkili teknikler kullanılmış ve bunlar bir coğrafi bilgi sistemi (CBS) ile entegre edilmiştir.

Projede, arkeolojik alanlar kolay edinilebilir teçhizatlar kullanılarak görüntülenmiş ve 3 boyutlu yapı stereo teknikler kullanılarak bu alınan görüntülerden elde edilmiştir. Renk ve doku bilgisi de oluşturulan 3 boyutlu modele eklendikten sonra sonuç geriçatım 3 boyutlu modelleme için sıkça kullanılan formatlardan birine (VRML) çevrilmiştir.

3B modeller arkeolojik kalıntıları andıran yüzeyler tanımlanarak ve desenler eklenerek sentetik olarak da oluşturulabilirdi. Ancak arkeolojik alanlar için böyle sentetik modeller kullanılması yapaylık hissini artıracağından biz gerçek imgeler kullanılarak 3B modellemeyi tercih ettik ve bu yönde çalışarak otomatik ve doğruluğu yüksek bir yöntem geliştirdik. Ayrıca, tümyönlü kameraların da normal kameralar ile beraber 3B geriçatım için kullanılmasını araştırdık ve bulgularımızı paylaştık.

Geriçatımı yapılan 3B modeller ve panoramik imge tabanlı sanal turlar Google Earth™ ile entegre edilip arkeolojik alan plan/krokileri de altlık olarak aynı ortama eklenmiştir.

Anahtar sözcükler: 3 Boyutlu geriçatım, Kültürel miras, Tümyönlü görüntüleme, Sanal gezi, Coğrafi bilgi sistemi

Abstract

The aim of this study is to build a Web-based virtual tour system, focused at the presentation of archaeological sites. The proposed approach is comprised of powerful techniques such as multiview 3D reconstruction, omnidirectional viewing based on panoramic images, and their integration with GIS technologies.

In the project, the scene is captured from multiple viewpoints utilizing off-the-shelf equipment and its 3D structure is extracted from the acquired images based on stereoscopic techniques. Color information is added to the generated 3D model of the scene and the result is converted to a common 3D scene modeling format such as VRML.

The 3D scene could be synthesized by replicating the surface and then adding texture information. However, such synthesized 3D models for cultural applications may cause in losing the feeling of reality. Therefore we preferred reconstructing the scene with real images and worked on an automatic procedure for accurate and photorealistic scene modeling. Moreover, we investigated the possible use of omnidirectional cameras together with perspective cameras for 3D reconstruction purposes and we shared our findings.

The reconstructed 3D models and interactive virtual tour tools based on panoramic imaging were integrated with Google Earth™ in which the excavation site plans can be added as detailed raster overlays.

Keywords: 3D Reconstruction, Cultural Heritage, Omnidirectional Vision, Virtual Tour, GIS

1. Giriş

İmge tabanlı 3 boyutlu geriçatımların Web tabanlı sanal geziler için kullanılması günümüzde görünüm kalitesinin son kullanıcı için yeterince çekici olmaması nedeniyle fazla yaygın değildir. Kültürel miras bu teknolojiyi kullanmak için en anlamlı alanlardan birisidir. Arkeolojik alanlar için örnek uygulamalar (Guarnieri ve diğ., 2004; Kadobayashi ve diğ., 2004; Conforti Andreoni ve Pinto, 2004) 3B lazer tarayıcı teknolojisi ile gerçekleştirilmiştir. 3B tarayıcı teknolojisi için geliştirilen aletler ortamı etkili bir şekilde tarayıp üzerine renk bilgisini ekleyip çıktı olarak 3B model verecek kadar ilerlemiştir. Ne var ki, bu yaklaşım için gerekli ekipman fiyatları son derece yüksektir ve veri kaydetme/işleme süresi de uzundur.

Proje kapsamında, otomatik ve imge-tabanlı bir 3B geriçatım yöntemi geliştirilmesi üzerine çalışılmıştır. Aynı amaçla Grün ve diğ. (2002) yıkılmış bir Buda heykelini turistik fotoğraflardan elde etmiştir ancak az sayıda ve düşük kalite fotoğraflar nedeniyle oluşan 3B model pek başarılı değildir. Pollefeys ve diğ. (1999) kontrollü bir ortamda video kamera ile alınan görüntülerden seçilen kareler ile 3B geriçatım üzerine çalışmışlar, daha sonra yöntemlerini Sagalassos arkeolojik sitesi için kullanarak video kamera ile 3B model oluşturmuşlardır (Pollefeys ve diğ., 2004).

3B modeller arkeolojik kalıntıları andıran yüzeyler tanımlanarak ve desenler eklenerek sanatkarlar tarafından sentetik olarak da oluşturulabilir. Ancak arkeolojik alanlar için böyle sentetik modeller kullanılması yapaylık hissini artıracığından biz gerçek imgeler kullanılarak 3B modellemeyi tercih ettik. Ayrıca sentetik olarak arkeolojik alanları oluşturma tecrübeli personel çalıştırılmasını gerektirmekte ve önemli miktarda bir maliyet oluşturmaktadır.

Otomatik, imge-tabanlı ve doğruluk derecesi yüksek 3B modelleri erişilebilir hesaplama gücü ve kaynakları ile oluşturmak zor bir hedeftir. Bu amaçla önerdiğimiz iş akışı aşağıda verilmiştir. Bilimsel açıdan mevcut yöntemlere göre getirdiğimiz yenilikler Adım 4 ve 5'tedir. Diğer adımlar için uygun yöntemler seçilerek kullanılmıştır/gerçeklenmiştir.

- 1) Arkeolojik alandan (tercihen yüksek çözünürlüklü) çok sayıda fotoğrafın alınması.
- 2) Kamera kalibrasyonu ile içsel parametrelerin tespiti.
- 3) İmgelerde objektiften kaynaklanan bozulmaların giderilmesi.
- 4) İmgelerin dış kalibrasyonu, yani her imge alınırken anlık kamera konum ve yönelimlerinin tespit edilmesi. Bu parametreler, imgelerde köşe noktaları bulunması, bu noktaların eşlenmesi ve kamera hareketinin bulunması ile çıkarılmaktadır.
- 5) İmgeler ve çıkarılan parametreler kullanılarak görüntülenen yerin 3B geriçatımının yapılması.
- 6) Geriçatım çıktısı olarak elde edilen noktaların imgelerden çıkarılan yüzey doku bilgileri ile birleştirilerek VRML veya benzeri formata çevrilmesi.

Pek çok stereo geriçatım yönteminde olduğu gibi sonuçların kalitesi büyük oranda kamera kalibrasyonunun (Adım 2-3) ve kamera hareketi tespitinin (Adım 4) ne derece başarılı yapıldığına bağlıdır. İçsel parametrelerin tespiti için kalibrasyon deseni ile alınmış görüntüler kullanan ve iyi sonuçlar veren araçlar mevcuttur, biz Bouguet (2007) MATLAB araç kutusunu kullandık. Otomatik hareket kestirimi problemi (Adım 4) için alıp kullanılabilecek düzeyde çözümler mevcut değildir. Bu problemi çözebilmek için şimdiye kadar önerilmiş yaklaşımları inceleyerek en uygun olanlarını örnek aldık (Beardsley ve diğ., 1997; Pollefeys ve diğ., 1999; Pollefeys ve diğ., 2004). Doğruluk derecesini daha da artırmak için bu yöntemlere bazı ekleme ve değişiklikler yaptık. Gürbüz nokta eşleme için SIFT (Lowe, 2004) kullanma ve kestirilen yapının toplu düzenlemesi için Seyrek Toplu Düzenleme (Sparse Bundle

Adjustment; Lourakis ve Argyros, 2004) bu deęişikliklerden en önemlileridir. Önerilen yöntemin detayları Bölüm 2'de verilmiştir.

Ardından 3B geriçatımı yüksek çözünürlüklü fotoęraflar kullanarak kolay erişilebilir bilgisayar kaynakları ile nasıl yaptığımız Bölüm 3'de anlatılmıştır. Geliştirilen teknik, geleneksel uzay-tarama (space sweeping) yaklaşımının (ör: Zabulis ve dię., 2003) geri-projeksiyon yüzeyi olarak düzlem deęil de küre yüzeyi kullanıldığı şeklindedir. Sonuçta doğruluk derecesi daha yüksek ve daha verimli hafıza kullanır hale gelinmiştir. Ayrıca, kabadan-inceye seviyeli bir geriçatım yapılmasını sağlayacak bir deęişiklik yapılmış, bu sayede daha hızlı sonuç almak mümkün hale gelmiştir.

Açıklanan yaklaşımla, piyasada bulunan tercihen yüksek çözünürlüklü fotoęraf makinaları ile alınan birkaç fotoęraf kullanılarak bir ortamın geriçatımını yapabilecektir. Bu şekilde kaliteli bir 3B model elde etmek için görüntü alımı işi çoklu kamera sistemleri kullanmaya veya yüksek kaliteli bir video kamera kullanıp uygun karelerin seçilmesine (Mordohai ve dię., 2007) göre basitleştirilmiştir.

Yoęun nokta kümesi olarak elde edilen geriçatım önce bir örgü (mesh) haline getirilir ve imgelerden oluşturulan doku (texture) bu örgü üzerine eklenir. Bu aşamalar için önerilmiş olan ve başarılı bulunduğumuz algoritmaları (sırasıyla Othake ve dię., 2005; Lempitsky ve Ivanov, 2007) hayata geçirdik. Sonuç olarak VRML (Virtual Reality Modeling Language) ya da doğrudan KML (Keyhole Markup Language) formatında bir 3B modele çevrim yapılır. KML çıktısı Google Earth™ platformu için doğrudan kullanılabilirdiğinden geriçatılmış modeller bir coęrafi bilgi sistemi ile kolayca entegre edilmiş olur. Ayrıca, Google Earth™ platformu modellerin sergilenmesi ve panoramik imge tabanlı sanal gezi gibi dięer geliştirilen uygulamaların entegre edilmesi için de çözümler sunmaktadır. Bu nedenle geliştirdiğimiz Web tabanlı pilot uygulama için Google Earth™ platformunu tercih ettik.

Panoramik imgelerin kullanıldığı sanal geziler son yıllarda bütün dünyada yaygın şekilde kullanılmaya başlanmıştır. Bu gezilerin etkinliği ve kullanılabilirliği Bastanlar (2007) ve Villaneuva ve dię. (2004) tarafından irdelenmiştir. Proje kapsamında gerçekleştirilen pilot uygulamada ilgili teknolojilerin hepsini bulundurmak adına bahsedilen panoramik imge tabanlı sanal geziden de bir örnek hazırladık ve Google Earth™ platformuna entegre ettik. Geliştirilen pilot uygulamanın detayları Bölüm 4'de verilmiştir.

Proje süresince ayrıca, tümyönlü (panoramik) kameraların da normal kameralar ile beraber 3B geriçatım için kullanılmasını araştırdık. Çalışmalar henüz perspektif kameralar ile olduğu gibi bir 3B model elde etme aşamasına gelmediğinden pilot uygulamaya dahil edemedik. Ancak bilimsel açıdan deęer taşıyan bu çalışmalarımızın bulgularını Bölüm 5'te raporladık.

2. Gürbüz kamera geometrisi saptama

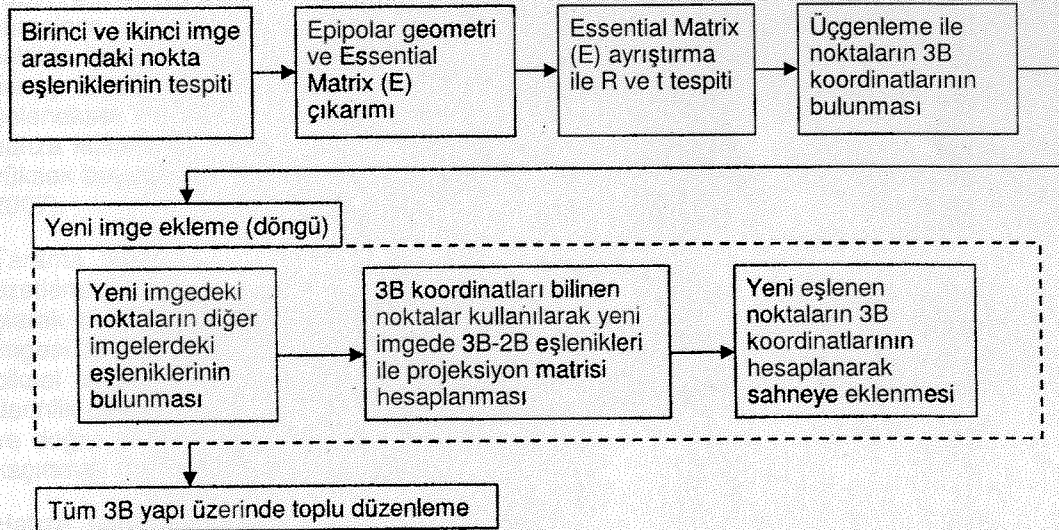
Bu bölümde her imge için kamera konum ve yönelimlerinin tespiti için uyguladığımız yöntem anlatılmaktadır. Kamera hareketinin yüksek doğruluk ile tespiti elde edilecek 3B geriçatımın kalitesini doğrudan etkilemektedir ve doğrudan alınıp kullanılabilir düzeyde araçlar mevcut deęildir. Bu nedenle literatürde geçen yaklaşımları inceleme ve önerilen yöntemlerin doğruluk derecesini daha da artırmak için bazı ekleme ve deęişiklikler yapma yoluna gittik.

Çalışmamızda kullandığımız yaklaşım ilk defa Beardsley ve dię. (1997) tarafından önerilmiş, Pollefeys ve dię. (1999, 2004) tarafından geliştirilerek video görüntülerinden 3B geriçatım için kullanılmıştır. Tola (2005) da tezinde bu yaklaşımı uygulamaya koymuştur. Şekil 2.1'de algoritmanın akışı özetlenmiştir.

Öncelikle aynı sahneyi gören sıralı imgeler nokta eşleniklerinin tespiti ile başlanır. Önceki gerçeklemede imgelerdeki noktalar Harris köşe dedektörü (Harris ve Stephens, 1988) ile tespit edilmiştir, ardından yanlış eşleştirmeleri elemek için noktaların yakınlığı ve simetrisi gibi özellikler kullanılmıştır (Tola, 2005). Biz öncelikle Harris yöntemi ile SIFT nokta bulma ve eşleştirme (Lowe, 2004) yöntemini karşılaştırdık ve SIFT ile elde edilen eşleştirme başarısının daha iyi olduğunu gözlemleyerek çalışmamızın devamında onu kullandık. Aradaki fark, özellikle kameralar arası mesafe arttığında daha önemli hale gelir ve köşe bulup iki imge arasında klasik bölge eşleştirmesi yapmaktansa SIFT gibi ölçek ve dönmeden bağımsız yöntemler kullanmak kesinlikle tercih edilmelidir. Yüksek çözünürlüklü imgelerde SIFT öznelikleri çıkartılırken işlemci hafızasının aşılması problemi ile karşı karşıya kalınabilir. Bunu aşmak için biz imgeyi parçalara bölerek öznelik saptama ve eşleştirme işini her bir parça için gerçekleştirdik. Arada öznelikleri ve eşleştirmeleri kaybetmemek için parçaları kenarlarından bir miktar örtüşecek şekilde böldük. Gerek bu örtüşme sonucu aynı eşlerin birden fazla saptanması gerekse SIFT'in kendi doğasından kaynaklanan aynı noktalar için farklı ölçeklerde eş saptanmasından dolayı oluşan tekrarları en son eş listesinden çıkardık.

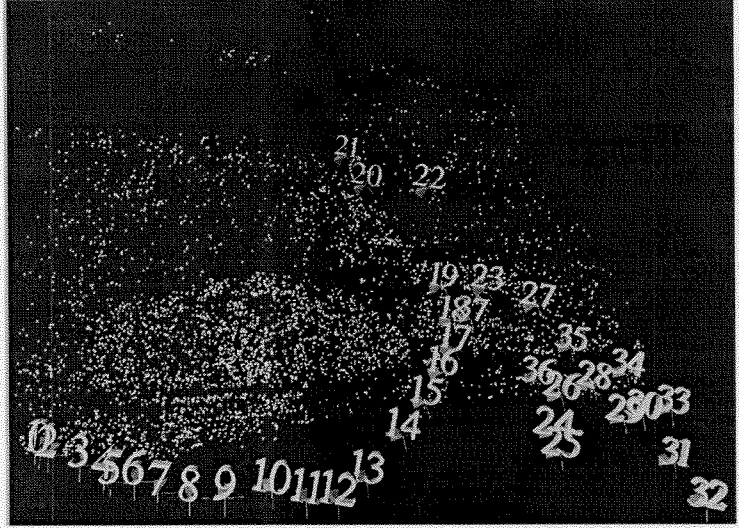
Sonraki adımda tespit edilen eşlenik noktalar ile epipolar geometri (temel matris F ya da öz matris E) tespit edilir ve RANSAC algoritması (Fischler ve Bolles, 1981) çalıştırılarak epipolar geometriye uymayan eşlenikler elenir. Sonuçta elde edilen E matrisinden ikinci imgenin ilk imgeye göre dönme matrisi (R) ve öteleme vektörü (t) elde edilir ve son olarak eşlenik noktaların 3B koordinatları doğrusal üçgenleme ile kestirilir. Bu adımlar için uyguladığımız yöntemler Hartley ve Zisserman (2000)'de ayrıntılı olarak anlatılmaktadır.

Bir dizi imge ile çalışıldığında, yukarıdaki yöntem ilk iki imge için uygulanır ve daha sonraki her imge için, nokta bulma ve eşleştirme kendisinden önceki imge ile yapılır. Ayrıca, bu eşlenik noktaları için daha önce tespit edilmiş olan 3B koordinatlar da belirlenir. Yine RANSAC tabanlı bir hesaplama ile yeni imgenin projeksiyon matrisi (P) hesaplanır. Bu arada ilk defa eşlenmiş olan noktalar için 3B koordinatlar hesaplanarak nokta kümesine eklenir. Her yeni imge ekleme işleminin ardından etkili bir algoritma olan 'seyrek toplu düzenleme' (Lourakis ve Argyros, 2004) kullanılarak çıkarılan 3B yapı tekrar-izdüşüm (reprojection) hatasını minimize edecek şekilde değiştirilmiştir.



Şekil 2.1: 3B yapı çıkarımı için iki imge ile başlayıp her defasında bir imge ekleme yaklaşımı

Bu işlemler elimizdeki imgelerin alındığı anlarda kameranın konum ve yönelimleri, bir sonraki adımda kullanılmak üzere tespit edilmiş olunur. Şekil 2.2'de Dion antik kenti için alınan 35 imge için tespit edilmiş konum ve yönelimler, ayrıca eşlenmiş olan noktaların 3B sahnedeki yerleri gösterilmiştir.



Şekil 2.2: 35 imge için tespit edilmiş kamera konum ve yönelimleri

3. 3B Geriçatım

Bu bölümde yüksek çözünürlüklü fotoğraflar kullanarak 3B geriçatım yapma metodumuz anlatılmaktadır. Geleneksel uzay-tarama (space sweeping) yöntemi daha az hesaplama gücü gerektirdiğinden ve grafik donanımı ile hızlandırılabilirdiğinden çok imgeli geriçatımlar için sıkça tercih edilen bir yöntemdir (Yang ve diğ., 2002; Li ve diğ., 2004). Ancak, görüntülenen yüzeyin yöneliminin neden olduğu izdüşümsel değişiklikleri hesaba katan diğer yaklaşımlara (Zabulis, 2007) göre doğruluğu daha düşük sonuçlar vermektedir. Bu nedenle çok yönlü tarama (Mordohai ve diğ., 2007) veya normal tarama ile elde edilen ilk geriçatımın sonradan detaylandırılması (Zabulis ve Kordelas, 2006) gibi yöntemler önerilmiştir.

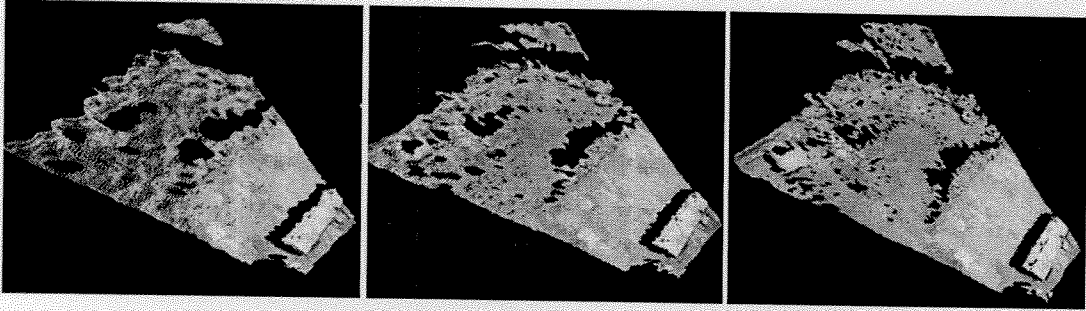
Uyguladığımız yaklaşım, geleneksel uzay-tarama yaklaşımının geri-izdüşüm (backprojection) yüzeyi olarak düzlem değil de küre yüzeyi kullanıldığı şeklindedir (Zabulis ve diğ., 2006). Proje süresince yapılan çalışmalar sonucu yöntem kabadan-inceye seviyeli bir geriçatım yapacak şekilde iyileştirilmiştir. İyileştirmeler neticesinde daha hızlı ve daha az işlemci hafızası ile sonuç almak mümkün kılınmıştır. Hafıza tutumluluğu iki açıdan önemlidir. Birinci neden, geleneksel PC'lerin yüksek çözünürlüklü imgeleri işlemekte yetersiz kalmaları ve sanal hafıza kullanarak işlemi çok yavaşlatmalarıdır. İkinci neden ise günümüz tekniklerinin daha yüksek boyutta veri işlemek için artık grafik donanımını da kullanmalarıdır (Mordohai ve diğ., 2007).

Tarama işlemlerinin akışı normal düzlemsel taramaya benzer şekildedir ve şu şekilde özetlenebilir. Her bir i derinliği için imgeler geri-izdüşüm yüzeyine geri-izdüşürülür ve yerel olarak birbirleriyle karşılaştırılır. Bu karşılaştırma sonucu her bir derinlik için bir benzeşim imgesi S_i elde edilir. Derinlik değiştirerek S_i 'deki pikseller S_{i+1} ve S_{i-1} 'de aynı konumdaki piksel değerleriyle karşılaştırılır. En güçlü yerel benzeşimin gözlemlendiği derinlik optimum derinlik olarak belirlenir. Korelasyonun en büyük olduğu bu durumda izdüşümler doğru yere ve doğru yönelim açılı yüzeye yapılmış demektir. Düz alanlarda yanlış derinlik tespitinden kaçınmak için 'yerel maksimum' şartı aranır.

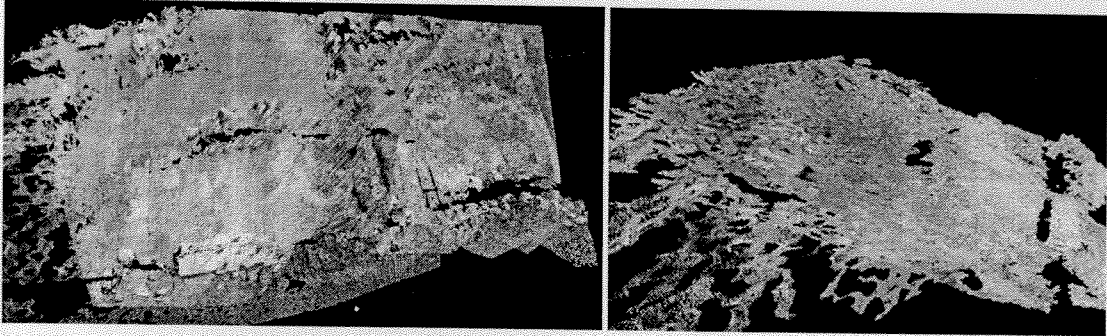
Hafıza tutumluluğu ise geri-izdüşüm imgesinin eşit büyüklükte parçalara bölünmesi ile sağlanır. Tarama algoritması her bir parça için birbirinden bağımsız olarak çalıştırılır. Bu parçalar, sınırlarda bozulma etkisi görülmemesi için birbirleriyle hafifçe örtüşmektedir.

Yöntemin hızlandırılması için ayrıca döngülü ve kabadan-inceye geriçatım tekniği gerçekleştirilmiştir. Her bir döngüdeki imge verisi, girdi imgenin klasik imge piramidinden alınır. Önce piramidin en tepesindeki küçük imgeden başlanır ve en son döngüde orijinal imge kullanılacak şekilde ilerlenir. Yine her bir döngüde geri-izdüşüm yüzeyi detayları alacak şekilde yoğunlaştırılır. Yukarıda açıklanan imgeleri parçalara bölme yöntemindeki parçalar da uyumlu olarak küçültülür. Her bir döngüde bir önceki döngüde tespit edilen derinlik değerinin komşuluğundaki derinliklerde tarama yapılır. Böylece detaylı taramalar küçük hacimler içerisinde kısıtlanmış olur ki bu büyük bir zaman kazanımı getirmektedir. Detaylandırma katsayısı 2 olan, yani her döngüde imgenin eni ve boyu iki katına, parça sayısının da dört katına çıkan bir uygulama için işlem süresinin yaklaşık 50 kat kısaldığını gözlemlenmiştir. Şekil 3.1'de bahsedilen detaylandırmanın üç aşaması görülmüştür.

Sonuçta elde edilen derinlik haritası, derinlik süreksizliklerindeki bozuklukları elemek için, Mulligan ve diğ. (2004) tarafından önerildiği gibi dikkatlice filtreledik. Bu şekilde bazı doğru bulunmuş derinlik değerleri de kaybedilir ancak uygulanan çok noktadan fotoğraf kullanılan yaklaşım ile o bölgelerin başka imge çiftlerinden geriçatılacağı varsayılır. Bölgesel geriçatımları birleştirmek ve pürüzleri yumuşatmak için aradeğerleme ile ince bir yüzey oturttuk. Bunun için Carr ve diğ. (2001)'ni örnek alarak RBF (Radial Basis Functions) kullandık. Bu aşamalardan sonra iyileştirilmiş yoğun bir 3B nokta kümesi vardır. Şekil 3.2'de Dion antik kenti için 2448 x 3264 piksel çözünürlükte 35 fotoğraf kullanılarak yapılmış bir geriçatımın tepeden ve yandan görünüşü verilmiştir.



Şekil 3.1: Kabadan inceye geriçatım yaklaşımı ile ilk üç iterasyonda elde edilen geriçatımlar

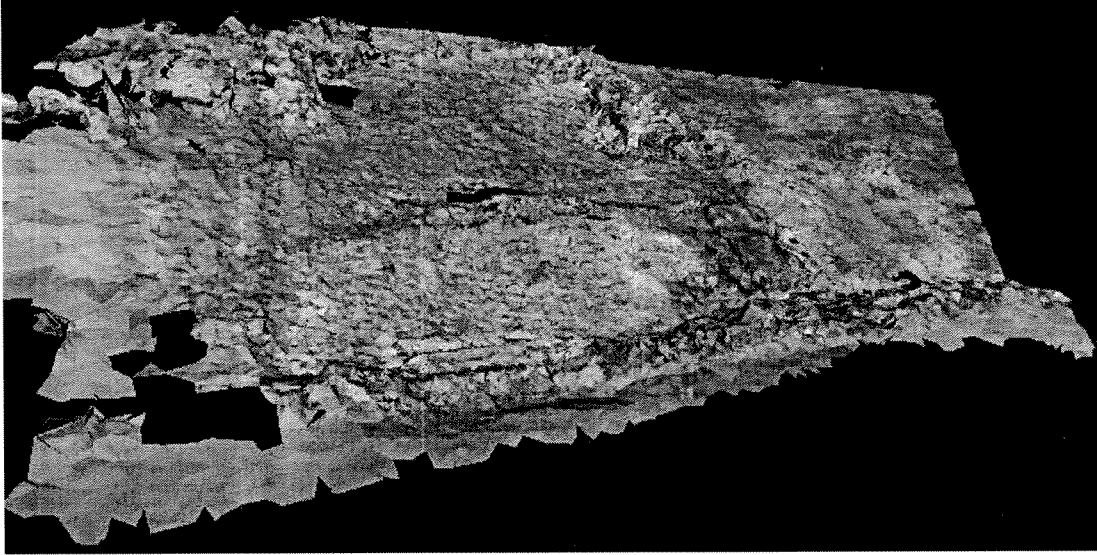


Şekil 3.2: Dion antik kenti için 35 imge ile elde edilen iyileştirilmiş yoğun nokta kümesi tepeden ve yandan görünüşü

Elimizdeki bu nokta kümesinden bir örgü elde etmeye ve arkasından üzerine doku oturtularak pilot uygulamamızda kullanmak üzere VRML ya da KML formatına dönüştürmeye ihtiyacımız vardır. Örgü elde etmek için biz Ohtake ve diğ. (2005) nin yöntemini ve sağladıkları yazılımı kullanmayı uygun bulduk. Dokuyu oluşturmak ise biraz daha karmaşıktır çünkü farklı bölgeler için farklı imgelerden elde edilen dokuların birleşme yerlerinin en az bozulma ile aktarılması gerekmektedir. Bu alanda halen araştırmalar sürmekte olup oturmuş bir yöntem mevcut değildir. Bu birleşme yerlerinde farklı imgelerden gelen dokuları en iyi

şekilde kullanabilmek için çizge kesit (graph cut) tabanlı bir yöntem (Lempitsky ve Ivanov, 2007) gerçekledik. Sonuçları görece iyi bulmamıza rağmen başarı büyük oranda geometrik modellemeye ve imgeler arasındaki renk değerlerinin tutarlı olmasına bağlıdır.

Belirtildiği gibi geriçatım için çok sayıda ve yüksek çözünürlüklü imgeler kullanarak detaylı geriçatımlar elde edebilmemize rağmen VRML halinin rahat kullanılabilmesi ve GoogleEarth gibi programlara entegre edilebilmesi için 3B modelin çözünürlüğü belli seviyede (<50000 poligon) tutulmuştur. Şekil 3.3'te VRML dosyasından alınan bir görüntü verilmiştir. Bu ve proje kapsamında elde edilen diğer 3B modellere aşağıdaki adreste ulaşılabilir.
http://www.ii.metu.edu.tr/~3daegean/3d_reconstruction

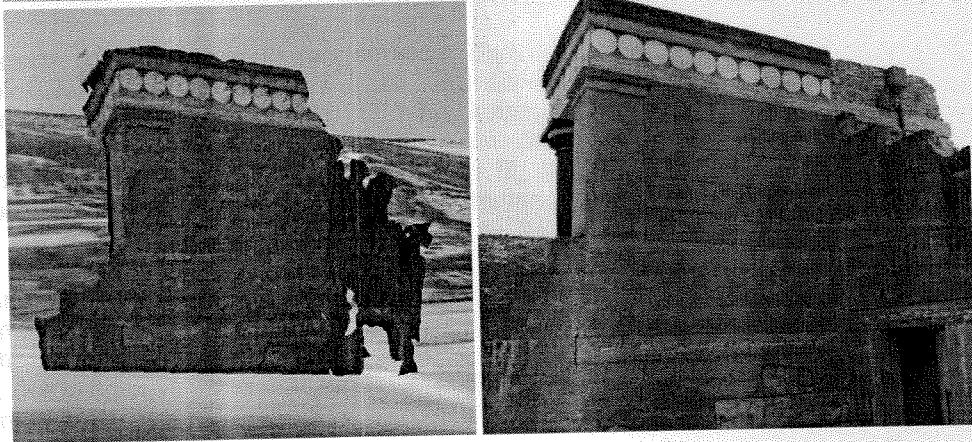
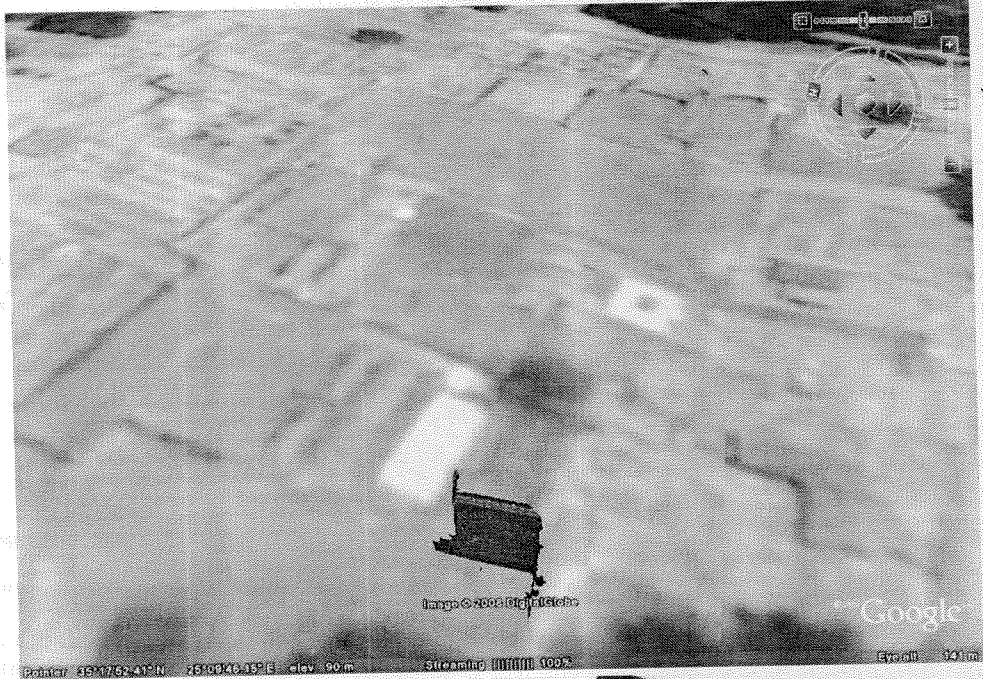


Şekil 3.3: Dion antik kentinden 35 imge kullanılarak yapılmış geriçatımdan elde edilen VRML'den bir görüntü

4. Sanal Gezi Pilot Uygulaması

Web tabanlı ve Google Earth™ (GE) platformunu kullandığımız pilot uygulamamız için, 3B geriçatım sonucu oluşturulan VRML modelini Collada (XML tabanlı 3B model) dosya formatına çevirerek CBS altlığı olarak üzerine yerleştirmekteyiz. Format dönüşümünü için PolyTrans isimli 3B model dönüşüm yazılımı ile modeli GE üzerine koymayı da Sketch-Up yazılımı ile yaptık. Şekil 4.1'de Knossos antik kenti için oluşturulan model ve GE üzerine eklenmiş hali ile örnek görüntü verilmektedir. Ayrıca TriDef™ Visualizer isimli bir GE eklentisi ile modeller otostereoskopik monitörler için 3B gerçekleştirilmekte daha etkileyici bir sunum yapılabilmektedir.

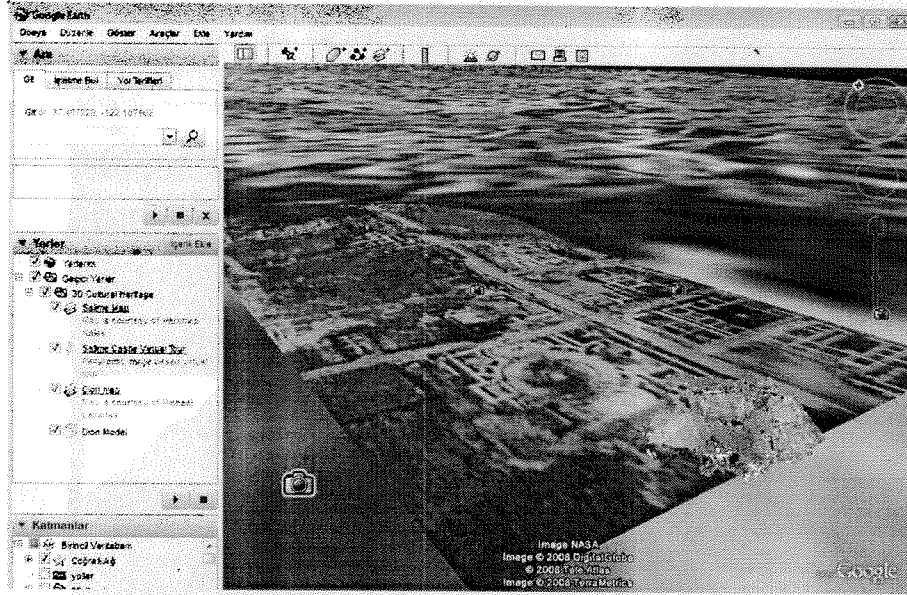
Ayrıca kazı alanlarına ait plan ve krokiler ölçekli bir şekilde dünya üzerindeki gerçek konumlarına oturtularak gösterilebilmektedir. Söz konusu bölgede GE çözünürlüğü düşük ise bu plan ve krokiler 3B modelin üzerine konduğu bir altlık görevi yaparlar. Bütün bu unsurlar bir KML (veya KMZ) dosyasında toplanarak GE'ye yüklenir. İlgili noktalara konulan yer imleçleri veya GE menüsündeki bağlantılar ile diğer uygulamalar (sanal gezi, ek görsellerin açılması v.b.) başlatılabilmektedir (Şekil 4.2).



Şekil 4.1: Knossos için oluşturulan 3B modelin Google Earth™ ile görüntülenmesi. Üstte: 3B gerçatımı yapılan duvar ile beraber alanın genel görünüşü. Altta solda: 3B modelin yakından görünüşü. Altta sağda: aynı duvarın gerçek fotoğrafı

Kullanıcının GE üzerindeki bağlantıya tıklayarak geçebildiği bir uygulama panoramik imge tabanlı sanal gezidir. Bir Web sayfasının üzerinden kullanıcıya sunulan bu gezinin esas unsuru kullanıcının kontrol edebildiği bir görüntüleme penceresidir. Bizim de kullandığımız Java Applet teknolojisi Web uygulamalarında sağladığı rahatlık nedeniyle uygun bir seçenektir.

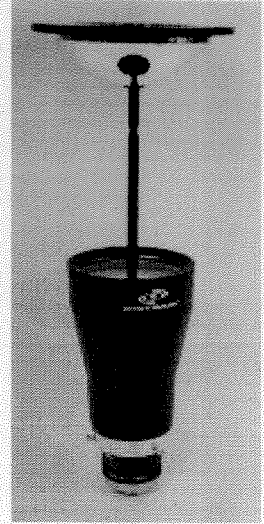
Görüntüleme penceresi sahadan alınan imgelerden elde edilen panoramik imgenin bir bölümünü kullanıcıya gösterir. Panoramik imgeler normal fotoğrafları birleştirerek oluşturulabileceği gibi 360° den görüntü alan tümyönlü kamera aparatları (Ör: Şekil 4.3) kullanılarak da elde edilebilir. Bizim de tercih ettiğimiz bu yöntem ile kameradan bir tümyönlü görüntü elde edilir ve sonrasında panoramik görüntüye çevrilir (Şekil 4.4).



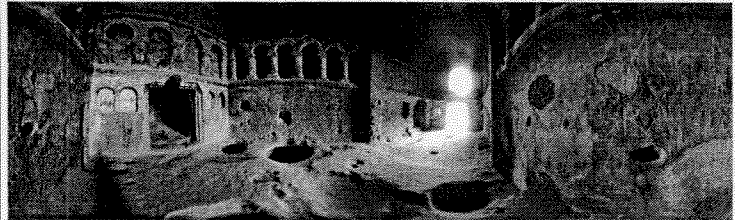
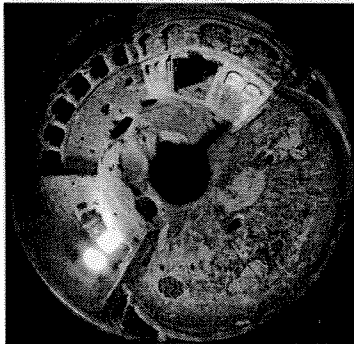
Şekil 4.2: Dion için bir hava fotoğrafını altlık olarak kullanılması, üzerine 3B geriçatımı yapılmış yüzeyin oturtulması ve bu öğelerin Google Earth™ ile görüntülenmesi

Bu sanal gezi sayfasında görüntülerin dışında başka Web teknolojileri kullanarak yazılı ve sözlü bilgi de verilebilmektedir. Görüntülenen arkeolojik alanın bir haritasını kullanmak kullanıcının yön algısını kuvvetlendirmektedir. Bir adım ötesi bu haritayı görüntüleme penceresi ile entegre etmektir. Bu tekniklerle daha çok bilgi ergonomik bir şekilde ulaştırılır (Bastanlar, 2007).

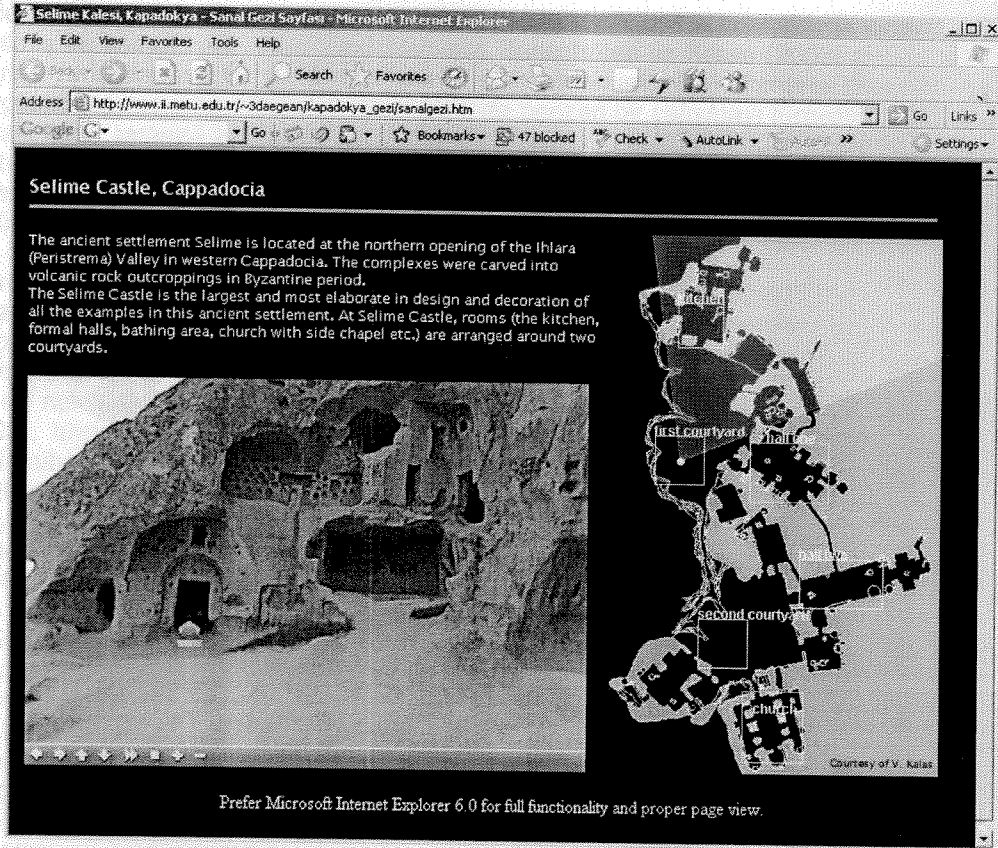
Şekil 4.5'te Kapadokya Bölgesi'ndeki Selime Kalesi için hazırlanan panoramik imge tabanlı sanal gezi sayfasından bir ekran görüntüsü yer almaktadır. Sol-altta görüntüleme penceresi sağ tarafta ise site planı bulunmaktadır. O anda görüntülenmekte olan alan, görüş açısı ve yönü plan üzerinde gösterilmekte ve kullanıcı görüntüyü oynattıkça güncellenmektedir.



Şekil 4.3: Tümyönlü görüntüleme aparatı



Şekil 4.4: Örnek tümyönlü görüntü ve ondan elde edilen panoramik görüntü



Şekil 4.5: Panoramik imge tabanlı sanal geziden bir ekran görüntüsü

Proje süresince çalıştığımız arkeolojik alanlar, Yunanistan'da Dion ve Knossos antik kentleri, Türkiye'de ise Selime Kalesi ve Priene antik kentidir. Bu alanlardan alınan görüntülerle yapılan gerçekleştirmeler GE üzerinde çalışan pilot uygulamada toplanmıştır. Pilot uygulamanın güncel haline aşağıdaki adreste erişilebilir:

<http://www.ii.metu.edu.tr/~3daegean/recent.htm>

5. Karma Sistemler ile 3B Geriçatım

Tümyönlü (omnidirectional) kameralar ile perspektif kameraların beraber kullanıldığı sistemlerle 3B geriçatım yapmak istenmesindeki temel motivasyon tümyönlü kameraların sağladığı geniş görüş açısı avantajını koruyup, ortaya çıkan düşük çözünürlük problemini normal kameralar ile gidermektir.

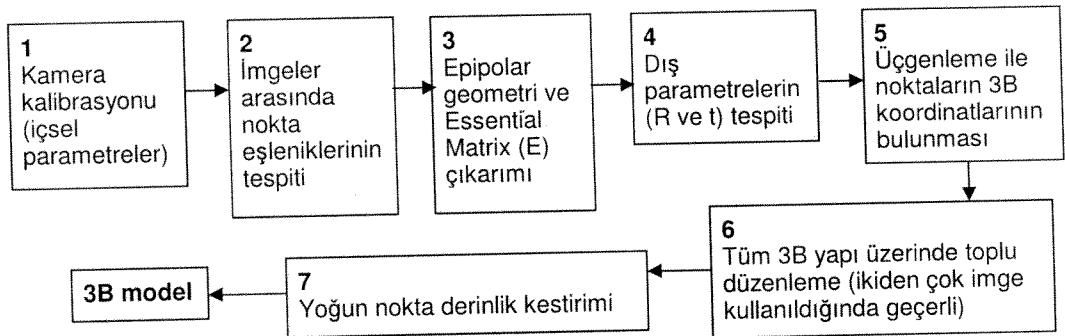
Karma kameralı sistemler sadece perspektif kamera veya sadece tümyönlü kamera kullanılan sistemlerden farklı sorunlar taşımaktadır. 3B geriçatım için ilk basamak olan farklı kamera görüntülerindeki noktaların eşlenmesi için imge bölgelerinin piksel değeri farklarının kareleri toplamı (SSD) gibi klasik yöntemler kullanılamaz çünkü görüntüler arasında kamera tipi değişmiştir. Epipolar geometri kestirimi için kullanılan formalizasyon da farklıdır. 3B yüzeyin (veya nokta kümesinin) oluşturulması için ise farklı imgelerden kullanılan bölgelerin kestirilen 3B yapıya uygun olarak bükülmesi gerekmektedir.

Sadece tümyönlü kameralar kullanılarak 3B geriçatım üzerinde son yıllarda çeşitli çalışmalar yapılmış ve somut ilerlemeler kaydedilmiştir. Bunschoten ve Krose (2003) gezici bir robotun üzerine yerleştirdikleri tümyönlü kamera ile elde edilen imgeler kullanarak bina içerisinde 3B geriçatım yapmışlardır. Fleck ve diğ. (2005) de benzeri bir çalışmayı tümyönlü görüntüleme olarak lazer tarayıcı kullanarak gerçekleştirmişlerdir. Lhuillier (2007) yakın zamanda aynı yöntemle nispeten daha başarılı sonuçlar elde etmiştir. Yine de, tümyönlü kameraların kullanılmasına bağlı olarak, kameraların kapsayamadığı bazı bölgelerde ve görüntülerin çözünürlük dezavantajı nedeniyle boşluksuz geriçatım yapılamamıştır.

Literatürdeki çalışmalardan sadece birkaç tanesi tümyönlü ve perspektif kameraların beraber kullanılmasını içermektedir. Chen ve Yang (2003) karma kameralar içeren bir sistemin kalibrasyonu üzerine çalışmışlardır. Öncelikle parabolik aynalı bir tümyönlü kamerayı kalibre etmişler, ardından aynı bölgeyi gören perspektif kamera görüntülerinde 3B koordinatları bilinen noktaları kullanarak perspektif kameraların kalibrasyonunu tamamlamışlardır. Sturm (2002), tümyönlü ve perspektif kameraları beraber kullanarak 3B yapı tespiti konusunda teorik altyapıyı tamamlayan çalışmalara imza atmış, farklı tür kamera görüntüleri kullanarak nasıl epipolar geometri tespit edilebileceğinin matematiksel çıkarımını yapmış, ayrıca tümyönlü kameraların perspektif kameralarla beraber temsil edildiği bir 'genelleşmiş kamera modeli' geliştirmiş ve hareketten yapı çıkarımı üzerinde çalışmışlardır (Ramalingam ve diğ. 2004).

Şekil 5.1'de 3B geriçatım basamakları perspektif kameralar için de geçerli olan genel haliyle verilmiştir. Önce kamera kalibrasyonu yapılarak içsel parametreler elde edilir (Adım 1). Sonrasında ise imgelerin dış kalibrasyonu yapılarak her imgeye ait dönme ve öteleme (R, t) değerleri kestirilir (Adım 2-4). Bu bilgiler kullanılarak 2. Adım'da eşlenmiş olan noktalar için 3B koordinatlar üçgenleme ile tespit edilebilir (Adım 5). Bu aşamada elimizde basit bir 3B geriçatım ve kamera pozisyonları mevcuttur. Eğer çok sayıda imge kullanıldıysa, tüm 3B yapı üzerinde toplu düzenleme (bundle adjustment) ve istenirse neredeyse tüm imge noktaları için yoğun derinlik kestirimi (dense depth estimation) yapılarak 3B yüzey/model elde edilir.

Proje süresince perspektif ve tümyönlü kamera içeren sistemler için hareketten yapı çıkarımının (structure from motion) kamera kalibrasyonu, nokta eşleştirme, epipolar geometri kestirimi ve üçgenleme aşamaları için teknikler geliştirilmiştir (Şekil 5.1'de ilk beş adım). Proje araştırmacılarından Yalın Baştanlar'ın doktora tezi olarak devam eden çalışmalarda 3B model oluşturma için kalan adımlar (Adım 6 ve sonrası) henüz tamamlanamamıştır. Proje süresince yapılan çalışmalar aşağıda alt başlıklar altında özetlenmektedir.



Şekil 5.1: 3B geriçatım için genel basamaklar

5.1 Kamera kalibrasyonu

Perspektif kameralarda genel olarak kabul gören yöntem doğrusal yolla odak sensör eğikliği/basıklığı gibi sistem parametrelerini tespit etmek, ardından objektif b parametrelerini de dahil ederek doğrusal olmayan eniyileme yapmaktır. Tümyönlü kamera için buna ayna parametrelerinin de dahil olması gerekir, ancak imge oluşumu olarak tanımlanamaz. Bu nedenle, tüm parametreler bir başlangıç değerinden sonra minimizasyonu yapılarak doğrusal olmayan yolla bulunur. Aynalı ve balıkgözü sistemi geliştirilmiş ve kullanıma açılmış doğrusal olmayan kalibrasyon algoritmaları m (Scaramuzza ve diğ. 2006; Mei ve Rives, 2007; Kannala ve Brandt, 2004).

Proje araştırmacısı Yalın Baştanlar 6 ay süre ile INRIA Rhone Alpes enstitüsünde bir Dr. Peter Sturm ile tümyönlü kameraları içerecek doğrusal bir kalibrasyon yöntemi üzerine çalışmıştır. Bu çalışmada küresel kamera modeli (Geyer ve Daniilidis, 2003) yükseltilmiş koordinatlar (lifted coordinates) kullanılmıştır. Aşağıda, geliştirilen algoritmanın anahatları verilmektedir. Okuyucu, hem ayrıntılar hem de küresel kamera modeli yükseltilmiş koordinatlar hakkında detaylı bilgi için çalışmanın ürünü olan makaleyi okumalıdır (Bastanlar ve diğ., 2008).

- Geliştirilen yöntem, küresel kamera modeli ile modellenebilen tüm kameralar için kullanılabilir. Tüm tek görüş noktalı (single-viewpoint) tümyönlü kameralar, hatta norm perspektif kameralar böyledir.

- Doğrusal olmayan geometrileri doğrusal denklemlerle ifade etmek için koordinatlar yükseltilmiştir. Örneğin parabolik aynalı bir sistemde imge koordinatları olarak 3-vektör $\mathbf{q}=(x,y,1)$ yerine 4-vektör $\hat{\mathbf{q}}=(x^2+y^2,x,y,1)$ kullanılarak 3x3 yerine 4x4'lük bir F matrisi epipolar geometri ifade edilebilir. Sadece parabolik aynalı değil de her çeşit tümyönlü kamerayı kapsamak istersek, 4 yerine 6 vektör yükseltilmiş koordinat kullanmak gerekir. Uzaydaki bir noktayı (\mathbf{Q}) ise 4-vektörden 10-vektöre yükselterek aynı sisteme dahil edebiliriz.

- Bu durumda standart haliyle 3x4'lük projeksiyon matrisi de 6x10'lük P_{cata} olarak yazılabilir. Bunun içerisinde iç parametreler matrisi K_m , dönme matrisi R ve öteleme matrisi T yükseltilmiş halleriyle (6x6, 6x6 ve 6x4) kaynaşmış şekilde bulunur.

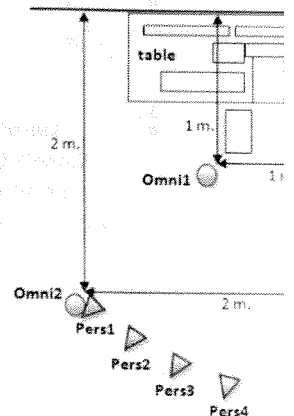
- Öncelikle bir kalibrasyon deseni ve yeterli sayıda 3B-2B nokta eşleniği kullanılarak 6x10'lük katadioptrik projeksiyon matrisi P_{cata} kestirilir. Bu model üç farklı düzleme dağınık en az 20 adet nokta eşleniği gerektirmektedir, dolayısıyla 3 boyutlu bir kalibrasyon denemesi gereklidir.

- İçsel parametreler bazı lineer cebir teknikleri kullanılarak P_{cata} içerisinde çekim geometrisi üzerinden reprojeksiyon hatası en küçük olacak şekilde tüm parametreler optimize edilir.

5.2 İmgeler arasında nokta eşleme

Farklı kameraların görüntüleri arasında otomatik nokta eşleme üzerine çalıştık. Bu aşamada, perspektif kameralar için kullanılan yöntemler doğrudan kullanılmadığından karma kameralı sistemler için genellikle elle eşleştirme yapılmıştır (Sturm, 2002; Ramalingam ve diğ. 2004). Puig ve diğ. (2008) tümyönlü görüntüleri panoramik görüntüye çevirerek perspektif görüntülerle SIFT nokta eşleştirmesi yapmışlar ve performansın yükseldiğini gözlemişlerdir.

Biz de çalışmalarımızda literatürde sık kullanılan ve etkin bir yöntem olan SIFT'i (Scale Invariant Feature Transform, Lowe 2004) kullandık. Şekil 5.2'de



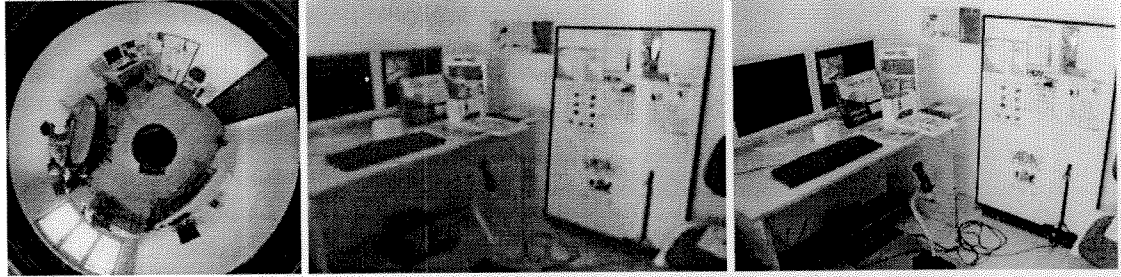
Şekil 5.2: Karma imgeler için deney kurulumu

resmedildiği gibi bir laboratuvar ortamı oluşturduk ve belirtilen noktalardan alınan perspektif ve tümyönlü görüntülerle deneyler yaptık. Tümyönlü-Perspektif SIFT eşlemesine ek olarak tümyönlü imgelerden VCP (virtual camera plane) görüntüleri elde edilmesi ve VCP-Perspective SIFT eşlemesini yapılmasını da araştırdık. VCP imgeler, yaklaşık kalibrasyon bilgisi ile tümyönlü imgelerden elde edilebilen sanal perspektif imgelerdir. Eğer kullanılan ayna tipi parabolik ise, ki bu durumda sistemi kısaca para-katadioptrik olarak adlandırabiliriz, kalibre edilmemiş tümyönlü imgeden çevrim daha da kolaydır. Şekil 5.3'te bir para-katadioptrik imge, ondan türetilmiş bir VCP imge ve aynı bölgeyi görüntüleyen bir perspektif imge örnek olarak verilmektedir.

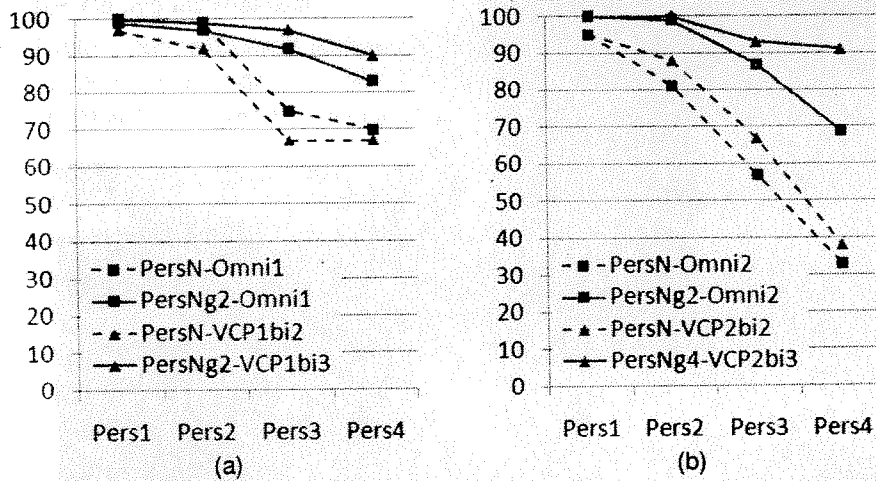
Tümyönlü-Perspektif SIFT eşlemesinin performansı ile ilgili birincil bulgumuz eğer perspektif imge bulanıklaştırılırsa SIFT performansının önemli ölçüde artması oldu. Şekil 5.4'te düz ve kesikli çizgileri karşılaştırınız. Bulanıklaştırmayı Gauss filtresi ile yaptık ve miktarını perspektif imgelerdeki detayları tümyönlü imgedeki seviyeye kadar kaybedecek şekilde ayarladık.

İkinci önemli bulgumuz VCP-Perspektif eşlemesinin artan kameralar arası mesafeyle beraber Tümyönlü-Perspektif eşlemesine göre daha başarılı kalması oldu. Şekil 5.4'te Pers1'den Pers4'e kadar kameralar arası mesafe (baseline) artmaktadır, düz üçgenli çizgi ile düz kareli çizgiyi karşılaştırınız.

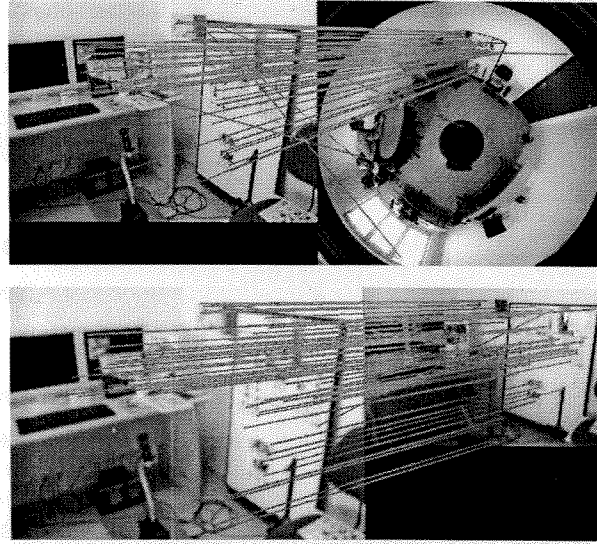
Yüksek oranda başarılı eşleştirmeler sonunda halen kalmış olan birkaç yanlış eşlemeyi elemek için karma epiolar geometriye dayalı RANSAC uygulanabilir.



Şekil 5.3: Örnek bir para-katadioptrik imge, ondan türetilmiş bir VCP imge ve aynı bölgeyi görüntüleyen bir perspektif imge.



Şekil 5.4: Omni1 (a) and Omni2 (b) tümyönlü imgeleri için farklı perspektif kameralar ile doğru/toplam eşleme oranları (yüzde olarak). PersNg2 ve PersNg4 bulanıklaştırılmış perspektif imgeler, sırasıyla $\sigma=2$ ve $\sigma=4$ Gauss filtresi ile.



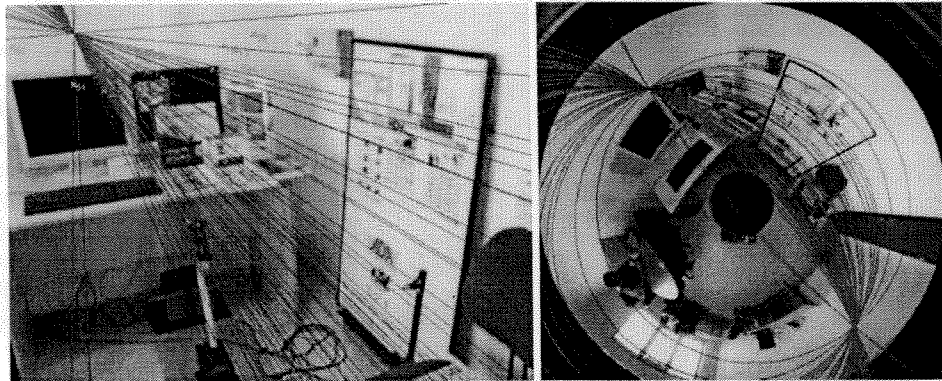
Şekil 5.5: Örnek Tümyönlü-Perspektif (üstte) ve VCP-Perspektif (altta) eşlemeler

5.3 Gürbüz epipolar geometri (Temel Matris F) çıkarımı

Literatürde karma kameralı sistemler için F matrisinin çıkarımını matematiksel olarak tanımlamış ancak gürbüz çıkarım için gerekli tekniklerin karma sistemlerde nasıl uygulanacağını tartışılmamıştır. Biz gürbüz çıkarım için başlıca iki nokta üzerine çalıştık:

1) Nokta koordinatlarının normalizasyonu

- Normalizasyon koordinatlar yükseltmeden önce veya yükseltildikten sonra uygulanabilir. İki türlü de olumlu sonuçlar aldık.
- Orijini noktaların ağırlık merkezine taşımak ve bu orijinden noktaların uzaklıklarının ortalamasını $\sqrt{2}$ olacak şekilde koordinatların ölçeğini ayarlamak uygun bir normalizasyon. Perspektif kameralar için de normalizasyon da ağırlıklı olarak bu şekilde uygulanmaktadır.
- Tablo 1 normalizasyon sonrası hata değerlerinin nasıl düştüğünü göstermektedir. Hata nokta-epipolar doğru ve nokta-epipolar-konik mesafelerinin toplamı olarak tanımlanmıştır (piksel cinsinden). Bu doğru ve koniklere örnek Şekil 5.6'da verilmiştir.



Şekil 5.6: Örnek karma epipolar geometri, epipolar doğru ve konikler

Tablo 5.1: F matrisinin farklı normalizasyon durumları ve farklı sayıda nokta ile hesaplandığında mesafe hatası ortanca değerleri. Hata, nokta-epipolar doğru ve nokta-epipolar konik mesafe toplamı olarak piksel cinsinden verilmiştir. Nokta seti değiştirilerek 30 defa tekrarlanmış ortalaması alınmıştır.

	12 nokta ile	24 nokta ile	36 nokta ile
Normalizasyon yok	239	216	193
Sadece ölçek normalizasyonu var	163	17.5	16.5
Ölçek ve orijin normalizasyonu var	10.3	2.25	2.13

2) F matrisinin doğrusal olmayan eniyilemesi için maliyet fonksiyonunun değiştirilmesi

Normal olarak mesafe üzerinden maliyet fonksiyonu nokta-epipolar doğru (p2l) + nokta-epipolar-konik (p2c) olarak tanımlanabilir. Eğer bu fonksiyon sadece p2l veya sadece p2c olursa eniyilemenin daha iyi netice verip vermeyeceğini araştırdık. Tablo 5.2'de görülebileceği gibi gerek epipolar çizgi hatası gerekse epipol pozisyon hatasında bir azalma gözlemleyemedik.

Tablo 5.2: Değişen maliyet fonksiyonu (p2l + p2c, p2l ya da p2c) için elde edilen sonuç hatalar. $\sigma=1$ piksel gürültülü, 20 nokta ile F kestirimli.

Sonuç hataları ↓	Maliyet fonksiyonu →	p2l + p2c	p2l	p2c
Nokta-epipolar doğru hatası		0.821	0.784	0.919
Nokta-epipolar konik hatası		0.770	0.798	0.753
Epipol konum hatası		22.65	25.93	23.13

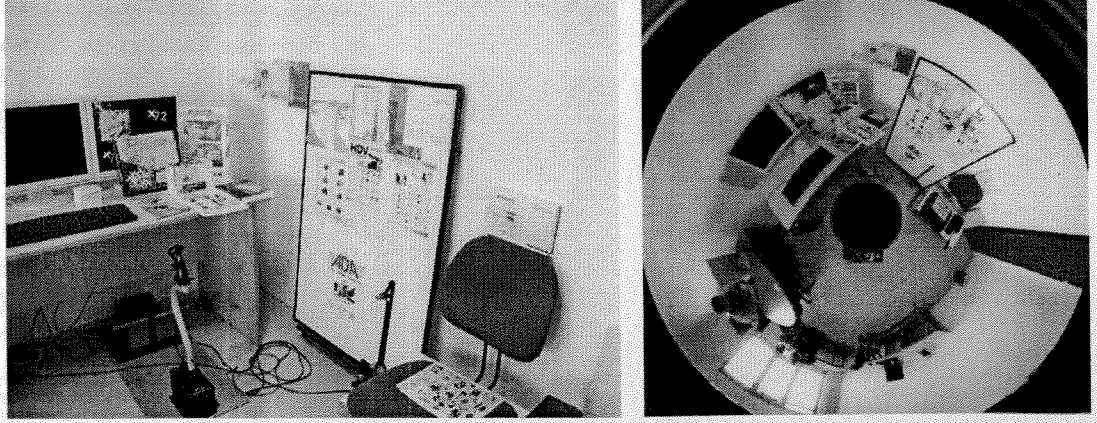
Dönme matrisi (R) ve öteleme vektörünü (t) bulmak için öncelikle E matrisini (Essential matrix) elde etmeliyiz. E matrisi kamera parametreleri kullanılarak F matrisinden elde edilebilir. Diğer bir yol da yine kalibrasyon bilgisi ile nokta koordinatlarından kalibre edilmiş ışınlara geçmek ve doğrudan E'yi kestirmektir. E bulunduktan sonra zaten bilinen yöntemlerle (Hartley ve Zisserman, 2000) R ve t elde edilir.

5.4 Üçgenleme

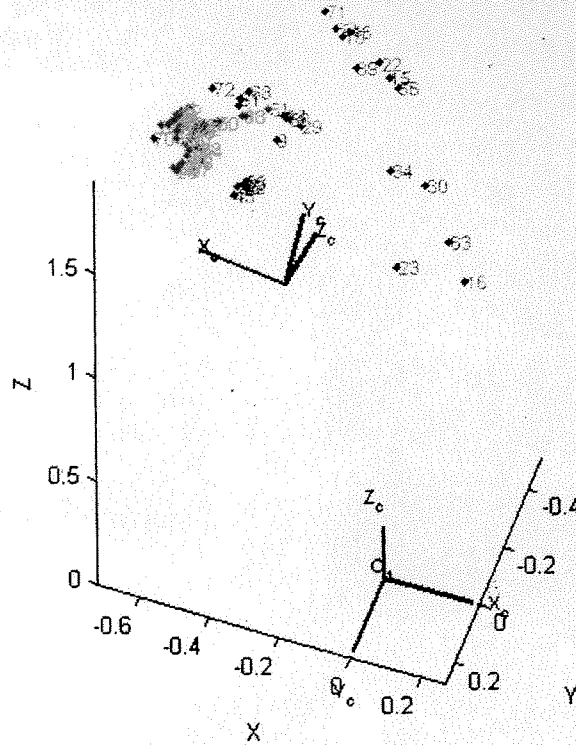
Doğrusal üçgenleme metodu için bir ağırlıklandırma stratejisi geliştirip, noktaların 3B pozisyon kestirimlerini iyileştirmeye çalıştık. Eğer belirli bir imgeden gelecek bilgilerin güvenilirliği daha fazla ise o imgeden gelecek denklemlerin ağırlığını artırarak çözümünüze katkısını artırabilirsiniz. Bizim durumumuzda perspektif imgelerin çözünürlüğü daha yüksek olduğundan o imgeden gelen denklemlerin ağırlığını artırdık. Simüle edilmiş ortamda yaptığımız deneyler sonucu, perspektif imgeden gelen denklemlerin ağırlığını 2-4 kat artırdığımızda eşit ağırlıklandırılmış duruma göre hatanın %14 dolayında azaldığını gözlemledik. Yinelemeli doğrusal üçgenleme için ise azalma %7 dolayında. Okuyucu üçgenleme metodları hakkında ayrıntılı bilgi almak için Hartley ve Sturm (1997)'u inceleyebilir.

5.5 Hareketten yapı çıkarımı

Bir tmynl bir de perspektif imge iin batan sona 'hareketten yapı ıkarımı' basamaklarını birletirdik. Œekil 5.2'deki Pers3 ve Omni1 imgeleri iin (Œekil 5.7'de gsterilmiŒtir), VCP yntemiyle nokta eŒleŒtirmesi yaparak bulunan 75 nokta epipolar geometri tabanlı RANSAC ile elenip 72'ye dŒrldkten sonra bu noktalar ile elde edilen E matrisinden hareket parametreleri ıkarılmıŒ ve yinelemeli dođrusal genleme ile noktaların 3B koordinatları kestirilmiŒtir. Kestirilen noktalar ve kameraların pozisyonları Œekil 5.8'de gsterilmektedir.



Œekil 5.7: Pers3 ve Omni1 (Œekil 5.2) kameralarının imgeleri ve zerlerinde tespit edilmiŒ eŒlenik noktalar



Œekil 5.8: Kestirilen noktalar ve kameraların konumları

Proje süresinde yetişmemekle beraber çalışmanın devamında 'karma kameralar ile 3B geriçatımı' tamamlamak için ele alınacak konular:

- Toplu düzenleme (Bundle adjustment)

Kamera sayısı arttığında önemli olan bu adım için, perspektif kameralar için sıkça kullanılmaya başlanan Seyrek Toplu Düzenleme (Lourakis ve Argyros, 2004) metodunu karma kameralar için gerçekleştirmeyi planlıyoruz.

- Yoğun nokta derinlik kestirimi (Dense depth estimation)

İmgedeki bütün piksellerin derinliklerini çıkarmak ve 3B yüzeyi oluşturarak üzerine renk/doku bilgisini eklemek için hem perspektif hem de tümyönlü kameralar ile kullanılmış yöntemler mevcuttur. Bu yöntemlerin karma kameralar için uygulanabilirliği araştırılacak ve gerekli noktalarda yöntemler modifiye edilecektir.

6. Sonuç ve Öneriler

Projede arkeolojik alanların sunumu odaklı olmak üzere Web tabanlı bir sanal gezi sistemi geliştirilmesine çalışılmış ve bir coğrafi bilgi sistemine (CBS) entegre ederek pilot bir uygulama hazırlanmıştır.

Çok sayıda ve yüksek çözünürlüklü imgeden 3 boyutlu geriçatım çalışmalarımız sonucunda imge-tabanlı ve doğruluk derecesi yüksek modeller erişilebilir hesaplama gücü ve kaynakları ile oluşturulmuştur. Arkeolojik alanlar kolay edinilebilir teçhizatlar ile görüntülenmiş ve 3 boyutlu yapı stereo teknikler kullanılarak bu alınan görüntülerden elde edilmiştir. Doku bilgisi de eklenerek 3B model elde edilmiştir. Bu yaklaşım bütünüyle, pahalı ve zahmetli olan 3B tarayıcı teknolojisine bir alternatif olarak görülmelidir. İleriye dönük amaçlarımız arasında geriçatımın çeşitli safhalarında kullanılan algoritmaların tek bir çatı altında kodlanıp bir fotoğrafı alıp 3B modeli veren bir programa dönüştürülmesi sayılabilir.

Ayrıca, tümyönlü kameraların da normal kameralar ile beraber 3B geriçatım için kullanılmasını araştırdık. Proje araştırmacılarından Yalın Baştanlar'ın doktora tezi olarak devam eden bu araştırma, proje süresinde 'hareketten yapı çıkarımı' safhasına kadar ilerlemiş ancak 3B model oluşturma henüz yapılamamıştır. Çalışmanın devamında 3B modeli de elde ederek geriçatımı tamamlamayı planlıyoruz.

Geriçatımı yapılan 3B modeller ve panoramik imge tabanlı sanal turlar Google Earth™ ile entegre edilip arkeolojik alan plan/krokileri de altlık olarak aynı ortama eklenmiştir. Bu projeden elde edilen bilgi birikiminin ileride CBS ile entegre bir arkeolojik görsel arşiv oluşmasına katkıda bulunarak bu tarz uygulamalar için bir taslak oluşturabilir.

Proje ortağımız olan ITI'nın kendi ülkelerinde sorumlu oldukları kuruma (GSRT) verdikleri rapor TÜBİTAK'ın uyguladığı yaklaşıma paralel olarak anadilde, yani Yunanca hazırlanmıştır. Eğer istenirse, proje web sayfamızda yayınlar kısmında görülebilir.

Referanslar

Bastanlar, Y., User Behaviour in Web-based Interactive Virtual Tours, In: *Proc. of 29th International Conference on Information Technology Interfaces*, Dubrovnik, Croatia, (2007).

Bastanlar, Y., Puig, L., Sturm, P., Guerrero, J., Barreto, J., DLT-like Calibration of Central Catadioptric Cameras, *Proc. of Workshop on Omnidirectional Vision*, (2008).

- Beardsley, P., Zisserman, A., Murray, D., Sequential Updating of Projective and Affine Structure from Motion, *IJCV*, 23(3), pp.235-259, 1997.
- Bouquet, J., *Camera Calibration Toolbox for Matlab*, (2007).
http://www.vision.caltech.edu/bouquet/calib_doc/ [28/02/2008]
- Bunschoten, R., Kröse, B., Robust scene reconstruction from an omnidirectional vision system, *IEEE Transactions on Robotics and Automation*, 19(2):351-357, (2003).
- Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R., 2001. Reconstruction and representation of 3D objects with radial basis functions, *Proc. SIGGRAPH*, (2001), pp.67-76.
- Chen, D., Yang, J., Calibration of a Hybrid Camera Network, *Proc. of IEEE International Conference on Computer Vision* (2003).
- Conforti Andreoni, D., Pinto, L., The Creation of The Digital Models for The Protection of Cultural Heritage: The Baptistery of Cremona, *ISPRS Comm. V Sym.*, (2004).
- Fischler, M.A., Bolles, R.C., Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, *Communications of the ACM*, vol.24, no.6, (1981).
- Fleck, S., Busch, F., Biber, P., Andreasson, H. and Strasser, W., Omnidirectional 3D Modeling on a Mobile Robot using Graph Cuts, *Proc. of IEEE International Conference on Robotics and Automation* (2005).
- Geyer, C., Daniilidis, K., Equivalence of Catadioptric Projections and Mappings of the Sphere. *Proc. of Workshop on Omnidirectional Vision*, (2000).
- Guarnieri, A., Vettore, A., El-Hakim, S., Gonzo, L., Digital Photogrammetry and Laser Scanning In Cultural Heritage Survey, *ISPRS Comm. V Sym.*, (2004).
- Grün, A., Remondino, F., Zhang, L., Reconstruction of the Great Buddha of Bamiyan, Afghanistan, *International Archives of Photogrammetry and Remote Sensing*, ISPRS (2002), 34(5), pp. 363-368.
- Harris, C.G., Stephens, M., A Combined Corner and Edge Detector, *Proc. of Fourth Alley Vision Conference*, Manchester, U.K., (1988), p.182-192.
- Hartley, R., Sturm, P., Triangulation, *Computer Vision and Image Understanding*, vol.68, no.2, pp.146-157, (1997).
- Hartley, R., Zisserman, A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, New York, (2000).
- Kadobayashi R., Kochi, N., Otani, H. Furukawa, R., Comparison and Evaluation of Laser Scanning and Photogrammetry and Their Combined Use for Digital Recording of Cultural Heritage, *ISPRS Comm. V Sym.*, (2004).
- Kannala, J., Brandt, S., A Generic Camera Calibration Method for Fish-Eye Lenses, *Proc. of International Conference on Pattern Recognition*, (2004).
- Lempitsky, V., Ivanov, D., Seamless Mosaicing of Image-Based Texture Maps, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, (2007).
- Lhuillier, M., Toward Flexible 3D Modeling using a Catadioptric Camera, *IEEE International Conference on Computer Vision and Pattern Recognition* (2007).
- Li, M., Magnor, M., Seidel, H.P., Hardware-accelerated rendering of photo hulls, *Eurographics*, 23(3), (2004).
- Lourakis, M., Argyros, A., The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm, TR #340, ICS-FORTH, (2004).

- Lowe, D., Distinctive image features from scale-invariant keypoints, *IJCV*, 60(2), pp.91-110, (2004).
- Mei, C., Rives, P., Single viewpoint omnidirectional camera calibration from planar grids. *Proc. of International Conference on Pattern Recognition*, (2007).
- Mordohai, P., Frahm, J., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., Nister, D., Real-time Video-based Reconstruction of Urban Environments, *Proc. of 3D-ARCH 2007 Workshop, 3D Virtual Reconstruction and Visualization of Complex Architectures*, ETH Zurich, Switzerland, (2007).
- Mulligan, J., Zabulis, X., Kelshikar, N., Daniilidis, K., Stereo-based Environment Scanning for Immersive Telepresence, *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3):304-320, (2004).
- Ohtake, Y., Belyaev, A., Seidel, H.P., An Integrating Approach to Meshing Scattered Point Data, *Proceedings of the ACM symposium on Solid and physical modelling*, Cambridge, Massachusetts, pp. 61 – 69., (2005).
- Pollefeys, M., Koch, R., Vergauwen, M., van Gool, L., Hand-held acquisition of 3D models with a video camera, In: *Proc. of the International Conference on 3-D Digital Imaging and Modeling* (1999).
- Pollefeys, M., van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R., Visual modeling with a hand-held camera, *IJCV*, 59(3), pp.207-232, (2004).
- Puig, L., Guerrero, J., Sturm, P., Matching of omnidirectional and perspective images using the hybrid fundamental matrix, *Proc. of Workshop on Omnidirectional Vision* (2008).
- Ramalingam, S., Lodha, S.K., Sturm, P., A generic structure-from-motion algorithm for cross-camera scenarios, *Proc. of IEEE Workshop on Omnidirectional Vision*, (2004).
- Scaramuzza, D., Martinelli, A., Siegwart, R., A Toolbox for Easily Calibrating Omnidirectional Cameras, *Proc. of Int. Conf. on Intelligent Robots and Systems*, (2006).
- Sturm, P., Mixing Catadioptric and Perspective Cameras, *Proc. of IEEE Workshop on Omnidirectional Vision*, (2002).
- Tola, E., *Multi-view 3D Reconstruction of a Scene Containing Independently Moving Objects*, Y.L. Tezi, Middle East Technical University, Ankara, Turkey, (2005).
- Villaneuva, R., Moore, A., Wong, B.L.W., Usability Evaluation of Non-immersive, Desktop, Photo-realistic Virtual Environments. In: *The 16th Annual Colloquium of the Spatial Information Research Centre*, (2004).
- Yang, R., Welch, G., Bishop, G., Real-time consensus-based scene reconstruction using commodity graphics hardware, *Proc. of Pacific Graphics*, Beijing, China, (2002).
- Zabulis, X., Patterson, A., Daniilidis, K., Digitizing Archaeological Excavations from Multiple Views. In: *Proc. of the International Conference on 3-D Digital Imaging and Modeling* (2003).
- Zabulis, X., Kordelas, G., Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo, *Proc. of 3DPVT*, (2006).
- Zabulis, X., Kordelas, G., Mueller, K., Smolic, A., Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces, *Proc. of ICIP*, Atlanta, USA, (2006).
- Zabulis, X., Utilization of the texture uniqueness cue in stereo. In *Three-Dimensional Television: Capture, Transmission, and Display*, Springer Verlag, (2007).

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje No: 105E187
Proje Başlığı: Ege Bölgesi'nin Kültürel Zenginlikleri için 3 Boyutlu Geriçatım ve Web-tabanlı Sanal Gezi
Proje Yürütücüsü ve Araştırmacılar: Prof. Dr. Yasemin Yardımcı, Yalın Baştanlar
Projenin Yürütüldüğü Kuruluş ve Adresi: ODTÜ Enformatik Enstitüsü, 06531, ODTÜ, Ankara
Destekleyen Kuruluş(ların) Adı ve Adresi: --
Projenin Başlangıç ve Bitiş Tarihleri: 01.11.2006 - 31.10.2008
Öz (en çok 70 kelime): Projenin amacı arkeolojik alanların sunumu odaklı olmak üzere Web tabanlı bir sanal gezi sistemi geliştirilmesidir. Arkeolojik alanlar kolay edinilebilir teçhizatlar kullanılarak görüntülenmiş ve 3 boyutlu yapı stereo teknikler kullanılarak bu alınan görüntülerden elde edilmiştir. Doku bilgisi de eklendikten sonra sonuç geriçatım 3B modele çevrilmiştir. 3B modeller, alan krokileri ve panoramik imge tabanlı sanal turlar ile birlikte Google Earth™ ile entegre edilerek ziyaret edilmeleri sağlanmıştır.
Anahtar Kelimeler: 3 Boyutlu geriçatım, Kültürel miras, Tümyönlü görüntüleme, Sanal gezi, Coğrafi bilgi sistemi
Projeden Yapılan Yayınlar: <ul style="list-style-type: none">- Yalın Baştanlar, Nikos Grammalidis, Xenophon Zabulis, Erdal Yılmaz, Yasemin Yardımcı, , Georgios Triantafyllidis, "3D Reconstruction for a Cultural Heritage Virtual Tour System", ISPRS 2008, http://www.isprs2008-beijing.org- Xenophon Zabulis, Nikos Grammalidis, Yalın Baştanlar, Erdal Yılmaz, Yasemin Yardımcı, "3D Scene Reconstruction Based on Robust Camera Motion Estimation and Space Sweeping", 3DTV-Conference: The True Vision: Capture, Transmission and Display of 3D Video, 28-30 May, Istanbul, Turkey.- Yalın Baştanlar, Alptekin Temizel, Yasemin Yardımcı, "3D Scene Reconstruction Based on Robust Camera Motion Estimation and Space Sweeping", ICIP 2009'da sunulmak üzere değerlendirmeye gönderilmiştir.- Nikos Grammalidis, Yalın Baştanlar, Xenophon Zabulis, Erdal Yılmaz, Yasemin Yardımcı, "Cultural Heritage Virtual Tour System with Image-based 3D Reconstruction and GIS Integration", Journal of Cultural Heritage (Elsevier)'a gönderilmek üzere hazırlık aşamasında.

3D SCENE RECONSTRUCTION BASED ON ROBUST CAMERA MOTION ESTIMATION AND SPACE SWEEPING FOR A CULTURAL HERITAGE VIRTUAL TOUR SYSTEM

Xenophon Zabulis¹, Nikos Grammalidis², Yalin Bastanlar³, Erdal Yilmaz³, Yasemin Yardimci Cetin³

Foundation for Research and Technology – Hellas, Institute of Computer Science, Heraklion, Greece¹
Informatics and Telematics Institute-CERTH, Thessaloniki, Greece²
Informatics Institute, Middle East Technical University, Ankara, Turkey³

ABSTRACT

An efficient 3D reconstruction technique based on robust camera motion estimation and an improved version of the space-sweeping stereo reconstruction approach is presented. The proposed approach is focused on generation of usable and fully automatic reconstruction of wide-area scenes with the computational resources of a conventional PC. The aim is to use this technique to capture such scenes in 3D utilizing off-the-shelf equipment. Color information is finally added to the derived 3D model of the scene and the result can be converted to common 3D scene modeling formats. The 3D models are integrated with GIS technologies within a web-based virtual tour system.

Index Terms—3D reconstruction, camera motion estimation, cultural heritage

1. INTRODUCTION

In this paper, automatic and photorealistic 3D scene reconstruction from images is used to create content for a cultural heritage virtual tour system. The application supports virtual walkthrough and animated events within photorealistic 3D models of the archaeological sites and also, registers these models onto the Google Earth application. The aimed contribution is a fully automatic procedure for accurate, wide-area, and photorealistic scene modeling that is efficient in terms of the computational resources utilized and may be used for 3DTV content creation.

As is the case for most multiview stereo reconstruction techniques, the accuracy of the final results greatly depends on the quality of both camera calibration and motion estimation. To efficiently tackle the problem of fully-automatic motion estimation, the proposed approach employs state-of-the-art techniques [2],[3] and a posteriori accuracy improvement through bundle adjustment [6]. In addition, the features derived from the Scale-Invariant Feature Transform (SIFT) [5] were investigated as to their performance against the, traditional, Harris features [18].

Computational efficiency is achieved by introducing an extension to the space-sweeping stereo reconstruction

approach. This approach is frequently used for multiview stereo reconstruction, due to its computational efficiency and its straightforward acceleration by graphics hardware [11][13]. However, it is less accurate than other approaches that account for the projective distortion due to the orientation of the imaged surface [14]. For this reason, approaches that employ sweeping in multiple directions [11] or refine an initial estimation obtained by space-sweeping have been proposed [17].

In the context of wide-area stereo, algorithmic complexity and memory capacity are important, because such scenes require more images to be fully reconstructed. Moreover, the need for memory conservation is reinforced by the recent growth of GPU-based software acceleration. The reason is that state-of-the-art graphics hardware has even less on-board memory than a conventional PC.

Usability was a primary concern when designing our system. The system offers to the user the ability to reconstruct a scene from a few snapshots acquired with an off-the-shelf camera, preferably of high resolution. This way, a few snapshots suffice for the reconstruction and the image acquisition process becomes much simpler than capturing the scene with a video camera or with a multicamera apparatus [11].

Under these conditions, the fully automatic multi-view reconstruction of a scene is not straightforward and, thus, a complete workpath for this task is proposed. This workpath consists of acquiring the images and then estimating the calibration and camera motion parameters as proposed in Sec. 2. The scene is then reconstructed as proposed in Sec. 3 and the final result is a textured mesh in either the Keyhole Markup Language (KML) or Virtual Reality Modeling Language (VRML) formats. The KML output allows integration to the GoogleEarth™ platform, thus the reconstructed sites can easily become a part of a large geographical information system (GIS) in the near future.

2. ROBUST CAMERA MOTION ESTIMATION BASED ON SIFT DETECTION AND MATCHING

Estimating robustly the camera motion is essential, since the accuracy of the produced 3D reconstruction is based on this information. Our work is based on the approach proposed

initially in [2] and, subsequently, extended in [3][4]. The approach establishes correspondences across consecutive images of a sequence to estimate camera motion.

Previous approaches used the Harris corner detector to extract point features in images. The matching procedure utilized similarity as well as proximity criteria [4], to avoid spurious matches. In this paper, an alternative procedure was tested, utilizing SIFT feature detection and matching [5]. In both cases (Harris/SIFT), a RANSAC framework is then utilized to remove spurious correspondences, followed by a Levenberg-Marquardt post-processing step to further improve the estimation. Intrinsic camera parameters are estimated a priori through a simple calibration procedure [15]. Besides reducing the unknowns in the following external calibration and bundle adjustment procedures, intrinsic calibration is used to compensate for radial distortion. As a result, the perspective camera model is better approximated and the system produces more accurate results. The output is an estimation of the essential matrix E , which is thereafter decomposed into view rotation and translation (R, T). Finally, triangulation is used to estimate the 3D coordinates of the corresponding features.

When a sequence of views is available, the above technique is applied for the first two views and for each new view i , the feature detection and matching approaches are applied to establish 2-D correspondences with the previous view $i-1$, which are then matched with the already established 3-D points, using a RANSAC-based technique that yields a robust estimate of the projection matrix P_i of the new view. We have used an efficient Bundle Adjustment procedure [6] as a final step at each addition of a new view. The procedure is illustrated in Fig. 1.

Although several error suppression and outlier removal steps are included, results show that the accuracy of the whole chain greatly relies on the success of the feature detection and matching. Despite the efficiency of the Harris corner detector and the neighborhood-based constraints utilized in correspondence establishment, SIFT yields better correspondences in terms of number and accuracy. Fig. 2 shows the 3D structure (as viewed from the side of the reconstruction) for a stereo pair obtained using both approaches. For this case, there were ~22% more correspondences with SIFT and they were more accurate yielding a better representation of the planar like scene).

The result above is interpreted as a requirement of wider baselines for more descriptive features. Also, while tracking Harris corners in videos provides accurate results [3], wide baselines may create problems to correlation-based matching. For our problem, robustness to large disparities or severe view angle changes is important because the scene is to be reconstructed from a few snapshots instead of a high-framerate video.

A technical issue encountered when high resolution images are utilized is that the computation of the SIFT features may require more memory than available. The

proposed treatment is to tessellate the image into blocks, compute the features independently in each, and merge the results. To avoid blocking artifacts, the blocks in the above tessellation are adequately overlapping. Duplicate features are often encountered, either due to block overlap or due to collocation of different SIFT that occur at different scales; they are all removed at the merging stage.

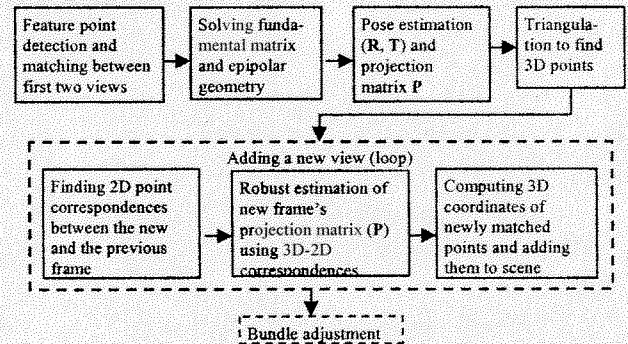


Figure 1: Illustration of the camera motion estimation procedure.

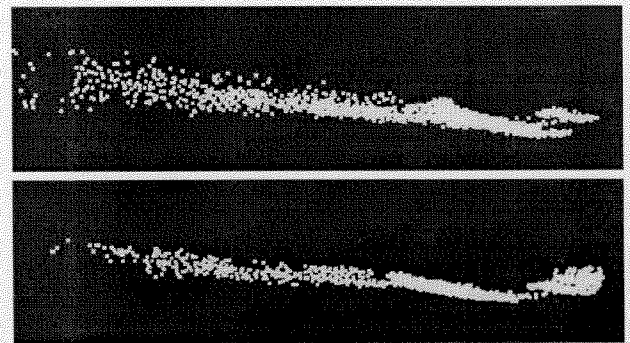


Figure 2: Comparison of the estimated structure when Harris (top) or SIFT (bottom) are used.

3. STEREO RECONSTRUCTION

In this section, an approach for reconstructing wide area-scenes from high-resolution images with the associated computational issues is proposed. In this context, a memory-conserving extension to the conventional space-sweeping approaches is proposed. Moreover, this extension facilitates the acceleration of the methods, based on a coarse-to-fine depth map computation. The importance of memory conservation is twofold. First, the memory of conventional PCs is insufficient to process high-resolution images and using virtual memory renders the process extremely slow. Second, state-of-the-art approaches to stereo reconstruction utilize the graphics hardware to process large amounts of data processing [11]. Since GPUs have even less on-board memory than PCs, RAM memory conservation is, therefore, an important property for GPU-accelerated techniques.

In the proposed approach, the space-sweeping approach is slightly modified to employ a sweeping spherical, instead of planar, backprojection surface (see [19] for an analytical

formulation). The technique provides higher reconstruction accuracy, especially in the periphery of the images (see [14] for an explanation) and, thus, the available images are more efficiently utilized. Otherwise, the sweeping procedure is similar to plane-sweeping and, for this reason, it is summarized here briefly. For each depth d_i , the images are backprojected on the, backprojection surface and locally compared. The output of this comparison is a *similarity image* S_i at each depth, whose size is equal to that of the backprojection surface. At each iteration i , the pixels in S_i are compared to their corresponding pixels in S_{i+1} and S_{i-1} . As depth increases, the values for a point in the similarity image correspond to locations along a ray of visibility from the cyclopean eye. The strongest *local* similarity maximum along each such a ray is selected as the optimum depth. The requirement for maxima to be local is used to avoid artifacts that may occur in the textureless areas of the input images.

Memory conservation is achieved by tessellating the backprojection image into, say, $k \times k$ equal spherical segments. This tessellation is parameterized along the two spherical coordinates that, also, correspond to image width and height. The sweeping algorithm is performed independently for each such partition. These partitions overlap slightly, in order to avoid "blocking artifacts" at their boundaries. The amount of overlap is exactly determined by the size of the comparison kernel so that a scene point is not reconstructed twice. Memory conservation is crucial, because even though sweeping is performed in a memory efficient manner (similar to [16]), the memory requirements are still large. The reason is that besides the buffer that stores the index of the optimal depth for each pixel in S_i , two additional such buffers are required to store S_{i+1} and S_{i-1} .

The acceleration of the space-sweeping approach is based on an iterative and coarse-to-fine approach that is combined with the above memory conservation technique. The image data in each iteration are obtained from traditional image pyramids of the input images, starting from the smallest image of the pyramid and advancing a layer in each iteration; at the last iteration the original image is utilized. Also in each iteration, the parameterization of the backprojection surface becomes denser. As described above, the backprojection surface is tessellated and the sweeping algorithm is executed independently for each segment. At each iteration, though, each spherical segment is re-segmented into $k \times k$ more segments. After the 2nd iteration, the range of evaluated depths (d_i) is drastically constrained, based on the reconstruction result previously obtained for the "parent" segment.

The obtained depth map is filtered very conservatively (as in [10]), to suppress artifacts at depth discontinuities and remove outliers. By doing so, some valid matches are indeed rejected; however, in the utilized multiview setup the corresponding points are most likely to be reconstructed from another binocular pair. The result is spatially quantized

as it is too large ($\approx 10^9$ points for 35 views of 8Mpix each, in this experiment) to fit in memory. To cope with the same limitations the merging process is performed volumetrically, by tessellating the reconstruction volume into cubical segments. Finally, a thin plate interpolating surface is fit [9], to yield a mesh outputted into the VRML or KML formats.

In Fig. 3, the proposed method is demonstrated. In the experiments presented in this paper, images were 2448 x 3264, 16-bit per layer, color images acquired with a Canon Powershot SLR camera, the number of iterations was 5 and the initial tessellation was 3 x 3. The coarse-to-fine refinement factor was 2, so that in each iteration: (a) the image rows and columns of the stereo and backprojection images were doubled and (b) the number of segments was increased by 4. The above scheme was measured to provide a speedup of ~ 50 for the scene of this experiment.

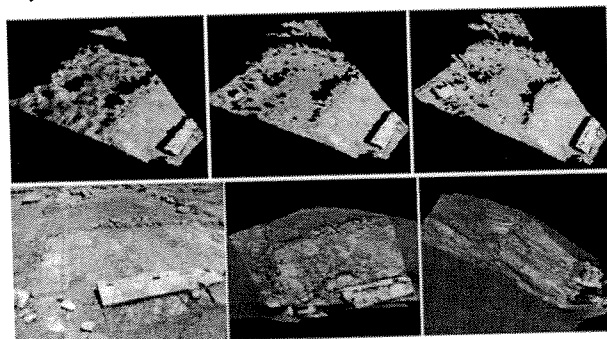


Figure 3: Coarse-to-fine acceleration scheme, for space-sweeping methods. Top row shows the reconstructions for the 3 first iterations of the proposed procedure. Bottom row shows original image from a ~ 40 cm baseline stereo pair (left) and two views of the RBF interpolated reconstruction with texture mapping.

Fig. 4 shows the result of an experiment that compares the reconstructions obtained from the proposed method in Harris and SIFT conditions of the previous section. The images in the first 2 rows shows the result of the reconstruction for an early frame (20 views): in the SIFT condition, a larger proportion of the scene is reconstructed ($\sim 18\%$, in this experiment). In addition, as already observed in Fig. 2, in the SIFT condition the resulting reconstruction is less noisy. Therefore, the robustness of camera motion estimation provided by SIFT features has a direct impact on the quality of the reconstruction. The last row, shows the result of the SIFT condition after 35 frames.

4. VIRTUAL TOUR APPLICATION-FUTURE WORK

The reconstructed VRML models are integrated with GIS technologies within a web-based virtual tour system, after first converting them to the XML-based Collada 3D file format, and then referencing to them in Keyhole Markup Language (KML), a format supported by the GoogleEarth™ GIS platform. Excavation site plans are used as detailed raster overlays, draped over terrain at the exact locations on the earth. When the reconstructed archaeological site is

placed on top of the site plan, users view the reconstruction together with the site plan. Additional information (e.g. site-related audio or text) can be presented to users via hot-spots. We added a hyperlink to the application described above, which directs users to a panoramic image based virtual-tour. Using a map of the archaeological site increases the comprehension of the tour and enhances the user's sense of orientation. With such tools, more information is communicated to the virtual tour users in an ergonomic fashion [8]. A larger level of immersiveness can be experienced by viewing the reconstructed 3D models on autostereoscopic displays, which can be achieved by using a special plug-in, TriDef™ Visualizer for GoogleEarth™, to render real-time 3D scenes.

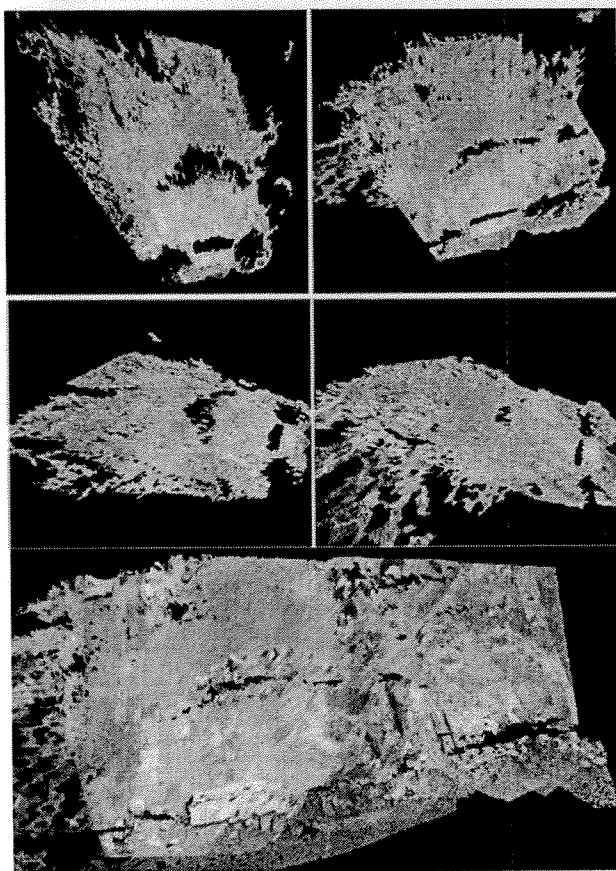


Figure 4: Comparing the reconstructions for the 2 conditions of the experiment in Sec. 3. Top 2 rows show the results for the Harris (left) and the SIFT (right) conditions, for two viewpoints (1st row is top view and 2nd row is side view). The last row of the figure shows a top view of the reconstruction of the same scene, from 35 views and the SIFT condition.

Once the scene has been extracted with reasonable accuracy, it can be passed as input to a more sophisticated, but also more computationally complex approach (e.g. [1],[3],[7]). The initial reconstruction of the scene is important in order to restrict the set of possible solutions

and, thus, significantly reduce the computational time. The high-accuracy reconstruction results could then facilitate scientific measurements on the 3D model (instead of the actual site) by archaeologists.

ACKNOWLEDGEMENT

The authors are grateful for support through the 3DTV European NoE, FP6 IST Programme and TUBITAK-GSRT Joint Research Project (105E187). They also thank Engin Tola, for providing the implementation of his structure-from-motion technique.

REFERENCES

- [1] X. Zabulis, A. Patterson, K. Daniilidis, *Digitizing Archaeological Excavations from Multiple Views*, 3DIM, 2003.
- [2] P. Beardsley, A. Zisserman, D. Murray, *Sequential Updating of Projective and Affine Structure from Motion*, IJCV, 23(3):235-259, 1997.
- [3] M. Pollefeys, R. Koch, M. Vergauwen, L. van Gool, *Hand-held acquisition of 3D models with a video camera*, 3DIM, 1999.
- [4] E. Tola, *Multi-view 3D Reconstruction of a Scene Containing Independently Moving Objects*, M.Sc. Thesis, Middle East Technical University, Ankara, Turkey, 2005.
- [5] D. Lowe, *Distinctive image features from scale-invariant keypoints*, IJCV, 60(2):91-110, 2004.
- [6] M. Lourakis, A. Argyros, *The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm*, TR #340, ICS-FORTH, 2004.
- [7] M. Pollefeys, L. van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, R. Koch, *Visual modeling with a hand-held camera*, IJCV, 59(3): 207-232, 2004.
- [8] Y. Bastanlar, *User Behaviour in Web-based Interactive Virtual Tours*, 29th International Conference on Information Technology Interfaces, p. 25-28 June, Dubrovnik, Croatia.
- [9] J.C. Carr, R.K. Beatson, J.B. Cherrie, T.J. Mitchell, W.R. Fright, B.C. McCallum, T. R. Evans, *Reconstruction and representation of 3D objects with radial basis functions*, Proc. SIGGRAPH, pp.67-76, 2001.
- [10] J. Mulligan, X. Zabulis, N. Kelshikar, and K. Daniilidis, *Stereo-based Environment Scanning for Immersive Telepresence*, IEEE Transactions on Circuits and Systems for Video Technology, 14(3):304-320, 2004.
- [11] P. Mordohai, J.Frahm, A. Akbarzadeh, B. Clipp, C.Engels, D. Gallup, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewenius, H. Towles, G. Welch, R. Yang, M. Pollefeys and D. Nister, *Real-time Video-based Reconstruction of Urban Environments*, 3D-ARCH 2007.
- [12] M. Li, M. Magnor, H. P. Seidel, *Hardware-accelerated rendering of photo hulls*, Eurographics, 23(3), 2004.
- [13] R. Yang, G. Welch, G. Bishop, *Real-time consensus-based scene reconstruction using commodity graphics hardware*, Pacific Graphics 02.
- [14] X. Zabulis, *Utilization of the texture uniqueness cue in stereo*. In Three-Dimensional Television: Capture, Transmission, and Display, Springer Verlag, 2007
- [15] J. Bouguet, *Camera Calibration Toolbox for Matlab*, http://www.vision.caltech.edu/bouguetj/calib_doc/
- [16] R. Collins, *A Space-Sweep Approach to True Multi-Image Matching*, CVPR, p. 358-363, 1996.
- [17] X. Zabulis, G. Kordelas, *Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo*, 3DPVT, 2006.
- [18] C.G. Harris, M. Stephens, *A Combined Corner and Edge Detector*, Proc. of Fourth Alley Vision Conference, Manchester, 1988, p.182-192.
- [19] X. Zabulis, G. Kordelas, K. Mueller, A. Smolic, "Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces", ICIP, Atlanta GA, 8-11 October 2006.

3D RECONSTRUCTION FOR A CULTURAL HERITAGE VIRTUAL TOUR SYSTEM

Y. Bastanlar^{a,*}, N. Grammalidis^b, X. Zabulis^c, E. Yilmaz^a, Y. Yardimci^a, G. Triantafyllidis^b

^a Informatics Institute, Middle East Technical University, Ankara, Turkey - (yalinb, eyilmaz, yardimy)@ii.metu.edu.tr

^b Informatics and Telematics Institute, CERTH, Thessaloniki, Greece - (ngramm, gatrian)@iti.gr

^c Foundation for Research and Technology - Hellas, Inst. of Computer Science, Heraklion, Greece - zabulis@ics.forth.gr

KEY WORDS: Cultural Heritage, Reconstruction, Web-based, Visualization, GIS, Virtual Reality

ABSTRACT:

The aim of this study is to build a Web-based virtual tour system, focused at the presentation of archaeological sites. The proposed approach is comprised of powerful techniques such as multiview 3D reconstruction, omnidirectional viewing based on panoramic images, and their integration with GIS technologies. In the proposed method, the scene is captured from multiple viewpoints utilizing off-the-shelf equipment and its 3D structure is extracted from the acquired images based on stereoscopic techniques. Color information is added to the generated 3D model of the scene and the result is converted to a common 3D scene modeling format. The 3D models and interactive virtual tour tools such as 360° viewing are integrated with GIS technologies in which the excavation site plans can be added as detailed raster overlays.

1. INTRODUCTION

Web-based virtual tour applications constructed by 360° panoramic images are started to being used extensively all over the world. Effectiveness and usability of these tours were discussed by Bastanlar (2007) and Villaneuva et al. (2004). Usage of image-based 3D reconstructions in virtual tours is limited due to their present day lower quality is not appealing yet. Cultural heritage is one of the most important application areas of these technologies. Example studies (Guarnieri et al., 2004; Kadobayashi et al., 2004; Conforti Andreoni and Pinto, 2004) on 3D reconstruction of cultural heritage were performed by merging image data with the output of 3D laser scanner. 3D scanner technology is efficiently developed to scan the environment and add color information to generate the 3D model. However, the necessary equipment is still very expensive and capturing the 3D data and post-processing is very time-consuming.

In this paper, automatic and photorealistic 3D scene reconstruction from images is used to create content for a cultural heritage virtual tour system. With the same aim, Grün et al. (2002) worked to generate 3D reconstruction of a demolished Buddha statue. However the 3D model is not impressive due to the usage of limited number and low quality photographs. Better results were obtained by Pollefeys et al. (1999) who used the recordings by a video camera. Later they applied their technique for the Sagalassos archaeological site (Pollefeys et al., 2004).

The 3D scene could be synthesized, e.g. by a 3D modeller, by performing surface modelling and then adding texture information. Some of the current WWW applications are composed of graphical textures which are displayed via a VRML plug-in. The problem with such synthesized 3D models for cultural applications is that the feeling of reality is lost and the procedure to generate them is tedious and requires highly-experienced personnel.

An automatic procedure for accurate and photorealistic scene modeling that is efficient in terms of the computational resources is not straightforward. We propose a pipeline for reconstruction and presentation of archaeological sites. In short, the steps are:

- 1) Acquisition of multiple high-resolution images or video-recording and subsequent selection of key frames.
- 2) Computation of internal camera calibration parameters.
- 3) Estimation of lens distortion and image rectification.
- 4) Extrinsic calibration of the acquired images, based on robust feature extraction, tracking and camera motion estimation techniques.
- 5) Multi-view stereo reconstruction of the scene using the acquired images and intrinsic and extrinsic calibration parameters.
- 6) Conversion of the reconstruction output to textured VRML format, which includes triangulation of points into a mesh, combination of textures from different images.
- 7) Generation of KML/KMZ file from VRML format.
- 8) Display of the reconstructed portion of the archeological site with the excavation site plans as detailed raster overlays, on the Google Earth™ system or other GIS tools that support KML/KMZ format.

As is the case for most multiview stereo reconstruction techniques, the accuracy of the final results greatly depends on the quality of both camera calibration and motion estimation (Steps 2 to 4). To efficiently tackle the problem of fully-automatic motion estimation, the proposed approach employs state-of-the-art techniques (Beardsley et al., 1997; Pollefeys et al., 1999; Pollefeys et al., 2004). Custom modifications were made to these techniques to improve accuracy of the calibration results, namely, robust feature point detection and matching using SIFT (Lowe, 2004) and bundle adjustment (Lourakis and Argyros, 2004). Details are presented in Section 2.

* Corresponding author.

The scene is then reconstructed with the technique proposed in Section 3. In this section, an approach for reconstructing wide area-scenes from high-resolution images with the associated computational issues is proposed. In our technique, the conventional space-sweeping approach (e.g. Zabulis et al. 2003) is slightly modified to employ a sweeping spherical, instead of planar, back-projection surface. Result is a more accurate and memory-conserving technique. Moreover, this extension facilitates the acceleration of the methods, based on a coarse-to-fine depth map computation.

The proposed approach offers to the user the ability to reconstruct a scene from a few snapshots acquired with an off-the-shelf camera, preferably of high resolution. This way, a few snapshots suffice for the reconstruction and the image acquisition process becomes much simpler than capturing the scene with a video camera or with a multicamera apparatus (Mordohai et al., 2007).

The final result is a textured mesh in either the Keyhole Markup Language (KML) or Virtual Reality Modeling Language (VRML) formats. The KML output allows integration to the Google Earth™ platform, thus the reconstructed 3D models and their virtual walkthrough applications can easily become a part of a large geographical information system (GIS) in the near future. Section 4 explains the Web-based virtual tour application developed.

2. ROBUST CAMERA MOTION ESTIMATION BASED ON SIFT DETECTION AND MATCHING

Robust estimation of the camera motion is essential, since the accuracy of the produced 3D reconstruction is based on this information. Our work is based on the approach proposed initially by Beardsley et al. (1997), and, subsequently, extended by Pollefeys et al. (1999, 2004) and Tola (2005). The approach establishes correspondences across consecutive images of a sequence to estimate camera motion.

Previous approaches used the Harris corner detector (Harris and Stephens, 1988) to extract point features in images. The matching procedure utilized similarity as well as proximity criteria (Tola, 2005) to avoid spurious matches. In this paper, an alternative procedure was tested, utilizing SIFT feature detection and matching (Lowe, 2004). In both cases (Harris/SIFT), a RANSAC framework is then utilized to remove spurious correspondences, followed by a Levenberg-Marquardt post-processing step to further improve the estimation. Intrinsic camera parameters are estimated a priori through a simple calibration procedure (Bouguet, 2007). Besides reducing the unknowns in the following external calibration and bundle adjustment procedures, intrinsic calibration is used to compensate for radial distortion. As a result, the perspective camera model is better approximated and the system produces more accurate results. The output is an estimation of the essential matrix E , which is thereafter decomposed into rotation matrix (R) and translation vector (t) of the new view. Finally, triangulation is used to estimate the 3D coordinates of the corresponding features.

When a sequence of views is available, the above technique is applied for the first two views and for each new view i , the feature detection and matching approaches are applied to establish 2-D correspondences with the previous view $i-1$, which are then matched with the already established 3-D points,

using a RANSAC-based technique that yields a robust estimate of the projection matrix P_i of the new view. We have used an efficient Bundle Adjustment procedure (Lourakis and Argyros, 2004) as a final step at each addition of a new view. The procedure is illustrated in Fig. 1.

Although several error suppression and outlier removal steps are included, results show that the accuracy of the whole chain greatly relies on the success of the feature detection and matching. Despite the efficiency of the Harris corner detector and the neighborhood-based constraints utilized in correspondence establishment, we observed that SIFT yields better correspondences in terms of number and accuracy. This is especially important for camera positions with wider baselines. For our problem, robustness to large disparities or severe view angle changes is important because the scene is to be reconstructed from a few snapshots instead of a high-frame rate video.

A technical issue encountered when high resolution images are utilized is that the computation of the SIFT features may require more memory than available. The proposed treatment is to tessellate the image into blocks, compute the features independently in each, and merge the results. To avoid blocking artifacts, the blocks in the above tessellation are adequately overlapping. Duplicate features are often encountered, either due to block overlap or due to collocation of different SIFT that occur at different scales; they are all removed at the merging stage.

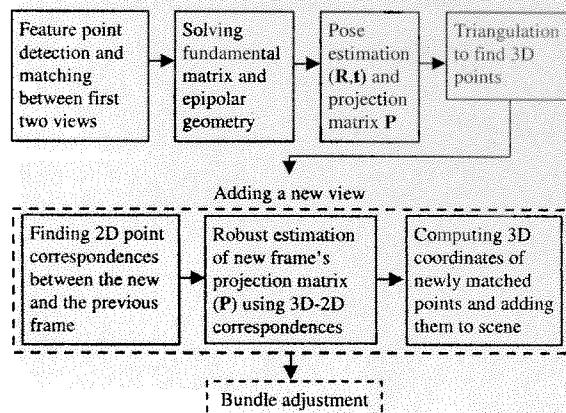


Figure 1: Illustration of the camera motion estimation procedure.

3. 3D RECONSTRUCTION

In this section, an approach for 3D scene reconstruction from high-resolution images is proposed and the associated computational issues are discussed. In the proposed method, the space-sweeping approach is slightly modified to employ a sweeping spherical, instead of planar, backprojection surface (see Zabulis, Kordelas et al. (2006) for an analytical formulation).

The conventional space-sweeping approach is frequently used for multiview stereo reconstruction, due to its computational efficiency and its straightforward acceleration by graphics hardware (Yang et al., 2002; Li et al., 2004). However, it is less accurate than other approaches that account for the projective

distortion due to the orientation of the imaged surface (Zabulis, 2007). For this reason, approaches that employ sweeping in multiple directions (Mordohai et al., 2007) or refine an initial estimation obtained by space-sweeping (Zabulis and Kordelas, 2006) have been proposed.

The proposed technique, based on spherical sweeping, provides higher reconstruction accuracy, especially in the periphery of the images (see Zabulis (2007) for an explanation) and, thus, the available images are more efficiently utilized. In addition, a memory-conserving extension is made to the conventional space-sweeping approaches. This extension also facilitates the acceleration of the methods, based on a coarse-to-fine depth map computation. The importance of memory conservation is twofold. First, the memory of conventional PCs is insufficient to process high-resolution images and using virtual memory renders the process extremely slow. Second, state-of-the-art approaches to stereo reconstruction utilize the graphics hardware to process large amounts of data processing (Mordohai et al., 2007).

The sweeping procedure, which is similar to plane-sweeping, is summarized here briefly. For each depth d_i , the images are backprojected on the, backprojection surface and locally compared. The output of this comparison is a *similarity image* S_i at each depth, whose size is equal to that of the backprojection surface. At each iteration i , the pixels in S_i are compared to their corresponding pixels in S_{i+1} and S_{i-1} . As depth increases, the values for a point in the similarity image correspond to locations along a ray of visibility from the cyclopean eye. The strongest *local* similarity maximum along each such a ray is selected as the optimum depth. The requirement for maxima to be local is used to avoid artifacts that may occur in the textureless areas of the input images.

Memory conservation is achieved by tessellating the backprojection image into, say, $k \times k$ equal spherical segments. This tessellation is parameterized along the two spherical coordinates that, also, correspond to image width and height. The sweeping algorithm is performed independently for each such partition. These partitions overlap slightly, in order to avoid "blocking artifacts" at their boundaries. The amount of overlap is exactly determined by the size of the comparison kernel so that a scene point is not reconstructed twice.

The acceleration of the space-sweeping approach is based on an iterative and coarse-to-fine approach that is combined with the above memory conservation technique. The image data in each iteration are obtained from traditional image pyramids of the input images, starting from the smallest image of the pyramid and advancing a layer in each iteration; at the last iteration the original image is utilized. Also in each iteration, the parameterization of the backprojection surface becomes denser. As described above, the backprojection surface is tessellated and the sweeping algorithm is executed independently for each segment. At each iteration, though, each spherical segment is re-segmented into $k \times k$ more segments. After the 2nd iteration, the range of evaluated depths (d_i) is drastically constrained, based on the reconstruction result previously obtained for the "parent" segment.

The obtained depth map is filtered very conservatively (as in Mulligan et al., 2004), to suppress artifacts at depth discontinuities and remove outliers. By doing so, some valid matches are indeed rejected; however, in the utilized multiview

setup the corresponding points are most likely to be reconstructed from another binocular pair. The result is spatially quantized as it is too large ($\approx 10^9$ points for 35 views of 8Mpix each, in this experiment) to fit in memory. To cope with the same limitations the merging process is performed volumetrically, by tessellating the reconstruction volume into cubical segments. Finally, a thin plate interpolating surface is fit (Carr et al., 2001), to yield a mesh outputted into the VRML or KML formats.

In Fig. 2, the proposed method is demonstrated for the Dion (Greece) archaeological site. In the experiments presented in this paper, images were 2448 x 3264, 16-bit per layer, color images acquired with a Canon Powershot SLR camera, the number of iterations was 5 and the initial tessellation was 3 x 3. The coarse-to-fine refinement factor was 2, so that in each iteration: (a) the image rows and columns of the stereo and backprojection images were doubled and (b) the number of segments was increased by 4. The above scheme was measured to provide a speedup of ~50 for the scene of this experiment.

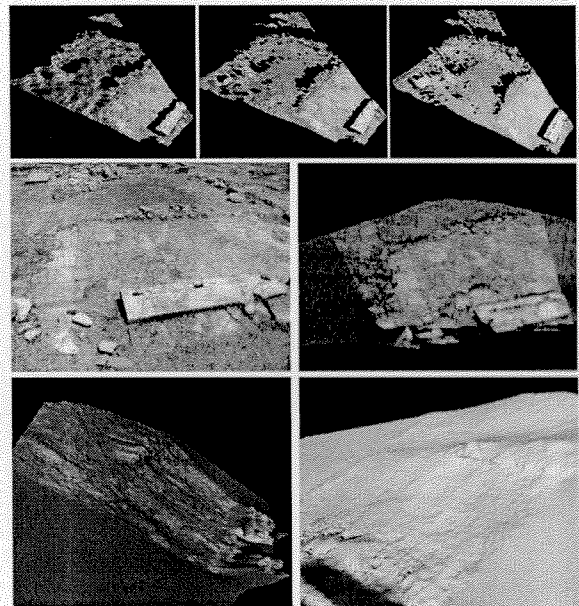


Figure 2: Coarse-to-fine acceleration scheme, for space-sweeping methods. Top row shows the reconstructions for the 3 first iterations of the proposed procedure. In the middle-left an original image from a -40cm baseline stereo pair (left) is shown. Others are the views of the RBF interpolated reconstruction with and without texture mapping.

Fig. 3 shows the result of an experiment that compares the reconstructions obtained from the proposed method in Harris and SIFT conditions of the previous section. The images in the first 2 rows show the result of the reconstruction for an early frame (20 views): in the SIFT condition, a larger proportion of the scene is reconstructed. The last row, shows the result of the SIFT condition after 35 frames.

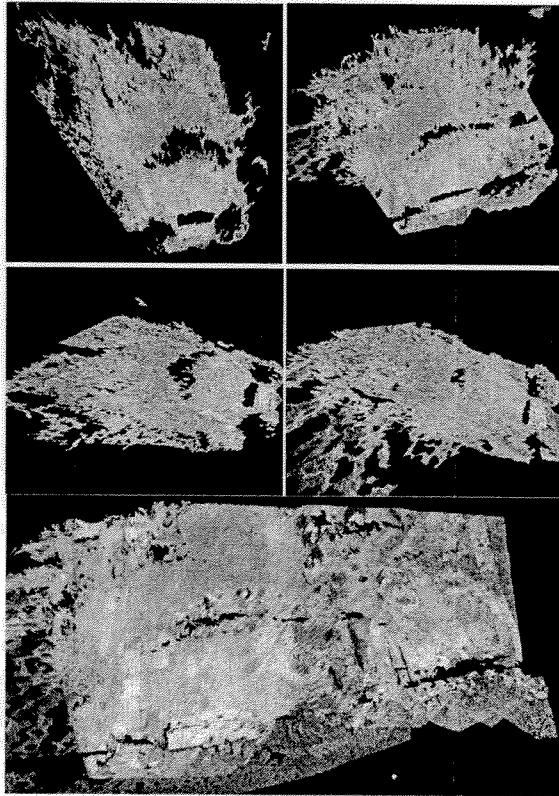


Figure 3: Comparing the reconstructions for the 2 conditions of the experiment in Sec. 3. Top 2 rows show the results for the Harris (left) and the SIFT (right) conditions, for two viewpoints (1st row is top view and 2nd row is side view). The last row of the figure shows a top view of the reconstruction of the same scene, from 35 views and the SIFT condition.

4. VIRTUAL TOUR APPLICATION

The reconstructed VRML models are integrated with GIS technologies within a Web-based virtual tour system, after converting them to the XML-based Collada 3D file format and then referencing to them in Keyhole Markup Language (KML), a format supported by the Google Earth™ GIS platform. Reconstructed part of the archaeological site is placed at its exact location on the terrain. Sample Google Earth™ views for Knossos (Greece) archaeological site are given in Fig. 4. If the resolution of Google Earth™ at that location is not satisfactory, excavation site plan can be used as detailed raster overlay, draped over the terrain. Then the reconstructed 3D model will be seen on the site plan.

We added a hyperlink to the application described above, which directs users to a panoramic image based virtual-tour. The main item in this a tour is a viewing window that the user can control. Using Java Applet technology is one proper way of creating such Web-based applications. In addition to the images, audio or textual information related to the site can be presented to the users with extra WWW tools. Using a map of the archaeological site increases the comprehension of the tour and enhances the user's sense of orientation. A step further is making this site plan interactive and integrated with the viewing window. With such tools, more information is communicated to the virtual tour users in an ergonomic fashion (Bastanlar, 2007).

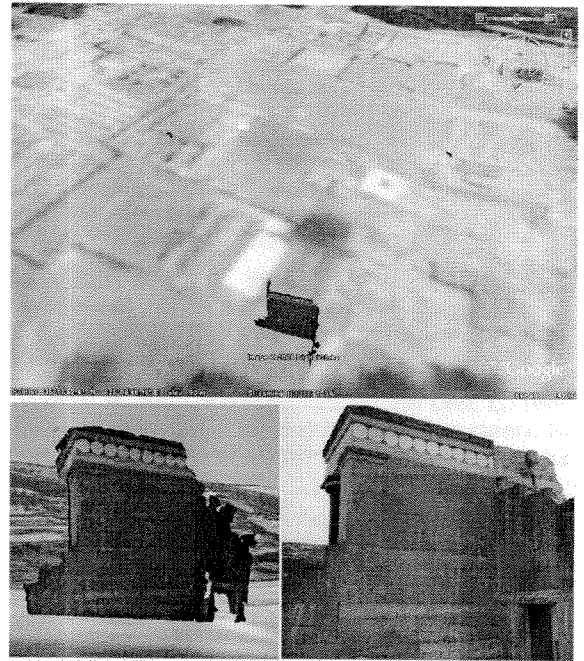


Figure 4: Viewing models in Google Earth™. At the top, overall view of the site together with the reconstructed wall. Bottom-left is the close view of the 3D model of the reconstructed section. The image at bottom-right is a real photograph taken from archaeological site.

In Fig. 5, a screenshot of the virtual tour page is shown, which is implemented for the ancient settlement Selime Castle in Cappadocia, Turkey. At bottom-left the viewing window and at the right the site plan are located. The section of the site that is currently presented in the viewing window, field of view (FOV) and direction of view are indicated in the floor plan. It is updated accordingly as the user changes these controls.

A larger degree of immersiveness can be experienced by viewing the reconstructed 3D models on autostereoscopic displays, which can be achieved by using a special plug-in, TriDef™ Visualizer for Google Earth™, to render real-time 3D scenes. We used this property to 3D render the scene for a stereoscopic notebook PC.

The pilot application implemented so far can be reached at <http://www.ii.metu.edu.tr/~3daegean/recent.htm>

5. CONCLUSIONS

In this study, a Web-based virtual tour system is built for the presentation of cultural heritage. In the proposed approach, the scene is captured from multiple viewpoints utilizing off-the-shelf equipment. We developed and presented the techniques to extract the 3D structure from the acquired images based on stereoscopic techniques. For presentation and 3D modeling of outdoor cultural heritage, the proposed approach as a whole constitutes an economic and practical alternative to the 3D scanner technology. Generated 3D model of the scene, detailed site plans and interactive virtual tour tools such as 360° viewing were integrated with GIS technologies.

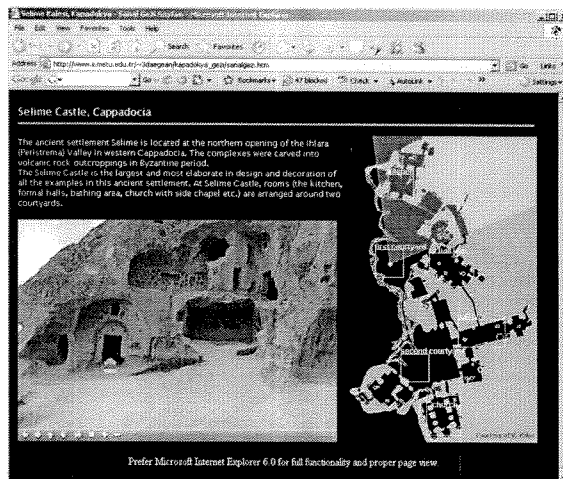


Figure 5: A screenshot from the panoramic image based virtual tour.

6. REFERENCES

- Bastanlar, Y., 2007. User Behaviour in Web-based Interactive Virtual Tours. In: *Proc. of 29th International Conference on Information Technology Interfaces*, Dubrovnik, Croatia.
- Beardsley, P., Zisserman, A., Murray, D., 1997. Sequential Updating of Projective and Affine Structure from Motion, *IJCV*, 23(3), pp.235-259.
- Bouguet, J., 2007. *Camera Calibration Toolbox for Matlab*, http://www.vision.caltech.edu/bouguetj/calib_doc/ [28/02/2008]
- Carr, J.C., Beatson, R.K., Cherrie, J.B., Mitchell, T.J., Fright, W.R., McCallum, B.C., Evans, T.R., 2001. Reconstruction and representation of 3D objects with radial basis functions, *Proc. SIGGRAPH*, pp.67-76.
- Conforti Andreoni, D., Pinto, L., 2004. The Creation of The Digital Models for The Protection of Cultural Heritage: The Baptistery of Cremona, *ISPRS Comm. V Sym.*
- Guarnieri, A., Vettore, A., El-Hakim, S., Gonzo, L., 2004. Digital Photogrammetry and Laser Scanning In Cultural Heritage Survey, *ISPRS Comm. V Sym.*
- Grün, A., Remondino, F., Zhang, L., 2002. Reconstruction of the Great Buddha of Bamiyan, Afghanistan, *International Archives of Photogrammetry and Remote Sensing*, 34(5), pp. 363-368.
- Harris, C.G., Stephens, M., 1988. A Combined Corner and Edge Detector, *Proc. of Fourth Alley Vision Conference*, Manchester, U.K., p.182-192.
- Kadobayashi R., Kochi, N., Otani, H. Furukawa, R., 2004. Comparison and Evaluation of Laser Scanning and Photogrammetry and Their Combined Use for Digital Recording of Cultural Heritage, *ISPRS Comm. V Sym.*
- Li, M., Magnor, M., Seidel, H.P., 2004. Hardware-accelerated rendering of photo hulls, *Eurographics*, 23(3).
- Lourakis, M., Argyros, A., 2004. The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm, TR #340, ICS-FORTH.
- Lowe, D., 2004. Distinctive image features from scale-invariant keypoints, *IJCV*, 60(2), pp.91-110.
- Mordohai, P., Frahm, J., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., Nister, D., 2007. Real-time Video-based Reconstruction of Urban Environments, *Proc. of 3D-ARCH 2007 Workshop, 3D Virtual Reconstruction and Visualization of Complex Architectures*, ETH Zurich, Switzerland.
- Mulligan, J., Zabulis, X., Kelshikar, N., Daniilidis, K., 2004. Stereo-based Environment Scanning for Immersive Telepresence, *IEEE Transactions on Circuits and Systems for Video Technology*, 14(3):304-320.
- Pollefeys, M., Koch, R., Vergauwen, M., van Gool, L., 1999. Hand-held acquisition of 3D models with a video camera. In: *Proc. of the International Conference on 3-D Digital Imaging and Modeling (3DIM 1999)*.
- Pollefeys, M., van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J., Koch, R., 2004. Visual modeling with a hand-held camera, *IJCV*, 59(3), pp.207-232.
- Tola, E., 2005. *Multi-view 3D Reconstruction of a Scene Containing Independently Moving Objects*, M.Sc. Thesis, Middle East Technical University, Ankara, Turkey.
- Villanueva, R., Moore, A., Wong, B.L.W., 2004. Usability Evaluation of Non-immersive, Desktop, Photo-realistic Virtual Environments. In: *The 16th Annual Colloquium of the Spatial Information Research Centre (SIRC 2004)*. http://www.business.otago.ac.nz/SIRC05/conferences/2004/28_Villanueva.pdf [accessed 28/02/2008]
- Yang, R., Welch, G., Bishop, G., 2002. Real-time consensus-based scene reconstruction using commodity graphics hardware, *Proc. of Pacific Graphics*, Beijing, China.
- Zabulis, X., Patterson, A., Daniilidis, K., 2003. Digitizing Archaeological Excavations from Multiple Views. In: *Proc. of the International Conference on 3-D Digital Imaging and Modeling (3DIM 2003)*.
- Zabulis, X., Kordelas, G., 2006. Efficient, Precise, and Accurate Utilization of the Uniqueness Constraint in Multi-View Stereo, *Proc. of 3DPVT*.
- Zabulis, X., Kordelas, G., Mueller, K., Smolic, A., 2006. Increasing the accuracy of the space-sweeping approach to stereo reconstruction, using spherical backprojection surfaces, *ICIP 2006*, Atlanta, USA.
- Zabulis, X., 2007. Utilization of the texture uniqueness cue in stereo. In *Three-Dimensional Television: Capture, Transmission, and Display*, Springer Verlag.

7. ACKNOWLEDGEMENT

The authors are grateful for support through the 3DTV European NoE, FP6 IST Programme and TUBITAK-GSRT Joint Research Project (105E187). They also thank Engin Tola, for providing the implementation of his structure-from-motion technique and Veronica Kalas, for her guidance and providing material for Selime Castle in Cappadocia. The images of Dion are courtesy of the Archeological Service of Greece at Dion, Katerini, Greece.

FEATURE MATCHING BETWEEN CATADIOPTRIC AND PERSPECTIVE CAMERA IMAGES USING SCALE-RESTRICTED SIFT

Yalin Bastanlar, Alptekin Temizel, Yasemin Yardimci Cetin

Informatics Institute
Middle East Technical University, Ankara, Turkey

ABSTRACT

We worked on robustly matching feature points in images taken by mixed camera types, namely catadioptric omnidirectional and perspective cameras. We used scale invariant feature transform (SIFT) as the matching method and proposed a scale restriction algorithm to reduce the number of false matches. We experimentally showed that, with the proposed approach, omnidirectional-perspective matching performance significantly increases. In addition, we evaluated the use of virtual camera plane (VCP) images, which are perspective images produced by unwarping a certain region of the omnidirectional image. We observed that VCP-to-perspective matching is more robust to increasing baseline length when compared to direct omnidirectional-perspective matching.

Index Terms— Catadioptric, omnidirectional, feature matching, SIFT, VCP, scale restriction

1. INTRODUCTION

Omnidirectional cameras provide 360° horizontal field of view in a single image, which is an important advantage in many application areas such as surveillance and 3D reconstruction [1, 2]. Major drawback of these images is that they have lower resolution than perspective images. Using perspective cameras together with omnidirectional ones could improve the resolution while preserving the enlarged view advantage. This would require effective feature matching between the images of different camera types. Widely accepted matching methods (eg. Scale Invariant Feature Transform: SIFT [3], Maximally Stable Extremal Regions: MSER [4]) do not perform well when they are directly employed for hybrid camera images [5, 6]. In this paper, we used SIFT for feature matching and developed an algorithm to obtain better feature matches between catadioptric and perspective images.

SIFT detects features at different scales and matches them regardless of their scales. However, we observed low matching accuracy when there is a major scale difference between the two images. This is especially important for perspective-omnidirectional hybrid imaging systems because there is an

inherent resolution difference. Moreover the resolution in catadioptric systems is non-uniform. We observed that most of the false matches in SIFT output are the ones with incorrect scale ratio. We propose a method to automatically detect the scale ratio and use it to obtain an improved SIFT matching output. Details are given in Section 2.

In addition to catadioptric-to-perspective matching we investigated VCP-to-perspective matching, where VCP stands for virtual camera plane. VCP image is produced by unwarping a certain region of the catadioptric image to generate a perspective image. Approximate calibration information can be used to unwarped catadioptric images (refer to Section 3).

Regarding the previous studies on hybrid systems, Ramalingam *et al.* [5] worked on structure-from-motion. They used manually selected feature point correspondences to estimate epipolar geometry and mentioned that directly applying SIFT did not provide good results for their fisheye-perspective image pairs. Chen and Yang [7] developed a region matching algorithm for hybrid views based on planar homography.

Although not for hybrid image pairs, an example study employing VCP is performed by Mauther *et al.* [8]. They performed matching between two images of a calibrated omnidirectional camera using a sequence of algorithms including MSER [4], SIFT [3] and Harris [9]. To our knowledge, the only study on automatic feature point matching technique between hybrid camera images is by Puig *et al.* [6]. They first applied a catadioptric-to-panoramic conversion and employed SIFT between panoramic and perspective views. To eliminate the false matches they employed RANSAC based on satisfying the epipolar constraint.

We demonstrated that the scale ratio detection and restriction algorithm proposed in this paper significantly increases the matching accuracy between catadioptric and perspective images even without catadioptric-to-panoramic or catadioptric-to-VCP conversion. The matches can be further improved, especially for increasing baseline, by generating VCP image from catadioptric image first. Proposed approach is much simpler than the framework proposed by Mauther *et al.* [8] and reaches higher success ratios when compared to panoramic-to-perspective matching presented by Puig *et al.* [6]. We present our experimental results in Section 4 and state final conclusions and comments in Section 5.

Authors thank to TUBITAK for support through the project 105E187.

2. SCALE RESTRICTION

SIFT detects features in the so-called *scale space* comprising levels and octaves which are obtained by low-pass filtering and downsampling the original image systematically [3]. Although this enables the detection of features at different scales, we observed that most of the false matches in SIFT output are due to matching a high-resolution feature in perspective image to a feature in catadioptric image. This is also valid for two perspective images with significant scale difference. Let us explain this phenomenon by an example. Table 1 shows the number of extracted features in different octaves for images Pers2 and Omni2 (Fig. 2). There is an approximate ratio between scales of true correspondences ($SR = \sigma_{pers} / \sigma_{omni}$), which is ≈ 3.6 for the given images. Corresponding octaves of true matches are indicated in the table with arrows. SIFT extracts many features (nearly 3000) at the first two (high-resolution) octaves as can be observed. When there is no scale restriction, it is quite likely that some of the many candidates from first two octaves of the perspective image are incorrectly selected as the best match of features in omnidirectional image. For the given image pair, there are 25 false matches out of 60 and 23 of these false matches have an $SR < 2.0$, whereas average SR of true matches is 3.57. Thus, applying a method to select matches with SR close to 3.6 eliminates a great proportion of false matches.

A similar idea to restrict the scale ratio has recently been mentioned by Yi *et al.* [10]. They formed the histogram of scale differences and defined a window around the peak of this histogram. Matches with scale differences outside this window were rejected.

In our case, the histogram of the example hybrid image pair is shown in Fig. 1a. The accumulation on the left (matches with $SR < 2.0$) is explained by false matches due to matching features in the first octaves of perspective image SIFT. False matches can be eliminated by defining a window around the peak [10], however we found out that it is possible to get rid of most of the false matches in the initial SIFT output by blurring and downsampling the perspective image. Blurring is achieved by low-pass filtering with a Gaussian filter. With this preprocessing, the scale ratio of matching features becomes close to 1 and the possibility of matching

Octave	Approximate scale(σ) in SIFT scale space	Omni 2 (1024x960)	Pers 3 (1100x300)	Pers 3, blurred and downsampled
-1	1	1365	2459	288
0	2	489	465	97
1	4	202	174	23
2	8	68	76	5
3	16	23	20	0
4	32	4	5	-

Table 1. Number of SIFT features detected in a catadioptric and perspective image pair.

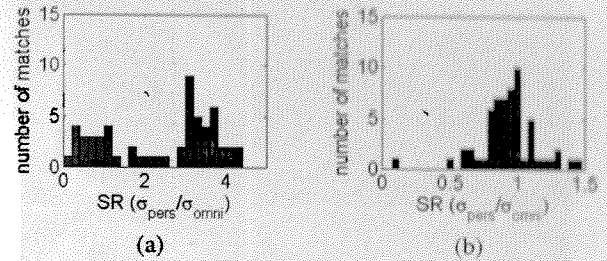


Fig. 1. Histogram of SR for the matches in the catadioptric-perspective image pair given in Table 1. (a) SIFT applied on original image pair, (b) SIFT applied on downsampled perspective image and original catadioptric image.

valuable features in omnidirectional image with features in the correct octaves of perspective image considerably increases. For the given example, last column of Table 1 shows the number of detected features where the perspective image is downsampled by 3.6 (both in horizontal and vertical axis) following a blurring operation.

We selected the downsampling factor from the histogram as the mean of the most dominant Gaussian in the mixture, because the SIFT scale space ratio also reveals the scale ratio of features in the images. The standard deviation of the Gaussian filter for initial blurring is selected as $\sigma = 2d/\pi$ where d is the downsampling factor. Fig. 1b shows the SR histogram when the perspective image is low-pass filtered and downsampled as explained. This matching resulted in a true/total ratio of 56/60.

We can further restrict the scale to remove a few false matches with improper SR as performed in [10]. To do this we define a window around the mean SR and discard the matches outside the window. For our experiments we chose the bounds of the restriction window as $0.6SR < SR < 1.4SR$. After this post-restriction, true/total ratio becomes 51/52 for the given example. If we directly apply scale restriction without preprocessing true/total ratio is 32/34. Since it is important to keep as many true matches as possible in most of the computer vision applications, especially for structure-from-motion, we preferred the downsampling approach but not the approach proposed by Yi *et al.* [10]. Another difference with their study is that we used scale ratio (SR) instead of scale difference (SD). As the scales in our case varies considerably and they are doubled between octaves of SIFT scale space, employing SR rather than SD becomes wiser. Results for all image pairs of our experiment are given in Section 4.

3. CREATING VIRTUAL CAMERA PLANE IMAGES

Another approach that can be employed in conjunction with scale restriction is to use so-called *virtual camera planes* (VCP) to create virtual perspective images from catadioptric images and perform matching between VCP and perspective images. Catadioptric-to-VCP conversion can easily be made

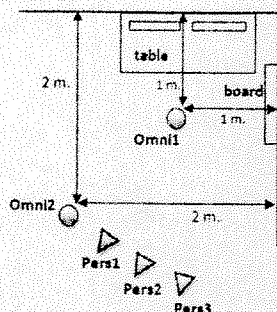


Fig. 2. Locations and orientations of cameras w.r.t. the scene for the experiment. See the resultant images in Fig.3.

if cameras were calibrated. If the calibration is not possible, approximate information regarding the mirror is adequate to create VCP images. This is demonstrated in the present paper for a camera with a paraboloidal mirror. The principal point can be assumed to be at the center of omnidirectional image circle. We can recover the mirror parameter h by assigning a position in the image which corresponds to an object with the same height with mirror focal point. The distance between the image center and that point gives us the parameter h in terms of pixels, for example the radius of white circle in the top-left image of Fig.3.

To generate virtual perspective image, a virtual camera plane with a certain viewing direction (azimuth), a vertical angle and a distance from the mirror focal point (origin) have to be defined. To find the intensity value of a pixel in the virtual perspective image (x_v, y_v) , corresponding pixel coordinates in the paraboloidal catadioptric image (x_i, y_i) are given by [11]:

$$x_i = \frac{h}{Z_S + \sqrt{d^2}} X_S \quad y_i = \frac{h}{Z_S + \sqrt{d^2}} Y_S \quad (1)$$

where, $d = \sqrt{X_S^2 + Y_S^2 + Z_S^2}$ and (X_S, Y_S, Z_S) are the 3D coordinates of the virtual image pixel w.r.t. the origin.

4. EXPERIMENTS

We conducted tests on real images captured in a controlled environment. The locations and orientations of the cameras and objects in our scene are given in Fig.2 and corresponding images are given in Fig.3.

To detect and match feature points, we used SIFT implementation of Andrea Vedaldi¹ with a modification to provide one-to-one matching. In the original algorithm it is possible to have many points in the first image matched to the same point in the second image. This causes inconsistencies depending on which image is defined as first. To eliminate this problem, we run the SIFT algorithm in both ways changing the order of images and declare a match only if it is found in both runs.

¹<http://vision.ucla.edu/~vedaldi/code/sift/sift.html>

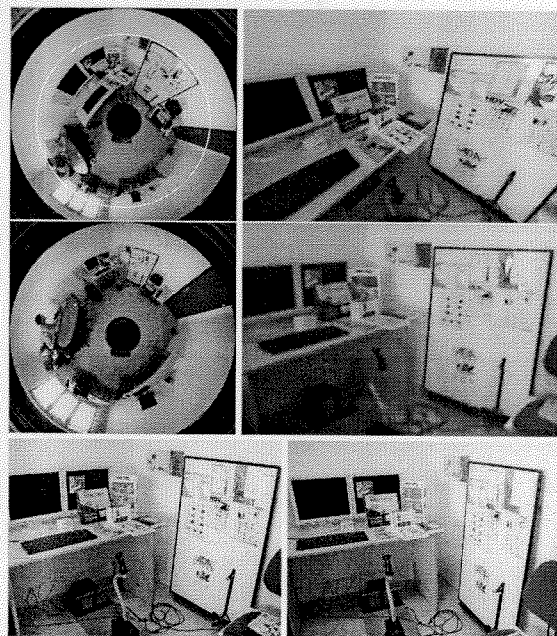


Fig. 3. Images used for the experiment. From top to bottom, first line is Omni1 and VCP image generated from it (VCP1), second line is Omni2 and its virtual image (VCP2), third line is two of the perspective images, Pers1 and Pers3, with small and wide lateral baseline respectively.

As the output of the catadioptric system is viewed through a mirror, the objects are *flipped* in the image. This is corrected by flipping the image w.r.t. a line passing through the center of catadioptric image circle.

Matching results are shown in Table 2. For different baseline scenarios, comparison between direct omnidirectional-perspective matching, matching after downsampling and VCP-to-perspective matching approaches can be observed. To keep the number of matched points same for different trials of an image pair, we adjusted the matching threshold of SIFT, which defines the strength of the matched point w.r.t. the second candidate match. As described in Section 2, to increase omnidirectional-to-perspective matching performance, we downsampled the perspective images after blurring with a Gaussian filter. In Table 2, PersN gA dB indicates that PersN image was blurred with $\sigma = A$ Gaussian filter and downsampled by a factor of B in each direction. B is extracted from SR histogram of the matches and A is chosen so that no aliasing will occur after downsampling.

VCP-to-perspective matching results are also given in Table 2. N in VCPN indicates the index of omnidirectional image it is generated from. Since we created the VCP images with a size close to the size of perspective images, SR of the true matches is already close to 1 and no downsampling is needed. However, we still have to blur the perspective image to prevent SIFT from extracting many features in first octaves.

Image pairs	no. of matches	T/F (T/total %)	T/F restricted
Pers1 - Omni1	100	97/3 (97%)	84/1
Pers1 g1.5 d1.65 - Omni1	100	99/1 (99%)	94/0
Pers1 g1.5 - VCP1	100	99/1 (99%)	99/0
Pers2 - Omni1	75	56/19 (75%)	51/6
Pers2 g1.5 d1.65 - Omni1	75	70/5 (93%)	67/3
Pers2 g1.5 - VCP1	75	73/2 (97%)	72/1
Pers3 - Omni1	60	42/18 (70%)	39/8
Pers3 g1.5 d1.65 - Omni1	60	50/10 (83%)	49/6
Pers3 g1.5 - VCP1	60	57/3 (95%)	57/1
Pers1 - Omni2	80	63/17 (79%)	62/0
Pers1 g2.5 d3.6 - Omni2	80	80/0 (100%)	80/0
Pers1 g2.5 - VCP2	80	80/0 (100%)	80/0
Pers2 - Omni2	60	35/25 (58%)	32/2
Pers2 g2.5 d3.6 - Omni2	60	56/4 (93%)	52/1
Pers2 g2.5 - VCP2	60	56/4 (93%)	56/3
Pers3 - Omni2	45	15/30 (33%)	15/1
Pers3 g2.0 d3.3 - Omni2	45	35/10 (78%)	30/6
Pers3 g2.0 - VCP2	45	37/8 (82%)	36/4

Table 2. Matching results for the image pairs given in Fig.3.

Since the VCP image is already blurry due to the conversion from omnidirectional image with interpolation, features detected in VCP are not accumulated in first octaves. Alternatively we could have produced small size VCP and downsampled the perspective image, which produce similar results.

We plotted the ratio of true/total matches in Fig.4 for different approaches. The table also shows, in the last column, the restricted true/false ratios obtained after applying final scale restriction. We did not plot these ratios because the number of true matches is also important along with the ratio (cf. Section 2). But please note that the success rates of both downsampling and VCP approaches in Fig.4 (solid lines) were further increased by scale restriction.

Pers1, Pers2 and Pers3 have approximately 30cm., 60 cm. and 90 cm. lateral baselines with omnidirectional cameras respectively. We observe the matching performance decreases with increasing baseline for all approaches and the number of false matches decreased significantly for both downsampling and VCP approaches. We also observe that VCP approach is more robust to increase in baseline at least for PersN-Omni1 pairs, thus we may suggest VCP approach. However downscaling approach has the advantage of being automatic since it does not require user interaction for generating VCP.

5. CONCLUSIONS

It had been stated that directly applying SIFT is not sufficient to obtain good results for mixed image pairs. In our study we showed that the performance of SIFT considerably increases with the proposed downsampling and scale restriction algorithm and it becomes a good choice for matching catadioptric omnidirectional and perspective images. Another approach

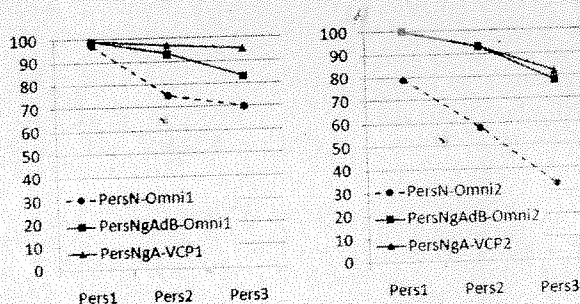


Fig. 4. The true/total match ratios (in percentage) for Omn1-PersN (on the left) and Omn2-PersN (on the right).

we proposed is generating VCP image from omnidirectional image first and then applying SIFT. We obtained satisfactory true/false ratios with both approaches and we observed that VCP approach is more robust to increasing baseline. On the other hand, downsampling with the estimated scale ratio (SR) has the advantage of being automatic. In both cases, to eliminate a few remaining false matches, RANSAC based on epipolar constraint can be applied [6].

Moreover, optimal SR detection and scale restriction approach proposed in this paper can be used for other applications employing SIFT where an approximate scale ratio exists in given images, including perspective image pairs.

6. REFERENCES

- [1] G. Scotti, L. Marcenaro, C. Coelho, F. Selvaggi, and C.S. Regazzoni, "Dual camera intelligent sensor for high definition 360 degrees surveillance," *IEE Proc.-Vision Image Signal Processing*, vol. 152(2), 2005.
- [2] M. Lhuillier, "Toward flexible 3d modeling using a catadioptric camera," in *CVPR*, 2007.
- [3] D.G. Lowe, "Distinctive image features from scale invariant keypoints," *IJCV*, vol. 60, pp. 91-110, 2004.
- [4] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *BMVC*, 2002.
- [5] S. Ramalingam, S.K. Lodha, and P. Sturm, "A generic structure-from-motion algorithm for cross-camera scenarios," in *Proc. of IEEE Workshop on Omnidirectional Vision*, 2004.
- [6] L. Puig, J.J. Guerrero, and P. Sturm, "Matching of omnidirectional and perspective images using the hybrid fundamental matrix," in *Workshop on Omnidirectional Vision*, 2008.
- [7] D. Chen and J. Yang, "Image registration with uncalibrated cameras in hybrid vision systems," in *Workshop on the Applications of Computer Vision*, 2005.
- [8] T. Mauthner, F. Fraundorfer, and H. Bischof, "Region matching for omnidirectional images using virtual camera planes," in *Proc. of Computer Vision Winter Workshop*, 2006.
- [9] C.G. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. of 4th Alley Vision Conf.*, 1988.
- [10] Z. Yi, C. Zhiguo, and X. Yang, "Multi-spectral remote image registration based on sift," *Electr. Letters*, vol. 44(2), 2008.
- [11] V. Peri and S.K. Nayar, "Catadioptric omnidirectional camera," in *Proc. of DARPA Image Understanding Workshop*, 1997.

TÜBİTAK
PROJE ÖZET BİLGİ FORMU

Proje No: 105E187
Proje Başlığı: Ege Bölgesi'nin Kültürel Zenginlikleri için 3 Boyutlu Geriçatım ve Web-tabanlı Sanal Gezi
Proje Yürütücüsü ve Araştırmacılar: Prof. Dr. Yasemin Yardımcı, Yalın Baştanlar
Projenin Yürütüldüğü Kuruluş ve Adresi: ODTÜ Enformatik Enstitüsü, 06531, ODTÜ, Ankara
Destekleyen Kuruluş(ların) Adı ve Adresi: --
Projenin Başlangıç ve Bitiş Tarihleri: 01.11.2006 - 31.10.2008
Öz (en çok 70 kelime): Projenin amacı arkeolojik alanların sunumu odaklı olmak üzere Web tabanlı bir sanal gezi sistemi geliştirilmesidir. Arkeolojik alanlar kolay edinilebilir teçhizatlar kullanılarak görüntülenmiş ve 3 boyutlu yapı stereo teknikler kullanılarak bu alınan görüntülerden elde edilmiştir. Doku bilgisi de eklendikten sonra sonuç geriçatım 3B modele çevrilmiştir. 3B modeller, alan krokileri ve panoramik imge tabanlı sanal turlar ile birlikte Google Earth™ ile entegre edilerek ziyaret edilmeleri sağlanmıştır.
Anahtar Kelimeler: 3 Boyutlu geriçatım, Kültürel miras, Tümyönlü görüntüleme, Sanal gezi, Coğrafi bilgi sistemi
Projeden Yapılan Yayınlar: <ul style="list-style-type: none">- Yalın Baştanlar, Nikos Grammalidis, Xenophon Zabulis, Erdal Yılmaz, Yasemin Yardımcı, , Georgios Triantafyllidis, "3D Reconstruction for a Cultural Heritage Virtual Tour System", ISPRS 2008, http://www.isprs2008-beijing.org- Xenophon Zabulis, Nikos Grammalidis; Yalın Baştanlar, Erdal Yılmaz, Yasemin Yardımcı, "3D Scene Reconstruction Based on Robust Camera Motion Estimation and Space Sweeping", 3DTV-Conference: The True Vision: Capture, Transmission and Display of 3D Video, 28-30 May, Istanbul, Turkey.- Yalın Baştanlar, Alptekin Temizel, Yasemin Yardımcı, "3D Scene Reconstruction Based on Robust Camera Motion Estimation and Space Sweeping", CVPR 2009'da sunulmak üzere değerlendirmeye gönderilmiştir.- Nikos Grammalidis, Yalın Baştanlar, Xenophon Zabulis, Erdal Yılmaz, Yasemin Yardımcı, "Robust Image-based 3D Reconstruction for Cultural Heritage", EURASIP Journal on Image and Video Processing, 'Image and Video Processing for Cultural Heritage' özel sayısına gönderilmek üzere hazırlık aşamasında.