

T.C.
GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

GERÇEK ZAMANLI KAPANMASIZ
GÖRÜNTÜ ÜRETİMİ İÇİN
ÇOK KAMERALI SİSTEM

Alparslan Ömer YILDIZ
YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

GEBZE
2009

T.C.
GEBZE YÜKSEK TEKNOLOJİ ENSTİTÜSÜ
MÜHENDİSLİK VE FEN BİLİMLERİ ENSTİTÜSÜ

GERÇEK ZAMANLI KAPANMASIZ
GÖRÜNTÜ ÜRETİMİ İÇİN
ÇOK KAMERALI SİSTEM

Alparslan Ömer YILDIZ
YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ
ANABİLİM DALI

DANIŞMANI
Yrd. Doç. Dr. Yusuf Sinan AKGÜL

GEBZE
2009



YÜKSEK LİSANS TEZİ JÜRİ ONAY SAYFASI

G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 20/07/2009 tarih ve 2009/16 sayılı kararıyla oluşturulan jüri tarafından 01/09/2009 tarihinde tez savunma sınavı yapılan Alparslan Ömer YILDIZ'ın tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Yrd. Doç. Dr. Yusuf Sinan AKGÜL

ÜYE

: Yrd. Doç. Dr. Fatih Erdoğan SEVİLGİN

ÜYE

: Yrd. Doç. Dr. Abdullah BAL

ONAY

G.Y.T.E. Mühendislik ve Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
...../...../20... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

ÖZET

TEZ BAŞLIĞI: GERÇEK ZAMANLI KAPANMASIZ GÖRÜNTÜ
ÜRETİMİ İÇİN ÇOK KAMERALI SİSTEM

YAZAR ADI: ALPARSLAN ÖMER YILDIZ

Bilgisayarla görme alanındaki önemli problemlerden birisi de gerçek zamanlı kapanmasız görüntü üretimidir. Kapanmasız görüntü üretimi, bir sahnedeki nesnelerin kamera görüntüsünde birbirlerini örterek oluşturdukları kapanmaların çözülerek sahnedeki hareketli nesnelerin sürekli olarak izlenebilmesini sağlar. Sunulan sistem temel olarak iki aşamadan oluşmaktadır; öğrenme aşaması ve çalışma zamanı aşaması. Öğrenme aşamasında sistemin çalışması için gerekli bilgiler toplanır. Bunlar sahne ve kameralar ile ilgili temel bilgilerdir. Öğrenme aşaması sistem başlatılmadan önce bir defaya mahsus gerçekleştirilir. Çalışma aşamasında ise öğrenilen sistem bilgileri ışığında sahnedeki nesneler ve kapanmalar tespit edilir, kapanmadaki nesnelerin görüntüleri tahmin edilir ve bu tahmin edilen görüntüler kapanmasız görüntü üretimi için kullanılır. Sistem kapanmaya uğrayan nesnelerin görüntülerini tahmin edebilmek için aynı nesnelerin daha önce kapanma olmadığı zamanlarda kaydettiği görüntülerini kullanır. Sistem gerçek zamanda çalışabilmektedir ve güvenlik, eğlence, aktivite analizi gibi alanlarda birçok uygulama için temel sistem niteliğindedir.

SUMMARY

**THESIS TITLE: A REALTIME MULTI-CAMERA SYSTEM FOR
THE GENERATION OF OCCLUSION-FREE
VIDEO**

AUTHOR: ALPARSLAN ÖMER YILDIZ

One of the important problems in the computer vision field is the generation of occlusion-free video in real-time. Occlusion-free video provides the opportunity of surveilling objects in a scene by resolving the occlusions caused by the overlapping of object images. The proposed system mainly consists of two phases: learning phase and runtime phase. In the learning phase, the system collects information necessary for the runtime phase. The information collected is about the geometry of the scene and the cameras. The learning phase is run only once before the system starts. In the runtime phase, with the help of the collected information, objects and occlusions are detected; appearances of the occluded objects are estimated and used for the generation of occlusion-free video. The system uses the previously recorded appearances of the objects to estimate the appearances of the occluded objects. The system can run in real-time rates and can be used in applications such as security, entertainment and activity analysis.

TEŞEKKÜR

Tezime başlamamdan bitirmeme kadar bana devamlı yol gösteren ve destek veren danışmanım Sayın Yrd. Doç. Dr. Yusuf Sinan AKGÜL'e ve bugünlere gelmemde çok büyük emekleri olan, eğitimimin her basamağında bana maddi ve manevi destekte bulunan sevgili aileme teşekkür ederim.

İÇİNDEKİLER DİZİNİ

	<u>Sayfa</u>
ÖZET	iv
SUMMARY	v
TEŞEKKÜR	vi
İÇİNDEKİLER DİZİNİ	vii
SİMGELER VE KISALTMALAR DİZİNİ	ix
ŞEKİLLER DİZİNİ	x
1 GİRİŞ	1
2 KAPANMASIZ GÖRÜNTÜ ÜRETİMİ	6
3 ÖĞRENME AŞAMASI	12
3.1 Arka plan modelleri	12
3.2 Yer düzlemi	12
3.3 Epipolar geometri	14
3.4 Perspektif sıfır noktaları ve doğruları	15
4 ÇALIŞMA ZAMANI AŞAMASI	18
4.1 Nesnelerin arka plandan ayrıştırılması	18

4.2 Nesne pozisyonları ve kapanma testi	25
4.3 Nesne Görüntüsünün Tahmini	29
5 DENEYLER	44
6 SONUÇLAR VE ÖNERİLER	38
KAYNAKLAR	39
ÖZGEÇMİŞ	41

SİMGELER VE KISALTMALAR DİZİNİ

3B : Üç boyutlu

2B : İki boyutlu

GKM : Gauss Karışım Modeli

ŞEKİLLER DİZİNİ

<u>Sekil</u>	<u>Sayfa</u>
1: Bir nesnenin 2B, 3B pozisyonları ve tepe noktaları	7
2: Sunulan sistemin özeti	10
3: Yer düzlemi için toplanan noktalar	14
4: Perpektif sıfır noktaları ve ufuk çizgisi	16
5: Ufuk çizgisinin bulunması	17
6: Piksel tabanlı ayrıştırma algoritması	19
7: Global ayrıştırma algoritması	19
8: Örnek bir üçlü maske resmi	22
9: Örnek bir resim ve hesaplanmış resim gradyanı. Resim gradyanı nesne sınırlarında yüksek değerlere sahiptir.	24
10: Nesne ayrıştırma sisteminin adımları	25
11: Kapanma testi	26
12: Kapalı alan deneyi	33
13: Sanal engel deneyi	35
14: Açık alan deneyi	36

15: Engel temizleme deneyi

36

1 GİRİŞ

Bilgisayarla görme alanındaki önemli problemlerden birisi de gerçek zamanlı kapanmasız görüntü üretimidir. Kapanma, bilgisayarla görme alanında, bir kamera tarafından izlenen nesnelerin kameradaki görüntülerinin birbirini örtmesi olarak tanımlanabilir. Kapanmasız görüntü üretimi için birçok farklı yöntem geliştirilmiştir. Bu yöntemlerden bazıları birden fazla kamera kullanarak kapanmaya uğrayan nesneleri farklı açılardan görmeyi amaçlar. Diğer yöntemler ise tek kamera kullanarak kapanmaya uğrayan nesnelerin görüntülerini aynı kamera ile farklı zamanlarda çekilmiş resimlerden elde etmeyi amaçlar. Bu tez çalışmasında birden fazla kamera kullanarak kapanmasız görüntü üretimi için yeni bir yöntem sunulmuştur.

Kapanma, birden fazla nesne bulunan bir sahnenin görüntüsünü çeken bir kamera için doğal bir olaydır. Nesneler zaman zaman kamera görüntüsünde birbirlerini tam ya da kısmi bir şekilde kapatabilir. Kapanmasız görüntü üretimi, bu gibi durumlarda görüntüsü kapanan, kamera tarafından görüntülenemeyen, nesnelerin görüntülerini izleyen kamera açısından tahmin etme işlemidir. Bu işlemin tam başarı ile yapılabilmesi için, izlenen nesnelerin ve sahnenin üç boyutlu yapıları ve pozisyonları bulunmalıdır. Ancak bilgisayarla görme alanında, maliyeti düşük olduğu için, sıklıkla iki boyutlu resimler kullanılmaktadır ve iki boyutlu resimler üç boyutlu dünya ile ilgili yeterince bilgi verememektedir. Üç boyutlu bilgiyi elde etmek için lazer tarayıcılar, derinlik kameraları v.s. kullanılır. Bu tip araçların edinilmesi ve kullanımı hem pratik değildir hem de oldukça pahalıdır. Ancak bazı bilgisayarla görme teknikleri ile iki boyutlu resimlerden üç boyutlu bilgi etmek mümkündür. Örneğin stereo; birden fazla kamera kullanılarak nesnelerin üç boyutlu dünyadaki konumları ile ilgili bilgi elde etmeyi amaçlayan bir yöntemdir.

Bazı bilgisayarla görme uygulamalarda ise üç boyutlu bilgiye doğrudan ihtiyaç yoktur. Üç boyutlu bilgiyi çıkartmak yerine, uygulamadaki üç boyutlu nesnelerin iki boyuttaki kamera görüntüleri üzerinden çıkarımlar yapılarak iki boyutlu resimler ile istenilen sonuçlar veya tatmin edici sonuçlar elde edilebilir. İki

boyutlu resim çeken kameralar, örneğin web kameralar, oldukça ucuzdur ve kullanımları da kolaydır. Bu tezde sunulan yöntem iki boyutlu resim çeken kameralar kullanarak kapanmasız görüntü üretimi sağlamaktadır.

Kapanmasız görüntü üretimi, sahnenin ve sahnedeki nesnelerin üç boyutlu yapıları ve pozisyonları ile ilgili doğrudan bilgi olmadan sadece iki boyutlu resimler kullanarak ve üç boyutlu yapıyı doğrudan çıkarmadan gerçekleştirilmesi zor bir uygulamadır. Bu uygulamanın gerçek zamanda gerçekleştirilebilmesi için çeşitli varsayımlarda bulunulmalı, basit ve pratik yöntemler geliştirip kullanılmalıdır. Literatürde kapanmaları ortadan kaldıracak çalışmalar mevcuttur. [2, 3] çalışmalarında onlarca kamera kullanılarak bir sahnedeki sabit nesnelerin arkalarındaki alan görülmeye çalışılmıştır. Böyle bir sistem çok masraflıdır ve ancak kısıtlı uygulaması vardır. [6] çalışmasında ise farklı kameralardan gelen bilgiler kullanıcının kamerası üzerine gerekli dönüşümler uygulandıktan sonra düşürülür. Bu sayede kullanıcının baktığı kameraya göre nesnelerin arkasında kalan alanlar kullanıcıya aktarılmış olur. Bu sistem farklı kameraların farklı açılardan oluşturduğu iki boyutlu görüntüleri kullanıcıya iletir. Bu şekilde üç boyutlu dünya ile ilgili farklı açılardan çekilmiş iki boyutlu resimlerin doğrudan kullanılması genel anlamda bilgi kaybına yol açar. Örneğin bu sistem, kapanma arkasındaki bir nesnenin pozisyonu veya büyüklüğü ile ilgili bilgi veremez, sadece yardımcı kameralardan nasıl görüldüğü bilgisini verebilir.

Bu tez çalışmasında sunulan yöntem iki boyutlu kameralar ile kapanmasız görüntü üretimini amaçlar. Günlük hayattaki sahnelerde cisimler herhangi bir kameraya göre farklı derinliklerde bulunduğundan dolayı kapanma sık rastlanan bir olaydır. Kapanmayı engellemek için birden fazla kamera farklı açılarla ve pozisyonlarla sahneye yerleştirilebilir. Birden fazla kamera kullanımının bir dezavantajı görüntü sayısının artmasıdır. Tam otomatik olmayan sistemlerde görüntü sayısının artması gereken insan gücünün de artması demektir. Mesela dört kamera izleyen bir güvenlik görevlisi kapanmalardan dolayı izleyemediği bazı bölgeleri de izleyebilmek için ekstra iki kamera daha kullanabilir ancak bu artırım bir noktadan sonra güvenlik görevlisinin işini zorlaştıracak ve bazı olayları gözden kaçırmaya veya tam olarak izleyememesine sebep olacaktır. Burada yapılması gereken, kapanmaları otomatik olarak çözmek ve sadece sahneye hâkim belirli kamera

görüşlerinden kapanmasız izleme yapmaktır. Bu sayede izlenmesi gereken kamera sayısı azaltılmış ve bir önceki örnekteki güvenlik görevlisinin işi kolaylaştırılarak daha verimli çalışması sağlanmış olur.

Sunulan sistemde çalışma prensibi olarak, kameralardan biri ana kamera seçilir ve diğer tüm kameralar yardımcı kamera olarak belirlenmiş olur. Ana kameradan alınan görüntüdeki kapanan bölgeler yardımcı kameralardan alınan bilgiler ile gerçek zamanlı olarak çözülür. Bu sayede ana kamera açısından kapanmasız bir görüş elde edilmiş olur. Günlük hayatta kullanılacak uygulamalarda, kameraların yerleri mükemmel hassasiyet ile bilinecek şekilde monte edilmesi zordur. Bu nedenle yardımcı kameralar keyfi şekilde farklı bakış açıları ile durabilmelidir. Yardımcı kameralardan alınan bilgiler ana kamera görüntüsü üzerine işlenmeden önce gerekli dönüşümlerden geçirilmelidir. Bu dönüşümlerin gerçekçi bir şekilde yapılabilmesi için sahnenin üç boyutlu yapısı çıkarılabilir. Üç boyutlu yapının çıkarılması her sahne için bir kere yapılacağı için bu ön çalışmanın alacağı zaman önemsizdir. Fakat üç boyutlu yapı çıkarılsa dahi, üç boyutlu bilgiyi işlemek iki boyutlu bilgiyi işlemekten daha uzun zaman alacağından gerçek zamanlı çalışacak bu uygulama için üç boyutlu bilgi kullanmak performans kaybına neden olabilir. Bu tez çalışmasında sunulan yöntem iki boyutlu resimler ile de yeterince tatmin edici ve pratikte uygulanabilecek derecede sonuçlar elde edilebileceğini göstermektedir.

Sistem, kapanmaları çözebilmek, kapanmadaki nesnelerin ana kamera görüntülerini tahmin edebilmek için daha önce gördüğü nesneleri tanıyan bir hafıza yapısı içerir. Bu hafıza sistemi kullanılarak iki boyutlu resimler ile gerçek zamanlı olarak üretilen kapanmasız görüntü, çoğu durumda gerçek, beklenen görüntü iken, bazı durumlarda da tam anlamıyla gerçekçi olmasa dahi tatmin edici görüntülerdir. Böyle bir sistem tamamen bilgisayar otomasyonu ile işlemeyen, insanlı uygulamalarda büyük kolaylıklar sağlar ve zaman kazandırır. Örneğin, askeri bir uygulamada farklı açılardan izlenebilen bir odaya farklı girişlerden giren askerler odanın bazı bölümlerini gerçekte göremeseler dahi odanın tamamından haberdar olarak hareket edebilirler. Bu çalışmada sunulan sistem, tek kamera ile tanıma ve izleme sistemlerine girdi üretecek kaynak sistem olarak da kullanılabilir.

Bu tez çalışmasında sunulan sistem, literatürdeki çeşitli yöntemleri etkili bir şekilde kullanarak ve literatürde derinlemesine araştırılmamış birçok konuda gelişmeler sağlayarak literatüre katkı sağlamaktadır. Arka plan ayrıştırma ile ilgili literatürdeki yöntemlerden daha hızlı çalışan ve daha başarılı sonuçlar üreten bir yöntem geliştirilmiş ve kullanılmıştır. Geliştirilen arka plan ayrıştırma yöntemi nesnelerin çerçevelerini kabaca tahmin ederek renk modellerinin daha anlamlı ve hızlı bir şekilde oluşturulmasını sağlamaktadır. Nesne takip etme ile ilgili literatürdeki yöntemlerin eksikleri incelenmiş ve birden çok kameradan gelen üç boyutlu bilgiyi kullanarak literatürdeki nesne takip etme yöntemlerinden çok daha hızlı çalışan ve daha başarılı bir nesne takip etme yöntemi geliştirilmiş ve kullanılmıştır. Geliştirilen yöntem epipolar geometrinin getirdiği bir özellik olan temel matrisleri kullanarak üç boyutlu bilgiyi kameralar arasında taşımakta ve bu sayede takip edilen nesneler ile ilgili birden çok kameradan gelen hem iki boyutlu hem de üç boyutlu bilgiyi bir arada kullanmaktadır. Nesne takip etme yönteminde kullanılan üç boyutlu bilgi örtüşen nesnelerin bulunduğu durumlarında daha başarılı sonuçlar vermektedir. Geliştirilen nesne takip yöntemi sadece nesne üzerindeki birkaç noktayı kullandığı için nesnelerin kamera görüşleri üzerindeki büyüklüklerinden bağımsız ve oldukça hızlı bir şekilde çalışabilmektedir. Bu tez çalışmasında sunulan sistemde ayrıca yer düzlemi homografilerinin ve epipolar geometrinin getirdiği temel matrislerin hesaplanması için otomatik bir yöntem geliştirilmiştir. Geliştirilen bu yöntem sistemin öğrenme aşamasında kontrollü bir şekilde tek bir nesne hareketini izleyerek ve nesnenin kamera görüşleri üzerinde oluşturduğu özel noktaları toplamakta ve sahne ile ilgili çalışma zamanında gerekli olacak bilgileri toplamaktadır.

Gerçek zamanlı kapanmasız görüntü üretimi için geliştirilen bu sistem Eylül 2007 tarihinde Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM) isimli uluslararası konferansta yayınlanmıştır [1]. Geliştirilen ve hızlandırılan sistem bir dergiye gönderilmek üzere hazırlanmaktadır. Bu çalışmanın farklı bir uygulaması ise uluslar arası bir konferansa gönderilmek üzere hazırlık aşamasındadır.

Kapanmasız görüntü üretimi yöntemi detaylı bir şekilde 2. Bölümde anlatılacaktır. 3. Bölümde sunulan yöntemin ilk aşaması olan öğrenme aşaması ve 4. Bölümde ikinci aşama olarak çalışma zamanı aşaması anlatılacaktır. 5. Bölümde yapılan deneyler ve 6. Bölümde genel sonuçlar ve öneriler verilecektir.

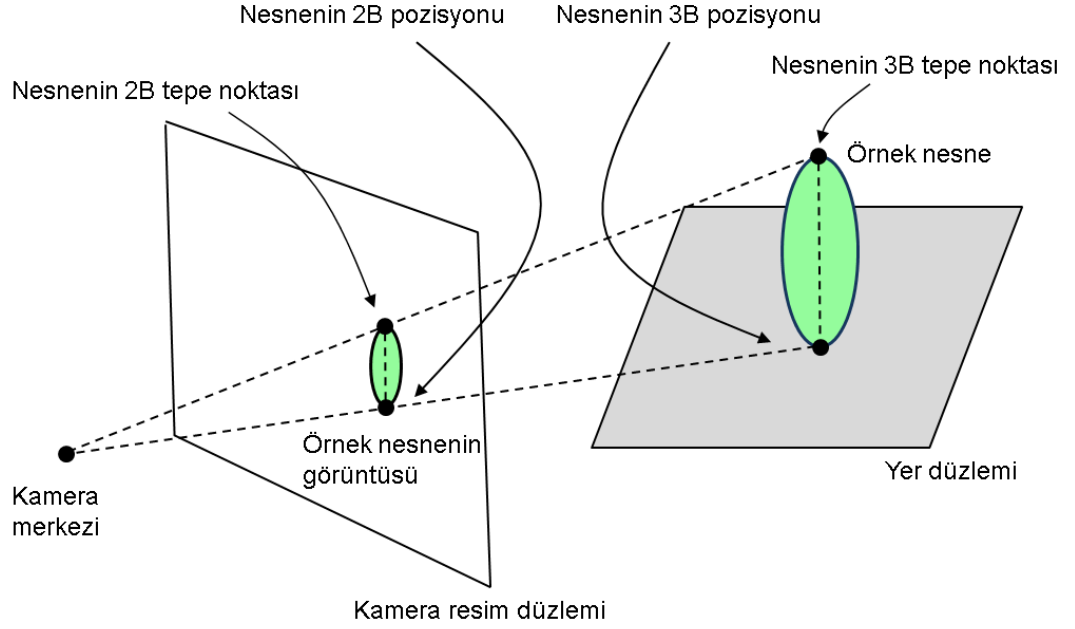
2 KAPANMASIZ GÖRÜNTÜ ÜRETİMİ

Kapanmasız görüntü üretimi, sahneyi görüntüleyen birden fazla kameradan alınan bilgilerin gerekli dönüşümler uygulandıktan sonra seçilen bir ana kamera görüntüsü üzerine düşürülmesi ile sağlanır. Ana kamera olarak seçilen kamera haricindeki kameraların tamamı yardımcı kamera olarak tanımlanır. Ana kamera seçimi öğrenme aşamasından önce bir kereye mahsus yapılır. Sistemin çalışabilmesi seçilen ana kameradan farklı olarak en az bir tane yardımcı kamera bulunmalıdır. Yardımcı kamera sayısı birden fazla olabilir. Bu durum işlem zamanını arttırarak sistemin çalışma hızını yavaşlatsa da birden fazla yardımcı kamera kullanımının sistemin başarısı üzerine olumlu etkisi vardır. Birden fazla ana kamera kullanımı bu tez çalışmasından sunulan yöntemin basit bir uzantısı olarak düşünülebilir. Birden fazla ana kamera kullananımı, sistemde var olan kameraları farklı şekillerde gruplayarak sağlanabilir. Bu durumda her kamera grubu için farklı bir ana kamera seçilir ve bu tez çalışmasında sunulan yöntemin bir kopyası o kamera grubu üzerinde çalıştırılır. Bu tez çalışmasında sunulan örneklerin tamamı tek ana kamera ve en az bir tane yardımcı kamera kullanılarak modellenmiş ve çözülmüştür.

Karmaşık sistemlerde, sistemi oluşturan öğeleri önceden tanımlamak sistemi ve çalışan yöntemi açıklamak için hayati öneme sahiptir. Çok kameralı kapanmasız görüntü üreten bir sistem için tanımlamalar bu kısımda verilecektir.

Kısaca *nesne*, 3B bir gerçek dünya nesnesinin herhangi bir 2B ifadesidir. 3B bir gerçek dünya nesnesine örnek olarak, sahnedeki bir masa, hareketli bir araç veya bir insan verilebilir. 3B nesnelerin 2B ifadeleri, kamera tarafından çekilmiş renkli bir resim veya nesnelerin hatlarını belli eden bir ikili harita olabilir. İkili harita sadece 0 veya 1 değeri alabilen piksellerden oluşan bir resimdir. Bu tez çalışmasında özel olarak *nesne*, 3B bir gerçek dünya nesnesinin kamera resim düzlemine projektif izdüşümünün bölümlenmiş, arka plandan ayrıştırılmış halidir. Bu şekilde tanımlanan bir nesne, 2B piksel bilgisi (renkli veya gri) ve 2B bir ikili harita (siluet, maske) bilgisi içerir.

Bir nesne ile ilgili çok önemli bir özellik de nesnenin *pozisyonu*dur. Bir nesnenin 3B pozisyonu, 3B ağırlık merkezinin 3B yer düzlemine dik izdüşümünden oluşan 3B bir noktadır. Bir nesnenin 3B pozisyonu nesne yer düzlemine temas etmese dahi daima 3B yer düzlemi üzerindedir. Bu özellik yer düzlemine temas etmeyen nesnelerin de takip edilebilmesini sağlar. Bir nesnenin 2B pozisyonu ise, o nesnenin 3B pozisyonunun kamera resim düzlemine projektif izdüşümü ile üretilen 2B resim noktası olarak tanımlanır. Aksi belirtilmedikçe nesne pozisyonu olarak ifade edilen bilgi nesnenin resim üzerindeki 2B pozisyonudur. Örnek olarak 2B ve 3B nesne pozisyonları Şekil 1’de gösterilmiştir.



Şekil 1: Bir nesnenin 2B, 3B pozisyonları ve tepe noktaları

Nesnelerin takip edilmesinde ve kapanmaların tespitinde kullanılan önemli bir nesne özelliği de *tepe noktası*dır. Bir nesnenin 3B tepe noktası, nesnenin 3B pozisyonundan yer düzlemine göre dik olarak çıkan doğru üzerinde ve yer düzleminden en uzak olan nesneye ait 3B olan. Bir nesnenin 2B tepe noktası ise, nesnenin kamera görüntüsünde, 2B nesne pozisyonundan kamera resim düzleminin y doğrultusunda çıkan doğru üzerinde ve 2B nesne pozisyonuna en uzak olan 2B nesne noktasıdır. Aksi belirtilmedikçe nesne tepe noktası olarak ifade edilen bilgi nesnenin resim üzerindeki 2B tepe noktasıdır. Örnek olarak 2B ve 3B nesne tepe noktaları Şekil 1’de gösterilmiştir.

Kamera görüşü, bir kameranın ürettiği resim olarak tanımlanır. Ana kamera ve yardımcı kameraların ürettiği resimler birbirinden farklıdır ve bu farkın sebebi her kameranın sahneye farklı bir bakış açısı ile bakmasıdır. Bu farkları belirtmek için *ana kamera görüşü* ve *yardımcı kamera görüşü* ifadeleri, kameraların ürettiği resimleri terimsel olarak temsil eder.

Nesne büyüklüğü, aksi belirtilmedikçe, bir nesnenin 2B görüntüsünün yüksekliği veya genişliğidir. Bu tez çalışmasında nesne büyüklüğü sadece göreceli bir tanım olarak kullanılmaktadır. Gerekli durumlarda bir nesnenin büyüklüğünün değişimi incelenmektedir. Örneğin, bir nesnenin görüntülendiği kameraya uzaklığına bağlı olarak büyüklüğü iki kat artmış veya azalmış olabilir. Nesne büyüklüğünün sayısal değer olarak yorumlanmasına gerek duyulmaması bir avantajdır.

Bu tanımlamalar ile birlikte kapanmasız görüntü üretimi kısaca şu şekilde özetlenebilir; bir nesne ana kamera görüşünde kapanmaya uğramış ise, o nesnenin ana kamera görüntüsü yardımcı kamera veya kameralardan alınan bilgiler kullanılarak tahmin edilir. Bunu başarabilmek için, bu nesnenin yardımcı kameralarda da kapanmaya uğrayıp uğramadığı tespit edilmelidir. En naif çözüm sahnenin 3B yapısı çıkarılarak nesnelerin pozisyonlarının ve büyüklüklerinin bulunması suretiyle nesnelerin herhangi bir kamera görüşü için kapanmaya uğrayıp uğramadığının tespitidir. Ancak sahnenin 3B yapısının çok farklı açılardan bakan kameralar ile çıkartılabilmesi için çok başarılı piksel eşlemesi yapılmalıdır. Bu zor bir işlemdir ve hataya çok açıktır. Literatürde kapanmaları çözmek için geliştirilmiş teknikler vardır [2, 3, 4], ancak bu teknikler birden fazla kameralı bir sistemde farklı açılardan bakan kameralar için uygun değildir. Bu tez çalışmasında çok hızlı ve etkin bir kapanma tespiti yöntemi geliştirilmiş ve tüm kamera görüşlerindeki kapanmaların tespiti için kullanılmıştır. Geliştirilen kapanma tespit yöntemi, farklı kamera görüşlerindeki nesnelerin pozisyonlarını ve epipolar geometri kullanarak nesne tepe noktalarını takip eder ve kamera görüşleri arasındaki tutarlılıklarını karşılaştırarak kapanmaları tespit eder.

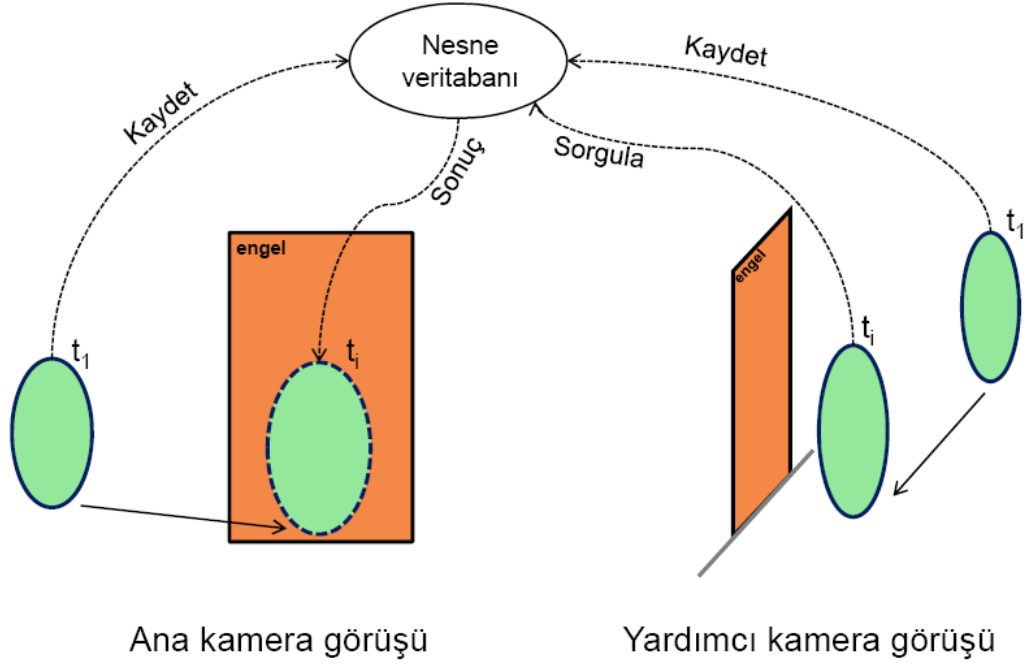
Kamera görüşlerindeki nesnelerin kapanma durumları tespit edildikten sonra ana kamera üzerinde kapanmaya uğramış nesnelerin ana kamera görüşü için olan görüntüleri tahmin edilmelidir. Bu tez çalışmasında sunulan yöntem, bu görüntü

tahminini efektif bir şekilde sadece 2B resimler kullanarak yapar. Buradaki ana fikir, nesnelerin kapanmaya uğramadan önceki görünür hallerinin görüntülerini tüm kamera görüşlerinden kayıt ederek dinamik bir veri tabanı oluşturmak ve kapanmaya uğrayan bir nesne için ana kamera görüntüsünü bu nesnenin veri tabanındaki kayıtlı görüntülerini kullanarak tahmin etmektir.

Kapanmasız görüntü üretimi temel olarak iki bölümden oluşmaktadır; öğrenme aşaması ve çalışma zamanı aşaması. Öğrenme aşamasında yer düzlemi homografisi, kameralar arasındaki epipolar geometri parametreleri, nesne büyüklüklerindeki değişimlerin hesaplanabilmesi için perspektif geometrinin sıfır noktaları ve nesnelerin arka plandan ayrıştırılabilmesi için her kamera için arka plan modelleri çıkartılır. Öğrenme aşamasında toplanan bu bilgiler ile ilgili detaylar 3. Bölümde verilecektir.

Öğrenme işlemi sistem başlatılmadan önce ve bir kereye mahsus olarak yapılır. Öğrenme aşamasında elde edilen bilgilerin tamamı kameraların pozisyonlarına bağlıdır. Örneğin bir kameranın yeri veya bakış açısı değişirse o kamera için arka plan modelinin güncellenmesi gerekir. Ayrıca yeri veya bakış açısı değişen bir kameranın diğer kameralar ile arasındaki yer düzlemi homografisi ve epipolar geometri parametreleri de değişeceği için böyle bir durumda öğrenme aşaması tekrar yapılmalıdır.

Çalışma zamanı aşamasında, her kamera görüşündeki nesneler arka plandan çıkartılarak nesne görüntülerinin sınırları belirlenir, nesnelerin pozisyonları hesaplanır ve kapanmaya uğrayan nesneler tespit edilir. Daha sonra, kapanmaya uğramamış nesnelerin görüntüleri daha sonra kullanılmak üzere her kamera görüşü için kaydedilir. Bu kayıt işlemi sistemin nesneleri öğrenme, tanıma işlemidir. Kapanmaya uğrayan nesnelerin görüntüleri ise daha önce kapanmaya uğramadığı zamanlardaki görüntülerinin kayıtları kullanılarak tahmin edilir. Tahmin edilen nesne görüntüleri doğru pozisyon ve büyüklükler hesaplanarak ana kamera görüşü üzerinde gerekli yere çizilir ve bu sayede kapanmasız görüntü üretimi gerçekleştirilmiş olur.



Şekil 2: Sunulan sistemin özeti

Şekil 2’de sunulan sistemin iki kamera için olan durumunun bir özeti şema şeklinde verilmiştir. t_1 anında kameralardan her ikisinde de görünür olan nesnenin pozisyonları, büyüklükleri, şekilleri ve görüntüleri her iki kamera açısından da kaydedilmektedir. Zaman t_1 anından t_i anına giderken, nesnenin kameralardan her ikisinde de görünür olduğu her video karesi için bu şekilde kayıt işlemi gerçekleştirilmektedir. Bu sayede nesne ile ilgili bilgi çeşitliliği sağlanmış ve sistem nesneyi görsel olarak tanımış olur. t_i anında nesne ana kamera görüşü için engelin arkasında kaldığından dolayı kapanmadır. Bu durumda nesnenin yardımcı kamera görüşündeki görüntüsü kullanılarak daha önce kayıt edilmiş görüntüler arasında ara yapılarak nesnenin ana kamera görüntüsü tahmin edilir. Bu aramadaki en önemli ölçüt, nesnenin t_i anındaki yardımcı kamera görüntüsünün bulunacak sonucun yardımcı kamera görüntüsüne benzemesidir. Bu sayede nesnenin kapanma anındaki pozuna benzer bir pozda kayıt bulunur ve bulunan kayıttaki ana kamera görüntüsü ana kamera görüşü üzerine yapıştırılır. Nesnenin ana kamera görüşü üzerindeki görüntüsünün pozisyonu ve büyüklüğü nesnenin yardımcı kamera görüşündeki pozisyonundan ve bulunan kaydın pozisyon ve büyüklük bilgilerinden yararlanılarak analitik olarak hesaplanır. Bu şekilde ana kamera görüşü üzerinde kapanmaya

uđramıř bir nesne iin grnt, pozisyon ve byklk bilgileri hesaplanarak kapanmasız grnt retimi gerekleřtirilir.

Kapanmasız grnt retimi yntemi daha detaylı bir řekilde anlatılmak iin alt problemlere ayrılmıřtır. Bu alt problemler sırası ile: nesnelerin arka plandan ayrıştırılarak nesne řekillerinin ve grntlerinin ıkartılması, nesne pozisyonlarının bulunması, nesnelerin takip edilmesi, nesneler iin kapanma testi ve kapanmadaki nesneler iin ana kamera grntsnn tahmin edilmesidir. Sıralanan alt problemler ve bunlara getirilen zmler takip eden blmlerde detaylı bir řekilde anlatılacaktır.

3 ÖĞRENME AŞAMASI

Öğrenme aşamasında, yer düzlemi homografisi, kameralar arasındaki epipolar geometri parametreleri, nesne büyüklüklerindeki değişimlerin hesaplanabilmesi için perspektif geometrinin sıfır noktaları ve nesnelerin arka plandan ayrıştırılabilmesi için her kamera için arka plan modelleri çıkartılır. Arka plan modelleri, arka plan resimlerini ve her piksel için kamera gürültüsünü temsil eden varyans değerlerini içerir. Öğrenme işlemi sistem başlatılmadan önce ve bir kereye mahsus olarak yapılır. Öğrenme aşamasında elde edilen bilgilerin tamamı kameraların pozisyonlarına bağlıdır. Örneğin bir kameranın yeri veya bakış açısı değişirse o kamera için arka plan modelinin güncellenmesi gerekir. Ayrıca yeri veya bakış açısı değişen bir kameranın diğer kameralar ile arasındaki yer düzlemi homografisi ve epipolar geometri parametreleri de değişeceği için böyle bir durumda öğrenme aşaması tekrar yapılmalıdır.

3.1 Arka plan modelleri

Öğrenme aşamasında ilk olarak her kamera için arka plan resimleri ve kamera gürültüsünü temsil eden piksellerin renk varyansları kaydedilir. Bu işlem her kamera için bağımsız olarak yapılır. Arka plan resimleri sahnenin hareketli nesne olmadan 30 resmi çekilerek ortalaması alınması ile elde edilir. Piksellerin renk varyansları da bu 30 resimden hesaplanır. Arka plan resmini temsil eden B ve piksellerin renk varyanslarını temsil eden Σ aşağıdaki formüller ile hesaplanır. Formüllerdeki toplama işlemleri her piksel için bağımsız olarak yapılır.

$$B = \frac{1}{30} \sum_{i=1}^{30} I_i \quad \Sigma = \frac{1}{30} \sum_{i=1}^{30} (I_i - B)^2$$

3.2 Yer düzlemi

Öğrenme aşamasında ikinci olarak kamera görüşlerine göre yer düzleminin çıkartılmasıdır. Yer düzlemi açıkça ifade edilmez ancak matematiksel olarak, kamera

çiftleri arasında doğrusal dönüşümler ile ifade edilebilir. Yer düzlemini ifade etmek için 3x3 lük bir homografi matrisi yeterlidir. Örneğin A kamerasından B kamerasına dönüşüm yapan bir H_{AB} homografi matrisi, A kamera görüşünde yer düzlemindeki 2B bir noktayı B kamera görüşüne taşır. Taşınan nokta yine yer düzlemi üzerindedir ve aynı 3B noktanın izdüşümüdür. A kamera görüşü üzerindeki 2B p^A noktası $[x^A \ y^A \ 1]^T$ ile homojen koordinatlarda ifade edilir ve aşağıdaki homografi denklemi ile B kamera görüşü üzerindeki 2B p^B noktasına taşınır.

$$p^B = H_{AB} p^A$$

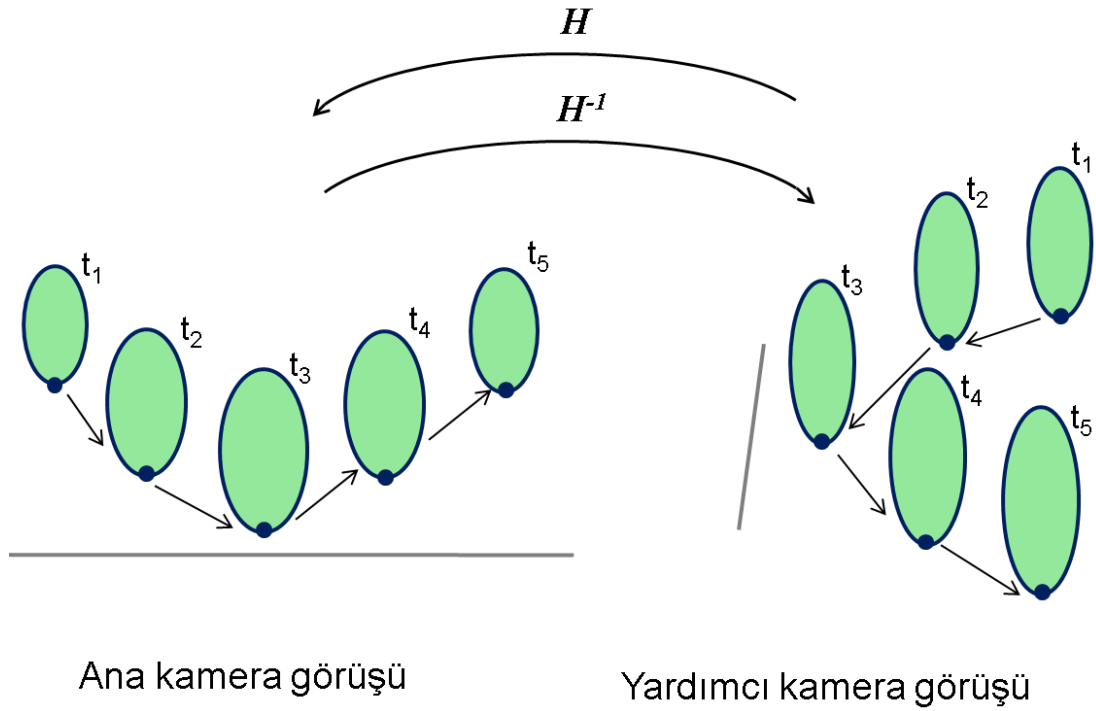
Öğrenme aşamasında her (A, B) kamera çifti için H_{AB} ve H_{BA} homografi matrisleri bulunur. Homografi matrisleri 3x3 lüktür ve pozitif bir çarpan ile genişletilebilirler. Örneğin yukarıdaki denklemin her iki tarafını da bir $k > 0$ katsayısı ile çarparsan sonuç değişmeyecektir ve denklemin sağ tarafında bu katsayı H_{AB} matrisi içine gizlenebilir. Bu özellik sayesinde H_{AB} matrisinin (3,3)'deki elemanı 1 olarak atanır ve geriye bulunması gereken 8 değişken kalır. Yukarıdaki homografi denklemi açıldığında

$$x^B = \frac{h_{11}x^A + h_{12}y^A + h_{13}}{h_{31}x^A + h_{32}y^A + 1}$$

$$y^B = \frac{h_{21}x^A + h_{22}y^A + h_{23}}{h_{31}x^A + h_{32}y^A + 1}$$

denklemleri elde edilir. Homografi matrisinin 8 bilinmeyen ($h_{11}, h_{12}, \dots, h_{32}$) hesaplanması için yer düzlemi üzerinde bulunan, her iki kameradan da görünen ve eşleştirilebilen noktalar gerekmektedir. Bu şekildeki bir nokta çifti yukarıdaki denklemde p^A ve p^B olarak yazılırsa 8 bilinmeyen için 2 denklem ortaya çıkar. Bu şekilde en az 4 nokta çifti ile en az 8 denklem oluşturulur ve doğrusal en küçük kareler yöntemi kullanılarak homografi matrisi hesaplanır. En küçük kareler yöntemine 4'ten fazla nokta verildiği takdirde daha başarılı sonuçlar vermektedir. Bu tez çalışmasında uygulanan yöntem bir aktörün sahnede hareket etmesi ve aktörün kamera görüşlerinde oluşturduğu nesnelerin pozisyonlarının en küçük kareler yöntemine verilmesi şeklindedir. Nesne pozisyonları yer düzleminde olduğu için bu

yöntem başarılı ve efektif bir şekilde her kamera çifti için yer düzlemini hesaplayabilmektedir. Şekil 3'te temsili olarak sahnede hareket eden bir nesne gösterilmiştir. Sahnede tek nesne olduğu için verilen herhangi bir anda nesne pozisyonları kameralar arasında otomatik olarak eşlenmektedir. Her iki kamerada da nesne pozisyonları alındıktan sonra en küçük kareler yöntemi ile sırasıyla yardımcı kamera görüşünden ana kamera görüşüne ve ana kamera görüşünden yardımcı kamera görüşüne dönüşümleri temsil eden H ve H^{-1} homografi matrisleri hesaplanır.



Şekil 3: Yer düzlemi için toplanan noktalar

3.3 Epipolar geometri

Öğrenme aşamasında üçüncü olarak epipolar geometriyi tanımlayan F temel matrisleri hesaplanır. A kamerasından B kamerasına dönüşümü temsi eden F_{AB} temel matrisi aşağıdaki denklem ile A kamera görüşü üzerindeki p^A noktasını B kamera görüşü üzerinde bir l^B doğrusuna taşır.

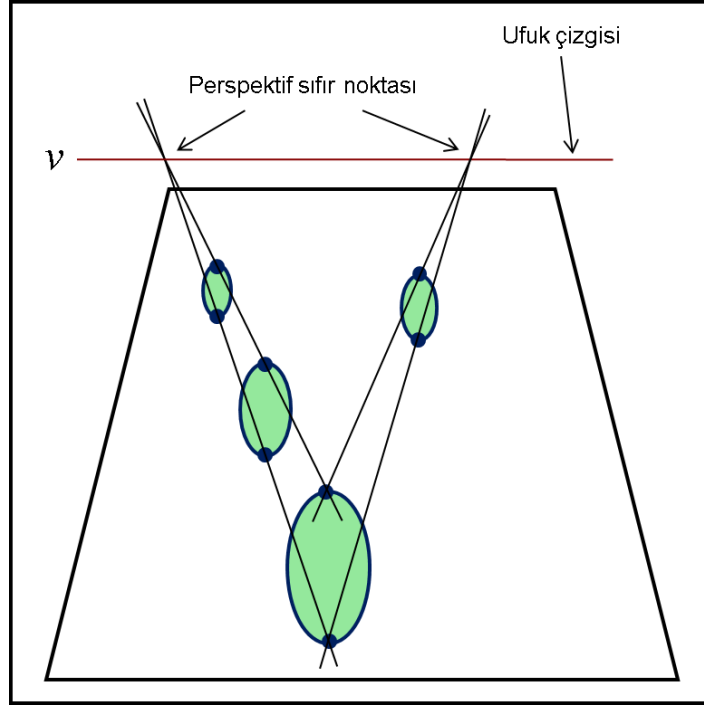
$$l^B = F_{AB} p^A$$

Burada $ax^B+by^B+c=0$ olacak şekilde $l^B = [a \ b \ c]^T$ ile ifade edilir. p^A noktasına B kamera görüşünde karşılık gelen nokta l^B doğrusu üzerindedir. Epipolar geometriyi tanımlayan F temel matrislerinin avantajı, H homografi matrislerinden farklı olarak tek bir düzlem için sınırlı olmamalarıdır. Düzlemsel olmayan noktalar için de çalışırlar.

Temel matrisler de yer düzlemi homografileri gibi her (A, B) kamera çifti için F_{AB} ve F_{BA} olarak hesaplanır. H matrislerinden farklı olarak F matrislerinin hesaplanabilmesi için en küçük kareler yöntemine verilecek 2B noktaları oluşturan 3B noktaların aynı 3B düzlem üzerinde olmaması gerekir. Bu gerekliliğin sebebi F matrisinin tanımı gereği düzlemsel olmayan noktalar için de çalışabilmesidir. Örneğin sadece yer düzlemi üzerinde bulunan noktalar ile ancak homografi matrislerinin vereceği kadar bilgi edinilebilir, F matrisleri hesaplanamaz. Bu sorunu aşmak için önerilen tez çalışmasında nesnelerin pozisyonlarının yanı sıra kamera görüşlerindeki nesneye ait ve yer düzleminden gelmeyen diğer noktalar da kullanılmaktadır. Örneğin nesne tepe noktaları.

3.4 Perspektif sıfır noktaları ve doğruları

Öğrenme aşamasında son olarak nesnelerin büyüklük değişimlerinin hesaplanabilmesi için gerekli olan sıfır noktaları ve doğruları hesaplanır. Perspektif geometrideki sıfır noktaları, paralel olan 3B doğruların 2B resim düzleminde birleştikleri noktalardır. Bu noktalar ile nesnelerin kameraya göre uzaklıklarına bağlı olarak görüntülerinin büyüklükleri bağıl olarak hesaplanabilir. Bir nesne kameradan uzaklaşırken kamera görünüşü üzerinde büyüklüğü azalacaktır. Benzer şekilde kameraya yaklaşan bir nesnenin büyüklüğü artacaktır. Bu bilgi ışığında bir nesnenin kameradan uzaklığındaki değişim o nesnenin görüntüsünün büyüklüğündeki değişim ile doğrusal olarak orantılıdır. Bu ilişki bir sonraki bölümde daha detaylı bir şekilde anlatılmaktadır. Bu ilişkinin kullanılabilmesi için kameraların ufuk çizgilerinin bulunması gerekmektedir. Nesne pozisyonlarının ufuk çizgilerine uzaklıkları ile büyüklük değişimleri hesaplanacaktır.

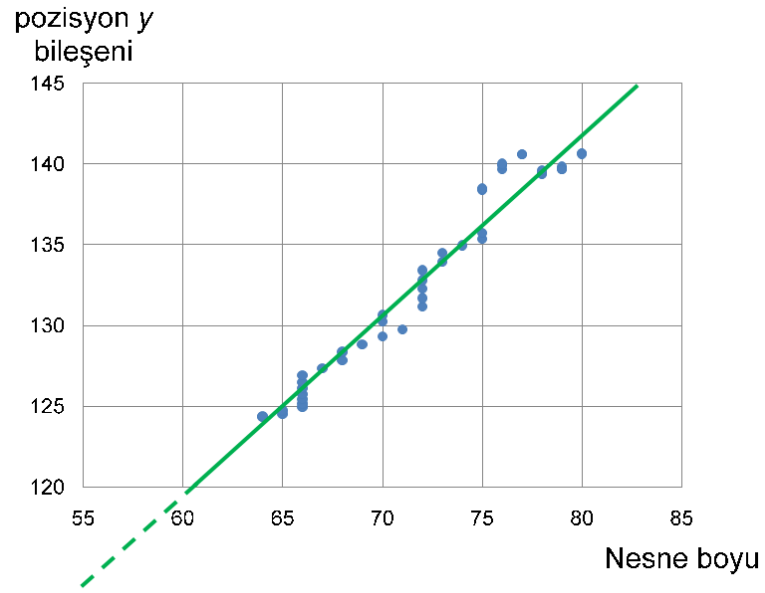


Şekil 4: Perspektif sıfır noktaları ve ufuk çizgisi

Sahne hareket eden aktörün kamera görüşünde oluşturduğu nesnenin pozisyonu sürekli yer düzleminden gelmektedir ve herhangi çakışık olmayan iki pozisyon yer düzlemi üzerinde bir doğru oluşturur. Benzer şekilde F matrislerinin hesaplanmasında kullanılan tepe noktaları da aynı düzlem üzerindedir. Herhangi iki nesne pozisyonuna denk gelen tepe noktalar alındığında yer düzleminde oluşan doğruya paralel ancak sanal düzlem üzerinde oluşan ikinci bir doğru elde edilir. Bu iki doğrunun paralellliği kullanılarak sıfır noktalar hesaplanmaktadır. Ufuk çizgisi bu sıfır noktalarının oluşturduğu çizgidir. Şekil 4’de bir kamera görüşü için perspektif sıfır noktaları ve ufuk çizgisi gösterilmiştir.

Ufuk çizgisinin hesaplanabilmesi için en az iki sıfır noktası yeterlidir ancak sunulan bu çalışmada sahnedeki hareketli aktörün her pozisyon değişimi alınmakta ve en az kareler yöntemi ile başarılı bir şekilde ufuk çizgileri hesaplanmaktadır. Şekil 5’de alınan her nesne yüksekliği ve pozisyonuna bağlı olarak ufuk çizgisinin hesaplanması gösterilmiştir. Her nesne görüntüsü için bir pozisyon ve kamera görüşü üzerindeki yüksekliği grafikte bir nokta oluşturur. Bu şekilde oluşturulan nokta kümesine en az kareler yöntemi ile uydurulan doğru grafikte gösterilmiştir. Grafikteki doğru grafiğin y eksenini ile kesiştirilirse nesne boyunun nerede 0 olacağını,

ufuk çizgisinin yerini, verir. Şekil 4’de de görüleceği gibi nesne boyunun 0 olacağı pozisyonun y bileşeni ufuk çizgisini gösterir.



Şekil 5: Ufuk çizgisinin bulunması

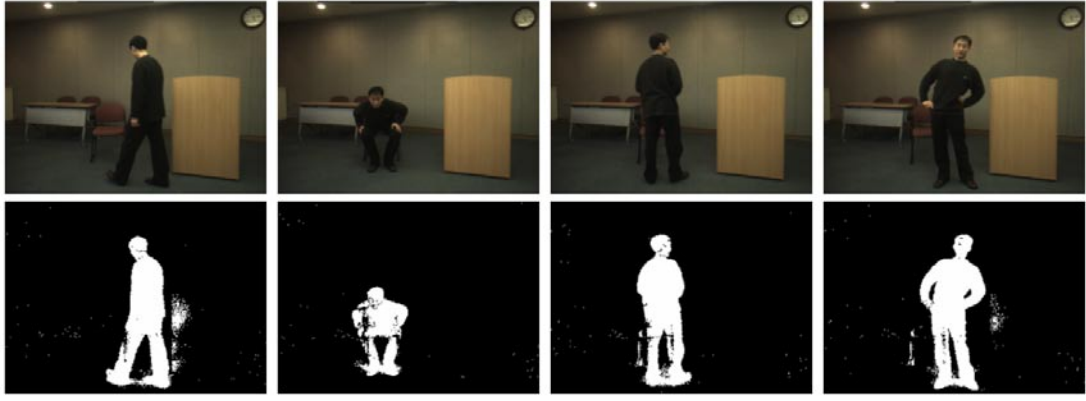
4 ÇALIŞMA ZAMANI AŞAMASI

4.1 Nesnelerin arka plandan ayrıştırılması

Nesnelerin arka plandan ayrıştırılması, kısaca nesne ayrıştırması, hem öğrenme hem de çalışma zamanı aşamaları için çok önemli bir işlemdir. Nesne ayrıştırma işlemi sahnedeki nesnelerin arka plandan çıkartılarak sadece nesnenin görüntüsünün elde edilmesi işlemidir. Başarılı pozisyon, büyüklük değişimi ve nesne görüntüsü tahminleri için nesne ayrıştırması da başarılı olmalıdır. Bu işlem sonucunda 0 veya 1 değeri almış piksellerden oluşan bir ikili nesne maskesi elde edilir. Şekil 6 ve Şekil 7’de örnek nesne ayrıştırma işlemleri ve ikili maskeler verilmiştir. Bu maske kullanılarak bir nesnenin görüntüsü kameranın çektiği görüntüdeki arka plandan ayrıştırılarak elde edilebilir.

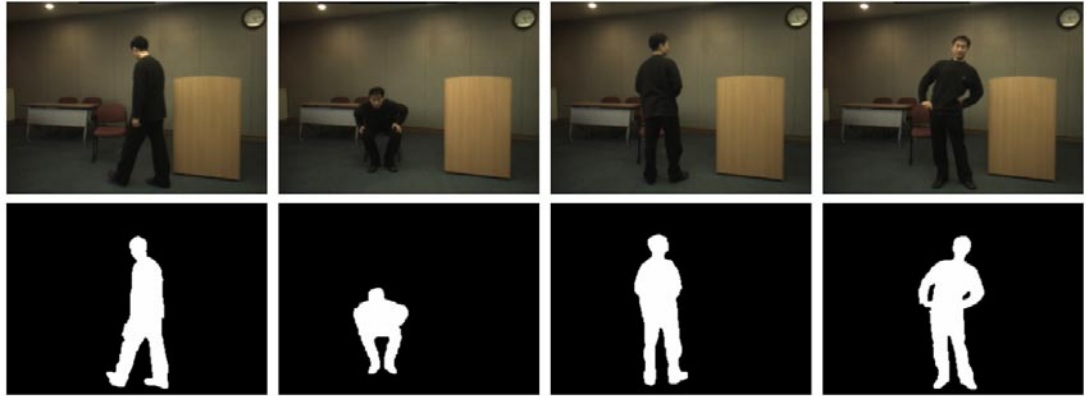
İki tür nesne ayrıştırma yöntemi vardır. Piksel tabanlı ve global nesne ayrıştırma yöntemleri. Piksel tabanlı nesne ayrıştırma yöntemleri [5, 6] oldukça hızlıdır ancak kolaylıkla hatalı ve gürültülü sonuçlar üretebilirler. Nesne görüntüsünde bazı kısımlar eksik çıkabilir veya nesne görüntüsüne arka plandan bazı detaylar karışmış olabilir. Piksel tabanlı nesne ayrıştırma yöntemleri renk değişimlerine ve kamera gürültüsüne oldukça hassastırlar. Piksel tabanlı nesne ayrıştırma yöntemlerinin dezavantajları genel olarak her piksel için bağımsız karar verilmesinden dolayıdır. Kamera gürültüsünden veya çok küçük ışık değişimlerinden kaynaklanan farklılıklar tek bir piksele bakıldığı zaman tespit edilemez. Bu nedenle piksel tabanlı nesne ayrıştırma yöntemleri gürültülü pikselleri nesne pikselleri olarak işaretlemektedir.

Global nesne ayrıştırma yöntemleri ise [7, 8, 9] çok başarılıdır fakat gerçek zamanlı işlemler için kullanıldığı zaman, özellikle birden fazla kamera kullanıldığı durumlarda, fazla zaman aldıkları için pratik değildirler. Global nesne ayrıştırma yöntemleri, nesne ayrıştırma problemini ikili etiketleme problemi olarak modellerler. Bu modelleme her pikselin nesne veya arka plan olma olasılığını ve her pikselin komşu piksellere göre durumunu içerdiğinden gürültüye karşı piksel tabanlı algoritmalara göre daha güvenilirdir.



Şekil 6: Piksel tabanlı ayırıştırma algoritması

Her bir kamera için nesne ayırıştırma işlemi birbirinden bağımsızdır. Bu nedenle daha hızlı çalışan bir sistem elde etmek için her kamera için nesne ayırıştırma işlemi ayrı bir işlemcide yaptırılabilir ancak bu işlem hem pahalıdır hem de farklı kameralar için video kareleri arasında çok kesin eş zamanlılık gerektirir.



Şekil 7: Global ayırıştırma algoritması

Bu tez çalışmasında başarılı ve hızlı bir nesne ayırıştırma yöntemi geliştirilmiş ve kullanılmıştır. Geliştirilen nesne ayırıştırma yöntemi tek bir işlemci ile 4 kamera için nesne ayırıştırma işlemini gerçek zamanda yapabilmektedir. Geliştirilen bu yöntem global bir yöntemdir ve [7, 8, 9] çalışmalarında olduğu gibi çizge kesme yöntemine dayanmaktadır [11]. Çizge kesme yöntemi nesne ayırıştırma işlemini piksellerin 0 veya 1 değeri alacağı ikili etiketleme problemi olarak modeller. İkili etiketleme problemi bilgisayarla görme alanında bir enerji denklemi olarak tanımlanır ve aranan etiketleme tanımlanan enerjiyi en az yapan etiketlemedir.

Çizge kesme yöntemine dayanan nesne ayrıştırma yöntemleri iki bileşenden oluşurlar; veri enerjisi ve düzgünleştirme enerjisi. Bu enerjiler her piksel için hesaplanır ve bu enerjiyi en düşük yapacak nesne ayrıştırması çizge kesme yöntemi ile bulunur. Önceki çalışmalardan farklı olarak burada geliştirilen yöntemde hem veri hem de düzgünleştirme enerjilerinin hesaplanması basitleştirilmiş ve hızlandırılmıştır. Ayrıca veri enerjisi daha anlamlı ve başarılı şekilde hesaplanmaktadır.

Çizge kesme yöntemi kullanan yaklaşımlarda nesne ayrıştırma yöntemi ikili etiketleme problemi olarak tanımlanmaktadır. Bu işlem sonunda her piksel iki etiketten birine sahip olacaktır. Arka plan olarak işaretlenen pikseller 0, nesne olarak işaretlenen pikseller ise 1 etiketine sahip olacaktır.

Nesne ayrıştırma yöntemini açıklamadan önce bazı tanımlamalar yapmak gerekir. B ve I ile temsil edilen değerler sırası ile arka plan resmi ve kameradan gelen en güncel resim karesidir. Bu durumda nesne ayrıştırma işlemi I üzerindeki pikseller için yapılacaktır. V , I resmi üzerindeki tüm piksellerin kümesi olsun. $\mathcal{E}$ ise I resmi üzerindeki tüm komşu piksellerin oluşturduğu ikili grupların kümesi olsun. Bir pikselin komşuları $2B$ resim düzleminde dikey ve yataydaki komşularıdır. İkili etiketleme problemi olarak nesne ayrıştırma yönteminde bir $f = \{f_r\}$ etiketlemesi aranmaktadır. Bu f etiketlemesi V kümesindeki her r pikseline bir $f_r \in \{0, 1\}$ etiketi atayacaktır ve aşağıdaki enerjiyi minimize ederek bulunur.

$$E(f) = \sum_{r \in V} E_1(f_r) + \lambda \sum_{(r,s) \in \mathcal{E}} E_2(f_r, f_s)$$

Burada $E_1(f_r)$ veri enerjisi veya renk enerjisi, $E_2(f_r, f_s)$ düzgünleştirme enerjisi veya karışıklık enerjisidir. Veri enerjisi r pikseline f_r etiketinin atanmasının maliyetini ölçer ve sadece r pikseline ve f_r etiketine bağlıdır. Örnek olarak, f_r etiketi 0 iken r pikselinin nesneye ait olma ihtimali yüksek ise $E_1(f_r)$ yüksek, aksi halde düşük olacaktır. Bu sayede E_1 enerjisini en düşük yapacak f aranmaktadır. Düzgünleştirme enerjisi ise komşu r ve s piksellerine sırası ile f_r ve f_s etiketlerinin atanmasının maliyetini ölçer ve hem r ve s piksellerine hem de f_r ve f_s etiketlerine bağlıdır. Örnek

olarak, birbirine benzemeyen r ve s pikselleri için f_r ve f_s farklı ise $E_2(f_r, f_s)$ düşük aksi halde yüksek olacaktır. Bu sayede f in değeri değiştiği yerlerin, diğer bir deyişle nesne sınırlarının, birbirine benzemeyen komşu piksellerin sınırları olması amaçlanır. Bulunacak f etiketlemesi toplam maliyeti veya enerjiyi minimize edecek olan etiketlemedir. Buradaki λ sabiti veri ve düzgünleştirme enerjilerinin birbirlerine göre bağıl önemliliklerini tanımlar. λ sabiti birçok deneme ile en az hataya sebep olacak şekilde deneylerden önce ortalama 50 olarak bulunmuştur ve tüm deneylerde bu şekilde kullanılmıştır.

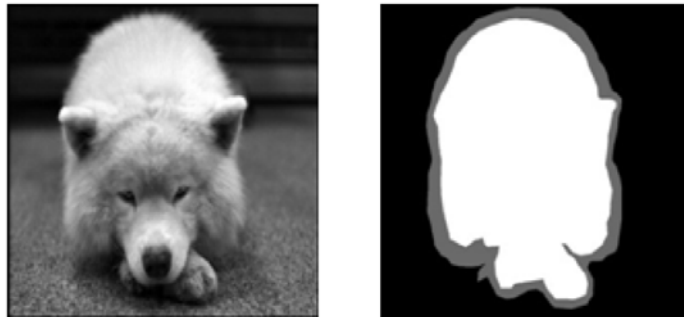
Veri/Renk Enerjisi

Renk enerjisi bir r pikselinin nesneye ($f_r = 1$) veya arka plana ($f_r = 0$) ait olma ihtimalini modeller. En basit şekilde renk enerjisi [10] çalışmasında olduğu gibi $\|I_r - B_r\|$ cinsinden ifade edilebilir. Bu yaklaşım çok basit ve kolay uygulanabilir olmasına karşın nesne ve arka plan ihtimalleri birbirinden bağımsız değildir. Bu modelde eğer bir pikselin nesne olma ihtimali $\|I_r - B_r\|$ cinsinden ifade edilirse, aynı pikselin arka plan olma ihtimali nesne olma ihtimalinin bir çeşit tümleyeni olacak şekilde tanımlanır. Renk enerjisinin bu tanımı nesne ve arka planın benzer renklerde olduğu bölgelerde çalışmamaktadır. Örneğin, kırmızı bir arka plan üzerindeki kırmızı bir nesne bu modele göre çok yüksek ihtimalle arka plandadır ve çok düşük ihtimalle nesnedir, oysa bu nesneye ait kırmızı piksellerin nesne olma ihtimalleri de arka plan olma ihtimalleri kadar yüksek olabilmelidir. Bu tez çalışmasında kullanılan nesne ayrıştırma işleminin başarı gereksinimi yüksek olduğu için renk enerjisi, [7, 8] çalışmalarında olduğu gibi, nesne ve arka plan ihtimalleri için uzaysal global Gauss Karışım Modelleri (GKM) ile modellenmiştir. Burada nesne ve arka plan için birbirinden bağımsız iki ayrı GKM kullanılmıştır. Bu yöntemin bir avantajı da nesne ve arka plandaki renk varyasyonlarını farkı şekilde modelleyebilmesidir. Örneğin bir sahnedeki arka plan çok çeşitli renklere sahip ise, arka plan için modellenen GKM daha büyük tutulabilir. Önceki çalışmalardan farklı olarak bu tez çalışmasında, GKM kullanan renk enerjisi tanımı basitleştirilmiş, kolay hesaplanabilir hale getirilmiş ve arka planda olup da nesne ile konum olarak alakasız olan pikseller atılarak GKM ler daha başarılı hale getirilmiştir.

Nesne renk modeli $p(I_r | f_r = 1)$ I_r pikselinin nesneye ait olma ihtimalini temsil eder ve I resmindeki kesinlikle nesneye ait olan pikseller kullanılarak öğrenilmektedir. Nesne renk modeline benzer şekilde arka plan renk modeli de $p(I_r | f_r = 0)$ I_r pikselinin arka plana ait olma ihtimalini temsil eder ve I resmindeki kesinlikle arka plana ait olan piksellerinden öğrenilmektedir. Kesinlikle nesneye ait veya kesinlikle arka plana ait pikselleri seçebilmek için renk modelleri öğrenilmeden önce I resmi üç bölgeye ayrılarak bir üçlü maske resmi elde edilir. Bu işlemde bir pikselin alabileceği değerler şunlardır: kesinlikle nesne, kesinlikle arka plan, belirsiz. Örnek bir üçlü maske Şekil 8’de verilmiştir. Üçlü maskenin hesaplanması aşağıdaki kural ile çok hızlı bir şekilde gerçekleştirilir.

$$T_r = \begin{cases} T_F & d_r > t_f \\ T_B & d_r < t_b \\ T_U & t_b < d_r < t_f \end{cases}$$

Burada $d_r = (I_r - B_r)^T \Sigma_r^{-1} (I_r - B_r)$ olarak (Mahalanobis uzaklığı [15]) tanımlıdır. t_f ve t_b sırası ile nesne ve arka plan testi için kullanılan eşik değerleridir. T_F , T_B ve T_U bir pikselin üçlü maskede alabileceği değerleri temsil eder ve sırası ile kesinlikle nesne, kesinlikle arka plan ve belirsiz bölgeleri temsil eder. I resmindeki her piksel bu teste tabi tutulur ve T_F , T_B , T_U değerlerinden birini alır. Σ_r matrisleri öğrenme aşamasında arka plan çıkarılırken bir kereye mahsus hesaplanırlar. Σ_r renkli bir resimde 3×3 ’lüktür ve r pikselinin kovaryansını temsil eder.



Şekil 8: Örnek bir üçlü maske resmi

Nesne ve arka plan renk modellerinin öğrenilmesi öğrenme işlemine katılan piksel sayısı arttıkça daha uzun sürmektedir. Öğrenme işlemi hızlandırmak için nesnelerin kabaca çerçeveleri çıkartılır ve sadece bu çerçeve içinde kalan pikseller kullanılarak renk modelleri öğrenilir. Nesne çerçeveleri, çok küçük olmayan birleşik piksel kümeleri için bu kümeleri kapsayan en küçük dikdörtgenler olarak hesaplanır. Bu işlem için seçilen piksel kümeleri, üçlü maskedeki kesinlikle nesne veya belirsiz piksellerden oluşmaktadır. Bu yöntem hem işlem zamanını kısaltmakta hem de ilgisiz pikselleri işleme almayarak başarıyı yükseltmektedir. Şekil 10'da, örnek bir üçlü maske ve nesne çerçevesi gösterilmiştir.

Nesne renk modeli matematiksel olarak bir pikselin nesneye ait olma olasılığını modeller ve şu şekilde tanımlanmıştır;

$$p(I_r | f_r = 1) = \sum_{k=1}^K w_k^f N(I_r | \mu_k^f, \Sigma_k^f)$$

Burada N bir Gauss dağılımıdır. w_k^f , μ_k^f ve Σ_k^f sırası ile nesne GKM sinin k -inci ağırlık, ortalama ve kovaryans bilgileridir. Arka plan renk modeli de benzer şekilde f yerine b üssü ile tanımlanır. Her iki modelde de bileşen sayısını temsil eden K sabiti 3 olarak seçilmiştir.

GKM lerin öğrenilmesi, standart EM algoritması [6] kullanılarak yapıldığında oldukça zaman alan bir işlemdir. [7]'deki çalışmada GKM lerin tekrarlamalı olarak öğrenilmesi önerilmiştir. Bu tez çalışmasında kullanılan yöntem de benzer şekilde çalışmaktadır. GKM leri öğrenmek için işleme girecek pikseller üzerinde k-means kümeleme algoritması [14] çalıştırılır ve her küme GKM'nin bir bileşenini oluşturur. Bulunan her kümenin ayrı ayrı ortalama ve kovaryansı hesaplanır ve GKM ortalama ve kovaryansları olarak kullanılırlar. GKM'deki ağırlıklar olarak ise her kümedeki piksel sayısının işleme giren tüm piksel sayısına oranı kullanılmaktadır. Bu şekilde öğrenilen GKM ler nesne çerçeveleri kullanıldığı için [7, 8] çalışmalarına göre daha az işlem gücü gerektirmektedir.

Nesne ve arka plan renk modelleri tanımlandıktan sonra veri/renk enerjisi şu şekilde tanımlanmaktadır:

$$E_1(f_r) = \begin{cases} -\log p(I_r | f_r = 0) & f_r = 0 \\ -\log p(I_r | f_r = 1) & f_r = 1 \end{cases}$$

Düzgünleştirme/Karşıtlık Enerjisi

Düzgünleştirme enerjisi genel olarak resim gradyanları cinsinden tanımlanmaktadır. Nesne sınırlarında resim gradyanları yüksek olacağından buraları tespit etmek kolaydır. Şekil 9’da örnek bir resim ve resim gradyanı gösterilmiştir. Ancak nesne ve arka plan renkleri birbirlerine benziyorsa veya arka planda nesne sınırları ile karışacak derecede gradyan var ise doğru nesne sınırlarını tespit etmek zorlaşmaktadır. Bu işlemdeki başarıyı arttırmak için B arka plan resminin gradyanları I resminden çıkartılır ve sadece nesneye ait gradyanlar kullanılmış olur.



Şekil 9: Örnek bir resim ve hesaplanmış resim gradyanı. Resim gradyanı nesne sınırlarında yüksek değerlere sahiptir.

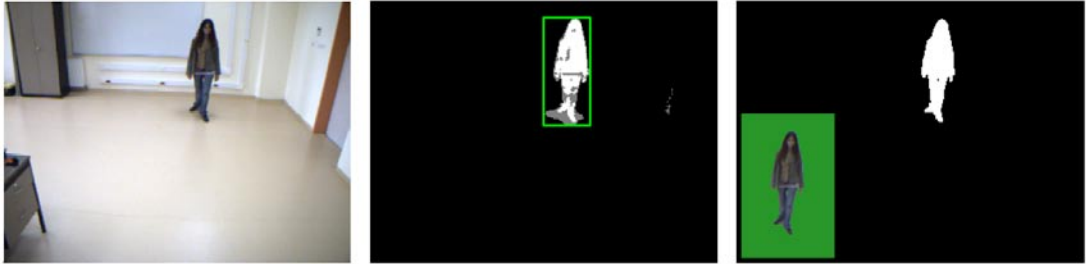
Bu tez çalışmasında sunulan yöntem düzgünleştirme enerjisi [8] çalışmasındakine benzer bir şekilde tanımlanır.

$$E_2(f_r, f_s) = |f_r - f_s| \cdot \exp \left\{ -\beta \left(\|I_r - I_s\|^2 + \frac{\|I_r - I_s\|^2}{1 + \|B_r - B_s\|^2} \right) \right\}$$

Buradaki β [9] çalışmasında $2 / \langle \|I_r - I_s\|^2 \rangle$ olarak tanımlanmıştır. $\langle . \rangle$ beklenti operatörüdür ve aşağıda verildiği şekilde tüm resim üzerinde hesaplanır.

$$\langle \|I_r - I_s\|^2 \rangle = \frac{\sum_{(r,s) \in \mathcal{E}} \|I_r - I_s\|^2}{\sum_{(r,s) \in \mathcal{E}} 1}$$

Hızlı çalışma zamanı için birçok doğal resim üzerinde nesne ayrıştırması yapılmış ve β değeri ortalama 0,0025 olarak bulunmuştur. β değeri bu tez çalışmasında gerçekleştirilen tüm deneylerde sabit olarak kullanılmıştır.



Şekil 10: Nesne ayrıştırma sisteminin adımları

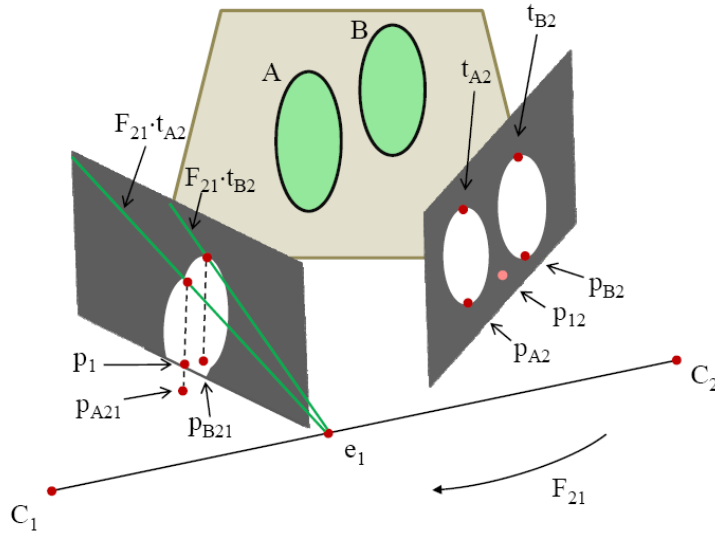
Renk ve düzgünleştirme enerjileri tanımlandıkta sonra enerji fonksiyonu [11] çalışmasında önerilen azami akış algoritması ile minimize edilerek f ve f' 'e karşılık gelen ikili maske bulunur.

4.2 Nesne pozisyonları ve kapanma testi

Bu tez çalışmasında sunulan sistem nesnelerin daima yer düzlemi üzerinde hareket ettiklerini varsaymaktadır. Bu varsayım [12, 13] çalışmalarında da kabul edilmiş ve doğal olarak sağlanmaktadır. Bu varsayım ile 2B resim düzlemindeki bir nesne pozisyonu 3B nesnenin kameraya olan uzaklığı ile ilişkilendirilebilir. Nesne

pozisyonunun daha önce verilen tanımına göre bir nesnenin kameraya olan uzaklığı, derinliği, 2B nesne pozisyonunun y bileşeni ile ters orantılıdır, bkz. Şekil 4. Burada resim düzleminin koordinat merkezi sol üst köşe olarak alınmıştır ve kameranın resim düzleminin x eksenine yer düzlemine kabaca paraleldir. Bu varsayımlar da kolaylıkla sağlanabilmektedir.

Nesne pozisyon bilgisi çıkarıldıktan sonra tek bir nesneyi bir ya da birden fazla kamera ile takip etmek oldukça kolaydır. [12] birden fazla kamera ile tek bir dinamik nesnenin pozisyonunu, nesnenin arka plandan ayrıştırılması çok başarılı olmasa dahi, başarılı bir şekilde bulabilmektedir. Ancak sahnede birden fazla dinamik nesne ile pozisyon takip işlemi birden fazla kameralı bir sistemde dahi kolay bir problem değildir. Bunun sebebi nesne görüntülerinin birbirleri ile örtüşmesi ve bazı nesnelerin bazı kamera görüşlerinde birleşerek tek bir nesne gibi görünmeleridir. Literatürde birden fazla nesneyi takip eden birden fazla kameralı sistemler mevcuttur [12, 13]. Bu sistemler yer düzlemi kısıtını kullanırlar ve yer düzlemi üzerinde hareket eden nesneleri takip etmeye çalışırlar. Bu yaklaşımın asıl dezavantajı sadece yer düzlemi kısıtını kullanmalarıdır. Bazı nesne durumlarında bu yaklaşım yeterli olmamaktadır.



Şekil 11: Kapanma testi

Şekil 11'deki durum sadece yer düzlemi kısıtının birden fazla nesnenin takibi için yeterli olmadığına bir örnektir. Şekildeki sahnede A ve B olarak adlandırılan iki

3B nesne, C_1 ve C_2 olarak adlandırılan iki kamera vardır. Bu kameraların sahneye bakış açıları oldukça farklıdır. C_1 kamerasının görüşünden bir ve C_2 kamerasının görüşünden ise iki nesne görüntüsü vardır. Bu durumda eğer sadece yer düzlemi kısıtını kullanılırsa nesnelerin pozisyonlarının tahmini için iki seçenek vardır. C_1 kamera görüşündeki nesne pozisyonunu yer düzlemi kısıtı ile C_2 kamera görüşüne taşımak ya da C_2 kamera görüşündeki nesne pozisyonlarını C_1 kamera görüşüne taşımak. İki durumda da taşınan noktalara denk gelen birer nesne olup olmadığı kontrol edilmelidir. Birinci durumda, eğer C_1 kamera görüşünden C_2 kamera görüşüne olan yer düzlemi dönüşümü uygulanırsa C_2 kamara görüşünde şekilde p_{12} ile gösterilen nokta elde edilir. Bu nokta C_2 kamera görüşündeki nesnelerin arasına denk gelmektedir ve problemi çözmemektedir. İkinci durumda, C_2 kamera görüşünden C_1 kamera görüşüne olan yer düzlemi dönüşümü kullanılırsa şekilde p_{B2} ve p_{A2} ile gösterilen noktalar C_1 kamera görüşünde p_{B21} ve p_{A21} ile gösterilen noktalara denk gelmektedir. Bu durumda da problem tam olarak çözülememiştir. Burada A nesnesinin C_1 kamera görüşünde görünür olup olmadığı kesin değildir.

Bu tez çalışmasında sadece yer düzlemi kısıtına dayanmayan ve birden fazla nesneyi başarılı bir şekilde takip edebilen birden fazla kameralı bir nesne takip yöntemi geliştirilmiş ve kullanılmıştır. Sunulan yöntem epipolar geometrideki F temel matrislerini kullanarak nesneler hakkında 3B bilgi ile daha başarılı sonuçlar elde etmektedir. F temel matrisi ile C_2 kamera görüşündeki nesnelerin tepe noktalarını C_1 kamera görüşüne taşınarak C_1 kamera görüşündeki tepe noktalar tahmin edilebilir. t_{A2} ve t_{B2} ile gösterilen tepe noktalar F_{21} ile dönüştürüldüklerinde C_2 kamera görüşünde epipolar doğrular oluştururlar. Bu epipolar doğrular ile yer düzleminden gelen nesne pozisyonları arasındaki dik doğrular kesiştirilerek yeni tepe noktalar tahmin edilmektedir, bkz. Şekil 11. Bu yöntem bir kamera görüşünde kapanmaya uğramış nesnelerin takibini ve resim düzlemindeki yükseklik değişimini incelemek için uygundur.

Epipolar geometrinin temel matrisini kullanmak nesne takibi işlemini başarılı bir şekilde gerçekleştirmeyi sağlar. Sadece yer düzlemi kısıtı, yer düzlemi sahneye ait olduğu için, sahne ile ilgili bilgi içermektedir. Öte yandan epipolar kısıtlar nesneler ile ilgili 3B bilginin kullanılmasını sağlamaktadır. Bu iki önemli bilginin bir

arada kullanımı birden fazla kameralı bir sistemde daha başarılı sonuçlar vereceği açıktır.

Yukarıda tanımlanan hali ile hem yer düzlemi hem de epipolar kısıtlar kullanılarak kapanma testi yapılmaktadır. Kapanma testi bir nesnenin hem pozisyonunun hem de tepe noktasının diğer kamera görüşlerine taşınarak kontrol edilmesi ile gerçekleştirilir. Kapanmaya uğramamış bir nesne için nesne pozisyonu ve tepe noktası tek bir nesne görüntüsü içerisinde yer almalıdır ve bu nesne görüntüsü içerisinde başa bir nesneye ait bir nokta yer almamalıdır. Eğer birden fazla nesneye ait noktalar tek bir nesne görüntüsü içerisinde yer alıyor ise, Şekil 11'da C_1 kamera görüşünde olduğu gibi, kameraya en yakın olan nesne kapanmaya sebep olan nesne, arkada kalan diğer nesne veya nesneler ise kapanmaya uğrayan nesne veya nesneler olarak işaretlenmektedir. Nesnelerin kameraya olan uzaklıkları nesne pozisyonlarının y bileşenleri ile ters orantılı olduğundan dolayı, nesne pozisyonunun y bileşeni en büyük olan nesne kameraya en yakın olan nesne olarak seçilir.

Sahnedeki nesnelerin genel olarak takip edilmesi ana kamera görüşündeki nesnelerin takibi üzerinden yapılmaktadır. Her video karesinde ana kamera üzerindeki nesneler takip edilir. Ana kamera üzerindeki bir nesne eğer kapanmaya uğramamış ise, yukarıda anlatılan yöntem gereğince, bu diğer kamera görüşlerinde de takip edilebiliyor anlamına gelir. Daha yüksek başarımlar için takip edilen nesnelerin daha önceki pozisyonları ve büyüklükleri ayrıca kontrol edilir. Eğer ana kamera görüşünde bir nesne kapanmaya uğramış ise bu nesnenin pozisyonu diğer kameralarda takip edilmeye devam edilir. Bu işlem nesnenin kapanmaya uğramadığı bir yardımcı kamera ana kamera olarak düşünülerek gerçekleştirilir.

Kapanma tespiti ve nesne takibi yer düzlemi homografilerinin ve epipolar geometri temel matrislerinin her kamera çifti için hesaplanmalarını gerektirir. Bu hesaplama işlemi sistemin kurulumu sırasında, sahnede bir nesnenin hareket etmesi ile sağlanır. Bu hareketli nesne tüm kamera görüşlerinde tespit edilir ve bu nesneye ait pozisyon ve tepe nokta bilgileri en küçük kareler yöntemi ile yer düzlemi homografilerinin ve epipolar geometri temel matrislerinin hesaplanması için kullanılır. Yaklaşık 5 saniyelik bir video ile 70 civarında nesne pozisyonu tüm kamera görüşleri üzerinde elde edilebilmektedir. Yer düzlemi homografileri bu

noktalar ile hesaplanmaktadır. Epipolar geometrinin temel matrislerinin hesaplanması için aynı 3B düzlemden gelmeyen noktalar gereklidir ve nesne pozisyonları aynı yer düzlemi üzerinde bulundukları için temel matrislerin hesaplanmasında doğrudan kullanılamazlar. Nesne pozisyonlarına ek olarak yer düzlemi üzerinde bulunmayan ancak kamera görüşleri arasında rahatlıkla eşleştirilebilecek noktalara örnek olarak nesnelerin tepe noktaları verilebilir. Öğrenme aşamasında sahnede tek bir nesne olduğu için kameralar arasında tepe noktaların ayrıca eşleştirilmesi gerekmez. Nesne pozisyonları ve tepe noktaları birlikte kullanılarak temel matrisler her kamera çifti için ayrı ayrı hesaplanır. Bu tez çalışmasında sunulan yöntemde %100 başarılı temel matrisler gerekmemektedir. Deneylerde görülmüştür ki; mükemmel olmayan bir temel matris ile hesaplanan kapanma testi için kabaca nesne çerçeve genişliğinin yarısı olacak bir t_{occl} eşik değeri yeterlidir. Eğer bir nesnenin pozisyonu diğer bir nesnenin pozisyonuna t_{occl} kadar veya daha yakındaysa ve nesne görüntüleri örtüşüyorsa, bu durumda kapanma vardır şeklinde yorumlanır.

Literatürde yer düzlemi homografilerini ve temel matrisleri sistem çalışırken eş zamanlı olarak hesaplayan ve güncelleyen yöntemler mevcuttur. Ancak burada anlatılan yöntem sunulan tez çalışması için yeterli ve oldukça başarılı bir yöntemdir.

4.3 Nesne Görüntüsünün Tahmini

Çalışma zamanı aşamasında, her kamera görüşünde nesneler arka plandan ayrıştırılarak tespit edildikten sonra ana kamera görüşü üzerindeki nesneler incelenir. Kamera görüşlerindeki nesneler zaman adımları boyunca O_i , $i=1, 2, \dots$ şeklinde isimlendirilsin. Her O_i için nesnenin kapanmaya uğrayıp uğramadığı kapanma testi ile tespit edilir. Deneylerde yardımcı kameralar, ana kameranın göremediği yerleri görecektir şekilde yerleştirilmiştir. Bu sayede ana kamerada kapanmaya uğrayan bir nesnenin yardımcı kamera veya kameralar tarafından görünmesi sağlanmıştır ve kapanma testinin başarısı arttırılmıştır. Kapanmaya uğramamış bir O_i nesnesi pozisyonu ve tüm kamera görüşlerindeki nesne görüntüleriyle birlikte daha sonra kullanılmak üzere kaydedilir.

Öte yandan, kapanmaya uğramış bir O_i nesnesi en az bir yardımcı kamera tarafından görülebiliyor olması gerekir ki kapanmaya uğradığı tespit edilmiş olsun. O_i nesnesinin hala görünür olduğu yardımcı kamera görüşündeki görüntüsü kullanılarak daha önceden kaydedilmiş nesne görüntüleri arasında aşağıdaki kurala göre bir arama yapılır.

$$O^*(t) = \arg \min_{i < t} \{S_h(O(i), O(t)) + \mu S_m(O(t-1), O(t))\}$$

Burada t zaman adımını, S_h ve S_m benzerlik fonksiyonu ve μ 0.01 olarak seçilen ağırlık katsayısıdır. S_h yardımcı kamera görüntülerini S_m ise ana kamera görüntülerini karşılaştırmaktadır ve aynı benzerlik fonksiyonu olarak gerçekleştirilmiştir. $O^*(t)$ arama sonucu bulunan en benzer kayıttır.

Bu arama kuralı ile yardımcı kamera görüntüsü şu anki yardımcı kamera görüntüsüne benzer ve ana kamera görüntüsü de bir önceki video karesindeki ana kamera görüntüsüne benzer olan bir nesne kaydı aranmaktadır. Ana kamera görüntülerinin benzerliği yardımcı kamera görüntülerine göre daha az önem taşımaktadır ve bu nedenle μ oldukça küçük seçilmiştir. Bir önceki video karesindeki ana kamera görüntüsü kapanmaya uğramamış nesneden geliyor olabileceği gibi, bir önceki arama sonucundan da geliyor olabilir.

Benzerlik fonksiyonu S_h ve S_m kareli farklar toplamı (SSD) ile gerçekleştirilmiştir. Arama işleminin efektif olabilmesi için nesne görüntüleri 64x64 lük karelere dönüştürüldükten sonra kaydedilmiştir. Bu sayede arama işlemi nesnenin büyüklüğünden bağımsız olur. Başarımı arttırmak için ise S fonksiyonu sadece ikili maskeleri örtüşen pikseller için çalıştırılmıştır. Bu sayede nesne çerçevesi içerisinde olup da nesne ile ilgisiz pikseller benzerlik için değerlendirilmemiştir.

Arama işlemi sonucunda en yüksek benzerliğe sahip kayıt bulunacaktır. Bulunan bu kaydın benzerliği belirli bir eşik değerinden daha düşük ise kayıt bulunmamış olarak kabul edilmektedir. Bu durumda O_i nesnesinin ana kamera görüntüsü tahmin edilememektedir ancak kullanıcıya O_i nesnesinin pozisyonu gösterilerek kapanmaya uğrayan nesne ile ilgili bilgi verilmektedir. Arama işlemi

sonucunda bulunan kayıt kabul edildiği zaman bu kayıt ile birlikte gelen ana kamera görüntüsü O_i nesnesinin şu anki ana kamera görüntüsü olarak kullanılmaktadır.

Arama sonucunda bulunan nesne ana kamera görüntüsü veya bir uyarı ana kamera görüşü üzerinde gösterilmektedir. Bu işlemin gerçekleştirilebilmesi için O_i nesnesinin ana kamera üzerindeki pozisyonunun bilinmesi gerekmektedir. O_i nesnesinin görünür olduğu yardımcı kamera görüşü üzerindeki pozisyonu yer düzlemi homografisi ile ana kamera görüşü üzerine taşınarak aranan pozisyon elde edilmiş olur.

Ana kamera görüşü üzerinde görüntülenecek olan nesne görüntüsü veya nesne varlığına dair bir uyarı herhangi bir büyüklükte gösterilemez. O_i kapanma arkasında kalmış ancak hala ana kameradan uzaklaşıyor veya ana kameraya yaklaşıyor olabilir. Bu durumda O_i 'nin ana kamera görüntüsünün büyüklüğü değişmektedir. Arama sonucu bulunan ana kamera görüntüsü de O_i 'nin şu anki ana kamera pozisyonu ve O_i için bulunan ana kamera görüntüsünün kaydedildiği zamanki pozisyonu kullanılarak boyutlandırılır. O_i 'nin ana kamera görüntüsü kaydedildiğindeki boyutu ile şu anki video karesinde olması gereken boyutu arasındaki ilişki aşağıdaki eşitlik ile sağlanır.

$$\alpha = \frac{p_{yeni}(y) - v}{p_{eski}(y) - v}$$

Burada α büyüklük değişimi için bulunan çarpandır ve v öğrenme aşamasında bulunan ana kameraya ait sıfır noktalarının oluşturduğu ufuk çizgisinin kamera resim düzleminde y bileşeninin konumudur. Bulunan kaydın ana kamera görüntüsü kaydedildiği sıradaki büyüklüğü α ile çarpılarak O_i şu anki ana kamera görüşünde hangi büyüklükte olması gerektiği hesaplanmış olur. Yeterli benzerliğe sahip olan bir kayıt bulunamadığı için kullanıcıya sadece nesne pozisyonu ile ilgili bir uyarı verilirken kayıttan gelen daha önceki pozisyon mevcut değildir. Bu nedenle kullanıcıya gösterilecek uyarının büyüklüğü kesin bir şekilde bilinemez. Burada izlenen yol, bir önceki ana kamera görüşündeki O_i büyüklüğünün kullanılması

yönündedir. Uyarı görüntüsü olarak da O_i 'nin bir önceki ana kamera görüntüsü şeffaf kırmızı ile maskelendikten sonra kullanılır.

5 DENEYLER

Sunulan yöntemin sonuçlarını görmek için çeşitli deneyler yapılmıştır. Tüm deneyler gigabit kameralar ve P4 3.0Ghz masaüstü bilgisayar ile yapılmıştır. Sunulan yöntemin kullanıcı müdahalesine ihtiyacı yoktur ve yapılan deneyler tatmin edici sonuçlar vermiştir.

Şekil 12’de kapalı alanda yapılan bir deneyden 8 kare verilmiştir. Bu deneyde oyuncu sahnedeki bir engelin arkasındaki iki defa geçmektedir ve bu sırada bazı hareketler yapmaktadır. Oyuncu 86. karede kapanmaya uğramıştır ve o adım için görüntü tahmini 4. kareden alınarak gerçekleştirilmiştir. Öte yandan 148. karede sistem oyuncuyu o pozda daha önce hiç görmediği için uygun bir görüntü tahmininde bulunamıyor. Bu adımda sistem kırmızı maske ile kullanıcıya bir uyarı gösteriyor, kapanmadaki nesnenin görüntüsünün doğru tahmin edilemediğini ancak pozisyonun ve büyüklüğünün kabaca bu şekilde olduğunu bildiriyor.



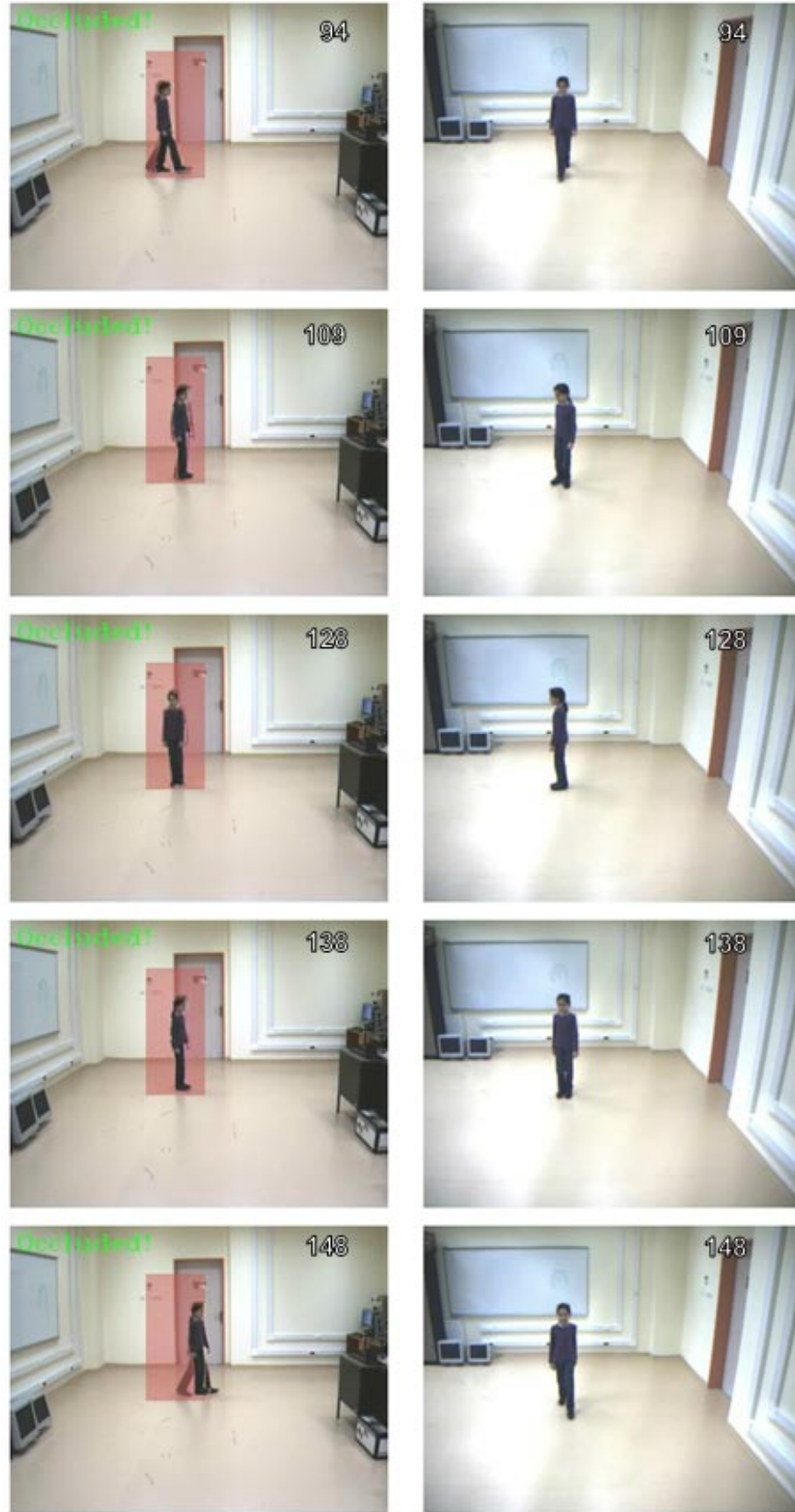
Şekil 12: Kapalı alan deneyi

Sistemin başarısını gerçek veriler ile karşılaştırabilmek için sanal bir engel ile de deneyler yapılmıştır. Sanal engel ile yapılan bir deneyin sonuçları Şekil 13’de verilmiştir. Sanal engel nesnelerin arka plandan ayırıştırma işlemi sırasında geçirgen olmayan bir maske uygulanarak sağlanmıştır. Bu sayede nesneler sabit bir maske arkasında kalmaktadır. Deneyin kalan kısımların aynı noktada şeffaf bir maske uygulanmış ve deney sonuçlarının maske arkasındaki gerçek sonuçlar ile karşılaştırılabilmesi sağlanmıştır. Şeffaf maske arkasındaki oyuncunun gerçek pozisyonu, büyüklüğü ve görüntüsü görülebilmektedir. Bu deneyde çoğunlukla doğru sonuçlar üreten sistem, Şekil 13’de verilen son karede görüntü tahminini doğru olarak yapamamaktadır. Bu durumun sebebi sistemin oyuncuyu o anki poz ile daha önce hiç görmemiş olması ancak çok benzer fakat yanlış bir pozda görmüş olmasıdır. Bu durumda dahi, sistem nesnenin pozisyonunu ve büyüklüğünü başarılı bir şekilde tahmin edebilmiştir.

Şekil 14’de açık alanda yapılan bir deneyin sonuçları verilmiştir. Bu deneyde bir otomobil büyük bir yapının arkasından geçmektedir. Bu deneyde görülüyor ki, kapanmaya neden olan yapının kapanmaya uğrayan nesneye göre büyüklüğü sistem açısından sorun teşkil etmemektedir. Yardımcı kamera ana kameranın sol tarafında oldukça uzağa yerleştirilmiştir. 37 numaralı video karesinde sistemin başarısı kısmen görülebilmektedir.

Şekil 15’de, dinamik kapanma ile yapılan bir deneyin sonuçları gösterilmektedir. Bu deneyde üç kamera kullanılmıştır. 3 numaralı video karesinde görülen hareketli insan izlenmek istenen nesne olarak seçilmiştir. Bu durumda diğer tüm nesneler, her video karesinde, kapanmaya neden olan statik nesneler olarak düşünülmüştür. Bu deneyde ana kamera görüşünün arka planı statik kabul edilen nesneler için de bilindiği için, bu nesneler tamamen ortadan kaldırmak mümkün olmuştur.

Zaman eksenini
↓



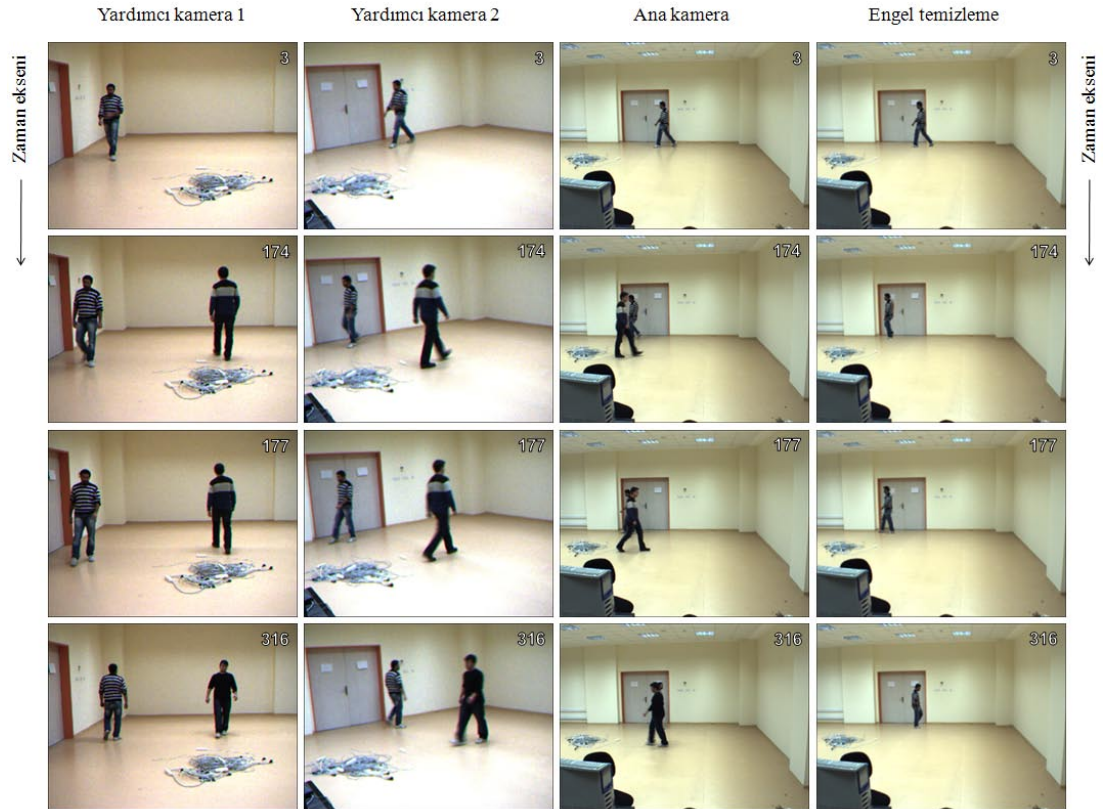
Ana kamera görüşü

Yardımcı kamera görüşü

Şekil 13: Sanal engel deneyi



Şekil 14: Açık alan deneyi



Şekil 15: Engel temizleme deneyi

Tüm deneylerde 320x240 çözünürlükte resimler kullanılmıştır. Sistemin en fazla zaman alan adımları nesnelerin arka plandan ayrıştırılmaları ve nesne görüntülerinin tahmini için kullanılan arama kuralıdır. Nesnelerin arka plandan ayrıştırılmaları işlemi üç kamera için yaklaşık olarak 15ms sürmektedir. Bu işlemde Bölüm 4.1’de anlatılan λ , t_f ve t_b değerleri sırası ile 35, 1000 ve 20000 olarak kullanılmıştır. t_f ve t_b eşik değerleri normalize oldukları için her deneyde farklı parametreler kullanılmasına gerek kalmamıştır. λ deneysel olarak bulunmuştur. Bu parametrelerin deneyler arasında değiştirilmesi belirgin bir iyileşmeye yol açmamıştır.

Nesne görüntülerinin tahmini işleminde S benzerlik fonksiyonu olarak SSD kullanılmış ve veritabanı arama işlemi 15 saniyelik bir video için yaklaşık olarak 10 milisaniye sürmektedir. Üç kameralı bir sistem yaklaşık 15 kare/saniye hız ile çalışabilmektedir. Her yeni kayıt ile birlikte veri tabanı büyümektedir ve arama için gereken zaman artmaktadır. Sistemin çalıştığı süre arttıkça, veri tabanı büyümekte ve arama işlemi yavaşlamaktadır. Ayrıca uzun çalışmalarda bellek problemleri de oluşabilir. Bu sorunları aşmak için belirli aralıklarla eski kayıtlar silinebilir veya akıllı bir veritabanı sistemi ile sadece birbirinden farklı kayıtlar tutularak benzer kayıtların kullanacağı bellekten tasarruf edilebilir.

6 SONUÇLAR VE ÖNERİLER

Bu tez çalışmasında gerçek zamanlı kapanmasız görüntü üretimi için temel bir yöntem önerilmiştir. Önerilen yöntem iki aşamadan oluşmaktadır; öğrenme aşaması ve çalışma zamanı aşaması. Öğrenme aşamasında yer düzlemi homografisi, kameralar arasındaki epipolar geometri parametreleri, nesne büyüklüklerindeki değişimlerin hesaplanabilmesi için perspektif geometrinin sıfır noktaları ve nesnelerin arka plandan ayrıştırılabilmesi için her kamera için arka plan modelleri çıkartılır. Çalışma zamanı aşamasında ise, her kamera görüşü resmi arka plandan çıkartılarak nesneler tespit edilir, nesnelerin pozisyonları hesaplanır ve kapanmaya uğrayan nesneler tespit edilir. Kapanmaya uğramamış nesnelerin görüntüleri her kamera görüşü için kaydedilir. Kapanmaya uğrayan nesnelerin görüntüleri ise daha önce kapanmaya uğramadığı zamanlardaki görüntülerinin kayıtları kullanılarak tahmin edilir. Tahmin edilen nesne görüntüleri doğru pozisyon ve büyüklükler hesaplanarak ana kamera görüşü üzerinde gerekli yere çizilir ve bu sayede sentetik bir şekilde kapanmasız görüntü üretimi gerçekleştirilmiş olur.

Sistem yapılan deneylerde tatmin edici sonuçlar vermiştir. Kapanmasız görüntü üretimi için temel bir sistem olarak geliştirilen bu yöntem bazı varsayımlar ile çalışmaktadır. Örnek olarak, nesnelerin takip edilebilmesi için en az bir kamerada görünüyor olması gerekmektedir. Kullanılan başka bir varsayım da sistemin kapanmadaki nesneyi daha önce görüp öğrenmiş olmasıdır. Bu varsayımların geçerli olmadığı durumda nesneler tabiatı gereği şu anki sistem ile takip edilemez ve kapanmadaki görüntüleri tahmin edilemez. Bu varsayımların da çözülmesi ve daha güçlü donanım ile geliştirilecek olan sistem daha genel bir hal alacaktır.

KAYNAKLAR

- [1] Yildiz A. and Akgul Y. S., “A Multi-View Camera System for The Generation of Real-Time Occlusion-Free Scene Video”, DAGM, 2007.
- [2] Vaibhav Vaish, Marc Levoy, Richard Szeliski, C. Lawrence Zitnick, and Sing Bing Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In IEEE Computer Vision and Pattern Recognition (2), pages 2331-2338, 2006.
- [3] Aaron Isaksen, Leonard McMillan, and Steven J. Gortler. Dynamically reparameterized light fields. In SIGGRAPH, pages 297-306, 2000.
- [4] Kameda Y., Takemasa T., Ohta Y., “Outdoor see-through vision utilizing surveillance cameras,” Mixed and Augmented Reality, 2004. ISMAR 2004. Third IEEE and ACM International Symposium on, vol., no., pp. 151-160, 2-5 Nov. 2004.
- [5] Horprasert T., Harwood D., Davis L. S., “A statistical approach for real-time robust background subtraction and shadow detection”, In Proc. IEEE Frame Rate Workshop 1-19, 1999.
- [6] Stauffer C., Grimson W.E.L., “Adaptive background mixture models for real-time tracking”, Int. Conf. Computer Vision and Pattern Recognition, Vol. 2, pp. 246-252, 1999.
- [7] Rother C., Kolmogorov V., Blake, A. “Grabcut - interactive foreground extraction using iterated graph cuts”, Proc. ACM Siggraph, 2004.
- [8] Sun J., Zhang W., Tang X., Shum H. Y., “Background cut”, In Proceedings of ECCV, volume II, pages 628-641, 2006.

- [9] Boykov Y. and Jolly M. Pi., “Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images”, In Proceedings of ICCV, pages 105-112, 2001.
- [10] Pritch Y., Rav-Acha A., Peleg S., “Non-Chronological Video Synopsis and Indexing”, IEEE Trans. PAMI, to appear Nov. 2008.
- [11] Boykov Y., Kolmogorov V., “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision”, In IEEE Transactions on PAMI, September 2004.
- [12] Kim K. and Davis L., “Multi-camera Tracking and Segmentation of Occluded People on Ground Plane Using Search-Guided Particle Filtering,” In Proc. ECCV, 2006.
- [13] S. M. Khan and M. Shah. “A multiview approach to tracking people in crowded scenes using a planar homography constraint,” In Proc. ECCV, 2006.
- [14] Arthur D., Vassilvitskii S., “k-means++: The advantages of careful seeding”, In Bay Area Theory Symposium, BATS, 2006.
- [15] De Maesschalck, R.; D. Jouan-Rimbaud, D.L. Massart “The Mahalanobis distance” Chemometrics and Intelligent Laboratory Systems 50:1–18, 2000.

ÖZGEÇMİŞ

1985 yılında Aydın’da doğdu. Orta ve lise öğrenimini Kartal Burak Bora Anadolu Lisesi’nde tamamladı. 2007 yılında Gebze Yüksek Teknoloji Enstitüsü, Bilgisayar Mühendisliği Bölümü’nden mezun oldu. 2007 yılında Yüksek Teknoloji Enstitüsü, Bilgisayar Mühendisliği Bölümü’nde yüksek lisans eğitimine başladı. Halen aynı üniversitede araştırma görevlisi olarak çalışmaktadır.