T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DİNAMİK ORTAMLARDA GÖRSEL SLAM BAŞARISININ İYİLEŞTİRİLMESİ

Alperen KÖSEM

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Dr. Öğr. Üyesi Erkan USLU

Mart, 2026

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DİNAMİK ORTAMLARDA GÖRSEL SLAM BAŞARISININ İYİLEŞTİRİLMESİ

Alperen KÖSEM tarafından hazırlanan tez çalışması 10.03.2026 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Dr. Öğr. Üyesi Erkan USLU
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Dr. Öğr. Üyesi Erkan USLU, Danışman
Yıldız Teknik Üniversitesi _____

Doç. Dr. Oğuz ALTUN, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üyesi Nihal ALTUNTAŞ, Üye
İstanbul Üniversitesi-Cerrahpaşa _____

Danışmanım Dr. Öğr. Üyesi Erkan USLU sorumluluğunda tarafımca hazırlanan Dinamik Ortamlarda Görsel SLAM Başarısının İyileştirilmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Alperen KÖSEM

İmza

Her zaman destekleriyle
yanımda olan aileme ve eşime

TEŞEKKÜR

"Dinamik Ortamlarda Görsel SLAM Başarısının İyileştirilmesi" başlıklı çalışmamın geliştirilmesi sürecinde değerli yönlendirmeleri ve sağladığı katkılar için danışmanım Dr. Öğr. Üyesi Erkan USLU'ya; hayatım boyunca her zaman maddi ve manevi desteklerini arkamda hissettiğim kıymetli annem, babam ve kardeşime; tez yazım sürecinde destekleriyle yanımda olan eşime ve iş arkadaşlarıma sonsuz teşekkür ederim.

Alperen KÖSEM

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ xi

ÖZET xiii

ABSTRACT xv

1 GİRİŞ 1

1.1 Literatür Özeti 1

1.2 Problem Tanımı 3

1.3 Tezin Amacı ve Kapsamı 3

2 GÖRSEL EŞ ZAMANLI KONUM BELİRLEME VE HARİTALAMA 5

2.1 Kamera Temelleri 5

2.2 Stereo Geometri ve Derinlik Gözlemlenebilirliği 7

2.3 Öznitelik Dedektörleri ve Betimleyiciler 8

2.3.1 Köşe ve Anahtar Nokta Dedektörleri 8

2.3.2 Ölçek Uzayı Tabanlı Dedektör ve Betimleyiciler 8

2.3.3 İkili Betimleyiciler 9

2.3.4 Çizgi Öznitelik Detektörleri ve Betimleyicileri 10

2.3.5 Temel Dedektörlerin ve Betimleyicilerin Karşılaştırılması . 10

2.4 Dolaylı (Öznitelik Tabanlı) Görsel SLAM Yöntemleri 11

2.5 Doğrudan (Direct) Görsel SLAM Yöntemleri 12

2.6 Yarı Doğrudan (Semi-Direct) Yaklaşımlar 13

2.7 Dinamik Ortamlarda Görsel SLAM Yaklaşımları 14

2.7.1 Semantik Bilgi Tabanlı Filtreleme 14

2.7.2 Optik Akış Tutarlılığı Tabanlı Filtreleme 15

2.7.3 Arka Planı Yeniden Yapılandırarak Statik Çevreyi Koruma . 16

3 YÖNTEM **18**
3.1 Genel Sistem Mimarisi 18
3.2 Stereo Derinlik Çıkarımı 19
3.3 Semantik Bilgi ile Dinamik Obje Adayı Çıkarımı 22
3.4 Statik Bölgelerden Kamera Hareketi Kestirimi 23
3.5 Statik Varsayım ile Beklenen Optik Akı Çıkarımı 25
3.6 Yoğun Optik Akı Çıkarımı 26
3.7 Dinamik Obje Kararı ve Maskeleme 27
3.8 ORB-SLAM3 ile Entegrasyon 29

4 DENEYSEL KURULUM **31**
4.1 Veri Seti: KITTI Odometry 31
4.2 Deney Kurgusu 32
4.3 Değerlendirme Metrikleri 35
4.3.1 Mutlak Yörünge Hatası (ATE) 35
4.3.2 Mutlak Dönme Hatası (ARE) 36
4.3.3 Bağıl Yörünge Hatası (RTE) 37
4.3.4 Bağıl Dönme Hatası (RRE) 37

5 DENEYLER VE BULGULAR **39**
5.1 Şehir İçi Düşük Trafikli Senaryolar 39
5.2 Yüksek Hızlı Otoyol Senaryosu 48
5.3 Statik ve Dinamik Nesnelerin Dengeli Dağılım Gösterdiği Senaryo . 51
5.4 Şehir İçi Çift Yönlü Trafik Senaryosu 55
5.5 Hesaplama Maliyeti ve Çalışma Zamanı Analizi 59
5.6 Sadece RAFT Tabanlı Dinamik Bölge Filtreleme 60
5.7 Tüm Senaryoların ve Yöntemlerin Toplu Değerlendirilmesi 63

6 SONUÇ **67**

KAYNAKÇA **69**

TEZDEN ÜRETİLMİŞ YAYINLAR **74**

SİMGE LİSTESİ

Ω	Aday Dinamik Bölgeler
$\mathbf{F}_t^{exp}(.)$	Beklenen Optik Akı
$Z(\cdot)$	Derinlik Haritası
p	İki Boyutlu Piksel Konumu
$\mathbf{K}$	Kamera İç Kalibrasyon Matrisi
f	Kamera Odak Uzaklığı
$LK(\cdot)$	Lucas–Kanade Optik Akı
$C(\cdot)$	Maliyet Fonksiyonu
D_p	Piksel p İçin Disparite Değeri
I_t^L	Sol Görüntü Karesi
P	Üç Boyutlu Nokta Konumu
$\mathbf{F}_t^{raft}(\cdot)$	Yoğun Optik Akı

KISALTMA LİSTESİ

ATE	Absolute Trajectory Error
ARE	Absolute Rotation Error
EKF	Extended Kalman Filter
EZKH	Eş Zamanlı Konumlama ve Haritalama
GPU	Graphics Processing Unit
GPS	Global Pozisyonlama Sistemi
IMU	Inertial Measurement Unit
İHA	İnsansız Hava Aracı
LiDAR	Light Detection and Ranging
ORB	Oriented FAST and Rotated BRIEF
PnP	Perspective-n-Point
RAFT	Recurrent All-Pairs Field Transforms
RANSAC	Random Sample Consensus
RRE	Relative Rotation Error
RTE	Relative Trajectory Error
RMSE	Root Mean Squared Error
SGBM	Semi-Global Block Matching
SGM	Semi-Global Matching
SLAM	Simultaneous Localization and Mapping
vSLAM	Visual SLAM
YOLO	You Only Look Once

ŞEKİL LİSTESİ

Şekil 2.1 Pinhole kamera modelinde perspektif izdüşüm: 3B noktaların kamera merkezinden geçerek odak düzlemine düşmesi 6
Şekil 2.2 Stereo geometrisi ile üçgenleme tabanlı derinlik kestirimi 7
Şekil 3.1 Önerilen çalışmanın genel mimarisi 19
Şekil 3.2 Örnek stereo disparite haritası 21
Şekil 3.3 Disparite haritasından elde edilen derinlik haritası 21
Şekil 3.4 Farklı YOLO ailelerinin COCO $\text{mAP}_{50:95}$ doğruluğu ile gecikme arasındaki karşılaştırma [50] 22
Şekil 3.5 Tespit edilen aday bölgeler 23
Şekil 3.6 RAFT modelinin temel mimarisi [51] 26
Şekil 3.7 Örnek RAFT optik akış çıktısı 27
Şekil 3.8 Aday nesneler için dinamik–statik kararının optik akış tutarsızlığına göre belirlenmesi 29
Şekil 4.1 Veri setinin oluşturulma aracı[52] 31
Şekil 4.2 Sensörlerin araç üzerindeki konumları ve eksenleri[52] 32
Şekil 4.3 Seçilen KITTI Odometry sekanslarının yer gerçeği (GT) yörüngeleri (x–z düzlemi). 33
Şekil 4.4 Sahnenin filtrelenmemiş ham görüntüsü (temel yöntem) 34
Şekil 4.5 Sahnenin semantik maskeleme ile maskelenmiş görüntüsü 34
Şekil 4.6 Sahnenin önerilen yöntem sonrası maskelenmiş görüntüsü 35
Şekil 5.1 Sekans 00: ardışık 6 kare için filtrelenmemiş görüntüler 40
Şekil 5.2 Sekans 00: ardışık 6 kare için semantik maskeleme ile filtrelenmiş görüntüler 40
Şekil 5.3 Sekans 00: ardışık 6 kare için önerilen yöntem ile filtrelenmiş görüntüler 41
Şekil 5.4 Sekans 00: seçilen sahnelerde ham görüntü, semantik maskeleme ve önerilen yöntem çıktısının görsel karşılaştırması 42
Şekil 5.5 Sekans 00 için x–z düzleminde yörünge karşılaştırması 42
Şekil 5.6 Sekans 00 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi . . 43
Şekil 5.7 Sekans 07: Üç farklı zaman anında ham görüntü, semantik maskeleme ve önerilen yöntem çıktısının karşılaştırması 44
Şekil 5.8 Sekans 07 için x–z düzleminde yörünge karşılaştırması 44

Şekil 5.9 Sekans 07 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi . . 45
Şekil 5.10 Sekans 01: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması 48
Şekil 5.11 Sekans 01 için x–z düzleminde yörünge karşılaştırması 49
Şekil 5.12 Sekans 01 için $x(t)$, $y(t)$ ve $z(t)$ eksenlerindeki konum değişimi . 49
Şekil 5.13 Sekans 03: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması 51
Şekil 5.14 Sekans 03 için x–z düzleminde yörünge karşılaştırması 52
Şekil 5.15 Sekans 03 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi . . 53
Şekil 5.16 Sekans 04: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması 55
Şekil 5.17 Sekans 04 için z–x düzleminde yörünge karşılaştırması 56
Şekil 5.18 Sekans 04 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi . . 57
Şekil 5.19 Sekans 00 üzerinde sadece RAFT tabanlı yaklaşım ile elde edilen örnek ardışık sahneler 62

TABLO LİSTESİ

Tablo 2.1 Öznitelik dedektörleri ve betimleyicileri 10
Tablo 5.1 Sekans 00 ve 07 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması. 46
Tablo 5.2 Sekans 00 ve 07 için bağıl yörünge hatalarının (RTE) karşılaştırılması 46
Tablo 5.3 Sekans 00 ve 07 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması. 47
Tablo 5.4 Sekans 00 ve 07 için bağıl dönme hatalarının (RRE) karşılaştırması 47
Tablo 5.5 Sekans 01 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması 50
Tablo 5.6 Sekans 01 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması 50
Tablo 5.7 Sekans 01 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması 51
Tablo 5.8 Sekans 01 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması 51
Tablo 5.9 Sekans 03 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması 53
Tablo 5.10 Sekans 03 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması 54
Tablo 5.11 Sekans 03 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması. 54
Tablo 5.12 Sekans 03 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması. 54
Tablo 5.13 Sekans 04 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması 57
Tablo 5.14 Sekans 04 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması 58
Tablo 5.15 Sekans 04 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması 58
Tablo 5.16 Sekans 04 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması 59

Tablo 5.17 Kare başına gecikme süresi (ort / $\pm$ std, ms) 60
Tablo 5.18 Sadece RAFT tabanlı yaklaşım için ATE ve RTE hata istatistikleri 62
Tablo 5.19 Sadece RAFT tabanlı yaklaşım için ARE ve RRE hata istatistikleri 62
Tablo 5.20 KITTI senaryoları için ham veri, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının ATE RMSE[m] karşılaştırması. 63
Tablo 5.21 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının RTE RMSE[m] karşılaştırması. 64
Tablo 5.22 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının ARE RMSE[$^\circ$] karşılaştırması. 65
Tablo 5.23 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının RRE RMSE[$^\circ$] karşılaştırması. 65

ÖZET

Dinamik Ortamlarda Görsel SLAM Başarısının İyileştirilmesi

Alperen KÖSEM

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Dr. Öğr. Üyesi Erkan USLU

Eş Zamanlı Konumlama ve Haritalama (EZKH, Simultaneous Localization and Mapping - SLAM) algoritmaları, GPS'in yetersiz kaldığı veya hiç bulunmadığı ortamlarda robotların hareket edebilmesi için gerekli olan poz kestirimini sağlamakta ve çevrenin harita bilgisini oluşturmaktadır. Özellikle kameralardan elde edilen görüntüler üzerinde çalışan görsel SLAM (vSLAM) yaklaşımları, düşük maliyetli sensörler ile zengin bilgi sunması nedeniyle robotik sistemlerde yaygın olarak tercih edilmektedir. Ancak güncel vSLAM sistemlerinin temel problemlerinden biri, dinamik ortamlarda gürbüzlüğünü kaybetmeleridir. Modern algoritmaların sahnenin statik olduğu varsayımıyla tasarlanmış olması, şehir içi trafik ya da yoğun hareketli ortamlarda hatalı poz kestirimine ve yörünge sapmalarına sebep olmaktadır. Bu sebeple, dinamik nesnelerin sahneden filtrelenmesi algoritmanın doğruluğunu iyileştirmede kritik rol oynamaktadır.

Literatürde dinamik sahnelerle başa çıkmak amacıyla hareket temelli ve semantik temelli yaklaşımlar önerilmektedir. Ancak bu iki yaklaşım tek başına kullanıldığında kamera hareketi ile nesnelerin hareketini ayrıştırmakta problem yaşamakta ya da semantik olarak tüm nesneleri maskeleyerek öznitelik sağlayabilecek faydalı bölgeleri de filtreleyerek algoritma başarısını düşürmektedir.

Bu çalışmada, stereo Oriented FAST and Rotated BRIEF SLAM3 (ORB-SLAM3) üzerinde çalışan, You Only Look Once (YOLO) sinir ağından elde edilen semantik bilgi ile Recurrent All-Pairs Field Transforms (RAFT) ile hesaplanan optik akı bilgisini birleştiren derin öğrenme tabanlı optik akış ve nesne tespiti füzyonuna

dayalı hibrit bir SLAM entegrasyonu önerilmektedir. Önerilen yöntemde YOLO sinir ağından gelen semantik bilgiler ile RAFT sinir ağından hesaplanan optik akı derinlik katsayılı bir eşikleme yapılmakta ve böylece fayda sağlayabilecek statik bölgeler korunurken, sahnedeki dinamik bölgelerin SLAM algoritmasına olumsuz etkisi giderilmektedir.

Önerilen yöntem, KITTI odometri veri seti üzerinde filtrelenmemiş görüntüler ve YOLO tabanlı semantik filtreleme yaklaşımı ile ORB-SLAM3 kullanılarak karşılaştırılmıştır. Mutlak yörünge hatası (absolute trajectory error - ATE), mutlak rotasyon hatası (absolute rotaion error - ARE), bağıl yörünge hatası (relative trajectory error - RTE) ve bağıl rotasyon hatası (relative rotation error - RRE) değerlendirme metrikleri üzerinden değerlendirmiştir. Filtrelenmemiş görüntüler ile karşılaştırıldığında, önerilen yöntem sekanslar genelinde ortalama %11.87 oranında iyileşme sağlamıştır. En yüksek kazanım Sekans 01 üzerinde elde edilmiştir ve ATE RMSE değeri 13.67 metreden 7.70 metreye düşerek %43.67 oranında azaltılmıştır.

Anahtar Kelimeler: Görsel Eş Zamanlı Konumlama ve Haritalama, Dinamik Obje Maskeleme, YOLO, RAFT

ABSTRACT

Enhancing Visual SLAM Accuracy in Dynamic Environments

Alperen KÖSEM

Department of Computer Engineering
Master of Science Thesis

Supervisor: Assist. Prof. Dr. Erkan USLU

Simultaneous Localization and Mapping (SLAM) algorithms provide required pose estimation for robots to navigate in environments where GPS is insufficient or unavailable and construct map information of the environment simultaneously. In particular, visual SLAM (vSLAM) approaches are widely preferred in robotic systems due to their low-cost sensors and the rich information they provide. However, one of the fundamental problems in current vSLAM systems is the loss of robustness in dynamic environments. The design of modern algorithms is based on the assumption that the scene is static cause inaccurate pose estimation and trajectory drift in urban traffic or intense dynamic environments. Therefore, filtering the dynamic objects from scene plays critical role in improving vSLAM accuracy.

In the literature, motion-based and semantic-based approaches are proposed for the purpose of handling dynamic scenes. However, when these two approaches are used individually, they degrade the algorithm's success by having difficulty distinguishing object motion from camera motion or by semantically masking all objects, thereby filtering out beneficial regions that can provide useful features.

In this work, we propose a hybrid SLAM integration for stereo vSLAM algorithms that fuses semantic information obtained from a YOLO network with optical flow computed by RAFT. In the proposed method, the semantic outputs from YOLO and the optical flow estimated by RAFT are combined through depth-weighted thresholding. In this way, potentially beneficial static regions are preserved while the negative impact of dynamic regions on the vSLAM algorithm is mitigated.

The proposed approach is evaluated on the KITTI Odometry dataset and compared against ORB-SLAM3 using unfiltered images and a YOLO-based semantic filtering. Error reduction has been demonstrated particually in the environments where dynamic and static objects coexist in same scene based on Absolute Trajectory Error (ATE), Absolute Rotation Error (ARE), Relative Trajectory Error (RTE) and Relative Rotation Error (RRE) evaluation metrics. Compared to unfiltered images, the proposed method achieves an average ATE RMSE reduction of 11.87% across the evaluated sequences. The largest improvement is observed on Sequence 01, where ATE RMSE decreases from 13.67 meter to 7.70 meter, corresponding to a 43.67% reduction.

Keywords: Visual Simultaneous Localization and Mapping, Dynamic Object Suppression, YOLO, RAFT

1
GİRİŞ

Teknolojinin gelişmesi ile birlikte mobil robotlar hayatımızın bir parçası haline gelmiştir. Mobil robotların çevrede güvenli bir şekilde hareket edebilmeleri için çevre farkındalığı ve konum bilgisi gibi temel bilgilere ihtiyaç duymaktadırlar. Konum bilgisinin doğrudan sağlanamadığı durumlarda eş zamanlı konumlama ve haritalama (SLAM) çözümleri geliştirilmiştir. SLAM, robotun üzerindeki sensörlerden elde edilen verilerle ortamın haritasını oluşturmasını ve aynı anda bu harita içerisindeki anlık konumunu kestirmesini sağlamaktadır.

Işık Tespiti ve Uzaklık Tayini (Light Detection and Ranging - LiDAR), kamera, stereo kamera ve Ataletsel Ölçüm Birimi (Inertial Measuremet Unit - IMU) tabanlı SLAM çözümleri literatürde yaygın olarak kullanılan yöntemler olmakla birlikte teker odometrisi[1] ile de sağlanan çözümler bulunmaktadır. Ancak teker odometrili çözümler kısa mesafelerde doğruluk sağlasa da teker sapmaları yüzünden hata kümülatif olarak büyümekte ve sistemin doğruluğu olumsuz etkileyebilmektedir [2]. Buna karşın, kameraların düşük maliyetli olması ve sahneden zengin bilgi sunması nedeniyle Görsel SLAM (Visual SLAM - vSLAM) yaklaşımları otonom araçlarda ve robotik sistemlerde oldukça tercih edilmektedir. Monoküler ve stereo çözümler olmak üzere ikiye ayrılan vSLAM yöntemlerinde monoküler çözümler ölçek belirsizliği nedeniyle ek varsayım ya da ek bilgiye ihtiyaç duyarken, stereo çözümler iki kamera arasındaki taban mesafesi sayesinde derinlik bilgisini doğrudan çıkararak ölçeği gözlenebilir hale getirebilmektedir[3].

1.1 Literatür Özeti

SLAM probleminin farklı koşullarında güvenilir biçimde çözülmesi için literatürde LiDAR, kamera ve IMU gibi çeşitli sensörlerin bir arada kullanıldığı yaklaşımlar yaygınlaşmıştır. Bu tür çoklu sensörlü sistemlerde, sensör verilerinin nasıl işlendiği ve hangi seviyede birleştirildiği, elde edilen konum kestiriminin doğruluğunu doğrudan etkilemektedir. Birden fazla sensörlü çözümler, sensörlerden gelen

verilerin işlenme ve birleştirme yöntemine göre sıkı bağlı (tightly-coupled) ve gevşek bağlı (loosely-coupled) olmak üzere iki temel mimariye ayrılmaktadır [4].

Gevşek bağlı mimaride, her sensörden gelen veri bağımsız olarak işlenmekte ve ayrı ayrı poz tahminleri üretilmektedir. Sonrasında oluşan poz tahminleri birleştirilmektedir. Bu yaklaşım sensörlerin birbirine bağımlılıkları az olduğu için modüler ve hızlı çözüm sağlamakla birlikte sensörler arasında tam korelasyon kurulmadığı için hata birikimlerine karşı dayanıksızdır. Chen ve arkadaşları tarafından geliştirilen nokta ve çizgi özniteliklerini kullanarak geliştirilen çalışma [5] ile Mourikis ve Roumeliotis tarafından gerçekleştirilen MSCKF [6]'de oluşturulan poz tahminleri Genişletilmiş Kalman Filtresi (EKF) ile gevşek mimari kullanılarak pozisyon kestirimi yapılmaktadır.

VINS-Mono [7] ve Xu ve ark. tarafından önerilen düşük ışıklı ortamlara yönelik nokta-çizgi füzyonlu odometri [8] gibi sıkı bağlı mimari tabanlı çalışmalarda, farklı sensörlerden gelen veriler ortak bir optimizasyon çatısı altında işlenerek poz kestirimi elde edilmektedir. Tüm sensörler optimizasyon problemine dahil edildiği için daha bütüncül ve gürbüz bir yaklaşım sergilenmektedir. EPLF-VINS [9], nokta-çizgi özniteliklerini KLT[10] optik akış yöntemi ile takip ederek sıkı bağlı yaklaşım ile düşük öznitelikli ortamlarda başarı sağlamaktadır. OKVIS2-X [11] ise OKVIS2[12] üzerine kurulan geleneksel görsel-ataletsel SLAM altyapısına LiDAR ve GNSS ile uyum sağlayabilen modüler bir yapıyı sıkı bağlı yöntemle kurarak esnek ve yüksek doğrulukta konum kestirimi gerçekleştirmektedir.

Bunun yanında düşük maliyet ve ağırlıkları sebebiyle mobil robotlarda ya da insansız hava araçlarında (İHA) genellikle vSLAM çözümleri tercih edilmektedir. vSLAM çözümleri, robotun bir veya birden fazla kamera verisiyle çevresinin haritasını oluşturmayı ve robotu konumlandırmayı çevredeki görsel değişimleri takip ederek sağlamaktadır. Dolaylı (indirect) ve doğrudan (direct) yöntemler çevre bilgisini takip etmek için literatürdeki temel yöntemlerdendir. Dolaylı yöntemler, görüntüdeki kenar ve köşe gibi ayırt edici öznitelik noktaları çıkarmakta ve bu noktaları ardışık görüntüler arasında eşleştirilerek kamera pozunu tahmin etmektedir[13]. Doğrudan yöntemler ise, öznitelik tespiti ve eşleştirmesi yapmaksızın görüntüdeki piksel yoğunluklarını kullanmaktadır. Tahmin edilen kamera hareketi altında bir kareden diğerine yeniden yansıtılan piksellerin yoğunluk farklarını ifade eden fotometrik hatayı minimize ederek poz kestirimi yapmaktadır. Ancak her iki yaklaşımda da çözümün temeli statik çevre yaklaşımına dayanmaktadır. Ortamdaki hareketli nesneler, fotogrametrik tutarlılığı bozarak SLAM çözümünde sapmalara neden olmaktadır. Bu durum, hareketli nesne içeren ortamlarda vSLAM algoritmalarının başarısını kritik noktada düşürmektedir. Bu

sebeple, dinamik nesnelerin tespiti ve filtrelenmesi vSLAM performansı gelişimi açısından büyük önem taşımaktadır. SLAM sistemlerinin bu varsayım üzerine kurulması sebebiyle literatürde hareketli nesnelerden gelebilecek olumsuz etkiyi azaltmak üzerine birçok çalışma yapılmıştır.

1.2 Problem Tanımı

vSLAM algoritmaları genellikle sahnenin statik olduğu varsayımıyla çalışması sebebiyle dinamik nesnelerin bulunduğu ortamlarda performans kaybı yaşamaktadır. İnsan ve araç gibi hareketli nesneler görüntüde geniş yer kapladığında, bu bölgelerden çıkarılan öznitelikler ardışık kareler arasında tutarsız hareket ederek hatalı eşleştirmelere ve dolayısıyla yanlış poz kestirimlerine neden olmaktadır. Bu etki, poz hatasının artmasına ve bazı durumlarda takip kaybına yol açabilmektedir. Sadece optik akış temelli karar verme ile filtreleme ise kamera hareketi ve düşük dokulu bölgeler gibi etkenlerden olumsuz etkilenebilmektedir.

Bu bağlamda dinamik bölgelerin bastırılmasına yönelik yaklaşımlarda temel zorluk, maskelemenin seçiciliğini doğru ayarlamaktır. Aşırı maskeleme, sahneden çok sayıda faydalı özniteliğin elenmesine bağlı olarak takip kararlılığını düşürürken, yetersiz maskeleme ise hareketli bölgelerden gelen tutarsız ölçümlerin poz kestiriminde hatalara neden olmaktadır. Bu bağlamda, semantik maskeleme yöntemleri "hareket edebilir" sınıfındaki nesneleri dinamik kabul ederek gereksiz eleme yapabilmekte; yalnızca optik akış temelli karar verme ise agresif kamera hareketlerinde gürbüzlüğünü kaybetmektedir.

1.3 Tezin Amacı ve Kapsamı

Bu tezde, stereo tabanlı vSLAM başarısını artırmak amacıyla hibrit bir dinamik bölge maskeleme yöntemi önerilmektedir. Önerilen yaklaşımda, semantik tespitler ile elde edilen aday dinamik bölgeler, hareket bilgisinden yararlanılarak doğrulanmakta ve yalnızca dinamik olduğu doğrulanan bölgeler maskeleme işlemine dahil edilmektedir. Böylece semantik maskelemenin bilinçsizce aşırı maskeleme eğilimi azaltılırken, optik akış ve semantik bilgi füzyonuyla objenin dinamik veya statik olduğu güvenilir biçimde belirlenerek sahnedeki dinamik bölgelerin filtrelenmesi ve faydalı öznitelik noktalarının korunması amaçlanmaktadır.

Önerilen yöntem, stereo tabanlı öznitelik temelli vSLAM yöntemler için bir ön işleme katmanı olarak tasarlanmıştır. ORB-SLAM3 ile entegre edilerek KITTI

Odometry veri seti üzerinde testleri gerçekleştirilmiştir. Yöntemlerin başarımı mutlak yörünge hatası (ATE), mutlak rotasyon hatası (ARE), bağıl yörünge hatası (RTE) ve bağıl rotasyon hatası (RRE) metrikleri üzerinden karşılaştırılmıştır.

Tezin devamı şu şekilde organize edilmiştir: Bölüm 2'de vSLAM konusunda temel kavramlar sunulmakta ve dinamik sahnelerde karşılaşılan zorluklara ilişkin çalışmalar özetlenmektedir. Bölüm 3'te önerilen hibrit maskeleme yöntemi ve sistem mimarisi ayrıntılı olarak açıklanmaktadır. Bölüm 4'te deneysel kurulum, veri seti, değerlendirme metrikleri ve karşılaştırma senaryoları verilmektedir. Bölüm 5'te elde edilen deneysel bulgular sunulmakta ve tartışılmaktadır. Bölüm 6'da sonuçlar özetlenerek gelecekte yapılabilecek çalışmalar belirtilmektedir.

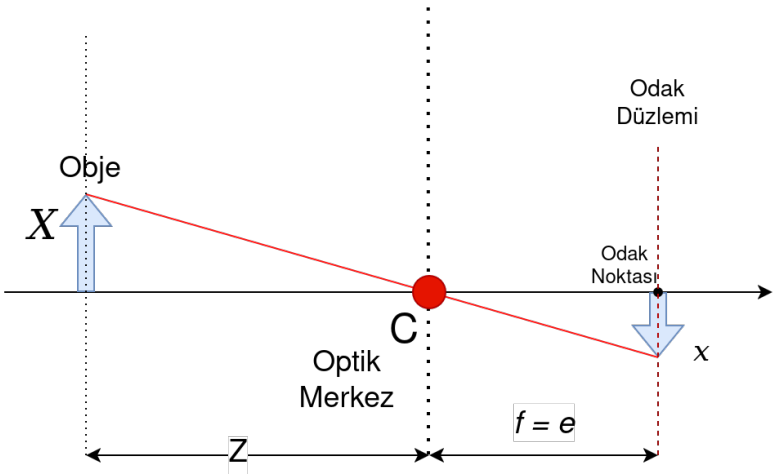
2
GÖRSEL EŞ ZAMANLI KONUM BELİRLEME VE HARİTALAMA

vSLAM probleminin temel hedefi, bir veya birden fazla kameradan elde edilen görüntü ölçümlerini kullanarak hem robotun zamanla değişen pozunu kestirmek hem de çevrenin haritasını eş zamanlı olarak oluşturmaktır. Bu amaç doğrultusunda literatürde görsel bilgiyi kullanma biçimine göre doğrudan (direct) ve dolaylı (indirect) olmak üzere başlıca iki yaklaşım öne çıkmaktadır. Dolaylı yöntemler, görüntüden çıkarılan ayırt edici öznitelik noktaları ve/veya çizgileri üzerinden eşleştirme yaparak kamera hareketini çoklu görüş geometrisi ve yeniden izdüşüm hatasının minimize edilmesiyle tahmin etmektedir. Doğrudan yöntemler ise piksel yoğunluklarını doğrudan kullanarak fotometrik hataları minimize etmeyi hedefler. Her iki yaklaşımın da farklı avantajları ve sınırlılıkları bulunduğundan, uygulama senaryosu ve veri koşullarına göre tercih edilen mimariler değişebilmektedir.

Bu bölümde önce kamera modeli ve stereo geometrinin temel kavramları verilmekte, ardından dolaylı/doğrudan/yarı-doğrudan vSLAM yöntemleri özetlenmekte ve dinamik sahnelerde vSLAM'i zorlayan etmenler literatür üzerinden tartışılmaktadır.

2.1 Kamera Temelleri

vSLAM sistemlerinin büyük bir bölümü, kamera modellemesinde ideal pinhole (iğne deliği) kamera varsayımını temel alır. Pinhole modelinde, 3B uzaydaki bir nokta ile görüntü düzlemi arasındaki ilişki, ışınların tek bir izdüşüm merkezinden geçerek odak düzlemine düşmesiyle tanımlanır. Şekil 2.1'de bu geometrinin iki boyutlu kesiti gösterilmektedir. Burada C kamera merkezi, odak düzlemi görüntünün oluştuğu düzlemi ve f odak uzaklığını ifade etmektedir. Z ekseni kameranın bakış doğrultusunu, X ekseni ise görüntü düzlemine paralel eksenlerden birini temsil eder. 3B uzayda X konumunda ve kameraya Z uzaklığındaki bir noktanın, odak düzlemindeki izdüşümü x olarak gösterilmiştir.



Şekil 2.1 Pinhole kamera modelinde perspektif izdüşüm: 3B noktaların kamera merkezinden geçerek odak düzlemine düşmesi

Pinhole modelinde benzer üçgenlerden yararlanılarak perspektif izdüşüm bağıntısı elde edilir. Kamera koordinat sisteminde bir 3B nokta $\mathbf{P}_c = [X,\ Y,\ Z]^T$ olsun. Bu noktanın odak düzlemindeki izdüşümü (x, y) Denklem (2.1) ile tanımlanır:

$$x = f\frac{X}{Z}, \qquad y = f\frac{Y}{Z}. \tag{2.1}$$

Bu ifadeler perspektif izdüşümün temel karakteristiğini ortaya koymaktadır. Aynı X, Y konumundaki bir nokta kameradan uzaklaştıkça görüntüde merkeze doğru yaklaşmakta ve küçülmektedir.

Gerçek sistemlerde görüntü koordinatları piksel cinsinden tutulmaktadır ve görüntü düzlemindeki (x, y) noktaları kamera iç (intrinsics) parametreleri kullanılarak piksel koordinatlarına Denklem (2.2) ile dönüştürülmektedir.

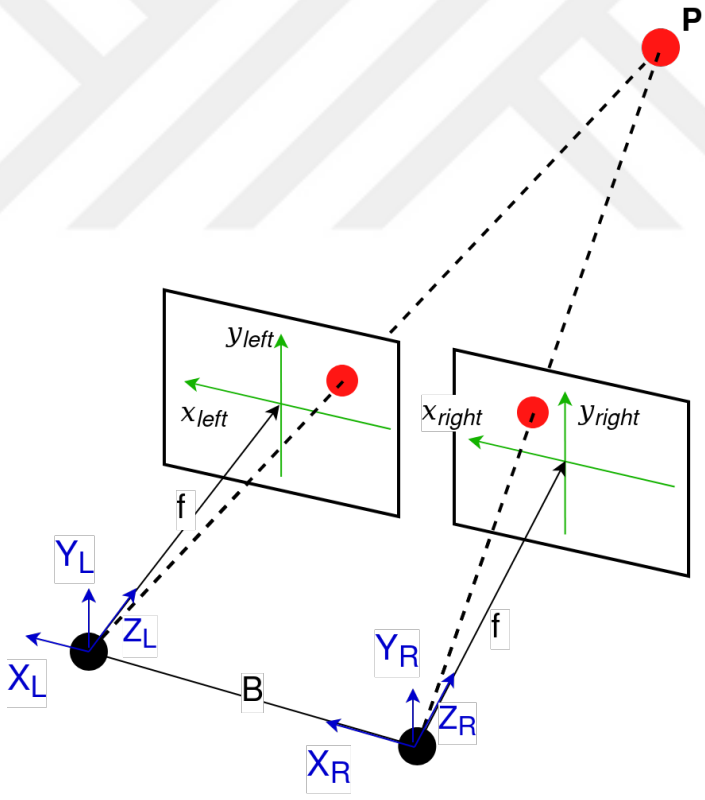
$$\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \qquad \mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2.2}$$

Denklem (2.2)'de f_x, f_y piksel biriminde odak uzunluklarını, $c_x, c_)$ ise ana bakış nokta (principal point) konumunu ifade etmektedir. İdeal bir pinhole kamera modelinde lens distorsiyonunun bulunmadığı ve projeksiyon modelinin doğrusal olduğu varsayılmaktadır. Bu varsayım altında iç parametre matrisi $\mathbf{K}$ sabittir. Kare pikselli sensörlerde en boy oranı bir kabul edilerek $f_x = f_y$ alınmakta ve piksel eksenlerinin birbirine dik olduğu varsayılmaktadır. İdeal durumda, ana bakış noktası (c_x, c_y) görüntünün tam ortasında kabul edilmektedir. Gerçek kameralarda ise lens distorsiyonu, otomatik odak/zoom değişimleri ve sensör hizasındaki

küçük sapmalar nedeniyle bu varsayımlar tam olarak sağlanamamaktadır. Bu nedenle iç(intrinsic) kalibrasyon ile $\mathbf{K}$ ve distorsiyon parametreleri kestirilerek görüntüler düzeltilmektedir. Lens distorsiyonu pinhole modelinin ideal varsayımına ters düştüğü için vSLAM yöntemlerinde genellikle kalibrasyon parametreleri ile distorsiyon giderimi uygulanarak düzeltmiş görüntüler kullanılmaktadır.

2.2 Stereo Geometri ve Derinlik Gözlemlenebilirliği

Monoküler sistemlerde tek bir görüntüden mutlak ölçek doğrudan gözlemlenemezken, stereo sistemlerde iki kamera arasındaki optik merkez uzaklığı (taban mesafesi) sayesinde derinlik doğrudan üçgenleme ile çıkarılabilir. Şekil 2.2 stereo üçgenleme geometrisi gösterilmektedir. Sol ve sağ kameraların optik merkezleri arasındaki taban mesafesi B ve her iki kameranın odak uzaklığı f bilindiğinde gerçek koordinat sistemindeki bir noktanın derinliği hesaplanabilmektedir.



Şekil 2.2 Stereo geometrisi ile üçgenleme tabanlı derinlik kestirimi

İç kalibrasyon ile düzeltilmiş stereo görüntülerde aynı 3B noktanın sol ve sağ görüntüdeki yatay piksel konumları arasındaki fark disparite Denklem (2.3)'te verilmiştir.

$$d = u_L - u_R. \tag{2.3}$$

Odak uzaklığı f ve taban mesafesi B kullanılarak derinlik yaklaşık olarak Denklem (2.4)'deki gibi hesaplanabilmektedir.

$$Z \approx \frac{fB}{d} \tag{2.4}$$

Bu denklem, stereo sistemlerin ölçeği gözlenebilir hale getirmesinin temel sebebidir. Öte yandan gerçek dünyada kameraya uzak olan bir noktanın disparitesi azaldığı için üçgenlemede derinlik hatasını arttırarak poz kestirimine olumsuz etki edebilmektedir.

2.3 Öznitelik Dedektörleri ve Betimleyiciler

Ardışık kareler arasında güvenilir geometrik bağlantılar oluşturabilmenin ön şartı, sahneden tekrar bulunabilir ve ayırt edici görsel öznitelikler çıkarabilmektir. Bu amaçla, bir öznitelik dedektörü görüntüde aday anahtar noktaları seçmektedir. Betimleyici (descriptor) ise seçilen her noktanın çevresini sayısal bir imza ile temsil ederek kareler arası eşleştirmeyi mümkün kılmaktadır. Dedektör ve betimleyici tasarımı; aydınlatma değişimleri, görüntü gürültüsü, bakış açısı farklılıkları, bulanıklık ve ölçek değişimi gibi etmenlere karşı dayanım ile hesaplama maliyeti arasında doğrudan bir ilişki içerisindedir [14, 15].

2.3.1 Köşe ve Anahtar Nokta Dedektörleri

Geleneksel köşe dedektörleri, yerel görüntü gradyanlarının hem yatay hem de düşey doğrultuda belirgin değiştiği bölgeleri seçmeyi hedeflemektedir. Harris köşe dedektörü [16], ikinci moment matrisi üzerinden köşe benzeri bölgeleri belirleyen öncü ve yaygın bir yaklaşımdır. Shi–Tomasi (Good Features to Track)[17] yöntemi ise Harris benzeri bir yapıdan yararlanıp özdeğer(eigenvalue) temelli bir seçim ölçütüyle takip edilebilir noktaları daha kararlı biçimde seçmeyi amaçlamaktadır . Rosten ve Drummond'un önerdiği FAST [18], piksel komşuluğunda hızlı bir parlaklık testi ile köşe adaylarını çok düşük maliyetle çıkarabildiğinden gerçek zamanlı uygulamalarda yaygın biçimde tercih edilmektedir.

2.3.2 Ölçek Uzayı Tabanlı Dedektör ve Betimleyiciler

Görüntüdeki anahtar noktaların farklı uzaklıklar ve yakınlaştırma seviyelerinde (ölçeklerde) tekrar bulunabilir olması, sistemin tutarlı bir şekilde çalışması için önemli bir gereksinimdir. Bu amaçla ölçek uzayı yaklaşımı, görüntüyü farklı ölçeklerde temsil ederek aynı yapının farklı boyutlarda

da algılanabilmesini hedeflemektedir. Lowe tarafından önerilen SIFT[14] yaklaşımı, Gauss bulanıklaştırma ile oluşturulan ölçek uzayında farkların farkı (Difference-of-Gaussians) üzerinden yerel maksimum ve minimum arayarak ölçekten bağımsız anahtar noktalar üretmektedir. Böylece anahtar noktaların hem konumu hem de karakteristik ölçeği birlikte tahmin edilmekte ve bu bilgi, eşleştirmede ölçek değişimlerine karşı dayanım sağlamaktadır.

SIFT, her öznitelik noktası için yerel gradyan dağılımını özetleyen yüksek boyutlu bir betimleyici tanımlayarak aydınlatma ve küçük bakış açısı değişimlerine karşı dayanıklılık sunmaktadır. Bununla birlikte, SIFT'in hesaplama maliyeti yüksek olması sebebiyle gerçek zamanlı uygulamalarda daha hızlı alternatiflere ihtiyaç duyulmuştur. Bay ve arkadaşları tarafından önerilen SURF[19], Gauss türevleri yerine kernel filtreleri ve integral görüntü yapısını kullanarak hem ölçek uzayı hesaplamasını hem de betimleyici çıkarımını hızlandırmayı amaçlamaktadır . Bu sayede SIFT benzeri bir sağlamlık hedeflenirken, pratikte daha düşük gecikmeyle çalışabilen bir öznitelik çıkarımı elde edilmektedir.

2.3.3 İkili Betimleyiciler

SIFT/SURF gibi kayan nokta tabanlı betimleyiciler yüksek ayırt edicilik sunarken eşleştirme maliyeti açısından gerçek zamanlı uygulamalarda daha hızlı eşleştirme ihtiyacı doğurmaktadır. Bu sebeple BRIEF[20], BRISK[21] ve ORB[22] gibi ikili betimleyiciler, piksel çiftleri üzerinden yapılan basit karşılaştırmalar ile bit dizileri üreterek Hamming uzaklığı ile hızlı bir eşleştirme imkanı sağlamaktadır.

BRIEF, ikili betimleyici yaklaşımının ilk örneklerinden biridir [23] ve yerel bir parça içindeki piksel çiftlerini karşılaştırarak bir bit dizisi üretmektedir. Bu sayede öznitelik eşleştirmesi, Hamming uzaklığı ile çok düşük maliyetle hesaplanabilmektedir. Ancak BRIEF tek başına dönme ve ölçek değişimlerine karşı dayanıksızdır.

ORB, bu eksiklikleri gidermek ve gerçek zamanlı uygulamalara uygun bir çözüm üretmek amacıyla FAST dedektörünü yönelim kestirimi ile birleştirmekte ve BRIEF betimleyiciyi yönlendirilmiş biçimde kullanarak hem hız hem de pratik dayanım hedeflemektedir.

Benzer biçimde BRISK, görüntü parçası üzerinde ölçekli dairesel bir örnekleme deseni kullanarak ikili karşılaştırmalar tanımlamakta ve bu karşılaştırmaların sonuçlarını bir bit dizisine dönüştürerek betimleyiciyi oluşturmaktadır. Böylece, örnekleme düzeninin farklı yarıçaplarda tanımlanması sayesinde ölçek değişimlerine karşı daha toleranslı bir temsil elde etmek hedeflenmektedir.

İkili betimleyiciler, hesaplama bütçesini korurken yeterli tekrar bulunabilirliği ve eşleşme kalitesini pratikte sağlamayı hedeflemektedir[24].

2.3.4 Çizgi Öznitelik Detektörleri ve Betimleyicileri

Nokta öznitelikleri, tekrar bulunabilirlik ve hızlı eşleştirme açısından pratikte yaygın olsa da, düşük dokulu sahnelerde yeterli öznitelik noktası tespit etmek zorlaşabilmektedir. Bu tür durumlarda çizgi öznitelikleri, kenar yapılarını temsil ederek ek geometrik bilgi sağlayabilmektedir. Literatürde güncel olarak, von Gioi ve arkadaşlarının önerdiği LSD [25] yöntemiyle çizgi öznitelikleri tutarlı bir şekilde tespit edilmesi amaçlanmaktadır. LSD, görüntüdeki kenar yönelimlerini istatistiksel anlamlılık temelli bir ölçütle değerlendirerek çizgi parçalarını öne çıkarmakta ve sahne yapısını temsil eden çizgileri gürültüden daha iyi ayırmayı hedeflemektedir.

Çizgi özniteliklerinin kareler arası eşleştirilmesi için, çizgi çevresindeki yerel görünümü temsil eden betimleyiciler kullanılmaktadır. Örneğin Zhang ve Koch tarafından önerilen LBD [26], çizgi boyunca gradyan düzenlerini özetleyerek çizgi eşleştirmeyi daha kararlı hâle getirmeyi amaçlamaktadır.

Nokta ve çizgi özniteliklerinin birlikte kullanımı, nokta tabanlı özniteliklerin yetersiz kaldığı durumlarda ek geometrik ipucu sağlayarak izleme ve eşleştirme adımlarına katkı sunabilmektedir.

2.3.5 Temel Dedektörlerin ve Betimleyicilerin Karşılaştırılması

Bu bölümde, incelenen öznitelik dedektörleri ve betimleyicileri ölçek ve rotasyon dayanıklılığı ile betimleyici yapıları açısından karşılaştırmalı olarak değerlendirilmiştir. Bu yöntemlerin temel özellikleri Tablo 2.1'te özetlenmiştir.

Tablo 2.1 Öznitelik dedektörleri ve betimleyicileri

Yöntem	Tür	Ölçek Dayanıklığı	Rotasyon Dayanıklığı	Betimleyici Tipi
Harris	Dedektör	×	×	-
Shi-Tomasi	Dedektör	×	×	-
FAST	Dedektör	×	×	-
SIFT	Dedektör + Betimleyici	✓	✓	128 Boyutlu Vektör
SURF	Dedektör + Betimleyici	✓	✓	64 Boyutlu Vektör
BRIEF	Betimleyici	×	×	Binary Vektör
ORB	Dedektör + Betimleyici	✓	✓	Binary Vektör

Tablodan görülebileceği üzere, klasik köşe dedektörleri (Harris, Shi–Tomasi, FAST) düşük hesaplama maliyetine sahip olmalarına rağmen ölçek ve rotasyon değişimlerine karşı sınırlı çözüm göstermektedir. Buna karşılık SIFT ve SURF gibi yöntemler daha yüksek hesaplama maliyeti ile birlikte sahne değişimine karşı daha

gürbüz özellikler sunmaktadır. ORB ise ikili betimleyici yapısı sayesinde gerçek zamanlı uygulamalar için dengeli ve stabil bir çözüm sağlamaktadır.

2.4 Dolaylı (Öznitelik Tabanlı) Görsel SLAM Yöntemleri

Dolaylı yöntemler, görüntüden ayırt edici özniteliklerin çıkarılmasına ve bu özniteliklerin kareler arasında eşleştirilmesine dayanmaktadır. Öznitelik temsili olarak köşe noktalar, çizgiler veya her ikisinin birleşimi kullanılabilmektedir. Özniteliklerin tekrar bulunabilirliği, aydınlatma değişimleri, görüntü gürültüsü ve bakış açısı farklılıklarına karşı dayanım açısından önemlidir. Bu yaklaşımlarda genel olarak işlemler aşağıdaki adımlardan oluşmaktadır.

1. Öznitelik Çıkarımı
2. Eşleştirme ve Takip
3. Poz Kestirimi

Öznitelik tabanlı yöntemler, her piksel yerine belirli bir ölçüm kümesiyle çalıştığı için uzun süre gerçek zamanlı uygulamalarda yaygın biçimde tercih edilmiştir.

Literatürdeki yaygın tercih edilen sistemler arasında, gerçek zamanlı çalışabilen ve döngü kapanımı ile büyük ölçekli haritalamayı destekleyen yöntemler özellikle öne çıkmaktadır. Bu bağlamda ORB tabanlı öznitelikleri kullanan ORB-SLAM mimarisi [27, 28], geniş ölçekli haritalama ve döngü kapanımı yetenekleri ile öne çıkan yöntemler arasında gösterilmektedir. Campos ve arkadaşları tarafından geliştirilen ORB-SLAM3[29] ise bu yaklaşımı görsel ataletsel ve çoklu harita senaryolarına genişleterek daha yüksek doğruluk ve kararlılık elde etmeyi amaçlamaktadır. ORB-SLAM ailesine ek olarak, stereo kamera kurulumları için öznitelik tabanlı bir çözüm sunan S-PTAM [30], seyrek öznitelik noktalarının tespiti ve eşleştirilmesine dayalı paralel izleme haritalama yapısını stereo ölçümlerle birleştirerek ölçek kestirimini ve 3B yeniden yapılandırmayı daha kararlı hâle getirmeyi hedeflemektedir. Bu çalışmalara ek olarak Ferrera ve arkadaşlarının sunduğu OV^{2}SLAM [31] ise, güçlü bir görüntü işleme ön işleme birimi ile grafik tabanlı optimizasyonu birleştirmeyi hedeflemektedir. Stereo kurulumlarda ölçeğin doğrudan gözlemlenebilmesini sayesinde ardışık kareler arasında güvenilir eşleşmeleri koruyacak şekilde görsel kısıtları kullanarak harita tutarlılığını desteklemektedir. Özellikle uzun sekanslarda hata birikiminin azaltılmasını ve ölçekli yörünge kestiriminin kararlı şekilde sürdürülmesini amaçlamaktadır. Bununla birlikte stereo tabanlı sistemlerde doğruluğu sınırlayan bir diğer önemli

etken, 3B noktaların stereo üçgenleme ile elde edilmesi sırasında ortaya çıkan derinlik belirsizliğininin poz kestirimine etkisidir. Zeng ve arkadaşları [32] stereo üçgenleme ile elde edilen 3B noktaların derinlik eksenindeki belirsizliğinin, klasik PnP tabanlı poz kestiriminde sistematik bir sapma üretebileceğini göstermektedir. Özellikle uzak mesafeli noktalar ve küçük disparite değerleri baskın olduğunda, derinlik dağılımı simetrik bir Gauss varsayımından sapmakta ve üçgenleme hatası 3B noktaların konumunu yönlü biçimde etkileyebilmektedir. Bu durumda, yeniden izdüşüm hatasını minimize eden standart PnP çözümleri, belirsizliği yüksek noktaları da benzer ağırlıkla kullandığı için kamera pozu üzerinde birikimli bir hata oluşabilmektedir. Önerilen yaklaşım, stereo ölçüm modelini ve üçgenleme belirsizliğini dikkate alarak PnP maliyet fonksiyonuna sistematik hata azaltıcı terimleri eklemekte ve belirsizliği yüksek gözlemlerin etkisini azaltacak şekilde ağırlıklandırma stratejileri uygulamaktadır. Böylece, özellikle uzak yapıların baskın olduğu sekanslarda, poz kestirimindeki sistematik sapmanın azaltılması ve daha tutarlı bir yörünge takibi hedeflenmektedir.

2.5 Doğrudan (Direct) Görsel SLAM Yöntemleri

Doğrudan yöntemlerin temel çalışma prensibi, iki ardışık görüntü arasında parlaklığın korunumu varsayımına dayanarak fotometrik hata çözümüne dayanmaktadır. Bu yaklaşıma göre, aynı dünya noktasına karşılık gelen piksellerin farklı karelerdeki yoğunluk değerlerinin aynı kalması beklenmektedir. Bu doğrultuda, ardışık kareler arasındaki hareket, fotometrik hatanın (e) minimize edilmesiyle Denklem (2.5) ile hesaplanmaktadır.

$$e(\mathbf{T}) = I_{ref}(\mathbf{u}) - I_{hed}(\omega(\mathbf{u}, d_{\mathbf{u}}, \mathbf{T})) \tag{2.5}$$

Burada I_{ref} referans görüntü karesini, I_{hed} hedef görüntü karesini, $\mathbf{T} \in SE(3)$ kareler arasındaki bağıl hareketi, $\mathbf{D}_p$ ise derinlik bilgisini temsil etmektedir. ω fonksiyonu ise bir pikselin derinlik ve hareket parametreleri kullanılarak diğer kareye izdüşümünü sağlayan görüntü dönüştürme (warping) fonksiyonu olarak ifade edilmektedir.

Denklem (2.5) ile hesaplanan fotometrik hata, optimizasyon problemi olarak ele alınmaktadır. Bu hatayı en aza indirecek uygun konum değerini bulmak amacıyla Gauss-Newton gibi iteratif yöntemler kullanılarak, kameranın bir kareden diğerine geçerken yaptığı hareket optimize edilmektedir.

Doğrudan yöntemler, işlenen veri yoğunluğuna göre yoğun, seyrek ve yarı yoğun

olarak sınıflandırılmaktadır. Yoğun yöntemlerin öncü çalışmalarından Newcombe ve arkadaşları tarafından geliştirilen DTAM [33], görüntünün belli bir kısmına odaklanmak yerine tüm pikseller için derinlik kestirimi yaparak çevrenin eksiksiz bir 3B modelini oluşturmayı hedeflemektedir. Ancak DTAM'in sunduğu bu yoğunluk, çok yüksek işlem gücü gerektirmekte ve genellikle GPU olmadan gerçek zamanlı çalışamamaktadır. Bu noktada Engel ve arkadaşları tarafından sunulan LSD-SLAM [34], yoğun yöntemlerin hesaplama yükünü hafifletmek amacıyla yarı yoğun bir yaklaşım önermektedir. Bu yöntemde, görüntünün tamamı yerine sadece belirgin renk geçişi içeren pikselleri işleyerek SLAM çözümünün hızlandırılması hedeflenmektedir. Seyrek yöntemlerde ise sadece belirli piksel gruplarını kullanarak yüksek hız sağlanması amaçlanmaktadır. Engel ve arkadaşları tarafından önerilen DSO [35], görsel odometride yenilikçi bir bakış açısı getirmiştir. DSO, diğer doğrudan yöntemlerin aksine pikseller arası tutarlılık varsayımını yok sayarak, görüntü genelinde homojen şekilde örneklenmiş gradyanı yüksek piksel kümeleri üzerinde çalışmaktadır. Bu sayede DSO, hem dolaylı yöntemlerin hızına yaklaşmakta hem de doğrudan yöntemlerin güçlü olduğu düşük dokulu ortamlarda gürbüzlüğünü korumaktadır.

2.6 Yarı Doğrudan (Semi-Direct) Yaklaşımlar

Yarı doğrudan yaklaşımlar, poz kestiriminde doğrudan yöntemlerin fotometrik hizalama avantajından yararlanırken, veri ilişkilendirme tarafında dolaylı yöntemlerden faydalanarak yönetilebilir bir yapı ortaya koymaktadır. Bu hibrit tasarım, yoğun yöntemlerin yüksek hesaplama maliyetini azaltırken, öznitelik eşleştirmeye dayalı sistemlerin zorluk yaşadığı dokusuz sahnelerde daha kararlı izleme sağlamayı amaçlamaktadır. Forster ve arkadaşlarının geliştirdiği SVO[36] ile gradyanı yüksek bilgi açısından zengin pikseller parça parça seçilmekte ve izleme adımında bu yamalar üzerinden fotometrik hata minimize edilerek kamera pozu güncellenmektedir. Haritalama tarafında ise yoğun bir yüzey yerine seyrek 3B noktalar tutulmakta ve yeni noktaların derinliği, çoklu görüş geometrisi altında epipolar arama benzeri süreçlerle kestirilerek yerel harita güncellenmektedir. Böylece sistem, fotometrik hizalamanın hızını korurken, seyrek geometrik harita sayesinde daha az işlem yüküyle çalışmaktadır.

Doğrudan ve dolaylı yöntemlere ek olarak Zhao ve arkadaşları [37], öznitelik tabanlı ön uçtan elde edilen kaba poz kestirimini, piksel yoğunluklarına dayalı doğrudan görüntü hizalama ile birleştiren yarı doğrudan bir yapı önermektedir. Bu yaklaşımda, öznitelik eşleştirmeleri büyük hareketlerde ve geniş görüş değişimlerinde sağlam bir başlangıç sağlamaktadır. Sonrasında, doğrudan

yöntemle, seçili görüntü bölgelerinde fotometrik hataların minimize edilmesiyle pozu iyileştirerek yerel doğruluğu arttırmayı hedeflemektedir. Böylece sistem, öznitelik tabanlı yöntemlerde görülebilen eşleşme gürültüsünü azaltabilirken, doğrudan yöntemlerde ortaya çıkabilen başlangıç duyarlılığına karşı da daha dayanıklı bir sonuç sunmaktadır. Özellikle kareler arası hareketin görece sınırlı olduğu ve fotogrametrik tutarlılığın yüksek olduğu koşullarda, bu tür yarı doğrudan adımı hatayı daha da düşürmeye katkı sağlayabilmektedir.

2.7 Dinamik Ortamlarda Görsel SLAM Yaklaşımları

Dinamik sahnelerdeki hareketli nesneler, eşleştirilen ölçümlerin statik dünya varsayımını ihlal etmesi sebebiyle konumlama ve haritalama doğruluğunu belirgin biçimde düşürebilmektedir. Bu motivasyonla literatürde hareketli nesnelerin etkisini bastırmaya yönelik yaklaşımlar genel olarak üç ana başlıkta toplanmaktadır:

1. Semantik Bilgi Tabanlı filtreleme
2. Optik akış tutarlılığı tabanlı filtreleme
3. Arka planı yeniden yapılandırarak statik çevreyi koruma

Bu bölümde ilgili yaklaşımlar incelenerek, güçlü ve zayıf yönleri tartışılmaktadır.

2.7.1 Semantik Bilgi Tabanlı Filtreleme

Semantik tabanlı yaklaşımlar, makine öğrenmesi tabanlı nesne tespiti veya segmentasyonu kullanarak potansiyel olarak hareketli sınıflara ait bölgeleri belirleyip bu alanlardan gelen ölçümlerin SLAM işlemi sırasında çıkarılmasını hedeflemektedir. Bu sınıfın temel avantajı, hareketli bölgeleri görüntü düzleminde doğrudan işaretleyebilmesi ve özellikle insan,araç gibi sık görülen hareketli sınıflarda yüksek seçicilik sağlayabilmesidir.

YOLOv8-SLAM [38] sistemde çıkarılan öznitelik noktalarını, tespit edilen dinamik bölgelerle örtüşmeleri durumunda öznitelik havuzundan çıkartarak bir filtreleme uygulamaktadır. Ancak maskelemenin öznitelikler tespit edildikten sonra yapılması, özellikle görüntüdeki öznitelik dağılımının dengesini bozabilmekte ve yerel izleme hatalarının büyümesine neden olabilmektedir.

Bu tür etkileri azaltmak amacıyla Gökçen ve Uslu [39] dinamik ortamlarda RGB-D tabanlı vSLAM'de hareketli nesnelerin ürettiği öznitelikleri aykırı değer olarak

ele alıp, sisteme girmeden önce görüntüyü filtreleyen bir yaklaşım önermektedir. Bu yöntemde RGB görüntülerde YOLOv3 [40] ile insan tespiti yapılmakta, tespit edilen bölgelerden çıkarılan ORB öznitelikleri maskelemeye tabi tutularak haritalama sürecinden çıkarılmaktadır. Böylece sahnede tespit edilen öznitelikler daha homojen dağılım göstermektedir.

2.7.2 Optik Akış Tutarlılığı Tabanlı Filtreleme

Optik akış analizi ile filtreleme yapan sistemler, dinamik bölgeleri yalnızca semantik sınıf bilgisinden değil, sahnedeki piksel hareketinin geometrik olarak beklenen hareketle ne kadar tutarlı olduğundan yararlanarak belirlemektedir. Temel fikir, statik dünya varsayımı altında kamera hareketiyle açıklanabilen görüntü hareketi ile gerçek gözlenen hareket arasındaki farkların hareketli nesneleri ortaya çıkarmasıdır. Bu sayede semantik olarak hareketli sınıfta olup o anda sabit duran nesneler daha doğru ele alınabilmektedir.

Yu ve arkadaşları tarafından önerilen DS-SLAM [41], ORB-SLAM2 tabanlı bir vSLAM sistemini dinamik ortamlara daha dayanıklı hale getirmek amacıyl SegNet [42] ile belirlenen nesne alanlarını optik akış temelli bir kontrol mekanizmasıyla birleştirerek hareketli bölgeleri daha doğru biçimde filtrelemeyi amaçlamaktadır. Sistem mimarisinde öznitelik takibi, semantik segmentasyon, yerel haritalama, döngü kapama ve harita üretimi paralel beş thread üzerinde çalışarak sistemin işlem hızını arttırmaktadır. Segmentasyon çıktısına göre sahnedeki potansiyel dinamik sınıflar tespit edildiğinde hareket tutarlılığı kontrolü adımıyla bu nesnenin gerçekten hareketli olup olmadığı tespit edilmektedir. Eğer nesne hareketli çıkarsa nesnenin sınırları içine düşen ORB öznitelikleri eşleştirmesi sırasında sistemden çıkarılarak dinamik nesnelerden gelen hatalı verilerin işlenmesi engellenmektedir.

Benzer şekilde He ve arkadaşlarının sunduğu DIO-SLAM [43], dinamik sahnelerde yalnızca nesne sınıfına bakarak maskeleme yapmanın gereksiz ölçüm kaybına yol açabilmesi probleminden yola çıkarak sahnedeki her nesneyi yoğun optik akış ile tutarlılığını kontrol ederek dinamik ya da statik kararını vermektedir. Bunun üstüne ORB-SLAM3 ile paralel çalıştırılan bir yapı dinamik kararı verilen objelerin üstünden çıkarılan öznitelik noktalarını eleyerek sistemin başarısını arttırmaktadır.

Fan ve arkadaşları tarafından geliştirilen Blitz-SLAM [44], iç mekan dinamik sahnelerde semantik bilgiyi geometrik doğrulama ile birleştirerek potansiyel dinamik bölgelerdeki aykırı ölçümleri bastırmayı amaçlamaktadır. Sistem, BlitzNet [45] ile elde edilen nesne sınıflandırma çıktılarından semantik maskeler

üretmektedir. Ardından çoklu görüş geometrisine dayalı epipolar kısıtlar ve derinlik bilgisini kullanarak bu maskeler içerisindeki ölçümlerin statik dünya varsayımıyla uyumunu test etmektedir. Bu sayede yalnızca sınıf temelli maskelemenin yol açabileceği gereksiz ölçüm kaybı azaltılırken gerçekten hareketli bölgelerde biriken geometrik aykırı noktaların yoğun haritalamaya taşınması engellenir. Yazarlar, bu seçici bastırma mekanizmasının nokta bulutu üzerinde "noise blocks" olarak adlandırılan gürültülü yapıların azalmasına katkı sağlayarak daha temiz ve tutarlı bir statik harita üretmeyi hedeflemektedir.

2.7.3 Arka Planı Yeniden Yapılandırarak Statik Çevreyi Koruma

Dinamik nesnelerin ölçümlerinin doğrudan elenmesi, izleme hatalarını azaltırken harita üzerinde boşluklar meydana getirmektedir. Bu nedenle bazı yöntemler, dinamik bölgeleri yalnızca maskelemek yerine, bu bölgelerdeki statik arka planı yeniden inşa ederek daha tutarlı ve uzun vadede kullanılabilir bir harita elde etmeyi hedeflemektedir. Özellikle yoğun haritalama yapan sistemlerde, hareketli nesnelerin nokta bulutu üzerinde biriktirdiği gürültülü yapılar zamanla kirli bir haritaya neden olabilmesi sebebiyle arka planı korumaya yönelik stratejiler gürültüsüz bir harita üretmek için önem arz etmektedir.

Arka planı yeniden yapılandırma amacıyla çalışan DynaSLAM [46], ORB-SLAM2 tabanlı bir yapı üzerine dinamik nesne tespiti ekleyerek hareketli bölgelerden gelen ölçümleri izleme ve haritalama süreçlerinden sistematik biçimde çıkartmayı hedeflemektedir. DynaSLAM'in en etkin yönü, dinamik nesnelerin maskeleme sonrası harita üzerinde oluşturduğu boşlukları azaltmak için arka plan tamamlama adımını sisteme entegre etmesidir. Bu sayede hareketli nesnelerle örtülen alanlarda statik arka planın görsel olarak tamamlanması hedeflenmekte ve uzun vadede boşluksuz bir statik harita üretimi mümkün kılınmaktadır. Özellikle yoğun haritalama gerektiren senaryolarda dinamik nesnelerin sahnede sıklıkla bulunmasına rağmen statik çevre modelinin korunabilmesi DynaSLAM'in güçlü yönlerinden biridir.

Zhang ve arkadaşları FlowFusion [47] ile dinamik sahnelerde RGB-D SLAM problemini, optik akış tabanlı dinamik statik ayrım ile statik arka planın yeniden inşasını eş zamanlı ele alan bir çalışma olarak sunmaktadır. Yöntemde ardışık iki RGB-D görüntü girdi olarak alınarak RGB görüntüler üzerinden bir optik akış ağı ile yoğun optik akış kestirilirken, derinlik ve yoğunluk bilgisi birlikte kullanılarak kamera hareketi fotogrametrik hataları minimize eden bir optimizasyonla tahmin edilmektedir. Elde edilen kamera hareketi kullanılarak statik dünya varsayımı altında beklenen görüntü hareketi hesaplanmakta ve bu beklenen hareket ile

gözlenen optik akış arasındaki farklardan hareketli bölgeler belirlenmektedir. Dinamik olarak işaretlenen alanlar haritalama güncellemelerinden çıkarılarak, hareketli nesnelerin hem konum kestirimi hem de yoğun arka plan rekonstrüksiyonu üzerindeki olumsuz etkisinin azaltılması hedeflenmektedir. Bu sayede yöntem, dinamik nesnelerle örtülen sahnelerde daha temiz ve tutarlı bir statik arka plan modeli üretmeyi amaçlamaktadır.

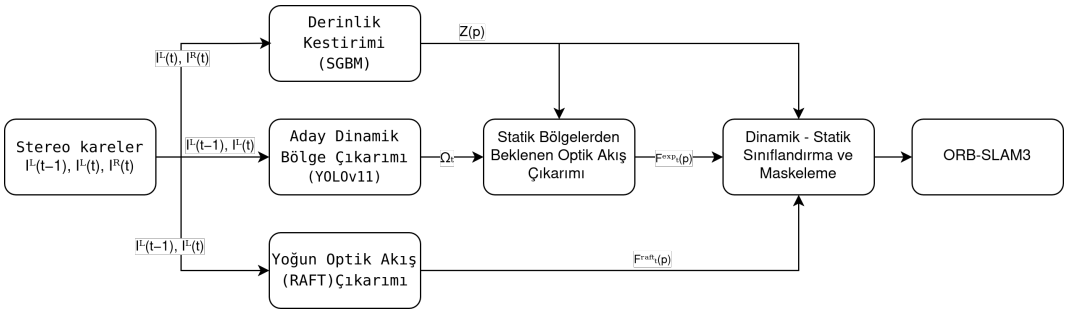
3
YÖNTEM

Bu bölümde, dinamik sahnelerde stereo kamera tabanlı öznitelik tabanlı vSLAM sistemlerinin doğruluğunu artırmak amacıyla geliştirilen önerilen yaklaşım sunulmaktadır. Yöntem, stereo ORB-SLAM3 vSLAM hattına eklenen bir ön işleme katmanı olarak tasarlanmış olup, sahnedeki hareketli bölgeleri maskeleyerek poz kestirimi aşamasında dinamik nesnelerden kaynaklanan hataları azaltmayı hedeflemektedir.

Bu kapsamda, ilk olarak Bölüm 3.1'de sistemin genel mimarisi ve veri akışı açıklanmaktadır. Ardından Bölüm 3.2'te stereo görüntülerden derinlik çıkarımı süreci verilmekte, Bölüm 3.3'te semantik bilgiler aracılığıyla dinamik obje adaylarının belirlenmesi anlatılmaktadır. Devamında, yalnızca statik bölgelerden seçilen ölçümler kullanılarak kamera hareketinin kestirildiği Bölüm 3.4'te sunulmaktadır. Bu hareket kestirimi altında statik sahne varsayımıyla beklenen optik akışın türetilmesi Bölüm 3.5'te açıklanmaktadır. Yoğun optik akışın derin öğrenme tabanlı hesaplanması Bölüm 3.6'te ele alınmaktadır. Gözlenen ve beklenen akışın karşılaştırılmasıyla dinamik veya statik karar mekanizması ve maskeleme adımı Bölüm 3.7'de tanımlanmaktadır. Son olarak, üretilen maskelerin ORB-SLAM3 hattına entegrasyonu Bölüm 3.8'te sunulmaktadır.

3.1 Genel Sistem Mimarisi

Önerilen yaklaşım, dinamik sahnelerde çalışan stereo kamera tabanlı sistemlerde, öznitelik tabanlı vSLAM'in dayanıklılığını artırmayı hedefleyen bir ön işleme katmanı olarak tasarlanmıştır. Bu katman, dinamik bölgeleri semantik bilgiler ve derin öğrenme tabanlı elde edilen optik akışı füzyonlayarak maskelemektedir. Şekil 3.1'de sistemin mimarisi verilmiştir. Önerilen ön işleme katmanı, her zaman adımı t için $(I^L_{t-1}, I^L_t, I^R_t)$ t anındaki sol ve sağ görüntü ile $t-1$ anındaki sol görüntü setini girdi olarak kullanmaktadır. Bu görüntüler işlenerek sol ve sağ görüntülerdeki dinamik bölgeleri ifade eden bir maske alanı oluşturulmaktadır.



Şekil 3.1 Önerilen çalışmanın genel mimarisi

Ön işleme katmanının ilk adımı, iç kamera kalibrasyonu ile lens bozulması giderilmiş stereo görüntüler (I_t^L, I_t^R) kullanılarak Semi-Global Block Matching (SGBM) [48, 49] ile disparite tahmini ve ardından derinlik çıkarımıdır. Böylece her piksel için metrik ölçekte bir derinlik haritası $Z(\mathbf{p})$ elde edilmektedir. Derinlik haritası, daha sonraki adımlarda beklenen optik akışın hesaplanması ve kamera hareketinin geometrik olarak ilişkilendirilmesi için önemli bir rol oynamaktadır.

İkinci olarak, Sol görüntü I_t^L üzerinde çalışan YOLOv11 yapay sinir ağı, hareket edebilecek sınıflara ait örnek segmentation çıktılarından aday dinamik bölge kümesi Ω_t üretmektedir. Bu aşamada, dinamik kararın doğrudan verilmesinden ziyade, sonraki geometrik doğrulama adımlarında incelenecek ve karar verilecek bölge sağlamayı hedefler.

Ardından, ardışık sol kareler (I_{t-1}^L, I_t^L) arasında RAFT sinir ağı ile yoğun optik akış hesaplanmaktadır. RAFT çıktısı her piksel için $\mathbf{F}_t^{raft}(\mathbf{p})$ vektörünü vermektedir. Sistem hareketli olduğu için RAFT'ın çıktısı statik sahnede de optik akı üretebilmektedir. Bu nedenle sistem, statik sahne varsayımı altında beklenen akışı da hesaplayarak karşılaştırmalı bir karar mekanizması kurmaktadır.

Son olarak, beklenen optik akış $\mathbf{F}_t^{exp}(\mathbf{p})$, yalnızca statik bölgelerden seçilen ölçümlerle kestirilen kamera hareketi ve derinlik $Z(\mathbf{p})$ kullanılarak üretilmektedir. Böylece, kamera hareketinin açıklayabildiği akış ile uyuşmayan akış bileşenleri ayrıştırılabilmekte ve aday bölgeler Ω_t içinde $\mathbf{F}_t^{raft}(\mathbf{p})$ ile $\mathbf{F}_t^{exp}(\mathbf{p})$ arasındaki tutarsızlık ölçülerek dinamik/statik sınıflandırma yapılmaktadır. Dinamik olarak doğrulanan bölgeler maskelenerek yeni görüntüler üretilmektedir.

3.2 Stereo Derinlik Çıkarımı

Bu çalışmada disparite haritası, Semi-Global Matching (SGM) yaklaşımını temel alan SGBM yöntemi ile hesaplanmaktadır. SGBM, yerel eşleşme maliyetini küçük bir pencere (blok) üzerinde hesaplayıp, çoklu doğrultularda maliyet hesabı yaparak

komşuluk boyunca parça parça sürekli bir disparite alanı üretmeyi hedeflemektedir. Böylece düşük dokulu bölgelerde yerel eşleştirmenin yol açtığı bilgiden yoksun sonuçlar azaltılırken, derinlik sürekliliği korunmaya çalışılmaktadır. SGBM'nin amacı, her piksel için dispariteyi yalnızca yerel benzerlik ölçüsüyle değil, aynı zamanda komşuluk tutarlılığını gözeten bir süreklilik varsayımıyla birlikte tahmin etmektir. Bu süreçte, komşu pikseller arasında küçük disparite değişimlerine düşük, büyük sıçramalara ise daha yüksek ceza verilerek yüzey sürekliliği ile kenar korunumu arasında denge kurulmaktadır.

Bu durum, yöntemsel olarak bir enerji minimizasyonu problemi şeklinde ifade edilebilmektedir. Öncelikle her piksel $p = (u, v)$ ve her aday disparite d için, sol görüntüdeki p pikselinin sağ görüntüde $(u-d, v)$ noktasına eşleştirilmesine karşılık gelen bir yerel eşleşme maliyeti Denklem (3.1)'deki gibi tanımlanmaktadır.

$$C(p, d) = \text{cost}\big(I^L(p),\, I^R(u-d, v)\big) \tag{3.1}$$

Denklem (3.1)'de $C(p, d)$, her piksel için olası disparite değerleri üzerinde oluşturulan maliyet fonksiyonunu temsil etmektedir.

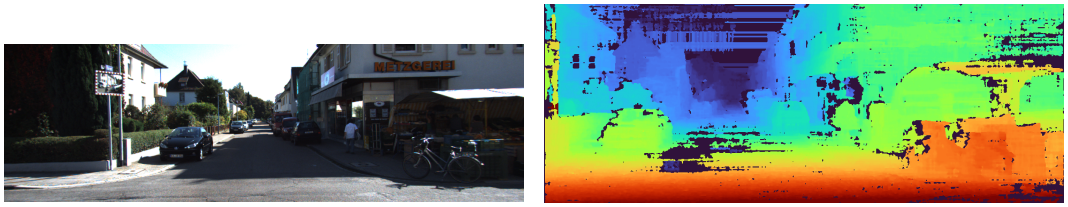
Sadece bu bilgi ile karar verilmesi durumunda, her piksel bağımsız biçimde en küçük maliyeti veren d değerini seçmektedir ve bu da özellikle düşük dokulu bölgelerde kararsız disparite haritalarına yol açmaktadır.

SGM yaklaşımı, bu yerel maliyeti komşuluk sürekliliğiyle birleştirerek tüm görüntü için Denklem (3.2)'deki gibi bir enerji fonksiyonunu minimize etmeyi hedeflemektedir.

$$E(D) = \sum_{p} \Big(C(p, D_p) + \sum_{q \in N_p} P_1\, T[|D_p - D_q| = 1] + \sum_{q \in N_p} P_2\, T[|D_p - D_q| > 1] \Big) \tag{3.2}$$

Denklem (3.2)'de D_p ise p pikselindeki disparite değerini göstermektedir. N_p, p pikselinin komşuluk kümesini ve $T[\cdot]$ ise koşul sağlandığında 1, aksi halde 0 değerini alan gösterge fonksiyonunu ifade etmektedir. P_1 ve P_2, sırasıyla küçük disparite değişimlerine ($|D_p - D_q| = 1$) ve daha büyük sıçramalara ($|D_p - D_q| > 1$) uygulanan ceza katsayılarıdır. P_2 ceza katsayısı ile büyük sıçralamalar daha çok cezalandırılarak tutarlılık güçlendirilmektedir. Ancak gerçek sahnede nesne sınırlarından oluşacak ani disparite değişimleri olabileceği için amaç bu ceza katsayılarını dengeli seçerek aşırı komşuluk birleştirme ile gürültü arasında bir uzlaşma sağlamaktır.

Tanımlanan enerji fonksiyonu, piksellerin tamamı arasında komşuluk ilişkileri içerdiği için iki boyutlu uzayda minimize edilmesi hesaplama açısından oldukça maliyetlidir. Bu nedenle SGM, çözümü yaklaşık olarak elde etmek için maliyetleri birden fazla doğrultu boyunca bir boyutlu yollar üzerinde birikimli olarak hesaplamaktadır. Farklı doğrultulardan gelen birikimli maliyetlerin toplanmasıyla her piksel ve her aday disparite için toplam maliyet elde edilmektedir. Son adımda ise disparite, toplam maliyeti en küçük yapan değer seçilerek belirlenmektedir.

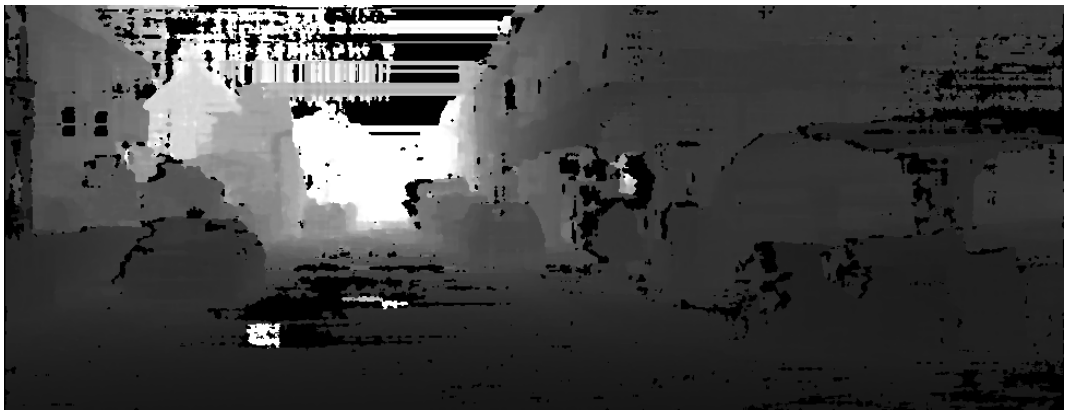


(a) Ham sol kamera görüntüsü I_t^L.

(b) SGBM ile hesaplanan disparite haritası

Şekil 3.2 Örnek stereo disparite haritası

Şekil 3.2(b), stereo görüntü çifti üzerinden SGBM ile elde edilen disparite haritasını göstermektedir. Disparite, sol görüntüdeki bir pikselin sağ görüntüdeki eşleşen noktasına yatay doğrultuda ne kadar kaydığına karşılık gelmektedir. Bu nedenle haritadaki büyük disparite değerleri kameraya daha yakın yüzeyleri, küçük disparite değerleri ise daha uzak bölgeleri temsil etmektedir. Görüntüde özellikle araçlar ve yol gibi belirgin geometrik yapılarda tutarlı disparite geçişleri gözlenirken, düşük dokulu alanlarda eşleşme belirsizliği nedeniyle daha gürültülü veya keskin geçişli bölgeler oluşabilmektedir.



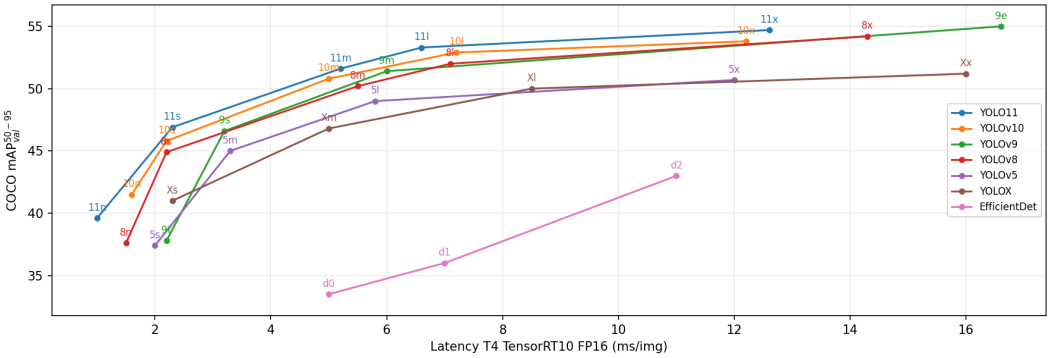
Şekil 3.3 Disparite haritasından elde edilen derinlik haritası

Elde edilen disparite haritası, Bölüm 2.2'de açıklandığı üzere stereo üçgenleme ilişkisi kullanılarak derinlik haritasına dönüştürülmektedir. Şekil 3.2(a) görüntüsünden oluşturulan derinlik haritası Şekil 3.3'te verilmiştir. Bu gösterimde her pikselin değeri, ilgili noktaya ait metrik derinliği (milimetre birimiyle)

temsil edecek şekilde 16-bit olarak kodlanmıştır. Ancak uzak ve dokusu az bölgelerde disparitenin çok küçük olması ve eşleşmenin belirsizleşmesi nedeniyle derinlik değerleri geçersiz pikseller oluşturabilmektedir. Kameradan uzaklaştıkça piksel değeri artmakta bu sebeple görüntü siyahtan beyaza doğru oluşmaktadır. Hesaplanamayan bölgeler ise geçersizliği ifade eden "0" değerini almaktadır.

3.3 Semantik Bilgi ile Dinamik Obje Adayı Çıkarımı

Önerilen yöntemde, hareketli nesneleri filtrelemek amacıyla semantik tabanlı bir aday bölge seçimi yapılmaktadır. Şekil 3.4'te, YOLO modellerinin doğruluk–gecikme karşılaştırması $\text{mAP}_{50:95}$–latency düzleminde verilmektedir. $\text{mAP}_{50:95}$ metriği, Kesişim–Birleşim Oranı (Intersection over Union, IoU) eşiklerinin 0.50–0.95 aralığında ortalaması alınarak hesaplanan ortalama kesinliği, latency ise tek görüntü için işleme süresini ifade etmektedir. IoU, modelin tahmin ettiği bölgenin yer doğruluğuna ne kadar uyduğunu ölçen standart bir örtüşme metriğidir. Bu karşılaştırmaya göre, doğruluk ve gecikme arasında en uygun dengeyi sağlayan YOLOv11 modeli seçilmiştir.



Şekil 3.4 Farklı YOLO ailelerinin COCO $\text{mAP}_{50:95}$ doğruluğu ile gecikme arasındaki karşılaştırma [50]

Her zaman adımında sol görüntü üzerinde YOLOv11 çalıştırılarak nesne tespiti/segmentasyonu elde edilmekte ve hareket edebilir sınıflara ait çıktılar aday dinamik bölgeler Ω_t olarak işaretlenmektedir. İşaretlenen Ω_t bölgeleri, segmentasyon maskeleri ile piksel düzeyinde tanımlanarak maskeleme işleminde kullanılmaktadır. Belirlenen Ω_t bölgeleri Şekil 3.5'te gösterilmiştir.

Elde edilen Ω_t bölgeleri üç amaç için kullanılmaktadır:

1. Kamera hareket kestiriminde kullanılacak statik ölçümlerin seçimi için maskeleme amacıyla



Şekil 3.5 Tespit edilen aday bölgeler

2. Beklenen akış ile gözlenen RAFT akışının karşılaştırılacağı alanın belirlenmesi için

3. Nihai olarak ORB-SLAM3 girdisinde filtrelenecek bölgelerin belirlenmesi için

3.4 Statik Bölgelerden Kamera Hareketi Kestirimi

Bu çalışmada kamera hareketi, hareketli bölgelerin poz kestirimini bozmasını önlemek amacıyla yalnızca statik kabul edilen alanlardan seçilen öznitelik noktaları üzerinden kestirilmektedir. Her zaman adımında, YOLOv11 ile elde edilen aday dinamik bölgeler Ω_{t-1} maskelenerek statik maske oluşturulur ve I^L_{t-1} üzerinde Shi–Tomasi köşe seçici ile ayırt edici noktalar belirlenmekte ve seçilen noktalar, Lucas–Kanade (LK) optik akış yöntemi ile I^L_t karesine taşınarak $p_{t-1} \rightarrow p_t$ biçiminde 2B karşılıklar elde edilir. İzlemenin kararlılığını artırmak için ileri–geri kontrol uygulanarak tutarsız eşleşmeler elenmektedir. Buna göre, önceki karede seçilen bir nokta p_{t-1}, Denklem (3.3) ile ileri yönde t anına izlenerek p_t elde edilmektedir. Ardından aynı nokta, Denklem (3.4) ile geri yönde tekrar $t-1$ karesine izlenerek $\hat{p}_{t-1}$ hesaplanmaktadır. İleri–geri izleme hatası Denklem (3.5) ile tanımlanmıştır ve e_{fb} değeri küçük olan eşleşmeler daha tutarlı kabul edilerek poz kestiriminde kullanılmaktadır.

$$p_t = \mathrm{LK}\big(I^L_{t-1}, I^L_t, p_{t-1}\big) \tag{3.3}$$

$$\hat{p}_{t-1} = \mathrm{LK}\big(I^L_t, I^L_{t-1}, p_t\big) \tag{3.4}$$

$$e_{\mathrm{fb}} = \|p_{t-1} - \hat{p}_{t-1}\| \tag{3.5}$$

Elde edilen 2B karşılıklardan kamera hareketinin kestirilebilmesi için, önceki karedeki her bir nokta $p_{t-1} = (u_{t-1}, v_{t-1})$ için stereo derinlik haritasından $Z_{t-1}(p_{t-1})$ uzaklık değeri okunarak Denklem (3.6) ile 3B düzleme geri izdüşürülmektedir.

$$P_{t-1} = \begin{bmatrix} X_{t-1} \\ Y_{t-1} \\ Z_{t-1} \end{bmatrix} = Z_{t-1}(p_{t-1})\, K^{-1} \begin{bmatrix} u_{t-1} \\ v_{t-1} \\ 1 \end{bmatrix} \tag{3.6}$$

Denklem (3.6)'da K kamera iç kalibrasyon matrisini ifade etmektedir ve K^{-1} ile $[u_{t-1}\ \ v_{t-1}\ \ 1]^T$ homojen piksel koordinatı normalize edilmiş bir bakış ışınına dönüştürülmektedir. Bu ışın $Z_{t-1}(p_{t-1})$ derinliği ile ölçeklenerek kamera koordinat sisteminde 3B nokta $P_{t-1} = [X_{t-1}\ \ Y_{t-1}\ \ Z_{t-1}]^T$ elde edilmektedir. Böylece, her bir takip edilen nokta için P_{t-1} (3B) ve p_t (2B) eşleştirmelerinden oluşan 3B–2B karşılıklar kümesi oluşturulmaktadır.

Elde edilen 3B–2B karşılıklar $\{(P_{t-1}^{(i)}, p_t^{(i)})\}_{i=1}^{N}$ üzerinden kamera hareketi, Perspektif-n-Nokta (PnP) problemi ile çözülerek kamera hareketi kestirilmektedir.

PnP çözümünde amaç, gözlenen $p_t^{(i)}$ noktaları ile Denklem (3.7) kullanılarak $P_{t-1}^{(i)}$ noktalarının görüntü düzlemine izdüşürülmesinden elde edilen tahmini 2B noktalar arasındaki yeniden-izdüşüm hatasını minimize eden R, t'yi bulmaktır.

$$\min_{R,t} \sum_{i=1}^{N} \left\| p_t^{(i)} - K\left(RP_{t-1}^{(i)} + t\right) \right\|^2 \tag{3.7}$$

Bu çalışmada PnP çözümü, gürbüzlüğü artırmak amacıyla RANSAC ile birlikte kullanılmaktadır. RANSAC, her iterasyonda 3B–2B karşılıklardan rastgele seçilen küçük bir alt küme ile aday bir kamera dönüşümü (R, t) üretmekte ve ardından bu aday dönüşüm, tüm karşılıklara uygulanarak her bir eşleşme için yeniden izdüşüm hatası hesaplanmaktadır.

Bir $P_{t-1}^{(i)}$ noktasının aday (R, t) altında görüntü düzlemine izdüşürülmesiyle elde edilen tahmini piksel konumu $\hat{p}_t^{(i)}(R, t)$ olarak tanımlanır. Buna göre yeniden izdüşüm hatası Denklem (3.8) olarak ifade edilebilir.

$$e_i = \left\| p_t^{(i)} - \hat{p}_t^{(i)}(R, t) \right\| \tag{3.8}$$

Hata eşiği $e_i < \tau$ koşulunu sağlamayan eşleşmeler aykırı değer olarak kabul edilmekte ve örneklem uzayından elenmektedir. İterasyonlar sonunda en fazla geçerli eşleşme üreten aday model seçilmekte ve nihai kamera hareketi olarak kabul edilmektedir.

3.5 Statik Varsayım ile Beklenen Optik Akı Çıkarımı

Önceki bölümde elde edilen kamera rotasyon matrisi, öteleme vektörü (R, t) ve önceki kareye ait derinlik haritası $Z_{t-1}(\cdot)$ kullanılarak, önceki karedeki bir pikselin t anında görüntü düzlemindeki hangi 2B konuma taşınacağı öngörülmektedir. Böylece beklenen akış vektörleri hesaplanabilmektedir.

Önceki karede seçilen bir piksel $p_{t-1} = (u_{t-1}, v_{t-1})$ için derinlik değeri $z = Z_{t-1}(p_{t-1})$ elde edilir. İç kamera kalibrasyon parametreleri K (f_x, f_y, c_x, c_y) ile ilgili pikselin kamera koordinatlarındaki 3B karşılığı Denklem (3.9) ile hesaplanmakta ve $P_{t-1} = [X_{t-1}\ Y_{t-1}\ Z_{t-1}]^T$ olarak tanımlanmaktadır.

$$X_{t-1} = \frac{(u_{t-1} - c_x)\, z}{f_x}, \qquad Y_{t-1} = \frac{(v_{t-1} - c_y)\, z}{f_y}, \qquad Z_{t-1} = z_{t-1} \tag{3.9}$$

Statik sahne varsayımında ilgili 3B nokta, kamera hareketi ile t anındaki kamera koordinat sistemine Denklem (3.10) ile taşınmakta ve 3B ilgili nokta konumu $P_t = [X_t\ Y_t\ Z_t]^T$ olarak tanımlanmaktadır.

$$P_t = R\, P_{t-1} + t \tag{3.10}$$

Sonrasında P_t noktasının görüntü düzlemindeki 2B konumu perspektif izdüşüm ile Denklem (3.11) ile elde edilmektedir.

$$u_t = f_x \frac{X_t}{Z_t} + c_x, \qquad v_t = f_y \frac{Y_t}{Z_t} + c_y \tag{3.11}$$

Bu durumda, p_{t-1} pikselinde beklenen optik akış vektörü Denklem (3.12) olarak hesaplanmaktadır.

$$F_t^{\exp}(p_{t-1}) = \begin{bmatrix} u_t - u_{t-1} \\ v_t - v_{t-1} \end{bmatrix} \tag{3.12}$$

Uygulamada bu hesap, yalnızca Ω_t aday bölgelerdeki pikseller için

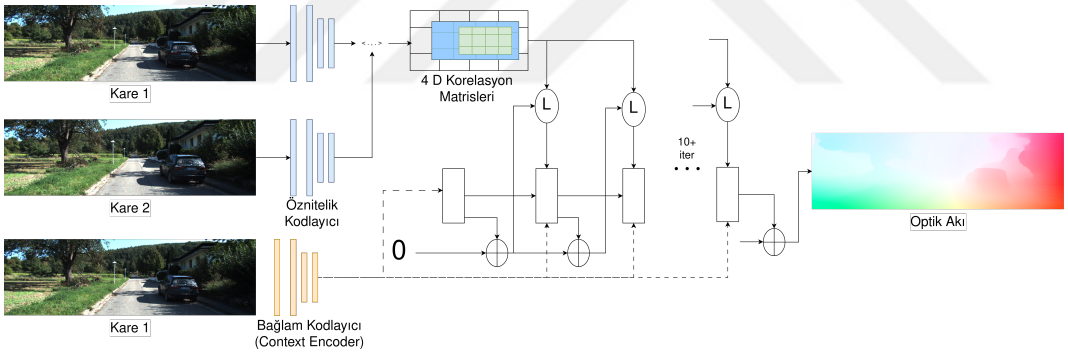
gerçekleştirilerek işlem yükünün hafifletilmesi amaçlanmaktadır.

3.6 Yoğun Optik Akı Çıkarımı

Bu aşamada ardışık iki sol görüntü kareleri (I^L_{t-1}, I^L_t) arasındaki piksel yer değiştirmelerini hesaplamak amacıyla, derin öğrenme tabanlı yoğun optik akı yöntemi RAFT[51] kullanılmaktadır. Şekil 3.6'te RAFT'in iki ardışık kare üzerinden yoğun optik akış çıkarmaya yönelik çalışma mimarisi gösterilmektedir.

İlk olarak iki ardışık kareyi ayrı birer öznitelik çıkarıcıdan geçirerek her piksel için birer temsil üretmektedir. Bu temsiller kullanılarak birinci karedeki her pikselin ikinci karedeki tüm piksellerle benzerliğini ifade eden bir eşleşme tablosu (4B korelasyon hacmi) oluşturulmaktadır. Bu sayede yöntem, yalnızca küçük yer değiştirmelerle sınırlı kalmadan daha büyük hareketleri de yakalayabilmektedir.

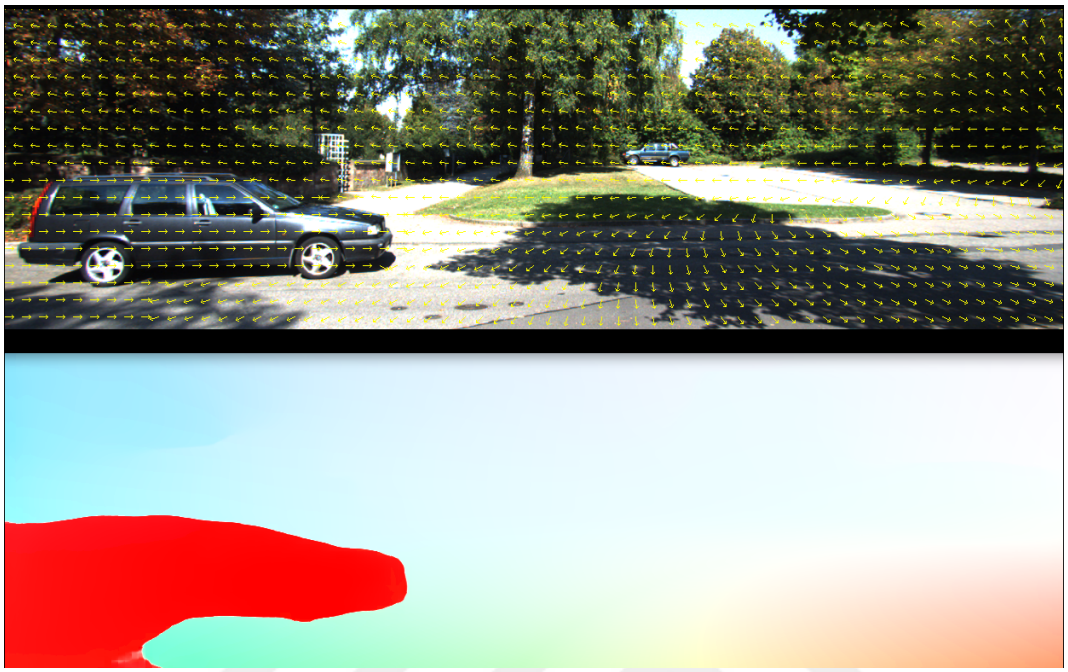
Bağlam kodlayıcı (Context Encoder) bloğu ise her iterasyonda güncelleme bloğuna kenar, doku, nesne sınırı gibi yerel sahne bilgisini sağlayarak hangi bölgelerin daha güvenilir olduğu bilgisi ile optik akışı tutarlı bir şekilde hesaplamaya yardımcı olmaktadır.



Şekil 3.6 RAFT modelinin temel mimarisi [51]

RAFT, optik akış vektörlerini tek adımda üretmek yerine, başlangıçta sıfır akış varsayımıyla başlamakta ve güncelleme bloğunu art arda çalıştırarak akışı iteratif biçimde iyileştirmektedir. Her iterasyonda model, mevcut akış tahminine göre korelasyon hacminden ilgili eşleşme bilgilerini okumaktadır. Ardından akışın ne kadar güncellenmesi gerektiğini ifade eden $(\Delta u, \Delta v)$ düzeltmesini hesaplamaktadır. Bu düzeltme mevcut akış bilgisi ile füzyon edilerek süreç birkaç kez tekrarlandıktan sonra sonuç olarak her piksel için iki bileşenli yoğun optik akış alanı elde Denklem (3.13) olarak elde edilmektedir.

$$\mathbf{F}_t^{\text{raft}}(p) = \begin{bmatrix} \Delta u(p) \\ \Delta v(p) \end{bmatrix}, \qquad p = (u, v) \tag{3.13}$$



Şekil 3.7 Örnek RAFT optik akış çıktısı

Şekil 3.7'de, hareketli kamera durumunda sahnede dinamik bir nesnenin (siyah park hâlindeki araç) yer aldığı bir örnek kare üzerinden RAFT optik akış çıktısının iki farklı görselleştirmesi sunulmaktadır. Üstteki görselde, ardışık iki kare arasındaki piksel yer değiştirmeleri orijinal görüntü üzerine oklar ile bindirilmiştir. Bu okların yönü optik akış vektörünün yönünü temsil etmektedir.

Alttaki renkli optik akış haritasında ise optik akış alanı yaygın olarak kullanılan HSV tabanlı bir gösterim ile görselleştirilmiştir. Bu gösterimde renk tonu (hue) akış yönünü temsil ederken, doygunluk (saturation) akış büyüklüğünü göstermektedir. Küçük büyüklükteki akışlar daha soluk, büyük büyüklükteki akışlar ise daha doygun renklerle ifade edilmektedir.

3.7 Dinamik Obje Kararı ve Maskeleme

Aday dinamik bölgeler için her bir nesnenin gerçekten hareketli olup olmadığı, statik-sahne varsayımı altında hesaplanan beklenen optik akış ile RAFT tarafından üretilen gözlenen optik akış arasındaki tutarsızlığa bakılarak belirlenmektedir.

Her bir aday nesne maskesi için maske içerisinden N_s adet piksel örneklenmektedir. Örneklenen bir piksel $p = (u, v)$ için RAFT akış vektörü $\mathbf{F}_t^{\text{raft}}(p) =$

$[\Delta u(p) \quad \Delta v(p)]^T$ olarak alınmaktadır. Aynı pikselde, önceki derinlik haritasından ve kestirilen kamera hareketi kullanılarak elde edilen beklenen akış $\mathbf{F}_t^{\exp}(p)$ ile elde edilmektedir.

Her örnek nokta için vektörel fark Denklem (3.14)'te tanımlanmıştır.

$$e(p) = \left\| \mathbf{F}_t^{\text{raft}}(p) - \mathbf{F}_t^{\exp}(p) \right\| \tag{3.14}$$

Her bir aday bölge içerisinden hesaplanan N_s nokta için vektörel farklar küçükten büyüğe sıralanarak medyanları alınmaktadır. Aday nesnenin bölge hatası $e_{\text{med}} = \text{median}\big(e(p)\big)$ şeklinde ifade edilerek her objenin optik akış farkı elde edilmektedir.

Optik akış farkı derinlik bilgisine karşı gürbüz olması amacıyla beklenen akış büyüklüğüne bağlı bir eşik ile normalize edilerek bir oran hesaplanmaktadır.

$$\theta_t \;=\; k_0 + k_1 \,\text{median}(\|F_t^{\exp}(p)\|)\,, \qquad r_t \;=\; \frac{\text{median}(e(p))}{\theta_t} \tag{3.15}$$

Denklem (3.15)'te k_0 ve k_1 sabit katsayılardır. k_0, taban tolerans gibi davranarak küçük optik akışların veya gürültünün gereksiz yere hareketli kararı üretmesini engellemeyi amaçlamaktadır. k_1 ise Optik akış büyüklüğüne bağlı esneklik sağlamaktadır. Sahnede statik varsayım altında beklenen akış büyüdükçe, örneğin hızlı kamera hareketi veya yakın nesneler nedeniyle, kabul edilen hata eşiği de k_1 ile orantılı biçimde artırılmaktadır. Böylece karar mekanizması, hem düşük optik akışlı durumlarda aşırı duyarlı olmaktan kaçınmakta hem de yüksek optik akışlı durumlarda hatayı daha gerçekçi bir toleransla değerlendirmektedir. Özetle r_t büyüdükçe, gözlenen akışın statik varsayımla uyumsuzluğunun ve bölgenin hareketli olma ihtimalinin arttığı kabul edilmektedir.

Tek karedeki gürültülü ölçümlerle sık karar değişimini önlemek amacıyla, her nesne için oran r_t üzerinde üstel hareketli ortalama (EMA) ile yumuşatma uygulanmaktadır.

$$\bar{r}_t = \alpha r_t + (1-\alpha)\bar{r}_{t-1} \tag{3.16}$$

Denklem (3.16) $\alpha \in (0, 1]$ yumuşatma katsayısıdır. Ardından iki eşikli histerezis ile hareketli veya statik kararı verilmektedir. $\bar{r}_t > \tau_{\text{high}}$ ise nesne hareketli, $\bar{r}_t < \tau_{\text{low}}$ ise statik olarak etiketlenmektedir. Ara bölgede ($\tau_{\text{low}} \leq \bar{r}_t \leq \tau_{\text{high}}$) önceki karar korunmaktadır. Bu yapı, küçük dalgalanmaların kararı sürekli değiştirmesini

engellemektedir.

Şekil 3.8'de, aday nesneler için gözlenen ve beklenen optik akış arasındaki farka dayalı karar mekanizmasının çıktısı gösterilmektedir. Statik sahne varsayımı ile uyumlu akış üreten bölgeler statik olarak sınıflandırılırken (kırmızı), bu varsayımla uyumsuzluk gösteren bölgeler dinamik (yeşil) olarak belirlenmektedir.



Şekil 3.8 Aday nesneler için dinamik–statik kararının optik akış tutarsızlığına göre belirlenmesi

Tüm bu mimari sonucunda sol görüntü üzerinde $\hat{\Omega}_t^L$ nihai maskeleme alanları elde edilmektedir. Sonrasında, sol görüntü üzerinde YOLOv11-seg ile elde edilen aday dinamik bölgelerden üretilen hareketli maske, stereo disparite bilgisinden yararlanılarak sağ görüntüye taşınmaktadır. Ancak sol–sağ izdüşüm farkları, örtüşme ve maske sınır hatalarını azaltmak amacıyla, `-yolo_on_right` parametresi ile sağ görüntü üzerinde de ek bir YOLOv11-seg çıkarımı gerçekleştirilmektedir.

Bu kapsamda, sağ görüntü I_t^R üzerinde YOLOv11-seg çalıştırılarak hareket edebilir sınıflara ait maske çıktıları elde edilmektedir. Ardından, sol maskeden disparite ile taşınarak oluşturulan sağ maske $\hat{\Omega}_t^R$ ile sağ YOLOv11 çıktıları arasında örtüşme hesaplanmaktadır. Örtüşmenin belirli bir örtüşmeyi sağlaması durumunda ilgili sağ YOLO maskesi kabul edilerek nihai sağ hareketli maske $\hat{\Omega}_t^R$ güncellenmektedir. Böylece sağ görüntü maskesi, yalnızca geometrik kaydırma ile değil, aynı zamanda sağ görüntü üzerinde semantik doğrulama ile daha tutarlı hale getirilmektedir.

3.8 ORB-SLAM3 ile Entegrasyon

ORB-SLAM3, dolaylı yöntem kullanan bir vSLAM yaklaşımı olup, görüntülerden ORB tabanlı öznitelik noktaları çıkararak kareler arası eşleştirme yapmaktadır. Stereo modunda, aynı zaman adımında sol ve sağ görüntü çifti kullanılarak eşleşen özniteliklerden anlık derinlik bilgisi elde edilebilmektedir. Bu sayede sistem, ölçek belirsizliği yaşamadan metrik ölçekte hareket kestirimi yapabilmektedir. ORB-SLAM3'ün görsel işleme tarafı, sahnede öznitelik tabanlı izleme gerçekleştirirken, optimizasyon tarafında yerel haritalama ve döngü kapanışı

ile birlikte grafik optimizasyonu çalıştırarak tutarlı bir harita ve kamera pozları üretmektedir.

Önerilen yöntemde, dinamik bölgelerin maskelenmesi ve statik sahne varsayımına daha uygun görüntü girdisi elde edilmesi hedeflenmektedir. Bu amaçla, önceki bölümlerde üretilen hareketli maske $\hat{\Omega}_t^L$ ve $\hat{\Omega}_t^R$ kullanılarak stereo görüntüler ORB-SLAM3'e verilmeden önce maskeleme uygulanmaktadır. Maskeleme işlemi ile hareketli nesnelere ait pikseller görüntüden çıkarılmak amacıyla bulanıklaştırılmaktadır. Böylece ORB-SLAM3'ün öznitelik çıkarımı ve eşleştirme adımlarında dinamik bölgelerden gelen hatalı ölçümlerin azaltılması amaçlanmaktadır.

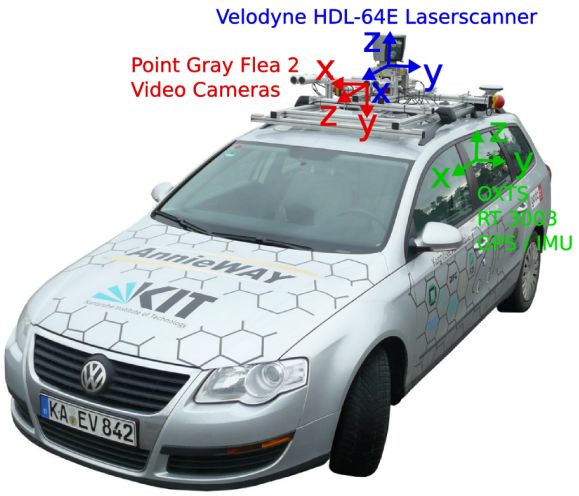
Bu entegrasyonda, ORB-SLAM3'ün iç optimizasyon yapısı değiştirilmemektedir. Sistemin girdisi olan stereo görüntüler, dinamik bölge maskeleme adımından geçirilerek ORB-SLAM3'e iletilmektedir. Böylece yöntem, ORB-SLAM3'ün standart izleme ve haritalama akışı ile uyumlu kalırken, özellikle dinamik sahnelerde öznitelik eşleştirmesini bozan hareketli nesne katkılarının azaltılmasıyla daha kararlı poz kestirimi elde edilmesi hedeflenmektedir.

4
DENEYSEL KURULUM

Bu bölümde, kullanılan veri seti ile birlikte önerilen hibrit yöntemin vSLAM sistemlerinde başarısı ve gürbüzlüğü üzerindeki etkileri değerlendirmek amacıyla yapılan deney çalışmaları açıklanmaktadır. Ardından, elde edilen yörüngelerin yer gerçeğiyle karşılaştırılmasında kullanılan değerlendirme metrikleri sunulmaktadır.

4.1 Veri Seti: KITTI Odometry

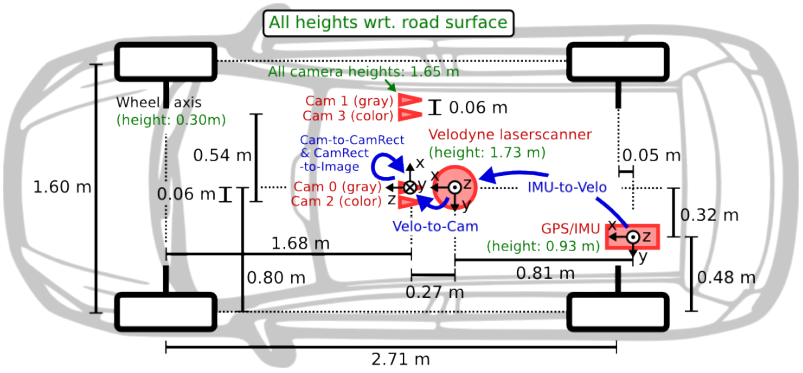
Deneyler, otonom sürüş senaryolarını gerçek dünya koşullarında içeren KITTI-Odometry veri seti üzerinde gerçekleştirilmiştir. Veri seti, şehir içi ve otoyol gibi farklı sürüş ortamlarında toplanmış stereo görüntüler ve bu görüntülere ait yer gerçeği yörünge bilgileri sağlayarak görsel odometri ve SLAM yöntemlerinin nicel olarak karşılaştırılmasına olanak tanımaktadır. Veri seti oluşturulurken kullanılan araç ve sensör dizilimi Şekil 4.1'de gösterilmiştir.



Şekil 4.1 Veri setinin oluşturulma aracı[52]

Veri toplama platformunda senkronize 1.4 megapiksel gri ve renkli olmak üzere iki stereo kamera sistemi, Velodyne HDL-64E 3B LIDAR sensörü, OXTS RT 3003 yüksek hassasiyetli ataletsel navigasyon sistemi birlikte kullanılmıştır. Sensörlerin

araç üzerindeki konumları Şekil 4.2 üzerinde gösterilmiştir. Yer gerçeği pozisyon bilgisi ise GPS/IMU tabanlı OXTS RT 3003 ataletsel navigasyon sisteminden elde edilen ölçümlerin füzyonu ile sağlanmaktadır.



Şekil 4.2 Sensörlerin araç üzerindeki konumları ve eksenleri[52]

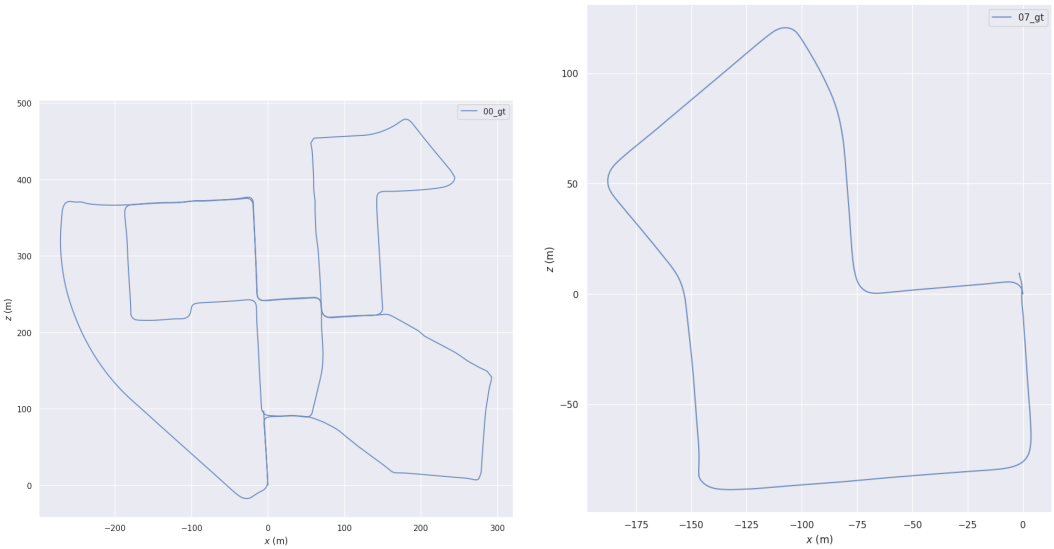
KITTI Odometry veri seti, PNG formatında saklanmış stereo görüntü dizilerinden oluşmakta ve toplam 22 sekans içermektedir. Bu sekansların 00–10 aralığındaki sekansları yer gerçeği yörünge bilgisi ile birlikte sağlanırken, 11–21 aralığındaki sekanslar başarı değerlendirmesi amacıyla yer gerçeği olmaksızın sunulmaktadır.

4.2 Deney Kurgusu

Bu çalışma kapsamında, dinamik nesne yoğunluğu ve sürüş profilleri açısından çeşitlilik içeren 00, 01, 03, 04 ve 07 sekansları seçilerek yöntemlerin hem statik hem de dinamik sahnelerdeki performansı analiz edilmiştir. Seçilen sekanslara ait yer gerçeği yörüngeleri, Şekil 4.3'te x–z düzlemine üstten bakış izdüşümü olarak gösterilmiştir. Bu görünüm ağırlıklı olarak düzlemsel hareketi vurgulasa da, sürüşler sırasında meydana gelen eğim ve yükseklik değişimleri nedeniyle y ekseni boyunca da belirgin değişimler gözlenmektedir. Sekansların yükseklik değişim aralıkları $\Delta y = y_{\max} - y_{\min}$ olarak tanımlandığında; Seq 00 için yaklaşık 26 m, Seq 01 için 68 m, Seq 03 için 45 m, Seq 04 için 8 m ve Seq 07 için 5 m'dir.

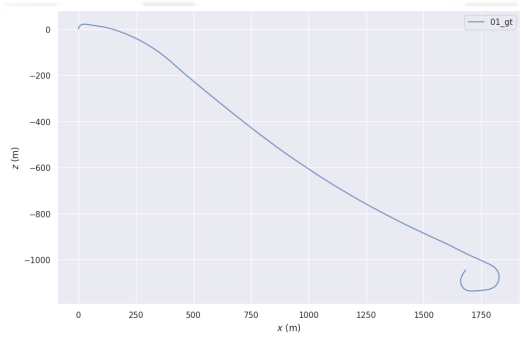
Seçilen sekansların içerikleri aşağıda verilmiştir:

- **Sekans 00 ve 07:** Genellikle statik nesnelerin baskın olduğu, düşük trafikli kentsel yol senaryolarını temsil etmektedir.
- **Sekans 01:** Yüksek hızlı araçların bulunduğu otoyol senaryosudur.
- **Sekans 03:** Statik çevre ile hareketli nesnelerin dağılımının dengeli olduğu bir ortam içermektedir.

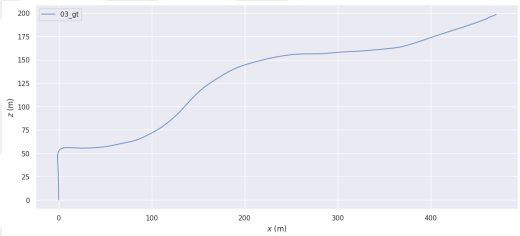


(a) Sekans 00

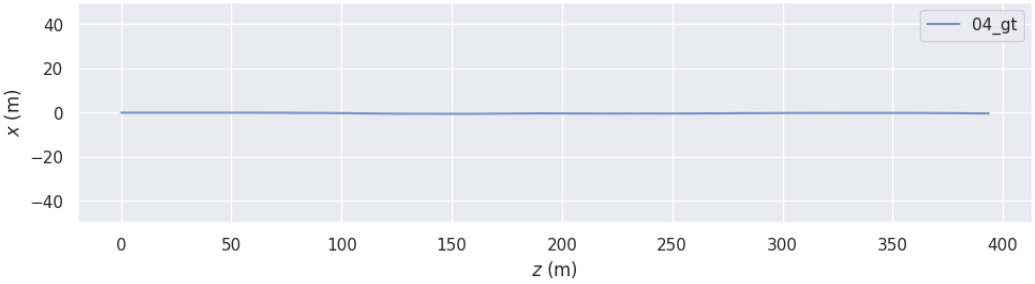
(b) Sekans 07



(c) Sekans 01



(d) Sekans 03



(e) Sekans 04

Şekil 4.3 Seçilen KITTI Odometry sekanslarının yer gerçeği (GT) yörüngeleri (x–z düzlemi).

- **Sekans 04:** Önde seyreden bir aracın bulunduğu ve karşı şeritte ters yönden gelen trafiğin gözlendiği, orta hızdaki sürüş senaryosunu içermektedir.

Tüm deneyler, NVIDIA RTX 5080 GPU ve 96 GB RAM donanımlı bir bilgisayar üzerinde gerçekleştirilmiştir. Yöntemin etkinliğini ölçmek amacıyla üç farklı senaryo karşılaştırmalı olarak analiz edilmiştir:

(i) **Temel Yöntem:** Herhangi bir filtreleme uygulanmadan kullanılan ham stereo görüntüler.

(ii) **Semantik Maskeleme:** Yalnızca YOLOv11 tarafından tespit edilen bölgelerin, segmentasyon tabanlı Gauss filtresi ile maskelendiği görüntüler.

(iii) **Önerilen Yöntem [YoRa-SLAM]:** Dinamik nesnelerin YOLO–RAFT füzyonu kullanılarak filtrelendiği görüntüler.

Şekil 4.4–4.5–4.6'da, veri setinden seçilen bir sahne üzerinden yapılan deneylerdeki giriş görüntüleri verilmiştir.



Şekil 4.4 Sahnenin filtrelenmemiş ham görüntüsü (temel yöntem)



Şekil 4.5 Sahnenin semantik maskeleme ile maskelenmiş görüntüsü

Ham görüntü olan Şekil 4.4'te sahnenin sol tarafında bisikletli kaynaklı hareket bulunmaktadır. Şekil 4.5'te YOLO tabanlı semantik maskeleme dinamik olma potansiyeli taşıyan sınıfları kapsayan tüm bölgeleri maskelemektedir. Bu durum da



Şekil 4.6 Sahnenin önerilen yöntem sonrası maskelenmiş görüntüsü

sisteme fayda sağlayacak park etmiş araçları da maskelediği için, sistemin öznitelik çıkarabileceği noktaları da elemektedir. Şekil 4.6'da YOLO'dan gelen semantik bilgiler ile hareket doğrulama kriterleri sonrasında maskeleme yapılarak SLAM'e girecek ölçümlerin gerçekten dinamik nesnelerden arındırılması hedeflenmektedir.

4.3 Değerlendirme Metrikleri

Bu çalışmada önerilen yöntemin doğruluğu, referans yörünge ile kestirilen yörünge arasındaki farklar üzerinden nicel olarak değerlendirilmiştir. Değerlendirmede hem mutlak hata metrikleri hem de yerel hata metrikleri kullanılmıştır. Mutlak metrikler yörüngenin genel doğruluğunu ölçerken, yerel metrikler zaman içinde biriken kaymayı daha iyi karakterize etmektedir.

4.3.1 Mutlak Yörünge Hatası (ATE)

Mutlak yörünge hatası, her zaman adımında kestirilen konum ile referans konum arasındaki öklid mesafesini ölçmektedir. Böylece yöntemin tüm yörünge boyunca konum doğruluğu değerlendirilmektedir. Bu çalışma kapsamında mutlak yörünge hatası kök ortalama karesel hata (RMSE) değeri ile raporlanmıştır.

$$\mathrm{ATE}_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\mathbf{p}_i - \hat{\mathbf{p}}_i\|_2^2}. \tag{4.1}$$

Denklem (4.1)'de $\mathrm{ATE}_{\mathrm{RMSE}}$ tanımı verilmiştir. Bu ifade, her zaman adımı için referans konum $\mathbf{p}_i$ ile kestirilen konum $\hat{\mathbf{p}}_i$ arasındaki öklid uzaklığı karesinin ortalamasının alınarak karekökünün hesaplanmasıyla elde edilmektedir. Böylece hata değerleri tek bir skalar metrik altında özetlenirken, büyük sapmalar karesel terim nedeniyle daha yüksek ağırlıkla cezalandırılmakta ve sonuç birim olarak yeniden konum birimine geri dönmektedir.

4.3.2 Mutlak Dönme Hatası (ARE)

Mutlak Dönme Hatası, her zaman adımında kestirilen yönelimin referans yönelimden ne kadar saptığını açısal olarak ölçmektedir. Bu amaçla, referans dönme matrisi $\mathbf{R}_i$ ile kestirilen dönme matrisi $\hat{\mathbf{R}}_i$ arasındaki dönme hatası Denklem (4.2) olarak tanımlanmaktadır.

$$\mathbf{R}_i^{\text{err}} = \mathbf{R}_i^{\top} \hat{\mathbf{R}}_i \tag{4.2}$$

$\mathbf{R}_i^{\text{err}}$ matrisi, referans yönelimden kestirilen yönelim arasındaki hatayı ifade eden bir dönme matrisidir. Rotasyon hatası ise bu dönmenin 3 eksendeki değeri teke indirgenerek Denklem (4.3) olarak elde edilmektedir.

$$e_i^{\text{rot}} = \angle(\mathbf{R}_i^{\text{err}})\,. \tag{4.3}$$

Denklem (4.3)'te $\angle(\cdot)$ fonksiyonu, iki yönelim arasındaki farkı temsil eden dönme matrisinin dönme açısını tek bir açı değeri ile ifade eder. Bu açı, dönme matrisinin diagonal elemanları toplanarak Denklem (4.4)'teki gibi açık şekilde ifade edilebilir.

$$e_i^{\text{rot}} = \cos^{-1}\left(\frac{\sum \text{diag}(\mathbf{R}_i^{\text{err}}) - 1}{2}\right) \tag{4.4}$$

Tüm zaman adımları için hesaplanan anlık dönme hataları e_i^{rot} değerleri, yönelim doğruluğunu tek bir skalar metrik ile özetlemek amacıyla RMSE istatistiği kullanılarak raporlanmıştır. Buna göre ARE_{RMSE} Denklem (4.5) olarak tanımlanmaktadır.

$$\text{ARE}_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(e_i^{\text{rot}}\right)^2} \tag{4.5}$$

Denklem (4.5)'te e_i^{rot}, i'inci zaman adımındaki anlık açısal hata değerini, N ise değerlendirilen toplam pozisyon sayısını ifade etmektedir. Kare alma işlemi büyük yönelim sapmalarını daha yüksek ağırlıkla hesaba katarken, karekök işlemi sonucun birimini tekrar açı birimine döndürmektedir.

4.3.3 Bağıl Yörünge Hatası (RTE)

Bağıl Yörünge Hatası, yörüngenin ardışık iki zaman adımı arasındaki bağıl öteleme hareketinin ne kadar doğru kestirildiğini ölçmektedir. Bu metrik, yerel doğruluğu ve zaman içinde biriken sürüklenme hatasını karakterize etmek için kullanılmaktadır.

Referans konumlar $\mathbf{p}_i$ ve kestirilen konumlar $\hat{\mathbf{p}}_i$ olmak üzere, ardışık iki zaman adımı arasındaki bağıl öteleme vektörleri Denklem (4.6) olarak tanımlanmaktadır.

$$\Delta\mathbf{p}_i = \mathbf{p}_{i+1} - \mathbf{p}_i, \qquad \Delta\hat{\mathbf{p}}_i = \hat{\mathbf{p}}_{i+1} - \hat{\mathbf{p}}_i \tag{4.6}$$

Bu iki bağıl öteleme arasındaki hata vektörü Denklem (4.7)'de verilmiştir.

$$\Delta\mathbf{p}_i^{\mathrm{err}} = \Delta\hat{\mathbf{p}}_i - \Delta\mathbf{p}_i \tag{4.7}$$

Bağıl öteleme hatası, bu hata vektörünün büyüklüğüdür ve Denklem (4.8) olarak ifade edilmektedir.

$$e_i^{\mathrm{RTE}} = \left\|\Delta\mathbf{p}_i^{\mathrm{err}}\right\|_2 . \tag{4.8}$$

Anlık bağıl öteleme hataları e_i^{RTE} elde edildikten sonra, bu hataların tek bir skaler değer ile özetlenmesi için RMSE istatistiği kullanılmıştır. Buna göre $\mathrm{RTE}_{\mathrm{RMSE}}$ aşağıdaki şekilde Denklem (4.9) olarak tanımlanmaktadır.

$$\mathrm{RTE}_{\mathrm{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(e_i^{\mathrm{RTE}}\right)^2}. \tag{4.9}$$

Denklem (4.9)'da N, değerlendirilen toplam bağıl adım sayısını ifade etmektedir.

4.3.4 Bağıl Dönme Hatası (RRE)

Bağıl Dönme Hatası, yörüngenin ardışık iki zaman adımı arasındaki bağıl açısal değişim başarısını ölçmektedir. Referans yönelimler $\mathbf{R}_i$ ve kestirilen yönelimler $\hat{\mathbf{R}}_i$ olmak üzere, ardışık iki zaman adımı arasındaki bağıl dönmeler Denklem (4.10)'da tanımlanmıştır.

$$\Delta\mathbf{R}_i = \mathbf{R}_i^{\top}\mathbf{R}_{i+1}, \qquad \Delta\hat{\mathbf{R}}_i = \hat{\mathbf{R}}_i^{\top}\hat{\mathbf{R}}_{i+1}. \tag{4.10}$$

Bu iki bağıl dönme arasındaki dönme hatası Denklem (4.11)'de verilmiştir.

$$\Delta\mathbf{R}_i^{\text{err}} = (\Delta\mathbf{R}_i)^\top \Delta\hat{\mathbf{R}}_i \tag{4.11}$$

Bağıl dönme hatası, Denklem (4.4)'daki gibi en küçük dönme açısı olarak Denklem (4.12) olarak ifade edilmektedir.

$$e_i^{\text{RRE}} = \angle(\Delta\mathbf{R}_i^{\text{err}})\,. \tag{4.12}$$

Anlık bağıl dönme hataları e_i^{RRE} elde edildikten sonra, bu hataların tek bir skalar değer ile özetlenmesi için RMSE istatistiği kullanılmaktadır. Buna göre RRE_{RMSE} Denklem (4.13) tanımlanmaktadır.

$$\text{RRE}_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(e_i^{\text{RRE}})^2}. \tag{4.13}$$

Denklem (4.13)'da N, değerlendirilen toplam bağıl adım sayısını ifade etmektedir.

5
DENEYLER VE BULGULAR

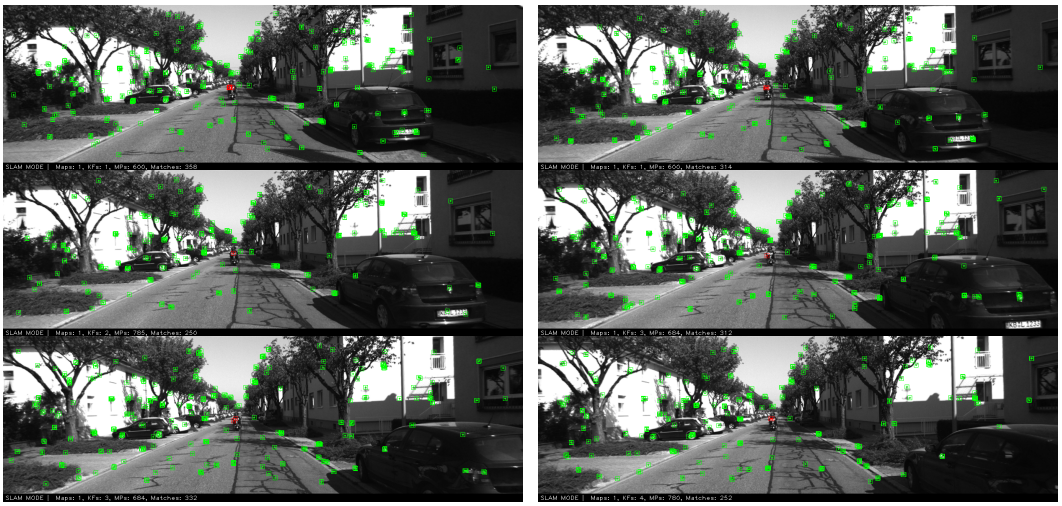
Bu bölümde, önerilen ön işleme katmanının dinamik sahnelerde stereo vSLAM performansına etkisi deneysel olarak değerlendirilmektedir. Bölüm 4'te anlatılan karşılaştırma senaryoları, değerlendirme metrikleri üzerinden nicel sonuçlar raporlanmaktadır. Son olarak, elde edilen sonuçlar sekans bazında odometri kestirimleri üzerinden tartışılmaktadır.

5.1 Şehir İçi Düşük Trafikli Senaryolar

Bu bölümde, önerilen yaklaşımın şehir içi düşük trafikli sahnelerdeki davranışı incelenmiştir. Bu senaryolar, dinamik nesne yoğunluğunun görece sınırlı olması nedeniyle, dinamik bastırma stratejilerinin statik sahneyi ne kadar koruduğunu ve aynı zamanda pozisyon kestirimine ne ölçüde katkı sağladığını analiz etmek için uygundur. Bu kapsamda KITTI Odometry veri setinde düşük trafik yoğunluğu ve baskın statik yapı içeren sekanslar üzerinde deneyler gerçekleştirilmiştir.

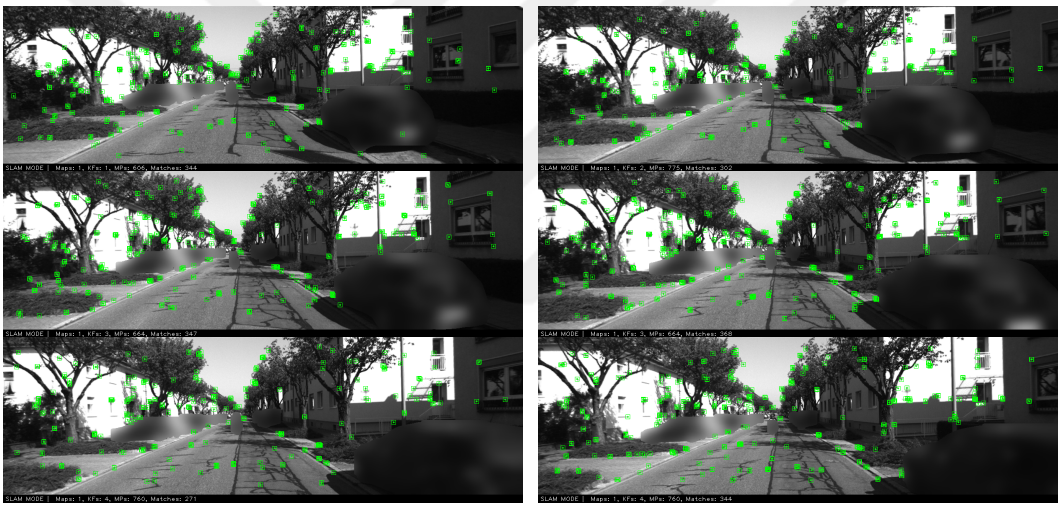
Düşük trafikli şehir içi sahnelerde temel hata kaynağı çoğunlukla ani ve yoğun dinamik nesne kümeleri değil, kısa süreli dinamik geçişler içeren sahnelerdir. Bu sebeple park halindeki araçlar gibi statik olan fakat semantik olarak dinamik sınıfta yer alabilen nesnelerin yanlış maskelemesi pozisyon kestirim başarısını düşürebilmektedir. Bu nedenle yalnızca semantik tespit tabanlı maskeleme, bazı karelerde gereksiz geniş maske üretip izleme için faydalı statik özniteliklerin kaybına yol açabilmektedir.

Sekans 00'ın başlangıcına ait ardışık altı filtrelenmemiş görüntüsü Şekil 5.1'de verilmiştir. Senaryo daha ilk karelerden itibaren aracın önünde hareketli bir motosiklet içermektedir ve bu nesne üzerinde izlenen öznitelikler kırmızı olarak işaretlenmiştir. Bu dinamik özniteliklerin sürekli şekilde izlenmesi pozisyon kestirimi süreçlerine bozucu etki olarak karşımıza çıkmaktadır. Düşük trafikli şehir içi sahnelerde dinamik nesne sayısı sınırlı olsa bile, özellikle ilk başlatma ve öznitelik takibi açısından kritik toplam kestirilen poz sayısında birikimli hatalara



Şekil 5.1 Sekans 00: ardışık 6 kare için filtrelenmemiş görüntüler

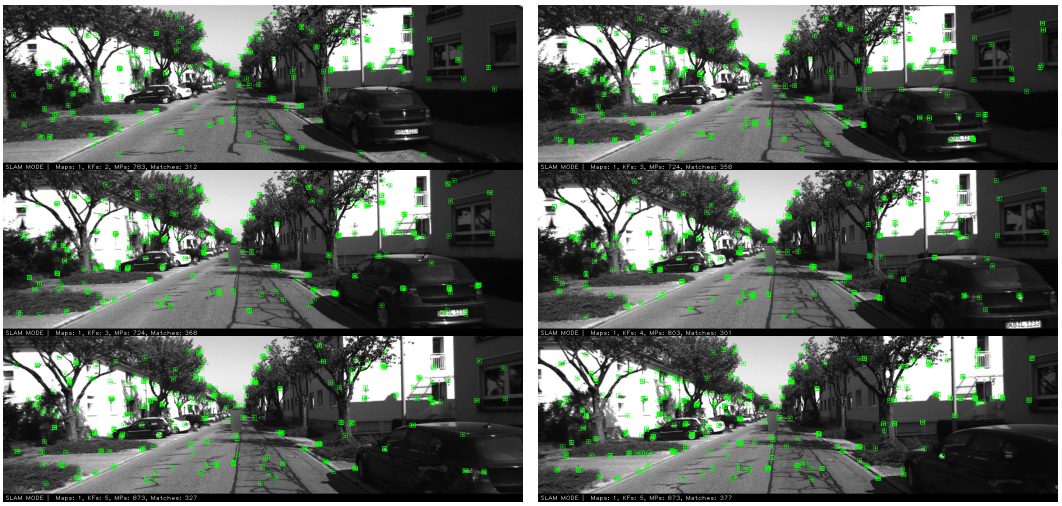
sebep olabilmektedir.



Şekil 5.2 Sekans 00: ardışık 6 kare için semantik maskeleme ile filtrelenmiş görüntüler

Bu duruma karşı en temel yaklaşım, semantik olarak dinamik kabul edilen sınıfları maskeleyerek sahneyi temizlemektir. Şekil 5.2'de yalnızca semantik maskeleme uygulandığında elde edilen ardışık kareler görülmektedir. Motosikletin bulunduğu bölgede dinamik kaynaklı özniteliklerin bastırılması genel olarak yararlı olmakla birlikte, semantik tabanlı maskelemenin agresif maskelemesi sebebiyle sahne dokusunun kısmen kaybedilmesine yol açtığı açıkça görülmektedir. Bunun sonucunda da pozisyon kestirimi için faydalı olabilecek statik referansların azalması gibi bir yan etki ortaya çıkmaktadır.

Önerilen yaklaşımın çıktıları sonucu bulunan öznitelik noktaları Şekil 5.3'te verilmiştir. Bu yöntemde semantik olarak aday seçilen bölgeler doğrudan maskelemek yerine, statik sahne varsayımı altında hesaplanan beklenen optik akış



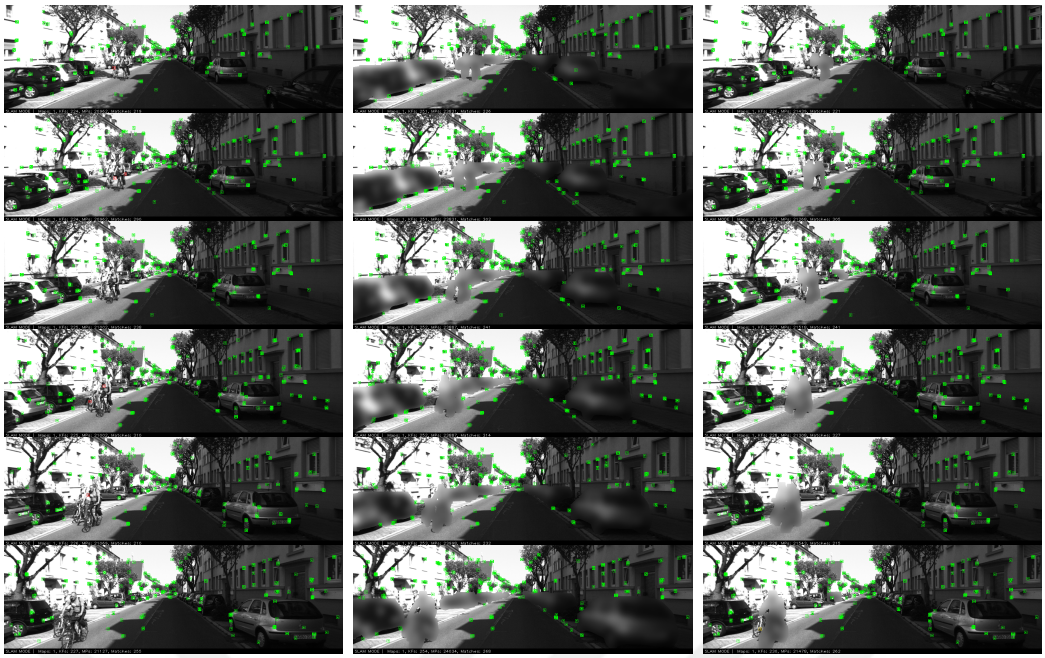
Şekil 5.3 Sekans 00: ardışık 6 kare için önerilen yöntem ile filtrelenmiş görüntüler

ile gözlenen optik akış arasındaki tutarlılığa göre doğrulanmaktadır. Dolayısıyla, gerçekten dinamik hareket sergileyen bölgeler daha seçici biçimde maskelenirken, semantik olarak dinamik sınıfta yer alsa dahi beklenen optik akışa uyumlu davranan bölgelerde gereksiz maskeleme önlenmektedir. Bu denge, düşük trafikli şehir içi senaryolarda yalnızca dinamik nesneleri bastırarak statik olan öznitelikleri mümkün olduğunca korumaktadır.

Şekil 5.4'te yine aynı senaryoda karşı şeritten gelen bisikletlilerin olduğu ardışık altı sahne incelenmiştir. Sütunlar sırasıyla (a) filtrelenmemiş görüntüleri, (b) semantik bilgi ile maskelenmiş görüntüleri ve (c) önerilen yöntem ile maskelenmiş görüntüleri içermektedir. Filtrelenmemiş görüntülerde hareketli olan öznitelik noktaları kırmızı, Maskelenmiş görüntülerde filtrelenmesi gerekirken filtrelenemeyen hareketli öznitelik noktaları turuncu ile işaretlenmiştir. Bu örnekte, önerilen yöntemin dinamik bölgeleri büyük ölçüde bastırsa da sahnedeki nesnelerin örtüşmesi gibi durumlarda bazı dinamik özniteliklerin kaçabildiği görülmektedir.

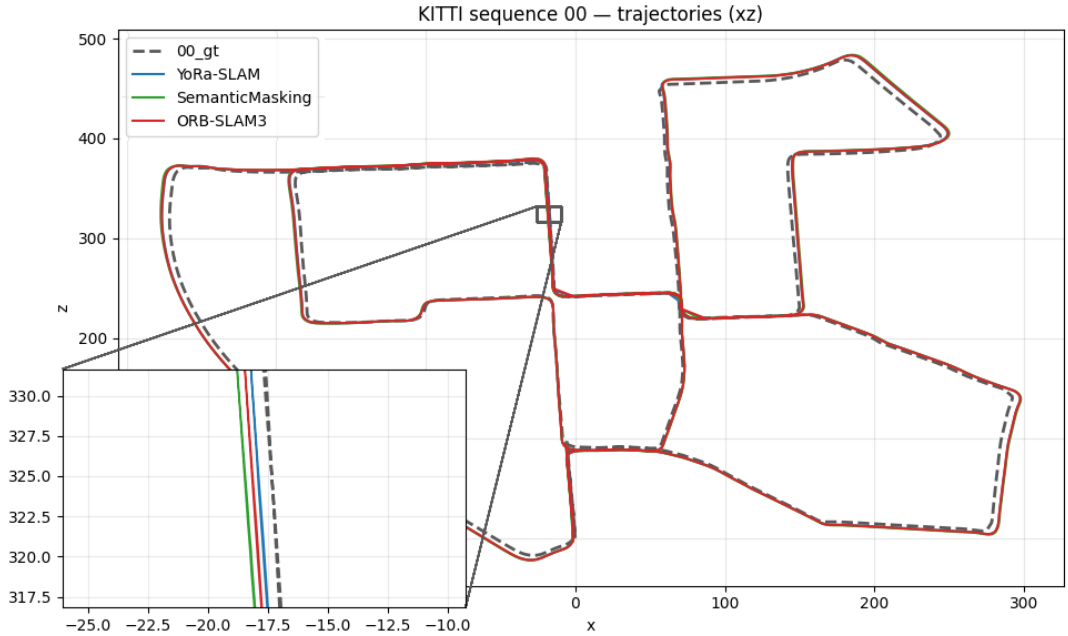
Şekil 5.4'de gösterilen bisikletli sahnesi, sekansın ilgili bölümünde sisteme bozucu etki olarak yansımaktadır. Bu gürültü, Şekil 5.5'te x–z düzlemindeki yörünge karşılaştırmasında ve Şekil 5.6'te eksen bazlı zaman serilerinde açıkça görülmektedir. Şekil 5.6 üzerinde kesikli mavi çizgi ile işaretlenen zaman aralığı etrafındaki yakınlaştırılmış grafikler, aynı eksen grafiğinin içinde küçük dikdörtgenler olarak verilmiştir. Önerilen yöntemin Ham ORB-SLAM3 çıktısı ile semantik maskeleme yaklaşımına göre yer gerçeğine daha yakın bir pozisyon kestirimi sergilediği gözlenmektedir.

Bisikletlilerin kameraya doğru geldiği sahnede, dinamik nesnelerden gelen hatalı ölçümlerin kısa süreli etkisini azaltırken statik arka plan özniteliklerini koruyarak

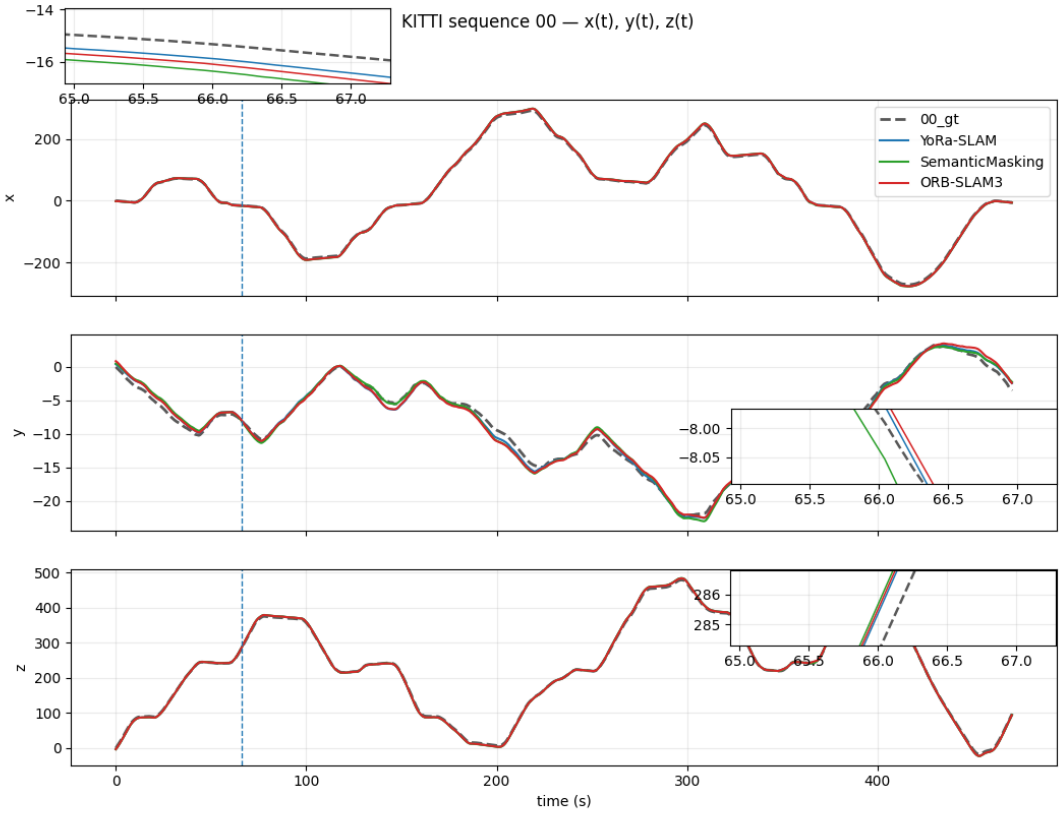


(a) Ham Görüntü (b) Semantik Maskeleme (c) Önerilen yöntem

Şekil 5.4 Sekans 00: seçilen sahnelerde ham görüntü, semantik maskeleme ve önerilen yöntem çıktısının görsel karşılaştırması



Şekil 5.5 Sekans 00 için x–z düzleminde yörünge karşılaştırması

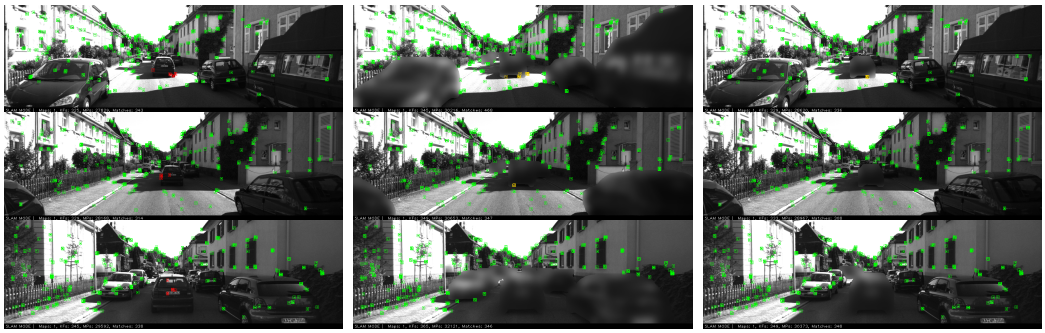


Şekil 5.6 Sekans 00 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi

izleme kararlılığını önerilen yöntemin temel katkısı açıkca görülmektedir. Bu davranış, hem x–z düzlemindeki yörünge yakınlığı (Şekil 5.5) hem de $x(t)$, $y(t)$ ve $z(t)$ eğrilerindeki yerel ayrışmanın azalması (Şekil 5.6) ile desteklenmektedir.

Şekil 5.7'te Sekans 07'den seçilen üç farklı zaman kesitinde (a) ham görüntüler, (b) semantik bilgi ile maskeleme görüntüleri ve (c) önerilen yöntemin görüntüleri karşılaştırılmıştır. Hareketli bölgenin üstüne düşen öznitelik noktaları kırmızı ile, hareket etmesine rağmen filtrelenemeyen öznitelik noktaları ise turuncu renk ile işaretlenmiştir.

Ham görüntülerde görüldüğü üzere sahne, park halindeki araçlar ve bina cepheleri gibi yapılara ek olarak önde hareket halinde olan araçlar içermektedir. Yalnız semantik maskeleme ile tüm hareket edebilir bölgeler filtrelendiği için statik öznitelik noktalarının bastırılmasına ve pozisyon kestirimi için faydalı özniteliklerin azalmasına neden olduğu Şekil 5.7.b'den açıkça gözükmektedir. Bununla birlikte sahnedeki dokunun bozulmasından dolayı tespit edilen öznitelik noktaları maskelenmiş alanların kenarında yoğunlaşarak sisteme zarar da vermektedir. Önerilen yaklaşımda ise sahnedeki dinamizm daha seçiçi bir şekilde karar verilmesi sebebiyle doğru maskeleme yaparak sahnedeki faydalı yapı korunmakla birlikte

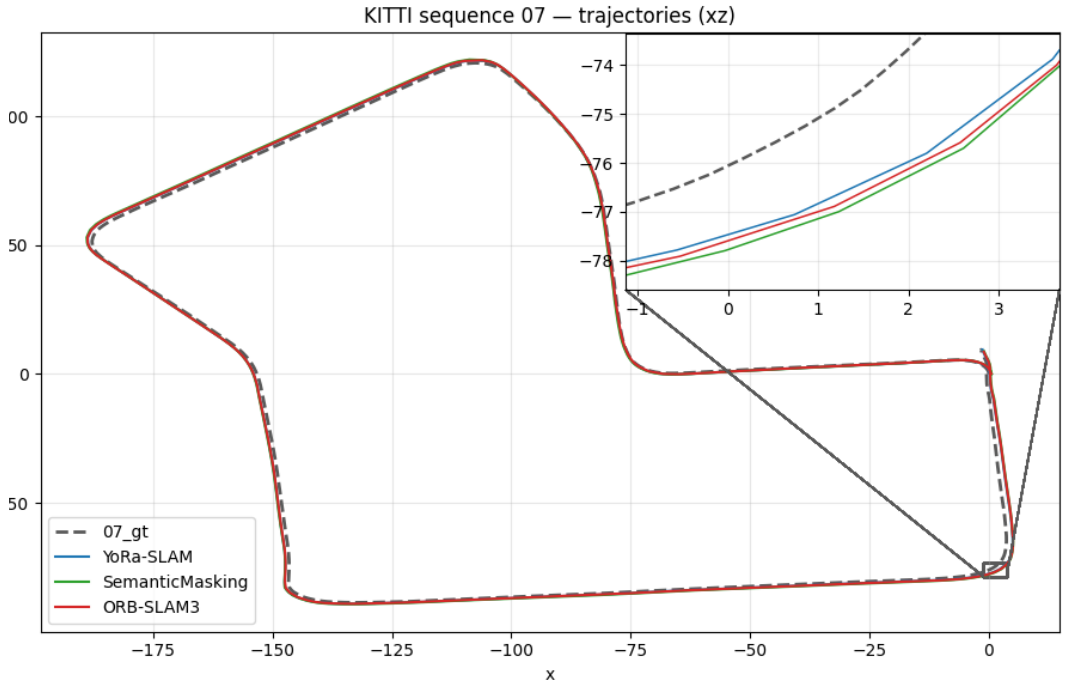


(a) Ham görüntü (b) Semantik maskeleme (c) Önerilen yöntem

Şekil 5.7 Sekans 07: Üç farklı zaman anında ham görüntü, semantik maskeleme ve önerilen yöntem çıktısının karşılaştırması

bozucu etki yaratan bölge de filtrelenmiştir.

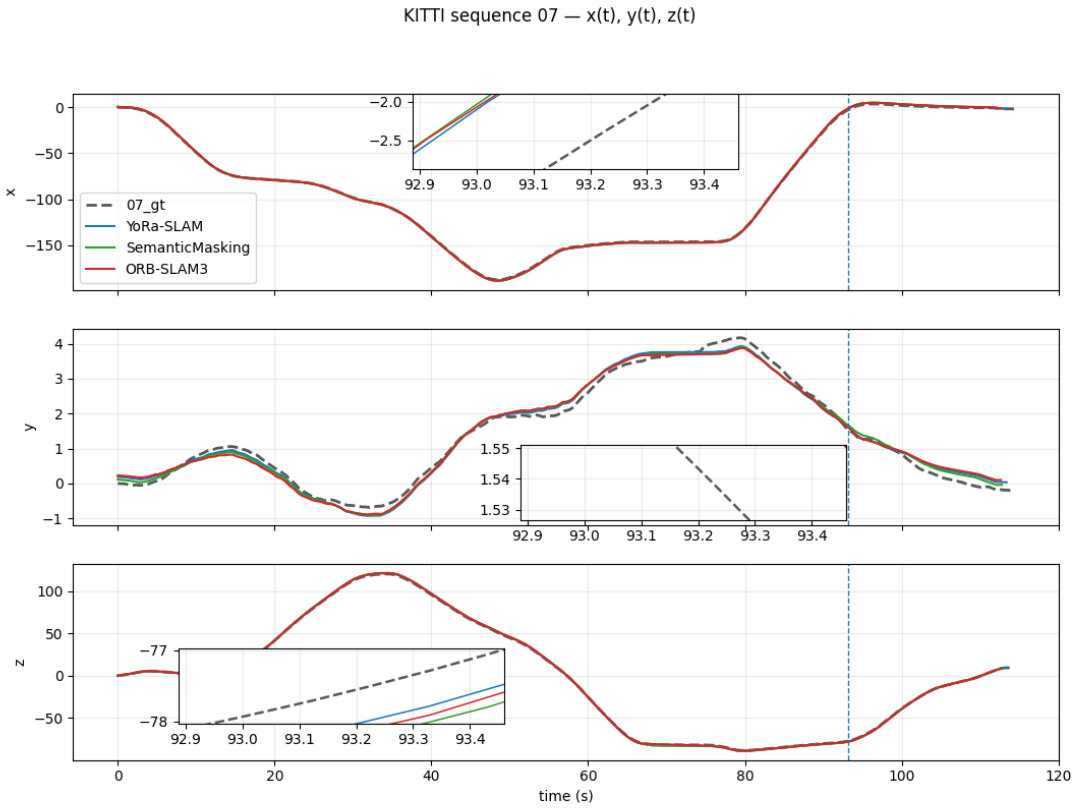
Şekil 5.8 ve Şekil 5.9, sekans 07 üzerinde elde edilen yörünge sonuçlarını sırasıyla x–z düzleminde ve zaman ekseninde ayrıntılı olarak göstermektedir. Görsellerdeki küçük dikdörtgenler ilgili bölgenin yaklaştırılmış halidir. Şekil 5.9 üzerinde kesikli mavi çizgi ile işaretlenen zaman aralığı etrafındaki yakınlaştırılmış grafikler, aynı eksen grafiğinin içinde küçük dikdörtgenler olarak verilmiştir. Şekillerdeki yakınlaştırılmış bölgeler Şekil 5.7'de incelenen sahnelerdeki bölgeyi göstermektedir.



Şekil 5.8 Sekans 07 için x–z düzleminde yörünge karşılaştırması

Genel yörünge profiline bakıldığında yöntemlerin hepsinin yer gerçeğine benzer

bir pozisyon kestirimi sağladığı görülmekle birlikte, yakınlaştırılmış bölgelerden de görüldüğü üzere hareketli nesnelerin olduğu bölgelerden kaynaklı küçük farklar meydana gelmektedir. Özellikle Şekil 5.8'deki yakınlaştırılmış bölgede, yöntemler arasındaki ayrışmanın en belirgin şekilde hareketli nesnelerin görüş alanında baskın olduğu kesitlerde ortaya çıkmasıyla birlikte sadece semantik maskeleme ile yapılan filtrelemenin sisteme verdiği zarar da açıkça görülmektedir.



Şekil 5.9 Sekans 07 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi

Bu karşılaştırmalar, düşük trafikli şehir içi senaryolarda yöntemin temel kazanımının yalnızca dinamik nesneleri elimine etmekle kalmayıp, aynı zamanda gereksiz maskelemenin önüne geçerek izleme kararlılığını da arttırdığını göstermektedir.

Tablo 5.1 ve Tablo 5.2, düşük trafikli şehir içi senaryoları temsil eden Sekans 00 ve 07 için sırasıyla mutlak ve bağıl yörünge hatalarının istatistiksel karşılaştırmasını sunmaktadır. En iyi başarıyı gösteren değerler kalın fontlu yazılmıştır.

Tablo 5.1 incelendiğinde Sekans 00 üzerinde ATE açısından önerilen yöntemin KOH ve ortalama değerlerinde en iyi sonucu verdiği, semantik maskelemenin ise bu iki ölçütte ham veriye kıyasla belirgin bri kötüleştirme sağladığı gözlenmektedir. Bu durum, Sekans 00'da dinamik nesnelerin sahneye kısa süreli ve sınırlı

Tablo 5.1 Sekans 00 ve 07 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması.

Sekans	İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
00	KOH (RMSE)	3.907	4.311	**3.797**
	Ortalama	3.638	4.008	**3.547**
	Std. Sapma	1.423	1.588	**1.355**
	En Küçük	0.137	**0.035**	0.136
	En Büyük	6.693	7.535	**6.569**
07	KOH (RMSE)	0.963	1.086	**0.920**
	Ortalama	0.899	1.015	**0.854**
	Std. Sapma	0.345	0.386	**0.343**
	En Küçük	0.262	0.345	**0.249**
	En Büyük	1.592	1.695	**1.502**

bölgelerde girmesi nedeniyle, yalnız semantik bilgi ile yapılan bastırmanın bazı karelerde faydalı statik öznitelikleri de azaltarak izleme kararlılığını olumsuz etkileyebilmesiyle uyumludur. Buna karşılık Sekans 07'de hem ATE KOH hem de ortalama değerlerinde önerilen yöntem en iyi sonucu üretmektedir. Bu da şehir içi düşük trafikli sahnelerde seçici dinamik bastırmanın yörünge doğruluğunu desteklediğini göstermektedir.

Tablo 5.2 Sekans 00 ve 07 için bağıl yörünge hatalarının (RTE) karşılaştırılması

Sekans	İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
00	KOH (RMSE)	0.088	**0.070**	**0.070**
	Ortalama	0.060	0.049	**0.048**
	Std. Sapma	0.064	**0.049**	0.052
	En Küçük	0.007	0.004	**0.002**
	En Büyük	1.027	**0.734**	0.803
07	KOH (RMSE)	0.051	**0.046**	**0.046**
	Ortalama	0.040	0.034	**0.033**
	Std. Sapma	**0.031**	**0.031**	**0.031**
	En Küçük	0.003	**0.002**	**0.002**
	En Büyük	**0.415**	0.520	0.513

Bağıl hata sonuçları Tablo 5.2 ile değerlendirildiğinde, semantik maskelemenin önerilen yönteme yakın sonuçlar verdiği gözlenmektedir. Bu durum semantik maskelemenin dinamik nesnelerden oluşan bozucu etkiyi giderse de park etmiş şekilde duran ve öznitelik takibi için değerli bölgeleri de yok ederek uzun vadede yörünge tahmininde sürüklenmeye neden olarak mutlak yörünge hatasını arttırmıştır.

Tablo 5.3 ve Tablo 5.4, düşük trafikli şehir içi senaryoları temsil eden Sekans 00 ve 07 için sırasıyla mutlak ve bağıl dönme hatalarının istatistiksel karşılaştırmasını sunmaktadır. En iyi başarıyı gösteren değerler kalın fontlu yazılmıştır.

Tablo 5.3 Sekans 00 ve 07 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması.

Sekans	İstatistik [°]	Ham veri	Semantik	Önerilen yöntem
00	KOH (RMSE)	0.7163	0.7941	**0.6974**
	Ortalama	0.6021	0.6732	**0.5776**
	Std. Sapma	**0.3880**	0.4211	0.3909
	En Küçük	0.0585	**0.0196**	0.0551
	En Büyük	6.7229	6.6874	**6.4434**
07	KOH (RMSE)	0.4305	0.4159	**0.3953**
	Ortalama	0.3981	0.3892	**0.3688**
	Std. Sapma	0.1638	0.1467	**0.1421**
	En Küçük	0.1236	**0.1002**	0.1029
	En Büyük	0.8170	0.7183	**0.6915**

Tablo 5.3 incelendiğinde, dönme hatası iki sekans için önerilen yöntem ile başarının arttığı gözlenmektedir. Sekans 00 ve 07'de KOH ve Ortalama değerlerde önerilen yöntem en düşük hatayı üretirken, semantik maskeleme aynı sekans üzerinde daha yüksek hatalara ulaşabilmektedir.

Tablo 5.4 Sekans 00 ve 07 için bağıl dönme hatalarının (RRE) karşılaştırması

Sekans	İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
00	KOH (RMSE)	0.2865	**0.2387**	0.2492
	Ortalama	0.0960	**0.0841**	0.0847
	Std. Sapma	0.2699	**0.2234**	0.2343
	En Küçük	0.0034	0.0033	**0.0031**
	En Büyük	**4.5820**	5.1415	4.9020
07	KOH (RMSE)	0.0706	**0.0607**	0.0637
	Ortalama	0.0554	**0.0486**	0.0497
	Std. Sapma	0.0438	**0.0363**	0.0398
	En Küçük	0.0047	0.0040	**0.0018**
	En Büyük	0.2691	**0.2525**	0.3110

Tablo 5.4 bağıl dönme hatası açısından değerlendirildiğinde, her iki sekans için de semantik maskeleme ile önerilen yaklaşımın ham veriye göre daha düşük hata seviyelerine indiği görülmektedir. Bununla birlikte, tabloda bazı metriklerde yöntemlerin birbirine çok yakın değerler üretmesi, dönme hareketinin kısa pencerelerde göreli olarak kararlı olmasına bağlanabilir. Burada kritik nokta, önerilen yöntemin yalnızca anlık bağıl hatayı düşürmekle kalmayıp, mutlak dönme

hatası tarafında da daha tutarlı bir iyileşme sunmasıdır.

5.2 Yüksek Hızlı Otoyol Senaryosu

Bu bölümde, önerilen yaklaşımın yüksek hızlı otoyol koşullarındaki performansı incelenmiştir. otoyol senaryoları, kamera hızının artması ve sahnedeki baskın hareket doğrultusunun belirginleşmesi nedeniyle, optik akış tabanlı karar mekanizmaları için daha zorlayıcı bir test ortamı sunmaktadır. Bununla birlikte şerit çizgileri ve bariyerler sahneye statik öznitelikler kazandırırken, özellikle yakın mesafeden geçen araçlar kısa süreli fakat yüksek bağıl hız kaynaklı dinamik öznitelik noktaları üretmektedir.

Şekil 5.10'da Sekans 01'den seçilen örnek karelerde (a) ham görüntü, (b) semantik maskeleme ve (c) önerilen yöntem çıktıları karşılaştırılmıştır. Bu karşılaştırma, semantik maskeleme ile önerilen hibrit yöntemin benzer maskeleme alanları yaptığını ortaya koymaktadır.



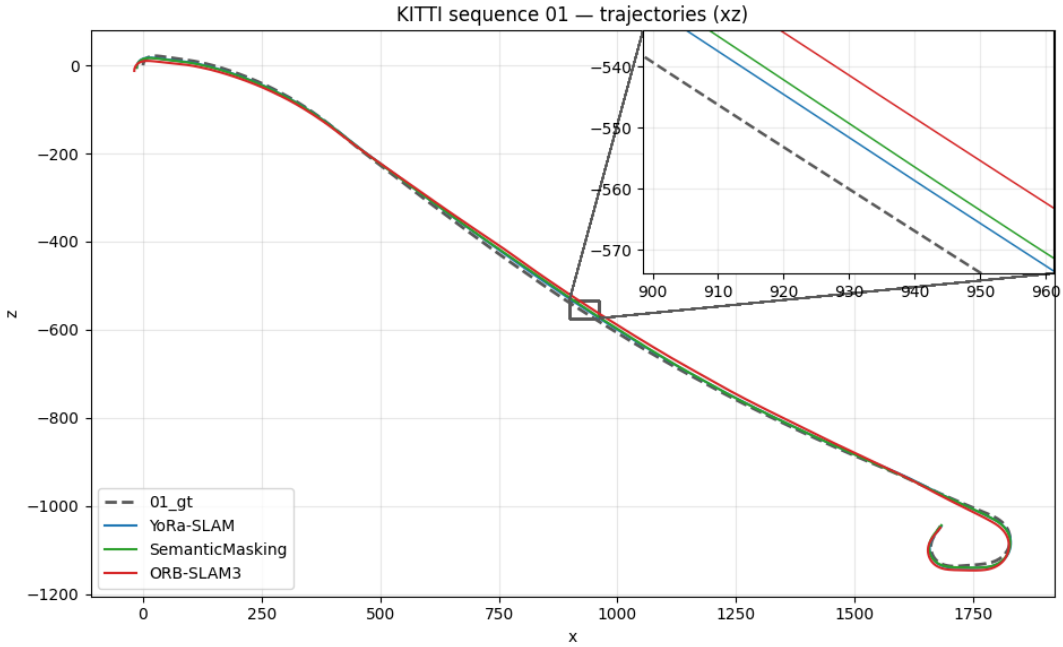
(a) Ham görüntü (b) Semantik maskeleme (c) Önerilen yöntem

Şekil 5.10 Sekans 01: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması

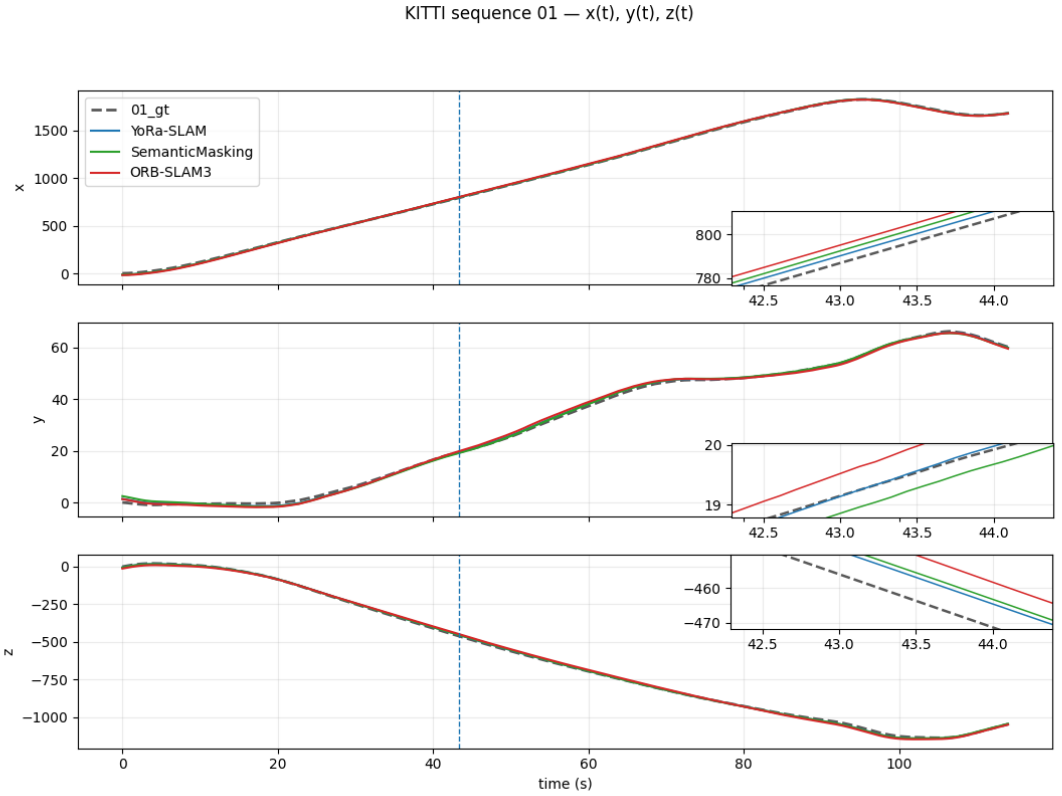
Şekil 5.11 ve Şekil 5.12, yüksek hızlı otoyol senaryosunu temsil eden Sekans 01 üzerinde elde edilen yörünge sonuçlarını sırasıyla x–z düzleminde ve zaman ekseninde göstermektedir.

Şekil 5.11'de gösterilen sağ üst köşedeki dikdörtgen alan ilgili bölgenin yakınlaştırılmasını göstermektedir. Şekil 5.12 mavi kesikli çizgi ile yakınlaştırılan zaman dilimi ise sahneye sağ taraftan hareketli iki aracın girdiği zamanı göstermektedir ve bu iki araç Şekil 5.10'un ilk satırında gösterilmektedir. Şekil 5.11'deki algoritma çıktıları incelendiğinde, ortamda sadece hareketli objeler olması sebebiyle önerilen yöntem ile semantik filtreleme başarıları benzer ve ham görüntüye göre, yer gerçeğine daha yakın olduğu bir sonuç elde edilmiştir.

Benzer biçimde, Şekil 5.12 üzerinde yakınlaştırılan zaman kesiti çevresinde $x(t)$, $y(t)$ ve $z(t)$ sahneye sol taraftan bir aracın girdiği ve bu sebeple algoritma çıktılarının yer doğruluğundan ayrıştığı gözükmektedir. Bu durum, dinamik



Şekil 5.11 Sekans 01 için x–z düzleminde yörünge karşılaştırması



Şekil 5.12 Sekans 01 için $x(t)$, $y(t)$ ve $z(t)$ eksenlerindeki konum değişimi

araçların kısa süreli görünür olması bile yüksek hız ve uzun doğrusal ilerleme hareketi sebebiyle birikimli hatalara sebep olduğunu göstermektedir.

Tablo 5.5 ve Tablo 5.6, yüksek hızlı otoyol senaryosunu temsil eden Sekans 01 üzerinde sırasıyla mutlak ve bağıl yörünge hata istatistiklerini sunmaktadır. En iyi başarıyı gösteren değerler kalın font ile yazılmıştır.

Tablo 5.5 Sekans 01 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması

İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	13.67	8.30	**7.70**
Ortalama	12.96	7.76	**7.18**
Std. Sapma	4.33	2.94	**2.78**
En Küçük	3.66	2.46	**0.39**
En Büyük	22.51	14.76	**14.25**

Tablo 5.5 incelendiğinde, önerilen yöntem ile semantik maskelemenin ham veriye kıyasla belirgin bir iyileşme sağladığı görülmektedir. Özellikle KOH ve ortalama hatalarda düşüş gözlenmesi, yüksek göreli hızın neden olduğu dinamik kaynaklı sapmaların daha etkin biçimde bastırıldığını ve bunun uzun vadeli birikimli sürüklenmeyi azalttığını göstermektedir.

Tablo 5.6 Sekans 01 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması

İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.049	0.048	**0.046**
Ortalama	0.043	0.043	**0.041**
Std. Sapma	0.022	0.022	**0.021**
En Küçük	**0.005**	0.008	0.007
En Büyük	0.229	0.225	**0.222**

Bağıl hata açısından Tablo 5.6 incelendiğinde, yöntemler arasındaki farkın ATE'ye göre daha az olduğunu görülmektedir. Otoyol senaryolarında araba yüksek hızlı ve doğrusal bir yol izlediği için küçük hatalar birikerek uzun vadede sisteme zarar vermektedir. Bu yüzden gürültülü bölgelerin sisteme verilmemesi oldukça önemlidir.

Tablo 5.7 ve Tablo 5.8, yüksek hızlı otoban senaryosunu temsil eden Sekans 01 için sırasıyla mutlak ve bağıl dönme hatalarının istatistiksel karşılaştırmasını sunmaktadır.

Tablo 5.7 incelendiğinde önerilen yöntem ile semantik maskeleme yönteminin, hem

Tablo 5.7 Sekans 01 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	3.758	3.475	**2.771**
Ortalama	3.359	2.975	**2.352**
Std. Sapma	1.686	1.796	**1.464**
En Küçük	1.750	1.608	**1.244**
En Büyük	6.719	6.663	**5.364**

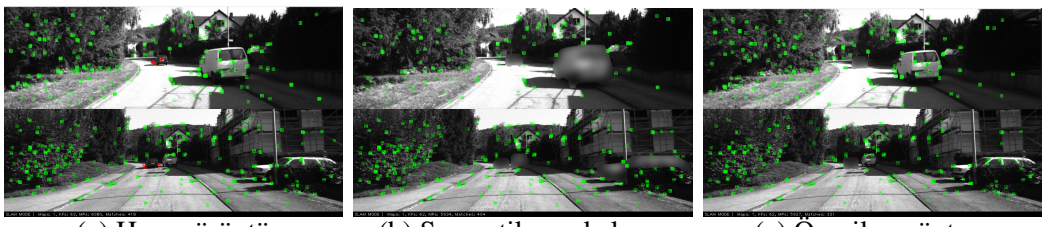
Tablo 5.8 Sekans 01 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.045	0.045	**0.040**
Ortalama	0.038	0.036	**0.034**
Std. Sapma	0.023	0.026	**0.021**
En Küçük	0.004	0.003	**0.003**
En Büyük	0.194	0.227	**0.177**

mutlak hem de bağıl dönme hatalarında ham veriye kıyasla üstün bir performans sergilediği görülmektedir. Yüksek hızlı otoyol senaryolarında, küçük açısal hatalar katedilen mesafe ile çarpılarak büyük yanal konum sapmalarına sebep olmaktadır. Dolayısıyla, önerilen yöntemle elde edilen açısal başarı, bir önceki bölümde sunulan mutlak yörünge başarısının temel nedeni olarak gösterilebilir.

5.3 Statik ve Dinamik Nesnelerin Dengeli Dağılım Gösterdiği Senaryo

Şekil 5.13'te sırasıyla; ham görüntü, semantik maskeleme ve önerilen hibrit yöntem ile elde edilen öznitelik noktaları verilmiştir.



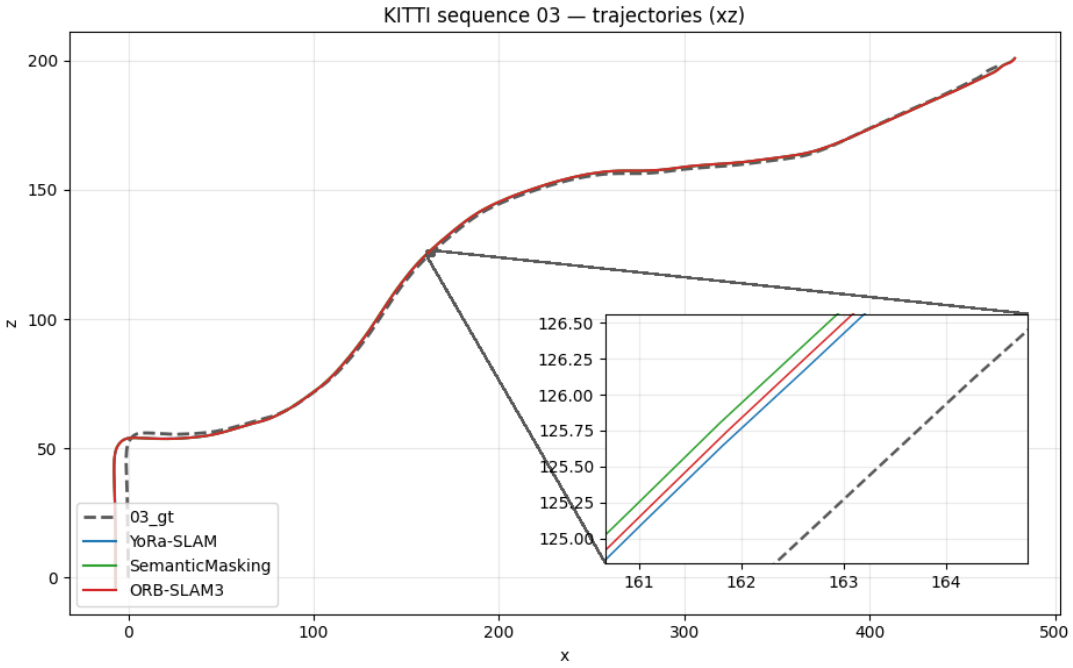
(a) Ham görüntü (b) Semantik maskeleme (c) Önerilen yöntem

Şekil 5.13 Sekans 03: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması

Ham görüntüde hem statik hem dinamik objeler üzerinde öznitelik bulunduğu görülmektedir. Bu özniteliklerden hareketli olanlar kırmızı ile işaretlenmiştir.

Doğrudan semantik maskeleme uygulanan sahnelerde dinamik nesneler ile birlikte park etmiş şekilde duran statik alanlar da maskelenerek faydalı olabilecek öznitelik noktaları yok edildiği gözükmektedir. Bu durum fayda sağlayabilecek bölgeleri azaltmakta ve başarıyı düşürmektedir. Önerilen hibrit yaklaşımda ise, sahnedeki hareketli nesneler daha başarılı şekilde sınıflandırılarak gerekli maskeleme uygulanarak gereksiz maskelemeden kaçınılmıştır.

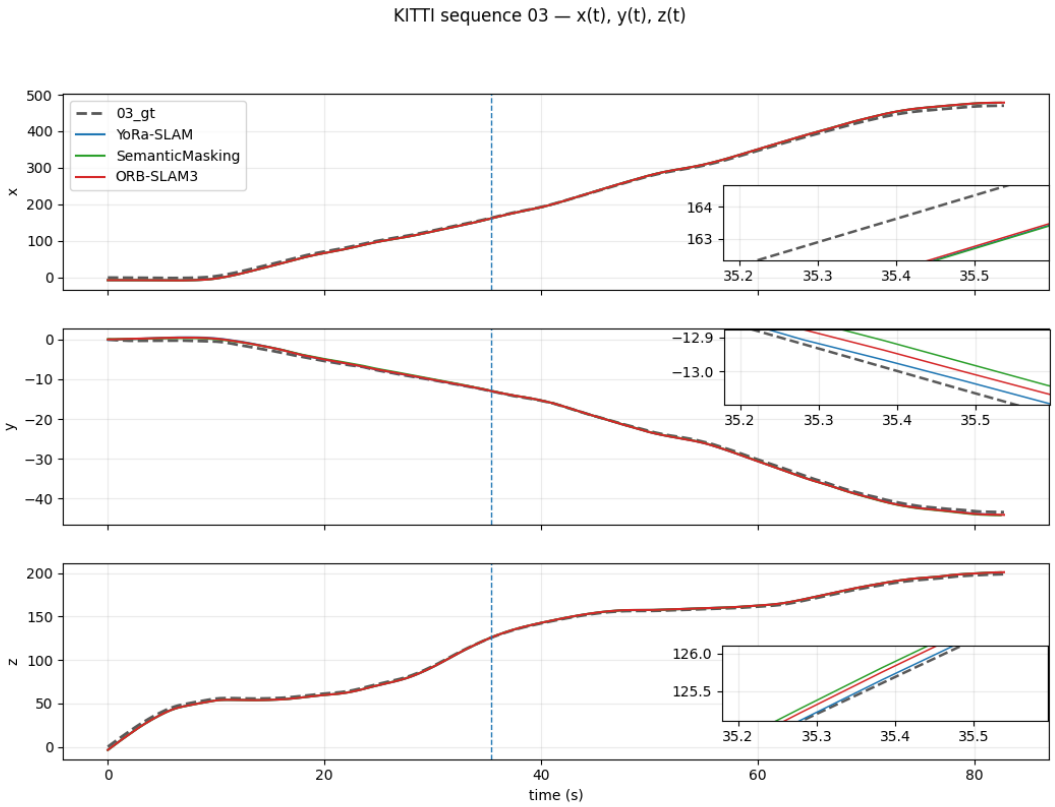
Şekil 5.14 ve Şekil 5.15, dinamizm açısından dengeli olan Sekans 03 üzerinde elde edilen yörünge sonuçlarını sırasıyla x–z düzleminde ve zaman ekseninde göstermektedir.



Şekil 5.14 Sekans 03 için x–z düzleminde yörünge karşılaştırması

Şekil 5.14'de gösterilen sağ taraftaki dikdörtgen alan ilgili bölgenin yakınlaştırılmasını göstermektedir. Şekil 5.15 mavi kesikli çizgi ile yakınlaştırılan zaman dilimi ise önde giden aracın olduğu durumdaki algoritma pozisyon kestirimlerini göstermektedir ve ilgili zaman aralığının yakınlaştırılmış görüntüsü aynı eksen içinde küçük dikdörtgenlerde verilmiştir.

Sahnenin dinamizm açısından daha dengeli olduğu durumlarda dinamik–statik kararının daha doğru veren yöntemlerin yer doğruluğuna daha yakın bir kestirim yapıldığı açıkça Şekil 5.15'te gözükmektedir. Dinamik kaynaklı bozucu etkiler azaltılırken, statik özniteliklerin korunma biçimindeki seçicilik yerel sapmaları azaltarak sistemin gürbüzlüğüne katkı sağlamaktadır.



Şekil 5.15 Sekans 03 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi

Tablo 5.9 ve Tablo 5.10, dinamizm açısından dengeli bir şehir içi sürüşü temsil eden Sekans 03 üzerinde sırasıyla mutlak ve bağıl yörünge hata istatistiklerini sunmaktadır.

Tablo 5.9 Sekans 03 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması

İstatistik [m]	**Ham veri**	**Semantik maskeleme**	**Önerilen yöntem**
KOH (RMSE)	4.92	4.87	**4.74**
Ortalama	4.27	4.22	**4.10**
Std. Sapma	2.44	2.43	**2.38**
En Küçük	0.77	0.78	**0.66**
En Büyük	8.68	8.53	**8.43**

Tablo 5.9 incelendiğinde, Sekans 03'te önerilen yöntemin hemen tüm ATE istatistiklerinde en düşük değerlere ulaştığı görülmektedir. Bu durum, dinamizmin tamamen baskın olmadığı ancak sahneye aralıklı olarak dinamik nesnelerin girdiği senaryolarda, dinamik–statik kararının seçici biçimde verilmesinin hatayı azalttığını net olarak göstermektedir. Özellikle KOH ve ortalama hatadaki düşüş, önerilen yöntemin uzun süreli izleme boyunca ürettiği toplam konum hatasının daha kararlı bir biçimde sınırlandığını ve sistemin yer gerçeğine daha yakın sonuçlar ürettiğini

işaret etmektedir.

Tablo 5.10 Sekans 03 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması

İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.068	0.068	**0.066**
Ortalama	0.063	0.064	**0.062**
Std. Sapma	**0.024**	**0.024**	**0.024**
En Küçük	**0.013**	0.019	0.017
En Büyük	0.188	**0.186**	0.206

Bağıl hata sonuçları Tablo 5.10 ile değerlendirildiğinde ise, yöntemler arasındaki farkın ATE tablosuna kıyasla daha benzer sonuçlar ürettiği gözlenmektedir. Önerilen yöntemin ortalama hatayı ve KOH değerini en düşük seviyeye çekmiş olması sebebiyle sağlanan küçük iyileştirmeler, birikimli hatanın oluşmasını engelleyerek sistemin uzun vadeli kararlılığını arttırmıştır.

Tablo 5.11 Sekans 03 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması.

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.521	**0.463**	0.567
Ortalama	0.498	**0.432**	0.538
Std. Sapma	**0.154**	0.168	0.179
En Küçük	0.249	**0.041**	0.250
En Büyük	**0.796**	0.833	0.925

Tablo 5.12 Sekans 03 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması.

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.052	0.059	**0.051**
Ortalama	0.044	0.046	**0.043**
Std. Sapma	0.028	0.037	**0.027**
En Küçük	**0.001**	0.004	0.002
En Büyük	0.217	0.389	**0.177**

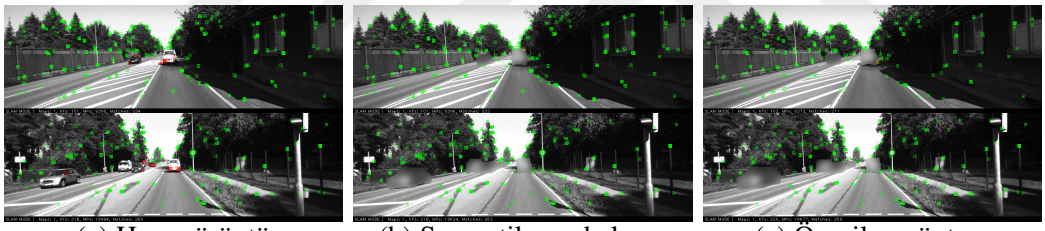
Sekans 03 için dönme hataları incelendiğinde, önerilen yöntemin bağıl dönme hatasında daha iyi sonuçlar verdiği, ancak mutlak dönme hatasında görece daha yüksek değerler ürettiği görülmektedir. Bu durum çelişkili gibi görünse de sekans 03'te aracın hareketinin büyük ölçüde doğrusal olması ve sahne boyunca dönüşlerin sınırlı kalması, mutlak yönelim hatasının etkin biçimde düzeltilmesini zorlaştırmaktadır. Dönüş hareketi sınırlı olduğunda, küçük bir yönelim ofseti yeniden gözlenebilir hale gelmemekte ve bu sapma zaman içinde birikerek mutlak

açı hatasında artışa neden olabilmektedir. Buna karşılık daha sık ve belirgin dönüşlerin bulunduğu senaryolarda, yönelim değişimleri daha güçlü gözlemsel kanıtlar ürettiğinden, optimizasyon sürecinin bu tür ofsetleri daha kolay düzeltmesi beklenmektedir.

5.4 Şehir İçi Çift Yönlü Trafik Senaryosu

Bu bölümde, şehir içi çift yönlü trafik akışına sahip senaryolarda önerilen yöntemin davranışı incelenmiştir. Bu tip sahnelerde kamera, yolda ilerlerken hem aynı yönde seyreden bir aracı takip etmekte hem de karşı şeritten gelen araçların oluşturduğu kısa süreli ve yüksek hız içeren dinamik etkileşimlere maruz kalmaktadır.

Şekil 5.16'da Sekans 04 için seçilen iki farklı zaman anında, ham görüntü ile yalnız semantik maskeleme ve önerilen hibrit yöntemin ürettiği çıktılar karşılaştırılmıştır. Sütunlar sırasıyla (a) ham görüntü, (b) semantik maskeleme ve (c) önerilen yöntem çıktısını göstermektedir. Ham görüntülerde dinamik nesnelerden çıkarılan öznitelikler kırmızı ile işaretlenirken, maskeleme sonrası filtrelenmesi gerekirken kalan hareketli öznitelikler turuncu ile gösterilmiştir.

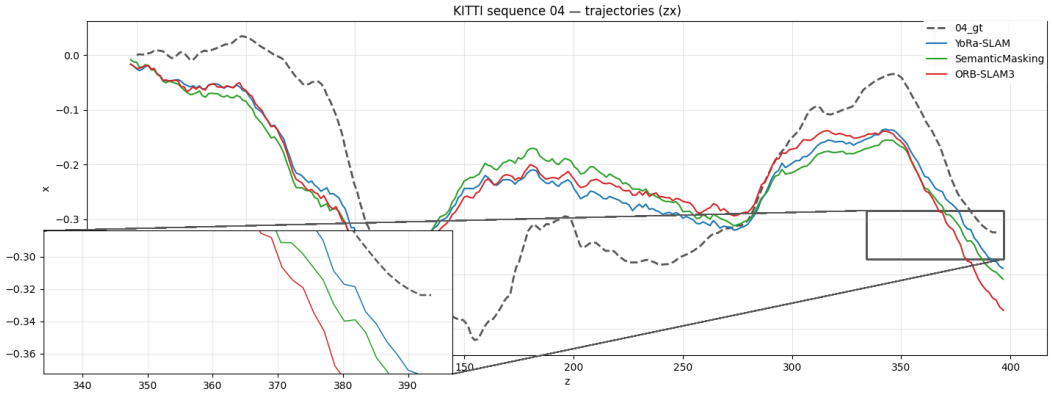


(a) Ham görüntü (b) Semantik maskeleme (c) Önerilen yöntem

Şekil 5.16 Sekans 04: seçilen örnek karelerde öznitelik dağılımı ve maskeleme çıktılarının görsel karşılaştırması

Şekil 5.16'da sunulan örnek kareler, dinamik araçların yoğun olduğu bölümlerde hareketli özniteliklerin izleme üzerinde baskın bir bozucu etki oluşturabildiğini göstermektedir. Sekans tamamen hareketli araçları içerdiği için semantik maskeleme ile önerilen yöntem aynı bölgeleri maskelemektedir. Bu sebeple algoritma çıktıları benzese de önerilen yöntem ile alınan kararlar daha tutarlılık göstermektedir.

Şekil 5.17 ve Şekil 5.18, Sekans 04 üzerinde elde edilen yörünge sonuçlarını sırasıyla z–x düzleminde ve zaman ekseninde göstermektedir. Şekil 5.17'deki dikdörtgen alan ilgili bölgenin yakınlaştırılmış görünümünü, Şekil 5.18'de mavi kesikli çizgi ile belirtilen zaman aralığı ise dinamik etkileşimin arttığı kesitteki davranışı vurgulamaktadır ve bu zaman aralığının yakınlaştırılmış görüntüsü aynı eksen içinde küçük dikdörtgenler ile sunulmuştur.



Şekil 5.17 Sekans 04 için z–x düzleminde yörünge karşılaştırması

Şekil 5.17'deki yakınlaştırılmış bölge, Şekil 5.16'te ikinci satırda verilen görsellerin olduğu sahnenin sonuna doğru karşı şeritten araçların geldiği zaman aralığını temsil etmektedir.

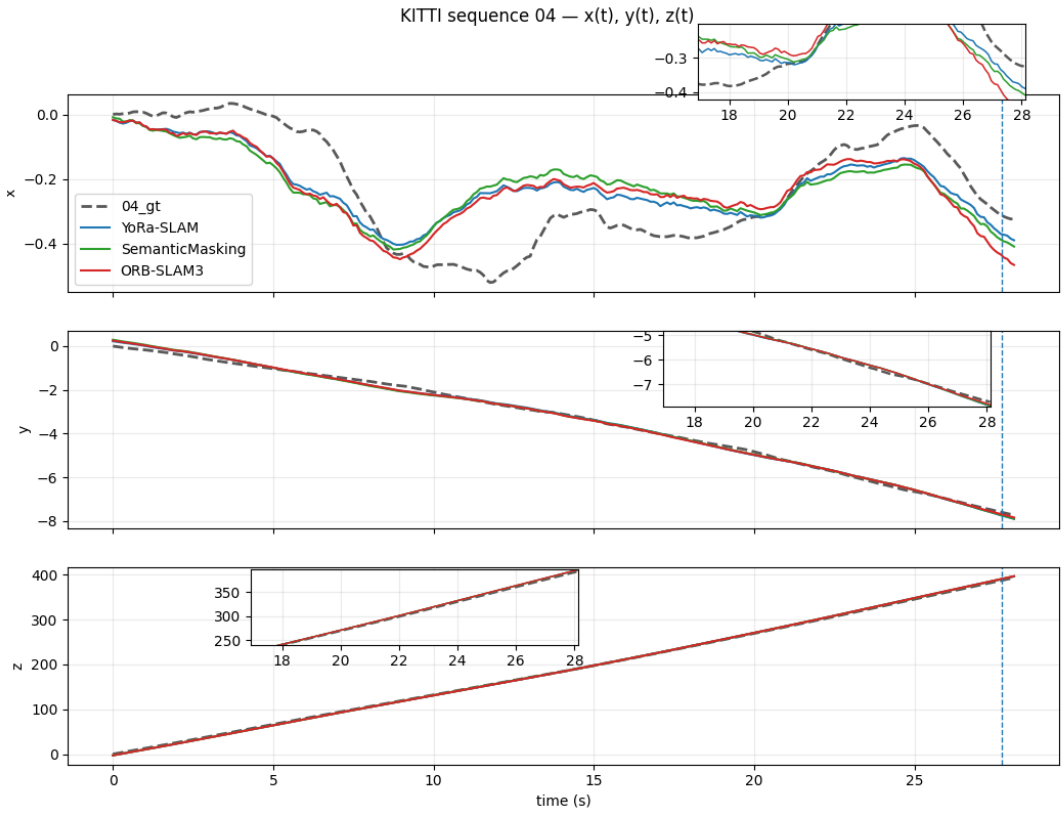
Bu zaman aralığında artan dinamik nesneler nedeniyle dinamik özniteliklerin etkisi belirginleşmekte ve yöntemlerin yer gerçeğe göre ürettiği pozisyon kestirimlerinde küçük yerel sapmalar oluşabilmektedir. Bu küçük sapmalara karşı dayanıklılık gösteren semantik maskeleme ve önerilen yöntem olduğu gözükmektedir.

Şekil 5.18'deki odaklanılmış zaman aralığı, sahnenin sonuna doğru karşı şeritten araçların geldiği zamanı dilimini göstermektedir. Burada araç düz ilerlediği için y ve z eksenleri neredeyse birebir uyuşsa da karşı şeritten gelen araçların verdiği etki büyük olmasa da küçük sapmalara sebep olduğu x ekseninde açıkça gözükmektedir. Aynı durumun sürekli yaşanması durumunda hatanın birikimli olarak artacağı ve uzun bir senaryoda bu durumun birikimli büyük hatalara sebep olacağı da muhtemeldir.

Dolayısıyla, bu tür durumlarda başarımı belirleyen kritik unsur, dinamik etkileşim anlarında bozucu ölçümleri baskılarken bariyer yol çizgisi gibi statik yol yapılarından gelen güvenilir özniteliklerin takibe yeterli katkıyı sürdürmesi olarak gösterilebilir.

Tablo 5.13 ve Tablo 5.14, şehir içi çift yönlü hızlı trafik senaryosunu temsil eden Sekans 04 üzerinde sırasıyla mutlak ve bağıl yörünge hata istatistiklerini sunmaktadır.

Tablo 5.13 incelendiğinde, Sekans 04'te hem semantik maskeleme hem de önerilen yöntemin ham veriye kıyasla ATE değerlerini düşürdüğü görülmektedir. Önerilen yöntem ve semantik maskeleme, KOH ve ortalama hatada en düşük değerlere



Şekil 5.18 Sekans 04 için $x(t)$, $y(t)$, $z(t)$ eksenlerindeki konum değişimi

Tablo 5.13 Sekans 04 için mutlak yörünge hatalarının (ATE) istatistiksel karşılaştırılması

İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	1.78	1.69	**1.69**
Ortalama	1.55	1.48	**1.47**
Std. Sapma	0.87	**0.81**	0.82
En Küçük	0.10	0.11	**0.07**
En Büyük	3.27	3.07	**3.03**

ulaşarak, dinamik kaynaklı yerel sapmaların birikimli sürüklenmeye dönüşmesini engellediği görülmektedir. Standart sapma değerlerinin de ham veriye göre azalması, yörünge hatasının zaman içinde daha kararlı bir düzeyde kaldığını göstermektedir.

Tablo 5.14 Sekans 04 için bağıl yörünge hatalarının (RTE) istatistiksel karşılaştırılması

İstatistik [m]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.0288	0.0285	**0.0280**
Ortalama	0.0268	0.0266	**0.0262**
Std. Sapma	0.0105	0.0100	**0.0099**
En Küçük	**0.0039**	0.0045	0.0052
En Büyük	0.0818	**0.0740**	0.0798

Bağıl hata sonuçları Tablo 5.14 üzerinden değerlendirildiğinde ise, üç yöntem arasında genel olarak yakın değerler elde edildiği, ancak önerilen yöntemin hem KOH hem de ortalama hatada en düşük seviyeye indiği gözlenmektedir. Bu durum, önerilen yaklaşımın dinamik bölgelerden gelen bozulmaları daha erken bastırarak ardışık poz tahminleri arasındaki tutarlılığı artırdığını işaret etmektedir.

Genel olarak Sekans 04 sonuçları, çift yönlü hızlı trafikte dinamik etkileşimlerin sıklaşmasının, küçük yerel sapmaları daha kritik hale getirdiğini ve bu nedenle kısa zamanlı iyileştirmelerin bile mutlak yörünge hatasına anlamlı biçimde yansıdığını ortaya koymaktadır.

Tablo 5.15 ve Tablo 5.16, Sekans 04 üzerinde sırasıyla mutlak ve bağıl dönme hatalarının istatistiksel karşılaştırmasını sunmaktadır.

Tablo 5.15 Sekans 04 için mutlak dönme hatalarının (ARE) istatistiksel karşılaştırılması

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	**25.7222**	33.7575	34.4695
Ortalama	**25.7216**	33.7574	34.4694
Std. Sapma	0.1716	0.0951	**0.0859**
En Küçük	**25.4445**	33.5961	34.3350
En Büyük	**26.0099**	33.9656	34.6587

Bağıl dönme hatası sonuçları değerlendirildiğinde, önerilen yöntemin dönme hareketini kareler arası daha tutarlı bir biçimde izlediği görülmektedir. Özellikle ani bozulmaların baskılanması sayesinde, kısa zaman aralıklarında yönelim değişimlerinin daha kararlı üretildiği ve dinamik kaynaklı yerel sapmaların azaltıldığı anlaşılmaktadır. Buna karşılık mutlak dönme hatası, tüm sahneyi

Tablo 5.16 Sekans 04 için bağıl dönme hatalarının (RRE) istatistiksel karşılaştırılması

İstatistik [°]	Ham veri	Semantik maskeleme	Önerilen yöntem
KOH (RMSE)	0.0333	0.0321	**0.0309**
Ortalama	0.0280	0.0279	**0.0272**
Std. Sapma	0.0181	0.0158	**0.0146**
En Küçük	**0.0014**	0.0033	0.0050
En Büyük	0.1850	0.1206	**0.0843**

kapsayan bir metrik olduğundan sahne boyunca taşınan yönelim ofsetini de yansıtmaktadır. Sekans 04'te hareketin büyük ölçüde doğrusal ilerlemesi ve dönüşlerin sınırlı kalması, mutlak yönelimde oluşabilecek küçük bir ofsetin yeniden gözlenebilir hale gelmesini zorlaştırmaktadır. Dolayısıyla bu ofsetin optimizasyon sürecinde etkin biçimde düzeltilmesini güçleştirmektedir. Bu nedenle, mutlak dönme hatası metriklerinde gözlemlenebilirlik kısıtından dolayı ham verinin hata değerleri daha düşük çıkmasına rağmen önerilen yöntemin standart sapmasının düşük çıkması aslında bir hata ofseti ile tutarlı bir sonuç izlediğini göstermektedir. Daha belirgin ve sık dönüş içeren senaryolarda ise yönelim değişimi daha güçlü geometrik kısıtlar oluşturacağından, mutlak açı sapmalarının daha tutarlı çıkması beklenmektedir.

5.5 Hesaplama Maliyeti ve Çalışma Zamanı Analizi

Bu bölümde, önerilen hibrit dinamik nesne filtreleme mimarisinin ORB-SLAM3 üzerine getirdiği ek hesaplama yükü incelenmiştir. Değerlendirme, KITTI Odometry sekansları üzerinde kare başına gecikme ölçülerek gerçekleştirilmiştir. Toplam uçtan uca süre ile birlikte her bir işlem bloğunun süreleri bileşen bazında raporlanmıştır. Tablo 5.17, her sekans için kare başına ortalama süreleri ve standart sapma değerlerini (ms) göstermektedir.

Tabloda yer alan sütunlar şu şekilde tanımlanmıştır:

- **Sekans**: KITTI Odometry sekans numarası
- **Toplam**: Önerilen yöntemin uçtan uca toplam kare başına ortalama işleme süresi
- **Derinlik**: Stereo eşleştirme ile derinlik haritası üretim süresi
- **YOLO**: YOLO ile nesne tespiti ve aday dinamik bölge çıkarım süresi
- **RAFT**: RAFT tabanlı yoğun optik akış çıkarımı süresi

Tablo 5.17 Kare başına gecikme süresi (ort / ± std, ms)

Sekans	Toplam	Derinlik	YOLO	RAFT	PnP	Karar	Maskeleme
00	197.69 ± 20.79	51.61 ± 1.91	15.20 ± 14.42	61.50 ± 8.73	44.93 ± 4.39	6.29 ± 3.12	18.00 ± 6.16
01	175.92 ± 34.56	50.31 ± 0.75	13.53 ± 28.80	55.35 ± 5.35	40.16 ± 4.75	0.96 ± 1.13	15.46 ± 7.81
03	181.58 ± 40.58	50.80 ± 1.73	13.73 ± 34.06	55.75 ± 4.48	44.65 ± 5.92	1.17 ± 1.36	15.33 ± 9.29
04	178.77 ± 64.85	49.60 ± 0.90	16.17 ± 55.16	51.43 ± 5.27	42.66 ± 9.63	2.06 ± 1.33	16.72 ± 14.48
07	184.91 ± 33.84	48.48 ± 0.54	15.55 ± 27.31	54.74 ± 7.30	42.16 ± 4.53	6.69 ± 3.29	17.14 ± 7.85

- **PnP**: PnP + RANSAC ile kamera hareketinin (R, t) kestirimi süresi
- **Karar**: Gözlenen akış ile beklenen akışın karşılaştırılması ve dinamik–statik kararının verilme süresi
- **Maskeleme**: Dinamik olarak sınıflandırılan bölgelerin görüntü üzerinde maskelenme süresi

Tablo 5.17 incelendiğinde, toplam süre bütçesinin ağırlıklı olarak derinlik, RAFT ve PnP aşamalarından oluştuğu görülmektedir. Buna karşılık karar ve maskeleme adımlarının ek yükü daha sınırlı olduğu gözlemlenmektedir. Öte yandan YOLO aşamasının standart sapmasının belirgin biçimde yüksek olması, çalışma zamanının sahnedeki tespit edilen örnek sayısına ve sahne karmaşıklığına duyarlı olduğuna işaret etmektedir.

Bu gözlem, özellikle yoğun dinamik nesne içeren karelerde semantik çıkarım maliyetinin dalgalanabildiğini belirtmektedir. Ancak genel olarak toplam gecikmenin asıl belirleyicisinin derinlik, optik akış ve PnP tabanlı hareket kestirimi olduğunu göstermektedir.

5.6 Sadece RAFT Tabanlı Dinamik Bölge Filtreleme

Bu bölümde, önerilen hibrit yöntemin bir bileşeni olan RAFT tabanlı yoğun optik akışın tek başına kullanıldığı bir yaklaşım değerlendirilmektedir.

Bu yaklaşımda, semantik bilgi ve geometrik tutarlılık kontrolü devre dışı bırakılmıştır. Bunun yerine, yalnızca ardışık kareler arasında hesaplanan yoğun optik akışın büyüklüğü kullanılarak dinamik bölgeler belirlenmiştir.

Her kare için RAFT ile elde edilen optik akış büyüklüğü Denklem (5.1) 'deki gibi hesaplanmaktadır.

$$m(p) = \sqrt{u(p)^2 + v(p)^2} \tag{5.1}$$

Burada $u(p)$ ve $v(p)$, sırasıyla p pikseli için optik akış vektörünün yatay ve düşey bileşenlerini ifade etmektedir.

Dinamik bölgeler, bu büyüklük değerinin kare bazlı ortalama değerinin üzerinde olduğu pikseller olarak Denklem (5.2)'taki gibi belirlenmiştir.

$$\mathcal{D} = \{p \mid m(p) > \alpha \cdot \bar{m}\} \tag{5.2}$$

Burada $\mathcal{D}$, dinamik olarak sınıflandırılan piksel kümesini ifade etmektedir. $m(p)$, Denklem (5.1) ile hesaplanan optik akış büyüklüğünü, $\bar{m}$ ise ilgili sahneye ait ortalama optik akış büyüklüğünü temsil etmektedir. α parametresi, dinamik bölge seçiminde kullanılan eşik ölçek katsayısı olup, deneysel olarak belirlenmiştir.

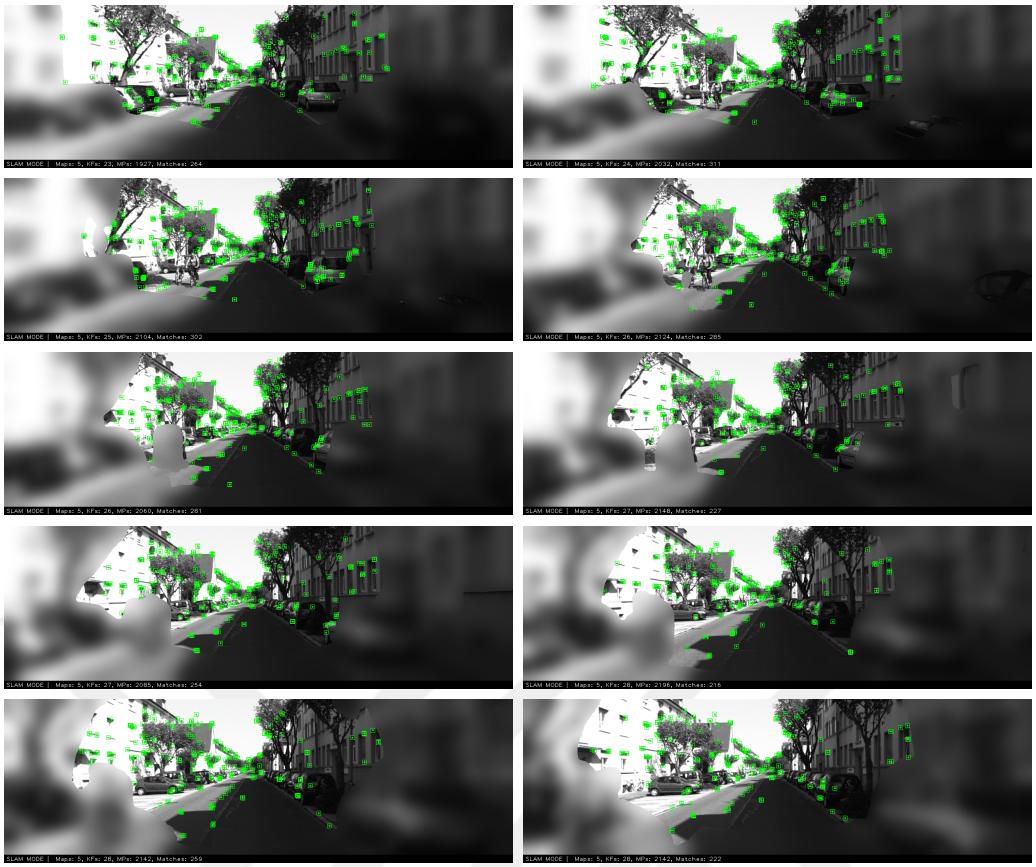
Elde edilen bölgeler Gauss filtresi ile bulanıklaştırılarak öznitelik çıkarımı engellenmiştir. Sekans 00 için ardışık sahneler Şekil 5.19'da verilmiştir.

Şekil 5.19 incelendiğinde, uzak bölgelerdeki optik akış büyüklüğünün, nesneler hareketli olsa dahi düşük kalması nedeniyle eşik değerin altında kalarak statik olarak sınıflandırıldığı görülmektedir. Buna karşılık, yakın mesafedeki bölgelerde optik akış büyüklüğünün yüksek olması, bu bölgelerin dinamik olarak sınıflandırılmasına neden olmaktadır.

Ayrıca optik akış büyüklüğüne dayalı bu yaklaşımın, aracın doğrusal ve açısal hareketine doğrudan bağlı olduğu ve bu nedenle sahne koşullarına göre değişkenlik gösterdiği açıkça gözlemlenmektedir.

Sadece optik akı yaklaşımı sonucunda elde edilen görüntüler, ORB-SLAM3'e giriş olarak verilerek sonuçlar Tablo 5.18'de ve Tablo 5.19'da gösterilmiştir.

Tablo 5.18 ve Tablo 5.19 incelendiğinde, yalnızca optik akış büyüklüğüne dayalı bu yaklaşımın dinamik nesneleri tespit etmede sınırlı kaldığı görülmektedir.



Şekil 5.19 Sekans 00 üzerinde sadece RAFT tabanlı yaklaşım ile elde edilen örnek ardışık sahneler

Tablo 5.18 Sadece RAFT tabanlı yaklaşım için ATE ve RTE hata istatistikleri

Sekans	ATE (metre)					RTE (metre)				
	En Küçük	En Büyük	Ortalama	Std	KOH	En Küçük	En Büyük	Ortalama	Std	KOH
00	6.99	129.45	47.16	21.58	51.86	0.01	132.34	0.66	3.61	3.67
01	1.74	66.00	38.54	14.57	41.21	0.01	3.46	0.30	0.37	0.48
03	0.91	12.26	5.20	3.15	6.08	0.02	0.63	0.10	0.08	0.13
04	0.26	3.36	1.77	0.93	2.00	0.01	0.32	0.08	0.05	0.10
07	0.95	14.66	5.85	2.92	6.54	0.01	4.64	0.12	0.53	0.55

Tablo 5.19 Sadece RAFT tabanlı yaklaşım için ARE ve RRE hata istatistikleri

Sekans	ARE (derece)					RRE (derece)				
	En Küçük	En Büyük	Ortalama	Std	KOH	En Küçük	En Büyük	Ortalama	Std	KOH
00	0.21	71.76	10.64	15.25	18.60	0.00	70.84	0.22	2.00	2.01
01	0.83	6.40	2.15	1.89	2.87	0.01	0.35	0.06	0.04	0.07
03	0.22	1.48	0.48	0.18	0.51	0.01	0.26	0.05	0.04	0.06
04	29.99	30.48	30.15	0.10	30.15	0.01	0.21	0.05	0.04	0.07
07	1.37	12.88	11.61	1.59	11.72	0.00	18.07	0.32	1.89	1.92

Elde edilen hata değerleri, özellikle hızlı hareketin olduğu sekanslarda belirgin biçimde artmakta ve yöntemin koşullara duyarlı olduğunu göstermektedir. Bununla birlikte uzak mesafedeki hareketli nesnelerin düşük optik akı üretmesi nedeniyle tespit edilememesi, yakın mesafedeki nesnelerin büyük optik akı üretmesi nedeniyle yanlış sınıflandırma yapılması bu yaklaşımın temel kısıtlarını ortaya koymaktadır.

Bu bulgular, yalnızca optik akış büyüklüğüne dayalı bir karar mekanizmasının sahne geometrisi ve kamera hareketi bilgisini içermediği için yeterli ayrıştırma performansına ulaşamadığını göstermektedir.

Buna karşılık, önerilen hibrit yöntemde kullanılan semantik bilgi ve beklenen akış modeli, söz konusu hatalı sınıflandırmaları önemli ölçüde azaltmaktadır.

5.7 Tüm Senaryoların ve Yöntemlerin Toplu Değerlendirilmesi

Bu bölümde, KITTI odometri veri kümesindeki tüm senaryolar için ham görüntü, semantik maskeleme, yalnızca RAFT tabanlı maskeleme ve önerilen yöntem kullanılarak elde edilen yörünge kestirim sonuçları karşılaştırılmıştır. Karşılaştırmalar, mutlak yörünge hatası (ATE) ve bağıl yörünge hatası (RTE) için hesaplanan RMSE ve standart sapma (STD) değerleri sırasıyla tablo 5.20'de ve tablo 5.21'de sunulmuştur.

Tablo 5.20 KITTI senaryoları için ham veri, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının ATE RMSE[m] karşılaştırması.

Sekans	Ham Görüntü		Semantik Maskeleme		Yalnızca RAFT		Önerilen Yöntem	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
00	3.91	1.42	4.31	1.59	51.86	21.58	**3.80**	**1.36**
01	13.67	4.33	8.30	2.94	41.21	14.57	**7.70**	**2.78**
02	6.89	2.63	7.30	2.97	7.20	2.43	**6.03**	**1.98**
03	4.92	2.44	4.87	2.43	6.08	3.15	**4.74**	**2.38**
04	1.78	0.87	**1.69**	**0.81**	2.00	0.93	**1.69**	0.82
05	2.09	**0.89**	2.14	0.95	43.05	18.17	**2.03**	0.95
06	2.22	**1.00**	2.47	1.10	2.79	1.35	**2.20**	**1.00**
07	0.96	0.35	1.09	0.39	6.54	2.92	**0.92**	**0.34**
08	**3.94**	2.02	4.31	2.11	–	–	**3.94**	**1.93**
09	3.36	0.74	5.17	1.54	5.14	1.98	**3.30**	0.81

ATE sonuçları incelendiğinde, önerilen yöntemin tüm senaryolarda en düşük hata değerlerini ürettiği görülmektedir. Bazı senaryolarda ham görüntü veya semantik maskeleme ile benzer sonuçları üretmiştir. Sahnedeki statik obje sayısı benzer ise ham görüntü yöntemine, dinamik obje sayısı benzer ise semantik

Tablo 5.21 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının RTE RMSE[m] karşılaştırması.

Sekans	Ham Görüntü		Semantik Maskeleme		Yalnızca RAFT		Önerilen Yöntem	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
00	0.0883	0.0644	**0.0701**	**0.0493**	3.6685	3.6078	0.0707	0.0520
01	0.0488	0.0222	0.0480	0.0222	0.4771	0.3682	**0.0464**	**0.0208**
02	0.0707	0.0588	**0.0478**	**0.0279**	0.0760	0.0444	0.0676	0.0363
03	0.0678	0.0242	0.0678	0.0236	0.1308	0.0798	**0.0664**	0.0243
04	0.0288	0.0105	0.0285	0.0100	0.0973	0.0481	**0.0280**	**0.0099**
05	0.0330	0.0208	**0.0320**	**0.0183**	1.1220	0.9495	0.0410	0.0238
06	0.0303	0.0147	0.0321	0.0148	0.6641	0.6492	**0.0296**	**0.0137**
07	0.0511	0.0316	0.0466	0.0311	0.5460	0.5332	**0.0460**	0.0318
08	0.0958	0.0848	**0.0951**	**0.0834**	–	–	0.0976	0.0869
09	0.0403	0.0188	**0.0317**	**0.0158**	0.0483	0.0318	0.0403	0.0191

maskeleme yöntemine benzer sonuçlar çıkarttığı senaryo 01, 04 ve 08'den açıkça gözlemlenmektedir.

ATE tablosunda dikkat çeken bir diğer nokta, yalnızca RAFT tabanlı yaklaşımın birçok sekans için belirgin biçimde yüksek hata üretmesidir. Özellikle 00, 01 ve 05 sekanslarında sadece RAFT yöntemi diğer tüm yöntemlere göre oldukça yüksek RMSE değerleri vermiştir. Senaryo 08 için karşılaştırabilecek bir harita çözümü bulamamıştır. Bu sonuç, optik akış bilgisinin tek başına güvenilir bir maskeleme kararı üretmek için yeterli olmadığını göstermektedir.

RTE sonuçları da genel eğilim açısından ATE bulgularını desteklemektedir. Önerilen yöntem 01, 03, 04, 06 ve 07 sekanslarında en iyi sonuçları verirken, bazı sekanslarda semantik maskeleme daha düşük RTE değerleri üretmiştir. Bu durum, semantik maskelemenin kısa vadeli göreli hareket kestiriminde bazı sahnelerde faydalı olabildiğini, ancak yönelim gibi diğer kestirimlerin yanlış olmasından dolayı genel sonuca bakıldığında sapma eğiliminin fazla olduğunu göstermektedir.

Genel olarak birlikte değerlendirildiğinde, önerilen yöntemin hem küresel yörünge hatasında hem de göreli hareket kestiriminde gürbüz ve çoğu durumda daha başarılı sonuçlar verdiği görülmektedir.

Yörünge doğruluğundaki bu genel eğilim, dönme hatası metriklerine de yansımaktadır. Bu kapsamda mutlak dönme hatası (ARE) ve bağıl dönme hatası (RRE) için elde edilen RMSE ve standart sapma (STD) değerleri sırasıyla Tablo 5.22 ve Tablo 5.23'de verilmiştir.

Tablo 5.22 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının ARE RMSE[°] karşılaştırması.

Sekans	Ham Görüntü		Semantik Maskeleme		Yalnızca RAFT		Önerilen Yöntem	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
00	0.7163	0.3880	0.7941	0.4211	18.6003	15.2545	**0.6974**	**0.3909**
01	3.7584	1.6856	3.4751	1.7958	2.8652	1.8898	**2.7706**	**1.4638**
02	1.4261	0.5496	1.5407	0.5507	**1.0453**	**0.3108**	1.1545	0.5951
03	0.5209	0.1536	**0.4634**	**0.1678**	0.5075	0.1769	0.5671	0.1791
04	**25.7222**	0.1716	33.7575	0.0951	30.1486	0.0957	34.4695	0.0859
05	0.4273	0.1765	0.4325	0.1798	11.9795	8.9321	**0.4013**	**0.1612**
06	**0.5413**	**0.1333**	1.0361	0.3311	0.9871	0.1116	0.8068	0.1623
07	0.4305	0.1638	0.4159	0.1467	11.7200	1.5856	**0.3953**	**0.1421**
08	1.1538	0.4019	1.3924	0.4368	–	–	**1.1535**	**0.3877**
09	**0.7378**	**0.2510**	1.1232	0.4624	1.4119	0.6480	0.7382	0.2544

ARE sonuçları incelendiğinde, önerilen yöntemin birçok sekansta dönmenin gözlemlenebilir olduğu durumlar için düşük hata değerleri ürettiği görülmektedir. Özellikle 00, 01, 05, 07 ve 08 sekanslarında önerilen yöntem en iyi ya da en iyiye çok yakın ARE RMSE sonuçları vermiştir. Yalnızca RAFT tabanlı yaklaşım ise özellikle 00, 05 ve 07 sekanslarında belirgin biçimde yüksek hata üretmiş, bu da optik akış bilgisinin tek başına güvenilir bir dönme kestirimi iyileştirmesi sağlamadığını ortaya koymuştur.

Tablo 5.23 KITTI sekansları için ham görüntü, semantik maskeleme, yalnızca RAFT ve önerilen yöntem sonuçlarının RRE RMSE[°] karşılaştırması.

Sekans	Ham Görüntü		Semantik Maskeleme		Yalnızca RAFT		Önerilen Yöntem	
	RMSE	STD	RMSE	STD	RMSE	STD	RMSE	STD
00	0.2865	0.2699	**0.2387**	**0.2234**	2.0136	2.0020	0.2492	0.2343
01	0.0447	0.0232	0.0447	0.0260	0.0743	0.0447	**0.0400**	**0.0206**
02	0.1016	0.0877	**0.0998**	**0.0856**	0.1238	0.1021	0.1220	0.1075
03	0.0523	0.0276	0.0592	0.0372	0.0642	0.0363	**0.0507**	**0.0272**
04	0.0333	0.0181	0.0321	0.0158	0.0674	0.0397	**0.0309**	**0.0146**
05	0.0684	0.0508	**0.0681**	**0.0495**	0.8127	0.7787	0.0881	0.0724
06	**0.0402**	**0.0240**	0.0431	0.0253	0.0725	0.0475	0.0411	0.0243
07	0.0706	0.0438	0.0607	0.0363	1.9160	1.8892	**0.0591**	**0.0355**
08	**0.0592**	**0.0350**	0.0607	0.0358	–	–	**0.0592**	**0.0350**
09	0.0529	0.0276	**0.0506**	**0.0266**	0.0617	0.0319	0.0532	0.0287

RRE sonuçları da genel olarak benzer bir eğilim göstermektedir. Önerilen yöntem 01, 03, 04 ve 07 sekanslarında en iyi sonuçları verirken, bazı sekanslarda semantik maskeleme daha düşük RRE değerleri üretmiştir. Bununla birlikte yalnızca

RAFT tabanlı yöntem, özellikle 00, 05 ve 07 sekanslarında diğer yöntemlere kıyasla oldukça yüksek hata vermiştir. Genel olarak değerlendirildiğinde, önerilen yöntemin dönme tabanlı hata ölçütlerinde de gürbüz ve dengeli bir performans sunduğu, semantik ve hareket tabanlı bilgilerin birlikte kullanılmasının yalnızca öteleme değil yönelim kestirimi açısından da olumlu katkı sağladığı görülmektedir. Önerilen yöntemin farklı sahne tiplerinde ham görüntü, yalnızca semantik maskeleme ve sadece optik akış tabanlı yaklaşımlara kıyasla genel olarak daha dengeli ve düşük hata ürettiğini göstermektedir.

6
SONUÇ

Bu çalışmada, dinamik ortamların vSLAM üzerindeki bozucu etkilerini azaltmak amacıyla, derin öğrenme tabanlı optik akış (RAFT) ve nesne tespiti (YOLO) bileşenlerini bir araya getiren gürbüz bir vSLAM mimarisi (YoRa-SLAM) önerilmiştir. Klasik semantik maskeleme yaklaşımlarının potansiyel olarak hareketli sınıfları ayrım gözetmeden maskelemesinin aksine, önerilen hibrit yöntem hareket tutarlılığı denetimi uygulayarak gerçekten hareket eden nesneleri park halindeki araçlar gibi statik örneklerden ayırt etmektedir. Bu sayede, dinamik bölgelerden gelen bozucu ölçümlerin izleme sürecine dahil edilmesi engellenirken, konum kestirimi için faydalı olan statik bölgeler mümkün olduğunca korunarak uzun yörüngelerde daha kararlı bir pozisyon tahmini hedeflenmektedir.

Önerilen hibrit dinamik maskeleme yaklaşımının farklı sürüş senaryolarındaki davranışı görsel ve nicel sonuçlar üzerinden değerlendirilmiştir. Sunulan sonuçlar, yöntemin temel kazanımının yalnızca dinamik nesneleri bastırmak değil, aynı zamanda statik bölgeleri koruyarak poz kestirim kararlılığını artırmak olduğunu göstermektedir. Bu bağlamda, ORB-SLAM3 kullanılarak üç farklı yöntem karşılaştırılmıştır. Filtrelenmemiş görüntüler, yalnız semantik bilgi ile maskelenmiş görüntüler ve önerilen hibrit yöntem sonucu maskelenmiş görüntüler, hem görsel örnekler hem de ATE/RTE/ARE/RRE istatistikleri ile karşılaştırılmıştır.

Genel olarak tüm senaryolar bir arada değerlendirildiğinde, önerilen yaklaşımın en güçlü yönünün "seçici dinamik–statik karar verme" olduğu sonucuna varılmaktadır. Semantik tespit, dinamik olma potansiyeli taşıyan bölgeleri hızlıca bulurken, optik akış tabanlı tutarlılık kontrolü, bu adayların gerçekten dinamik olup olmadığını sahnenin hareket ipuçları üzerinden denetleyerek gerçeğe yakın bilgi vermektedir. Bu sayede, dinamik kaynaklı bozucu etkiler giderilirken, izlemeye katkı sağlayan statik öznitelikler korunabilmekte ve özellikle uzun sekanslarda birikimli hatanın büyümesi sınırlandırılabilmektedir. Sonuç olarak önerilen yöntem, dinamik sahnelerde vSLAM doğruluğunu artırmak için daha doğru ve seçici maskeleme yaklaşımını benimseyerek hem öznitelik takibi kararlılığına hem de yörünge

doğruluğuna katkı sağlamaktadır.

Gelecek çalışmalarda, önerilen hibrit filtreleme mimarisi daha etkin hale getirilerek gerçek zamanlı çalışabilirlik açısından geliştirilebilir. Ayrıca bu yapı, mevcut vSLAM sistemlerini yalnızca ön işleme seviyesinde beslemek yerine, vSLAM hattına doğrudan entegre edilerek pozisyon kestirimi ve haritalama adımlarında kullanılabilir. Bu entegrasyon sayesinde, statik sahne varsayımı altında beklenen optik akışın türetilmesi için gerekli kamera hareketi bilgisi doğrudan vSLAM çıktılarından beslenebilir ve dinamik–statik ayrımı optimizasyon döngüsünün bir parçası haline getirilebilir.

KAYNAKÇA

[1] Y. Wu, J. Kuang, X. Niu, J. Behley, L. Klingbeil, H. Kuhlmann, "Wheel-SLAM: Simultaneous Localization and Terrain Mapping Using One Wheel-Mounted IMU," *IEEE Robotics and Automation Letters*, c. 8, no. 1, ss. 280–287, 2023. doi: 10.1109/LRA.2022.3226071.

[2] J. Borenstein, L. Feng, "Measurement and correction of systematic odometry errors in mobile robots," *IEEE Transactions on Robotics and Automation*, c. 12, no. 6, ss. 869–880, 1996. doi: 10.1109/70.544770.

[3] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2. bs. Cambridge University Press, 2004.

[4] C. Cadena ve diğ., "Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age," *IEEE Transactions on Robotics*, c. 32, no. 6, ss. 1309–1332, 2016. doi: 10.1109/TRO.2016.2624754.

[5] Z. Chen, Z. Miao, M. Liu, C. Wu, Y. Wang, "A Fast and Accurate Visual Inertial Odometry Using Hybrid Point-Line Features," *IEEE Robotics and Automation Letters*, c. 9, no. 12, ss. 11 345–11 352, 2024. doi: 10.1109/LRA.2024.3490406.

[6] A. I. Mourikis, S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," *Proceedings 2007 IEEE International Conference on Robotics and Automation*, 2007, ss. 3565–3572. doi: 10.1109/ROBOT.2007.364024.

[7] T. Qin, P. Li, S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, c. 34, no. 4, ss. 1004–1020, 2018. doi: 10.1109/TRO.2018.2853729.

[8] H. Xu, X. Liu, Y. Liu, "Visual Inertial Odometry with Fusion of Point and Line Features in Low Illumination Environments," *2024 12th International Conference on Traffic and Logistic Engineering (ICTLE)*, 2024, ss. 52–56. doi: 10.1109/ICTLE62418.2024.10703877.

[9] L. Xu, H. Yin, T. Shi, D. Jiang, B. Huang, "EPLF-VINS: Real-Time Monocular Visual-Inertial SLAM With Efficient Point-Line Flow Features," *IEEE Robotics and Automation Letters*, c. 8, no. 2, ss. 752–759, 2023. doi: 10.1109/LRA.2022.3231983.

[10] J.-Y. Bouguet, "Pyramidal implementation of the lucas kanade feature tracker," 1999. erişim adresi: https://api.semanticscholar.org/CorpusID:9350588.

[11] S. Boche, J. Jung, S. B. Laina, S. Leutenegger, "OKVIS2-X: Open Keyframe-Based Visual-Inertial SLAM Configurable With Dense Depth or LiDAR, and GNSS," *IEEE Transactions on Robotics*, c. 41, ss. 6064–6083, 2025. doi: 10.1109/TRO.2025.3619051.

[12] S. Leutenegger, "OKVIS2: Realtime Scalable Visual-Inertial SLAM with Loop Closure," 2022. erişim adresi: https://api.semanticscholar.org/CorpusID:246996555.

[13] G. Klein, D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces," *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, 2007, ss. 225–234. doi: 10.1109/ISMAR.2007.4538852.

[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, c. 60, no. 2, ss. 91–110, 2004.

[15] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, "ORB: An Efficient Alternative to SIFT or SURF," *International Conference on Computer Vision (ICCV)*, 2011.

[16] C. Harris, M. Stephens, "A Combined Corner and Edge Detector," *Proceedings of the Alvey Vision Conference*, 1988.

[17] J. Shi, Tomasi, "Good features to track," *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1994, ss. 593–600. doi: 10.1109/CVPR.1994.323794.

[18] E. Rosten, T. Drummond, "Machine Learning for High-Speed Corner Detection," *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, A. Pinz, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, ss. 430–443, isbn: 978-3-540-33833-8.

[19] H. Bay, T. Tuytelaars, L. Van Gool, "SURF: Speeded Up Robust Features," *Computer Vision – ECCV 2006*, A. Leonardis, H. Bischof, A. Pinz, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, ss. 404–417, isbn: 978-3-540-33833-8.

[20] M. Calonder, V. Lepetit, C. Strecha, P. Fua, "BRIEF: Binary Robust Independent Elementary Features," *Computer Vision – ECCV 2010*, K. Daniilidis, P. Maragos, N. Paragios, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, ss. 778–792, isbn: 978-3-642-15561-1.

[21] S. Leutenegger, M. Chli, R. Y. Siegwart, "BRISK: Binary Robust invariant scalable keypoints," *2011 International Conference on Computer Vision*, ss. 2548–2555, 2011. erişim adresi: https://api.semanticscholar.org/CorpusID:1211102.

[22] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *2011 International Conference on Computer Vision*, 2011, ss. 2564–2571. doi: 10.1109/ICCV.2011.6126544.

[23] M. Calonder, V. Lepetit, C. Strecha, P. Fua, "BRIEF: Binary Robust Independent Elementary Features," *European Conference on Computer Vision (ECCV)*, 2010.

[24] J. Heinly, E. Dunn, J.-M. Frahm, "Comparative Evaluation of Binary Features," *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid, ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, ss. 759–773, isbn: 978-3-642-33709-3.

[25] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, G. Randall, "LSD: A Fast Line Segment Detector with a False Detection Control," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 32, no. 4, ss. 722–732, 2010. doi: 10.1109/TPAMI.2008.300.

[26] L. Zhang, R. Koch, "An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency," *Journal of Visual Communication and Image Representation*, c. 24, no. 7, ss. 794–805, 2013, issn: 1047-3203. doi: 10.1016/j.jvcir.2013.05.006. erişim adresi: https://www.sciencedirect.com/science/article/pii/S1047320313000874.

[27] R. Mur-Artal, J. M. M. Montiel, J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, c. 31, no. 5, ss. 1147–1163, 2015. doi: 10.1109/TRO.2015.2463671.

[28] R. Mur-Artal, J. D. Tardós, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, c. 33, no. 5, ss. 1255–1262, 2017. doi: 10.1109/TRO.2017.2705103.

[29] C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel, J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," *IEEE Transactions on Robotics*, c. 37, no. 6, ss. 1874–1890, 2021. doi: 10.1109/TRO.2021.3075644.

[30] T. Pire, T. Fischer, G. Castro, P. D. Cristóforis, J. Civera, J. J. Berlles, "S-PTAM: Stereo Parallel Tracking and Mapping," *Robotics and Autonomous Systems*, c. 93, ss. 27–42, 2017, issn: 0921-8890. doi: 10.1016/j.robot.2017.03.019. erişim adresi: https://www.sciencedirect.com/science/article/pii/S0921889015302955.

[31] M. Ferrera, A. Eudes, J. Moras, M. Sanfourche, G. Le Besnerais, "OV^{2}SLAM: A Fully Online and Versatile Visual SLAM for Real-Time Applications," *IEEE Robotics and Automation Letters*, c. 6, no. 2, ss. 1399–1406, 2021. doi: 10.1109/LRA.2021.3058069.

[32] G. Zeng ve diğ., "Bias-Eliminated PnP for Stereo Visual Odometry: Provably Consistent and Large-Scale Localization," *IEEE Robotics and Automation Letters*, c. 10, no. 11, ss. 11 840–11 847, 2025. doi: 10.1109/LRA.2025.3614050.

[33] R. A. Newcombe, S. J. Lovegrove, A. J. Davison, "DTAM: Dense tracking and mapping in real-time," *2011 International Conference on Computer Vision*, 2011, ss. 2320–2327. doi: 10.1109/ICCV.2011.6126513.

[34] J. Engel, T. Schöps, D. Cremers, "LSD-SLAM: Large-Scale Direct Monocular SLAM," *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars, ed., Cham: Springer International Publishing, 2014, ss. 834–849, isbn: 978-3-319-10605-2.

[35] J. Engel, V. Koltun, D. Cremers, "Direct Sparse Odometry," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 40, no. 3, ss. 611–625, 2018. doi: 10.1109/TPAMI.2017.2658577.

[36] C. Forster, M. Pizzoli, D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014, ss. 15–22. doi: 10.1109/ICRA.2014.6906584.

[37] X. Zhao, L. Liu, R. Zheng, W. Ye, Y. Liu, "A Robust Stereo Feature-aided Semi-direct SLAM System," *Robotics and Autonomous Systems*, c. 132, s. 103 597, 2020, issn: 0921-8890. doi: 10.1016/j.robot.2020.103597. erişim adresi: https://www.sciencedirect.com/science/article/pii/S0921889020304371.

[38] Q. U. Islam, H. Ibrahim, P. K. Chin, K. Lim, M. Z. Abdullah, "Dynamic Object-Aware SLAM: YoloV8 Integration and Enhanced Multi-View Geometry for Reliable Mapping in Dynamic Environment," *2023 7th International Conference on Imaging, Signal Processing and Communications (ICISPC)*, 2023, ss. 22–31. doi: 10.1109/ICISPC59567.2023.00013.

[39] B. Gökcen, E. Uslu, "Object Aware RGBD SLAM in Dynamic Environments," *2022 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, 2022, ss. 1–6. doi: 10.1109/INISTA55318.2022.9894245.

[40] K. J. Oguine, O. C. Oguine, H. I. Bisallah, "YOLO v3: Visual and Real-Time Object Detection Model for Smart Surveillance Systems(3s)," *2022 5th Information Technology for Education and Development (ITED)*, 2022, ss. 1–8. doi: 10.1109/ITED56637.2022.10051233.

[41] C. Yu ve diğ., "DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments," *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, ss. 1168–1174. doi: 10.1109/IROS.2018.8593691.

[42] V. Badrinarayanan, A. Kendall, R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 39, no. 12, ss. 2481–2495, 2017. doi: 10.1109/TPAMI.2016.2644615.

[43] L. He, S. Li, J. Qiu, C. Zhang, "DIO-SLAM: A Dynamic RGB-D SLAM Method Combining Instance Segmentation and Optical Flow," *Sensors*, c. 24, no. 18, 2024, issn: 1424-8220. doi: 10.3390/s24185929. erişim adresi: https://www.mdpi.com/1424-8220/24/18/5929.

[44] Y. Fan, Q. Zhang, Y. Tang, S. Liu, H. Han, "Blitz-SLAM: A semantic SLAM in dynamic environments," *Pattern Recognition*, c. 121, s. 108 225, 2022, issn: 0031-3203. doi: 10.1016/j.patcog.2021.108225. erişim adresi: https://www.sciencedirect.com/science/article/pii/S0031320321004064.

[45] N. Dvornik, K. Shmelkov, J. Mairal, C. Schmid, "BlitzNet: A Real-Time Deep Network for Scene Understanding," *2017 IEEE International Conference on Computer Vision (ICCV)*, ss. 4174–4182, 2017. erişim adresi: https://api.semanticscholar.org/CorpusID:215806888.

[46] B. Bescos, J. M. Fácil, J. Civera, J. Neira, "DynaSLAM: Tracking, Mapping, and Inpainting in Dynamic Scenes," *IEEE Robotics and Automation Letters*, c. 3, no. 4, ss. 4076–4083, 2018. doi: 10.1109/LRA.2018.2860039.

[47] T. Zhang, H. Zhang, Y. Li, Y. Nakamura, L. Zhang, "FlowFusion: Dynamic Dense RGB-D SLAM Based on Optical Flow," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, ss. 7322–7328. doi: 10.1109/ICRA40945.2020.9197349.

[48] H. Hirschmuller, "Accurate and efficient stereo processing by semi-global matching and mutual information," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, c. 2, 2005, 807–814 vol. 2. doi: 10.1109/CVPR.2005.56.

[49] H. Hirschmuller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, c. 30, no. 2, ss. 328–341, 2008. doi: 10.1109/TPAMI.2007.1166.

[50] Ultralytics, *Model Benchmarking with Ultralytics YOLO*, Ultralytics YOLO Documentation. erişim tarihi 13 Ara. 2025. erişim adresi: https://docs.ultralytics.com/modes/benchmark/.

[51] Z. Teed, J. Deng, "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow," *Computer Vision – ECCV 2020*, A. Vedaldi, H. Bischof, T. Brox, J.-M. Frahm, ed., Cham: Springer International Publishing, 2020, ss. 402–419, isbn: 978-3-030-58536-5.

[52] A. Geiger, P. Lenz, R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, ss. 3354–3361. doi: 10.1109/CVPR.2012.6248074.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. A. Kösem ve E. Uslu, "Hareketli Özellik Noktalarının Görsel Eşzamanlı Konumlandırma ve Haritalama Doğruluğuna Etkisi: ORB-SLAM3 ile Karşılaştırma," *Grad Symposium 2025 – Mühendislikte Sürdürülebilir Çözümler*, Yıldız Teknik Üniversitesi, Davutpaşa Kongre ve Kültür Merkezi, İstanbul, Türkiye, 15 Mayıs 2025.