



Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ TEKNOLOJİ FAKÜLTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

Düşük Maliyetli Stereo Kamera Sistemleri ile Taşınabilir Nesne SLAM

PROJE YAZARI

Ali Asghar

DANIŞMAN

Dr. Öğr. Üyesi Gözde Karataş Baydoğmuş

İSTANBUL, 2024



Teknoloji Fakültesi

MARMARA ÜNİVERSİTESİ TEKNOLOJİ FAKÜLTESİ BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME PROJESİ

Düşük Maliyetli Stereo Kamera Sistemleri ile Taşınabilir Nesne SLAM

PROJE YAZARI

Ali Asghar

170419926

DANIŞMAN

Dr. Öğr. Üyesi Gözde Karataş Baydoğmuş

İSTANBUL, 2024

MARMARA ÜNİVERSİTESİ
TEKNOLOJİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği Öğrencisi Ali Asghar’ın “Düşük Maliyetli Stereo Kamera Sistemleri ile Taşınabilir Nesne SLAM ” başlıklı bitirme projesi çalışması, 03/06/2024 tarihinde sunulmuş ve juri üyeleri tarafından başarılı bulunmuştur.

Jüri Üyeleri

Dr. Öğr. Üyesi Gözde Karataş Baydoğmuş (Danışman)

Marmara Üniversitesi (İMZА)

Doç. Dr. Buket Doğan (Üye)

Marmara Üniversitesi (İMZА)

Prof. Dr. Ali Buldu (Üye)

Marmara Üniversitesi (İMZА)

İÇİNDEKİLER

	Sayfa
SEMBOLLER/SYMBOLS.....	i
KISALTMALAR/ABBREVIATIONS.....	ii
ŞEKİL LİSTESİ.....	iii
ÖZET.....	iv
1. GİRİŞ.....	v
1.1. Genel Bilgiler.....	1
1.2. Literatür Taraması.....	1
1.3. Özgünlük.....	3
1.4. Materyal ve Yöntem.....	4
2. TEORİK ALTYAPI.....	6
2.1. Kamera Kalibrasyon.....	6
2.2. SLAM.....	8
3. Deneysel Çalışmalar.....	12
3.1. Kamera Kalibrasyon ve Kurumu.....	12
3.2. Veri Toplama.....	15
3.3. Performans Analizi.....	19
3.4. Zorluklar ve Sınırlamalar.....	22
4. SONUÇLAR.....	23
4.1. Genel Açıklamalar.....	23
4.2. Gelecek İyileştirmeler.....	23
4.3. Yaygın Etki.....	25
5. KAYNAKLAR.....	26

SEMBOLLER/SYMBOLS

t : Zaman (s)

u : 3D harita noktasının 2D izdüşümü

K : Kamera iç matrisi

R : Döndürme matrisi

t : Öteleme vektörü

π : Projeksiyon fonksiyonu

T_i : Ana karenin pozu

X_j : 3D harita noktası

KISALTMALAR/ABBREVIATIONS

SLAM : Simultaneous Localization and Mapping

VO : Visual Odometry

GPU : Graphics Processing Unit

CPU : Central Processing Unit

ORB : Oriented FAST and Rotated BRIEF

FAST : Features from Accelerated Segment Test

BRIEF : Binary Robust Independent Elementary Features

RANSAC : Random Sample Consensus

PnP : Perspective-n-Point

BOW : Bag Of Words

EKF : Extended Kalman Filter

ŞEKİL LİSTESİ

	Sayfa
Şekil 1. SLAM sisteminin genel mimarisi.....	2
Şekil 2. Özel Stereo Kamera Sistem.....	4
Şekil 3. Monoküler Sistemde Ölçek Belirsizliği ve Star Wars.....	4
Şekil 4. Apriltag Kalibrasyon hedefi.....	5
Şekil 5. Kalibrasyon örnek.....	6
Şekil 6. Distorsiyonun kamera ile çekilen görüntüler üzerindeki etkisi.....	8
Şekil 7. İmaj düzeyindeki noktanın gerçek dünyaya 3D konumu bulma görseli.....	9
Şekil 8. Yerel Bundle Adjustment Görsel Açıklaması.....	10
Şekil 9. Kalibrasyon için stereo kameramızdan çekilen örnek çift kareler.....	12
Şekil 10. Bulanıklığından dolayı elenmiş resimler.....	13
Şekil 11. Polar dereceye göre kameranın piksel bazlı reprojection hatası.....	13
Şekil 12. Düzeltip birleştirilmiş sonuç stereo görüntü.....	14
Şekil 13. EuroC dataset gösterme.....	15
Şekil 14. Çektiğimiz Çalışma Oda veri kümesinden kare çiftler.....	16
Şekil 15. Çektiğimi Koridor veri kümesinden kare çiftler.....	17
Şekil 16. Bodrum Kat/Yemekhane veri kümesinden kare çiftler.....	18
Şekil 17. Yurt önü veri kümesinden kare çiftler.....	18
Şekil 18. vSLAM çalıştırılması.....	19
Şekil 19. Işık seviyesi tutarlı olmadığından sistem doğru konumlandıramaz.....	20
Şekil 20. Fazla hızlı hareketten dolayı sistem doğru konumlandıramaz.....	21
Şekil 21. Başarılı izleme örneği.....	22
Şekil 22. Kameranın fazla hızlı dönmesi nedeniyle izleme başarısızlığı.....	22
Şekil 23. DxSLAM gösterimi.....	24
Şekil 24. Düşük Maliyetli vSLAM'in olası uygulamaları.....	25

ÖZET

Konumlandırma, otomasyon ve robotik de dahil olmak üzere birçok alanda önemli bir problemdir. Bir otomatın hareket talimatlarını tam olarak takip edebilmesi için, çevresine ve son konumuna göre nerede durduğuna dair kesin bilgiye sahip olması gereklidir. GPS gibi yerleşik yerelleştirme yöntemleri kısa mesafelerde düşük doğruluğa sahiptir ve kapalı ortamlarda sinyal düşüşünden muzdariptir. Bu makalede önerilen vSLAM boru hattı, yerelleştirme sorununa düşük maliyetli ve verimli bir çözüm sunmaktadır. Ucuz, kolayca tekrarlanabilir bir stereo kamera sistemi oluşturmak için iki hazır web kamerasını birleştirir. ORB (Yönlendirilmiş HIZLI ve Döndürülmüş BRIEF) ve BOW (Kelime Torbası) konum çıkarma ve ORB-SLAM3 tarafından sağlanan Döngü Kapatma, Haritalama ve Görselleştirme özellikleri ile özellik çıkarımı kullanarak nesneyi yerelleştirebilir, hareket yörüngesini izleyebilir ve 5 m'nin altındaki mesafeler için 5,96 mm içinde doğru çevrenin bir nokta bulutu temsilini sağlayabilir.

ABSTRACT

Localization is an important problem in many different fields, including automation and robotics. In order for an automaton to follow movement instructions precisely, it needs to have precise knowledge of where it stands relative to its surroundings and the final position. Established localization methods like GPS have low accuracy over short distances and suffer from signal drop in covered environments. The vSLAM pipeline proposed in this paper provides a low-cost and efficient solution to the localization problem. It combines two off-the-shelf webcams to create an inexpensive, easily reproducible stereo camera system. Using ORB (Orientated FAST and Rotated BRIEF) and feature extraction with BOW (Bag of Words) location extraction and the Loop Closing, Mapping and Visualization features provided by ORB-SLAM3 it is able to localize the object, track its movement trajectory and provide a point-cloud representation of the surroundings accurate to within 5.96 mm for distances under 5 m.

1. GİRİŞ

1.1. Genel Bilgiler

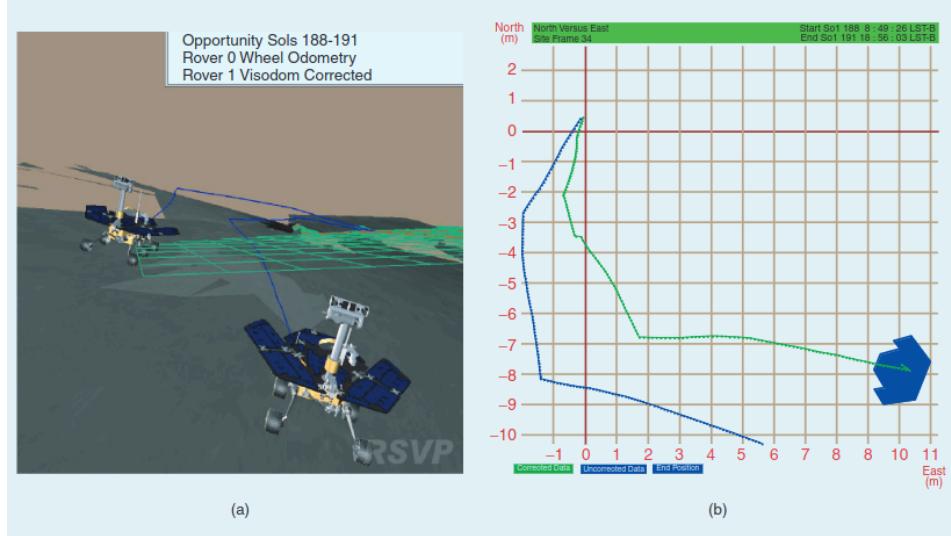
Bir ajanın doğru konumlandırma otomasyon ve robotikte temel bir problemdir. Otomatın çevresi içindeki ve hedefine göre konumu, hareket gerektiren görevleri tamamlamak için temel bir gereklilikdir. Örnekler arasında otomatik sürücüler ve hizmet robotları yer alır. GPS gibi geleneksel yerelleştirme yöntemleri, küçük mesafelerde düşük doğrulukları ve iç mekan/su altı/karasal ortamlarda güvenilmezlikleri nedeniyle bu amaç için yetersizdir [1]. Bu sorunu çözmek için LiDAR, radyo dalgaları ve görsel bilgiler gibi birçok farklı yaklaşım kullanılmıştır. Bu makalede, kamera edinme ve bunlardan gelen verileri işleme kolaylığı nedeniyle görsel yaklaşımı incelenecaktır.

Bu makale, yerelleştirme sorununa bir çözüm olarak Görsel Eşzamanlı Yerelleştirme ve Haritalamayı (vSLAM) inceleyecektir. vSLAM, görsel bilgileri kullanarak çevrenin bir haritasını aşamalı olarak oluşturarak ajanı yerelleştirmeyi ve aynı zamanda bu haritayı harita içinde ajanı yerelleştirmek için kullanmayı amaçlamaktadır. SLAM'a yönelik farklı yaklaşımların teorik yönleri üzerine kapsamlı çalışmalar yapılmıştır, ancak pratik uygulamalar genellikle daha kapsamlı ve maliyetli Görsel-IMU sistemlerine göre ayarlanmıştır. Bu makalenin temel amacı, ucuz ve kolaylıkla temin edilebilen hazır kamera sistemleri kullanarak bir vSLAM işlem hattı geliştirmek ve optimum olmayan kamera sistemleri ve ortamlar için yerelleştirme doğruluğunu optimize etmek üzere nasıl değiştirilebileceğine ilişkin teknikleri keşfetmektir.

1.2. Literatür Taraması

Konumlandırma gerçekleştirmek için görsel verilerin kullanılması Görsel Odometri olarak bilinir. Görsel Odometri problemi son birkaç on yılda büyük bir ilerleme kaydetmiştir. Stereo çiftlerin akışının özellik tabanlı karşılaştırmasını kullanan odometri ilk olarak 80'lerde Matthies tarafından ortaya atılmıştır, ancak ilk büyük ölçekli uygulama 2005 yılında NASA'da Y. Cheng ve arkadaşları tarafından Mars Exploration Rover [2] üzerindeki çalışmalarının bir parçası olarak yapılmıştır. Tekerlekler büyük kayalarda ve dik yamaçlarda çekiş gücünü kaybettiginde ortaya çıkan ilk tekerlek odometrisi tabanlı hareket tahminindeki hataları düzeltmek için kamera tabanlı Görsel Odometri kullandılar. Bu

yaklaşımın, engebeli Mars arazisine rağmen gezginin görevlerini yerine getirmesine yardımcı olma konusunda oldukça başarılı olduğu kanıtlandı.



Şekil 1. Opportunity'nin Sol 188'den Sol 191'e kadar olan 19 m'lik sürüşünün görüntüleri. (b)'deki mavi tepegoz yolu, yolunun yalnızca IMU ve tekerlek kodlayıcılarından nasıl tahmin edilebileceğini göstermektedir. Her hücre bir metre kareyi temsil etmektedir. (b)'deki yeşil yol doğru görsel odometri ile güncellenmiş konumu göstermektedir [2].

SLAM'in ilk teorik temelleri 1986 yılında San Francisco'da düzenlenen IEEE Robotik ve Otomasyon Konferansı'nda atılmıştır. Bu konferansta, lokalizasyon problemlerine yönelik tahmin teorik yöntemleri araştıran birçok onde gelen araştırmacı bir araya gelmiş ve bulgularını birbirleriyle tartışmıştır. Sonraki birkaç yıl içinde Smith ve Cheesman [3] ve Durrant-Whyte tarafından robotikte uzamsal belirsizliğin kestiriminin uygulanmasına ilişkin bir dizi önemli makale yayınlandı. Bu çalışma, SLAM problemine ilk çözüm olan EKF-SLAM'ı ortaya çıkarılan Extended Kalman Filtresi yöntemine yol açtı. 2006-08 yılları arasında EKF'yi kullanan ancak doğruluğunu artırmak için üzerine inşa edilen çeşitli FastSLAM yaklaşımıları geliştirilmiştir [4].

2010'ların başında, görüntüler arasında görsel özelliklerin çıkarılmasına ve eşleştirilmesine dayanan özellik tabanlı SLAM yöntemlerinde bir atılım yaşandı. Mur-Artal ve arkadaşları [5], izleme ve haritalama için ORB özelliklerini ve harita optimizasyonu için ana kare tabanlı bir demet ayarlama tekniğini kullanan etkili bir sistem olan ORB-SLAM'ı önerdi. Engel ve arkadaşları [6], çıkarılan özellikler yerine ham görüntü yoğunlukları üzerinde çalışan ve zorlu ortamlarda daha yüksek doğruluk ve sağlamlık elde eden doğrudan bir yöntem olan LSDO'y'u (Büyük Ölçekli Doğrudan Monoküler SLAM) tanıttı.

Son yıllarda derin öğrenme tekniklerinin görsel SLAM'e entegrasyonu önemli gelişmelere yol açmıştır. Bloesch ve arkadaşları [7], özellik algılama ve eşleştirme için derin

öğrenmeyi kullanan ve geleneksel el yapımı özelliklerden daha iyi performans gösteren CodeSLAM'ı tanıtmıştır. Bavle ve arkadaşları [8] tarafından geliştirilen VPS-SLAM gibi anlamsal SLAM yöntemleri, harita tutarlığını ve sağlamlığını iyileştirmek için anlamsal bilgileri de dahil ederek sahneyi anlamak için derin öğrenmeden yararlanmıştır. D. Li ve arkadaşlarının [9] DeepSLAM'i gibi uçtan uca derin öğrenme yaklaşımları ortaya çıkmıştır; burada tüm SLAM işlem hattı doğrudan verilerden öğrenilmektedir.

Şu anda, son teknoloji görsel SLAM sistemleri geleneksel geometrik teknikler ve derin öğrenme bileşenlerinin bir kombinasyonunu içermektedir. Kayda değer örnekler arasında derin öğrenme tabanlı özellik çıkarma ve demet ayarlamayı birleştiren doğrudan bir SLAM sistemi olan DXSLAM [10] ve gelişmiş 3D yeniden yapılandırma ve sahne anlayışı için anlamsal segmentasyondan yararlanan gerçek zamanlı yoğun bir SLAM sistemi olan SemanticFusion [11] bulunmaktadır [12].

1.3. Özgünlük

Bu makale boyunca tartışılan ve ekteki projede uygulanan SLAM algoritmaları ve teknikleri orijinal değildir; son on yılda bunlar titizlikle teorik olarak geliştirilmiş ve uygulamaya konulmuştur. Buna bir örnek olarak, yaklaşımı bu makalenin uygulamasının temelini oluşturan ORB-SLAM verilebilir. Yazar, bu makalenin Görüntü İşleme Araştırmaları dünyasına katkısının, özellikle bu amaçla üretilmemiş genel kameralar için bir vSLAM ardışık düzeninin geliştirilmesinde, sonuçların incelenmesinde ve bu sonuçları iyileştiren tekniklerin ve ardışık düzende yapılan değişikliklerin keşfedilmesinde yattığına inanmaktadır.

1.4. Materyal ve Yöntem

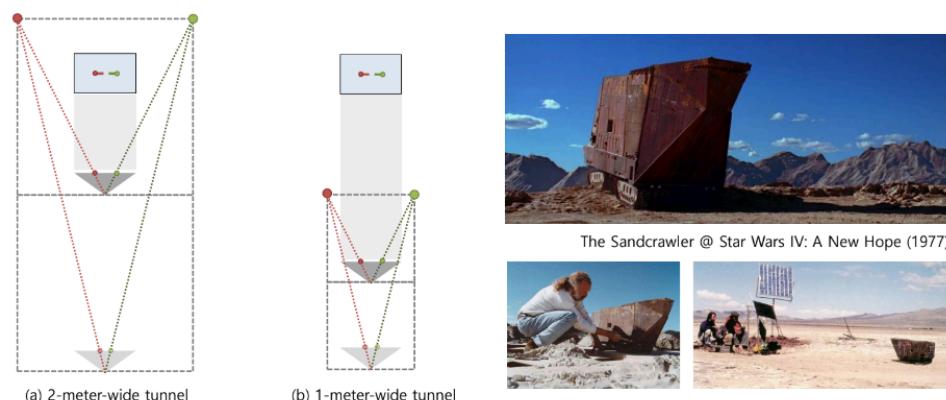
1.4.1. Donanım:

1. İki Trust Trino HD Web Kamerası.



Şekil 2. Monte edilmiş stereo kameramız.

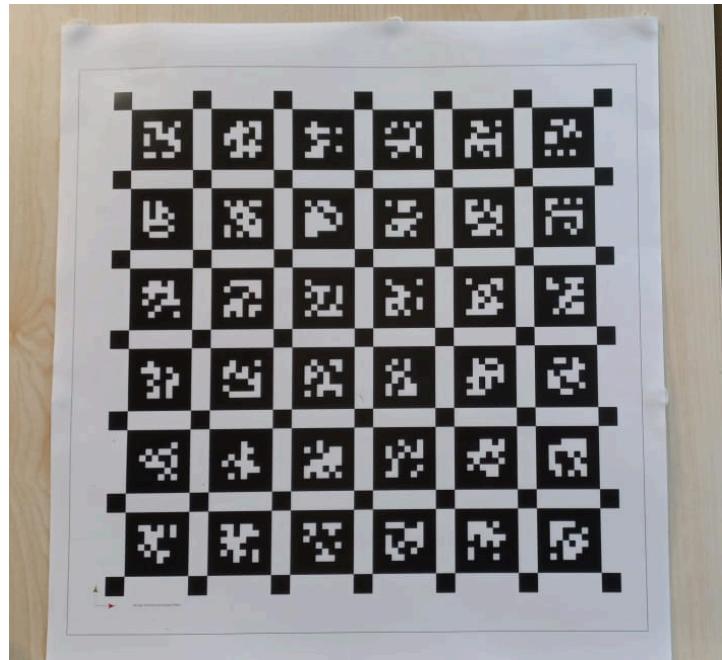
Stereo kamera sisteminin monoküler bir sisteme tercih edilmesinin birkaç nedeni vardı. Stereo kamera sistemleri ölçek belirsizliğinden muzdarip değildir ve bu nedenle derinliği doğru bir şekilde algılamak ve hesaplamak için kullanılabilir.



Şekil 3. Monoküler kamera sistemde derinliği hesaplamasının imkansız olduğunu gösteren bir örnek (Sol). Star Wars film setinde monoküler kamera sistemdeki ölçek belirsizliğin ilginç kullanımı. .

Birbirini tamamlayan görüşler sağlayan iki kamera ile stereo görüş tipik olarak daha doğru özellik eşleştirme ve 3D yeniden yapılandırma sunar. Bu da kamera hareketi ve sahne yapısı hakkında daha hassas tahminler yapılmasını sağlayarak görsel odometrinin genel doğruluğunu artırır. Tüm bunlara ek olarak, birden fazla bakış açısından yakalanan bilgilerdeki fazlalık, oklüzyonlara, dokusuz bölgelere ve zorlu aydınlatma koşullarına karşı sağlanlığı artırır. Monoküler sistemler, tek bir bakış açısına dayandıkları için bu zorluklara karşı daha hassastır.

2. Intel Core i5 10. Nesil işlemci ve Nvidia MX130 GPU içeren HP Probook Dizüstü Bilgisayar.
3. 6x6 36h11 Apriltag kalibrasyon target tahtası Şekil 4'te.



Şekil 4. 6x6 36h11 desenli A3 kağıt üzerinde basılmış Apriltag kalibrasyon tahtası.

1.4.2. Yazılım

1. Ubuntu 22.04.4 LTS.
2. C++ ve Python Programlama Dilleri.
3. Görüntüleri yakalamak ve işlemek için OpenCV.
4. Stereo kamera dışsal ve içsel parametrelerinin kalibrasyonu için ROS ve Matlab ile Kalibr.
5. GPU ile arayüz oluşturmak ve 3D ortamlar oluşturmak için OpenGL'i kullanan Pangolin.
6. Yardımcı SLAM izleme ve haritalama algoritmaları için ORB-SLAM3. ORB-SLAM3'in kendisi, Bag Of Words konum takibi için DBoW2, 3D matematiksel hesaplamalar için Sophus ve grafik optimizasyonu için g2o kullanır[13]

1.4.3. Yöntem

Geliştirilen sistemin etkinliği, birkaç farklı iç ve dış mekândan veri yakalayarak ve sistemin ajanın konumunu ne kadar iyi takip ettiğini değerlendirerek test edilecektir. Veri toplama sırasında kamera aparatının hareketinin doğası da yavaştan hızlıya ve

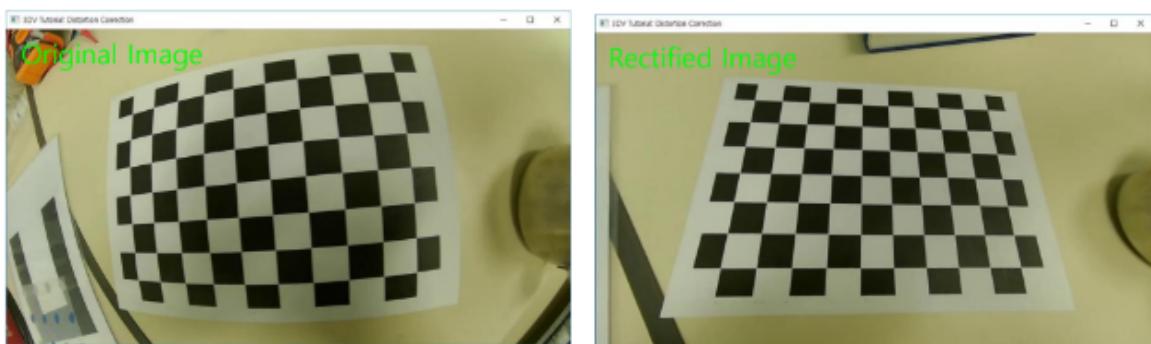
düzensiz/sarsıntılarından sabit olana kadar değiştirilerek bunun lokalizasyon ve haritalama sonuçlarını nasıl etkilediği inceleneciktir. SLAM sistemi, giriş görüntü akışında yeterli özelliklerini izleyemediğinde izleme iş parçacığını durdurur ve bir başarısızlık mesajı yazdırır, bu da bize sistemin başarısı veya başarısızlığı hakkında iyi bir ölçü verir. Sistemin parametreleri daha sonra buna yanıt olarak değiştirilir. Bu parametreler arasında, bunlarla sınırlı olmamakla birlikte, kamerasın giriş çözünürlüğü ve renk derinliği, ORB çıkarıcılarının görüntü başına özellik sayısı, ölçek faktörü, seviye sayısı ve başlangıç ve minimum Hızlı eşik değerleri yer almaktadır.

Testlerden sonra, en iyi keşfedilen teknikler ve parametreler kaydedildi.

2. TEORİK ALTYAPI

2.1. Kamera Kalibrasyon

Kameralar, çekikleri görüntülerin doğruluğunu etkileyen kusurlara sahiptir. İzleme ve Haritalama Konularına doğru gerçek dünya verileri sağlamak için bu parametrelerin görüntü işleme sırasında doğru bir şekilde kalibre edilmesi ve düzeltilmesi gereklidir.



Şekil 5: Kameradan çekilen görüntüler üzerinde distorsyonun etkisi (sol) ve düzeltmesi (sağ). Bu örnek doğru kalibrasyonun ne kadar önemli olduğunu gösterir.

Stereo kamera sisteminin kalibrasyonu için gerçekleştirilen sürec aşağıda detaylandırılmıştır:

- 1. Görüntü alma:** Bilinen bir kalibrasyon nesnesinin görüntülerini birçok farklı açı ve mesafeden çekilir. Bu proje için kullanılan kalibrasyon nesnesi 36h11 6x6 april etiketidir. Her iki kameradan görüntülerini yakalayan ve bunları görüntü veri toplama için kullanılan geleneksel formatta kaydeden bir C++ komut dosyası yazılmıştır. Kalibrasyon nesnesini kurarken ve görüntülerini yakalarken en yüksek doğruluk derecesini sağlamak için ekstra özen gösterilmiştir.

2. **Kalibrasyon modelinin tespiti:** Bu aşama, bilgisayarla görme teknikleri ve kalibrasyon nesnesinin hücrelerinin verilen gerçek dünya ölçümleri kullanılarak her görüntüdeki kalibrasyon modelinin köşelerinin veya noktalarının tam olarak tespit edilmesini ve doğru bir şekilde konumlandırılmasını içerir.
3. **Başlangıç Parametrelerinin Tahmini:** Kalibrasyon sistemi her iki kameranın içsel parametreleri içeren K kamera matrisinin parametreler için bir başlangıç tahmini oluşturur.

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

Odak uzaklığı f_x, f_y ve prensip nokta c_x, c_y ile temsil edilir.

Stereo kamera sistemlerinde dışsal parametreler, yani orijine göre dönme ve öteleme (sol kameranın orijin olduğu varsayılmış) de bu aşamada tahmin edilir.

$$\mathbf{T} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_1 \\ R_{21} & R_{22} & R_{23} & t_2 \\ R_{31} & R_{32} & R_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

R değerler kamera 1'e göre kamera 2'nin 3×3 rotasyon matrisi ve t değerler 3×1 öteleme vektörü temsil etmektedir.

4: **Optimizasyon:** Optimizasyon algoritmaları bu parametreleri iyileştirmek için kullanılır. En yaygın yöntem, reprojection hatasını, yani gözlenen görüntü noktaları ile tahmin edilen parametrelerden hesaplanan yansıtılmış noktalar arasındaki farkı en aza indirmektir. Levenberg-Marquardt algoritması gibi doğrusal olmayan optimizasyon teknikleri optimum bir minimum elde etmek için kullanılır. Kamera tarafından üretilen görüntüleri en doğru şekilde tanımlayan iğne deliği kamera modeli, dolayısıyla yeniden projeksiyon için formül (3) kullanılır:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

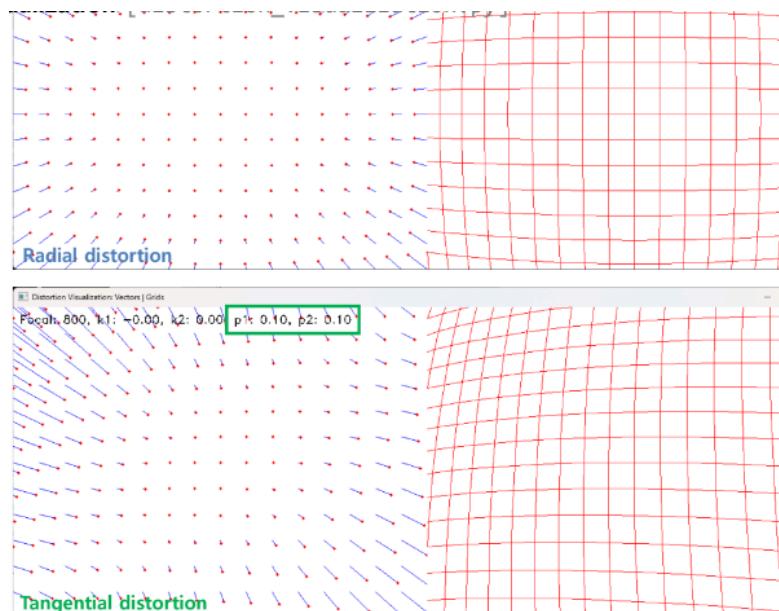
Eğim ve Lens bozulma katsayıları da bu sırada hesaplanır. Lens bozulmaların kamera görüntüyü nasıl etki ettiğini Şekil 6'de gösterilmiştir.

$$\begin{aligned}x_{\text{distorted}} &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\y_{\text{distorted}} &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)\end{aligned}\quad (4)$$

k_1, k_2, k_3 değerleri radyal bozulma katsayıları, r^2 distorsiyon yuvarluğun yarıçapının katsayısı, ve x, y orijinal noktalardır.

$$\begin{aligned}x_{\text{distorted}} &= x + [2p_1 xy + p_2(r^2 + 2x^2)] \\y_{\text{distorted}} &= y + [p_1(r^2 + 2y^2) + 2p_2xy]\end{aligned}\quad (5)$$

p_1, p_2 değerleri tangential bozulma katsayıları, r^2 distorsiyon yuvarluğun yarıçapının katsayısı, ve x, y orijinal noktalardır.



Şekil 6. Radyal distorsonun kamera ile çekilen görüntüler üzerindeki etkisi.

5: Doğrulama: Bilinen 3 boyutlu noktalar tahmini parametreler kullanılarak görüntüye yansıtılır ve tahmin, yansıtılan noktalar gerçek görüntü noktaları ile karşılaştırılarak doğrulanır. Tahminin geçerliliğini değerlendirmek için nihai bir yeniden yansıtma hatası hesaplanır. Stereo kamera çiftimizi kalibre etmek için ROS kullanan açık kaynaklı Kalibr kalibrasyon araç zinciri kullanılmış ve bağımsız doğrulama için Matlab kullanılmıştır.

2.2. SLAM

SLAM işlem hattı birkaç temel aşamadan oluşur: Başlatma, İzleme, Haritalama, Döngü Kapatma, Yer Tanıma.

1. Başlatma: SLAM süreci, tüm görüntü dosyalarının yollarının toplandığı ve saklandığı başlatma aşamasıyla başlar. Kelime Torbası (BoW) özellik tespiti için gerekli olan ORB Kelime dosyası belleğe yüklenir. Üç ana iş parçacığı (İzleme, Yerel Haritalama ve Döngü Kapatma) başlatılır. İlk harita, ORB özelliklerini tespit ederek, kareler arasındaki hareketi tahmin ederek, ilk 3D nokta bulutunu oluşturmak için bir hareket modeli kullanarak ve dünya koordinatlarında bir başlangıç pozu ayarlayarak ilk birkaç kare kullanılarak oluşturulur.

2. İzleme: İzleme iş parçacığı gerçek zamanlı poz tahmini için çok önemlidir ve birkaç alt görev içerir:

Feature Algılama ve Mapping

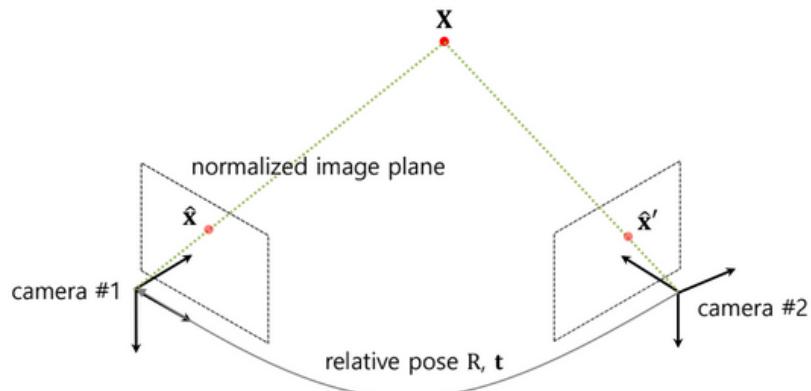
ORB feature tespiti için FAST (Hızlandırılmış Segment Testinden Özellikler) ve descriptor hesaplaması için BRIEF (İkili Sağlam Bağımsız Temel Özellikler) kullanılarak gelen karede tespit edilir. Tespit edilen özellikler, yazışmalar oluşturmak için mevcut harita noktalarıyla eşleştirilir[14].

Poz Tahmini

Kamera pozu, sağlamlık için RANSAC (Random Sample Consensus) ile geliştirilmiş bir Perspective-n-Point (PnP) çözümü kullanılarak tahmin edilir. PnP problemi aşağıdaki gibi formül (6) ile elde edilebilir

$$\mathbf{u} = \mathbf{K}[\mathbf{R}|\mathbf{t}]\mathbf{X} \quad (6)$$

Burada \mathbf{u} bir 3D harita noktası \mathbf{X} 'nin 2D izdüşümünü, \mathbf{K} kamera iç matrisini, \mathbf{R} döndürme matrisini ve \mathbf{t} öteleme vektörünü temsil eder.



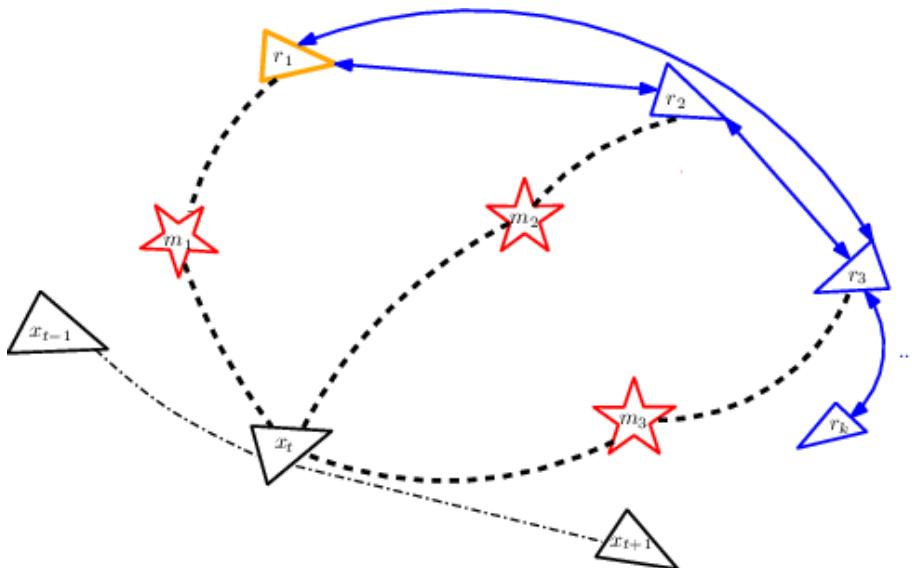
Şekil 7. Stereo kamera sisteminde imaj düzeyinde x ve x' noktalarında olan noktanın gerçek dünya 3D konumu bulma görseli

Yerel Bundle Adjustment

Yerel Bundle Adjustment (BA), yerel bir ana kare penceresi içinde yeniden yansıtma hatasını en aza indirerek poz ve harita noktalarını iyileştirmek için gerçekleştirilir [15]. Görsel anlatım Şekil 8. de verilmiştir. BA için amaç fonksiyonu şöyledir:

$$\min_{\mathbf{T}_i, \mathbf{X}_j} \sum_{i,j} \| \mathbf{u}_{ij} - \pi(\mathbf{T}_i, \mathbf{X}_j) \|^2 \quad (7)$$

Burada π projeksiyon fonksiyonu, T_i ana karesinin pozu ve X_j 3D harita noktasıdır.



Şekil 8. SLAM'de Yerel Paket Ayarlamasının Gösterimi. Bu şekil, yakındaki çerçevelerin $(r_1, r_2, r_3, \dots, r_k)$ ve gözlemlenen yer işaretlerinin (m_1, m_2, m_3) pozlarını optimize ederek aracın tahmini küresel konumunu (x_t) ayarlama sürecini göstermektedir. Kesikli çizgiler gözlemleri, düz çizgiler ise aracın yörüngesini temsil etmektedir. Ayarlama, aracın yörüngesindeki ve yer işaret konumlarındaki genel hatayı en aza indirmeyi amaçlamaktadır.

Yeniden konumlandırma

İzleme kaybedilirse, yer tanıma modülü mevcut kareyi daha önce ziyaret edilen yerlerle eşleştirir ve benzer BoW vektörlerine sahip ana kareleri tanımlayarak yeniden konumlandırmayı sağlar.

3. **Eşleme:** Yerel Haritalama iş parçası, geçerli kamera pozu etrafındaki yerel haritayı yönetir. Önemli hareket veya sahne değişikliklerine göre haritaya yeni anahtar kareler eklenir.

Bu, haritanın yeni bilgilerle gelişmesini sağlar. Bu, yeni harita noktaları oluşturmayı, gereksiz noktaları ayıklamayı ve pozları ve harita noktalarını iyileştirmek için yerel harita üzerinde yerel BA gerçekleştirmeyi içerir.

Düğümlerin ana kareleri, kenarların ise paylaşılan harita noktalarını temsil ettiği bir eşgörülebilirlik grafiği tutulur. Bu grafik verimli yerel eşleme ve optimizasyonu kolaylaştırır.

4. Döngü Kapatma

Döngü kapatma, zaman içinde biriken sapmayı düzeltmek ve küresel harita tutarlığını sağlamak için gereklidir.

Algılama

Döngü kapanışları, BoW vektörleri kullanılarak mevcut ana karelerin daha önce ziyaret edilen ana karelerle eşleştirilmesiyle tespit edilir. Doğrudan özellik eşleştirme, döngü kapatma adaylarını doğrular.

Poz Grafiği Optimizasyonu

Poz grafiği optimizasyonu, sapmayı düzeltmek ve global tutarlılığı sağlamak için ana karelerin ve harita noktalarının pozlarını ayarlar. Optimizasyon problemi şu şekilde formüle edilir:

$$\min_{\mathbf{T}_i} \sum_{(i,j) \in \mathcal{E}} \left\| \mathbf{T}_j^{-1} \mathbf{T}_i \mathbf{T}_{ij}^{-1} \right\|^2 \quad (8)$$

Burada E döngü kapanışlarını temsil eden kenarlar kümesidir ve T_{ij} i ve j ana kareleri arasındaki görelî dönüşümüdür.

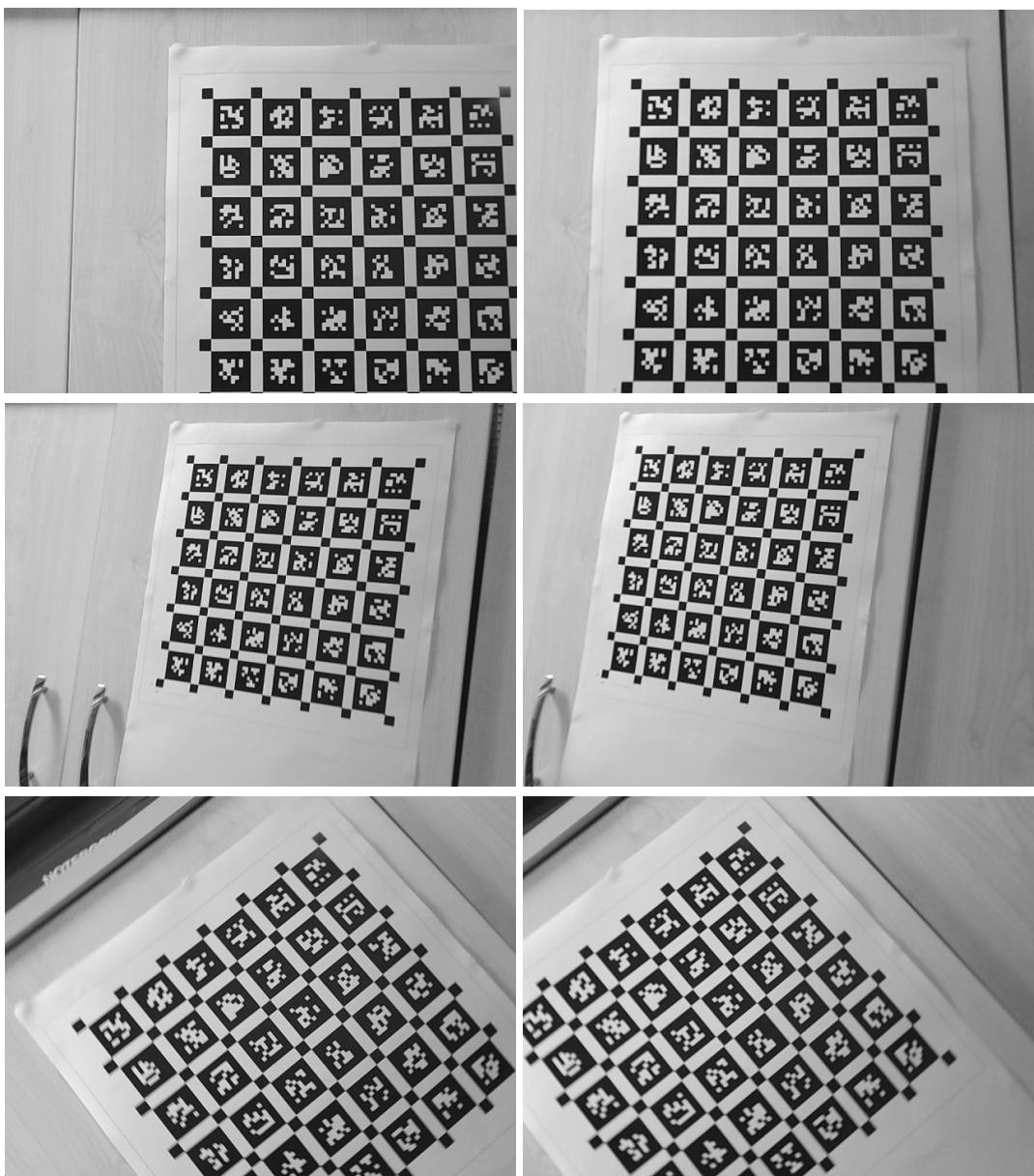
5. Yer Tanıma

Yer tanıma, BoW ve doğrudan özellik eşleştirme kullanarak verimli döngü tespiti ve yeniden konumlandırma sağlar. Doğru poz tahminini sağlamak için bir döngü kapanması veya yeniden konumlandırma tespit edildiğinde doğrudan özellik eşleştirme gerçekleştirilir.

3. Deneysel Çalışmalar

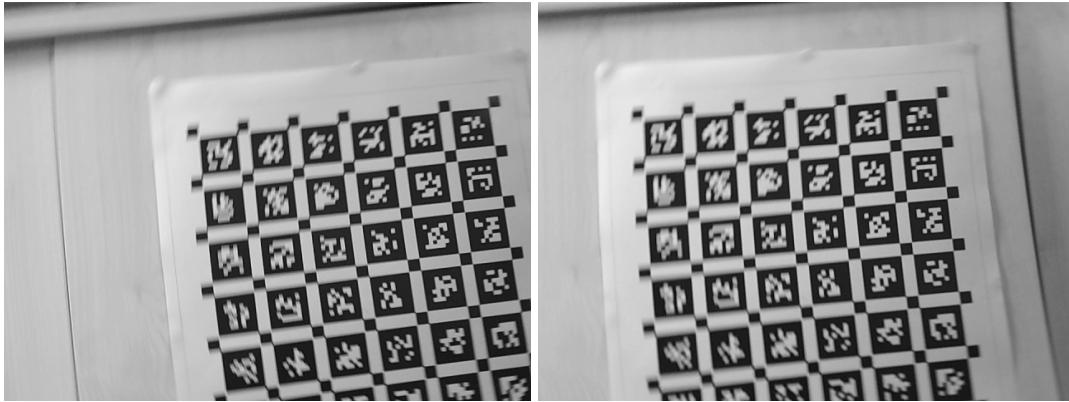
3.1. Kamera Kalibrasyon ve Kurumu

Kamera sistemimizin sınırlamalarını aşmak için, optimum kalibrasyon ve veri toplamayı sağlamak üzere çeşitli stratejiler uygulandı. İlk olarak, dama tahtası deseni yerine 36h11 6x6 AprilTag deseni kullanıldı ve hedef mümkün olan en büyük boyutta basıldı. Bu yaklaşım, düşük çözünürlüklü kameralarımızın desendeki ayrıntıları daha kolay tespit etmesini sağladığı için kalibrasyonun doğruluğunu korumak açısından çok önemliydi.



Şekil 9. Kalibrasyon için stereo kameramızdan çekilen örnek çift kareler.

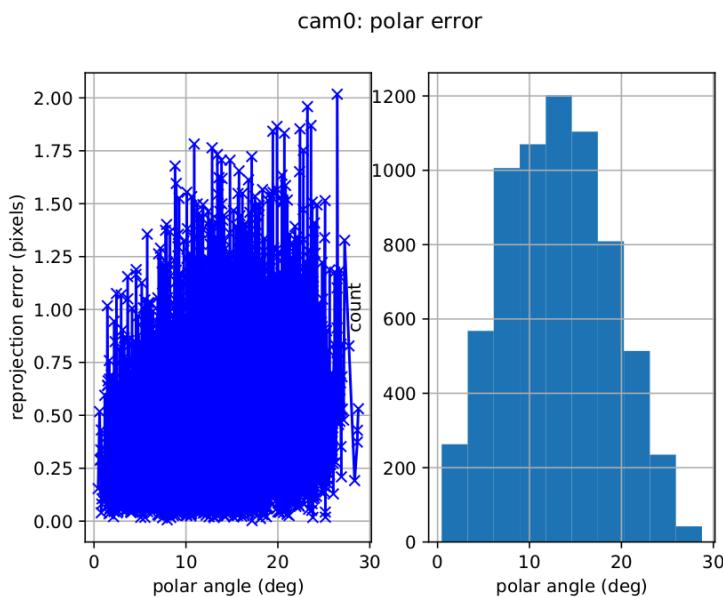
Desenin düz ve aynı seviyede kalmasını sağlamak için de büyük çaba harcandı. Ek olarak, kamera hareketi dikkatle kontrol edildi; hareket bulanıklığını önlemek için ani hareketlerden kaçınarak kameraları yavaş ve istikrarlı bir şekilde hareket ettirdik.

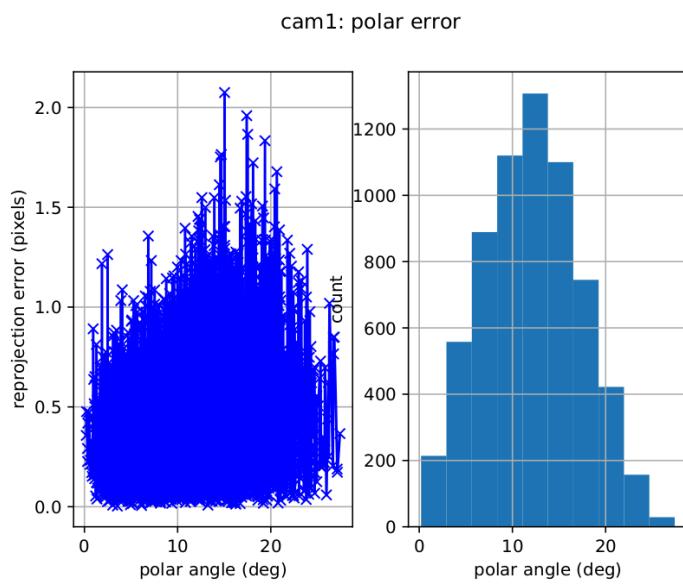


Şekil 10. Hareket bulanıklığından dolayı kalibrasyon için kabul edemez çift kare. Bu tür kareler elendi.

Hedef, tamamen çerçeveye içinde tutulurken çeşitli mesafelerden ve yönlerden çekildi, bu da kalibrasyon için çeşitli veri noktaları toplamamızı olanak sağladı. Ayrıca, yakalanan görüntü akışının sıklığı, aradaki karelerin işlemeyi gereksiz yere yavaşlatmasını önlemek için azaltıldı. İki kamerasının dönüş ve ötelemesinin deneyler boyunca sabit kalmasını sağlamak, verilerin bütünlüğünü korumak için çok önemliydi. Yeterli aydınlatma ve iki kamera tarafından çekilen görüntüler arasındaki gecikmenin en aza indirilmesi de sistemin performansını artırmak için önceliklendirildi.

Kalibrasyon sonuç değerlerimiz aşağıdaki Şekillerde verilmiştir. Hata değerleri en aza indirmek için kalibrasyon defalarca yapıldı ortalama hatası 0.5'e düşmesine kadar.





Şekil 11. Kalibrasyon sonucunda kameralarımızın odak noktaya bakarken polar dereceye göre piksel bazlı reprojection hatası (Yukarı).



Şekil 12. Toplanan veriden örnek kare çift (Yukarı). Düzeltip birleştirilmiş sonuç görüntü (Aşağı).

3.2. Veri Toplama

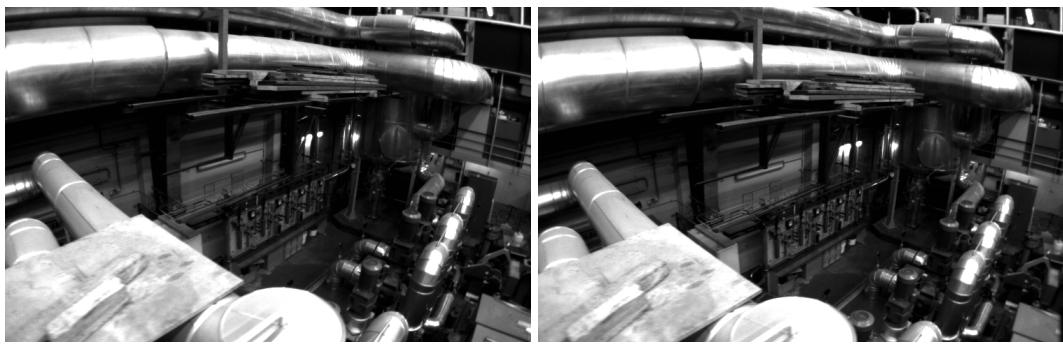
3.2.1 Süreç

Öncelikle geliştirilen vSLAM sistemi ETH Zürich tarafından derlenen EuroC veri seti [16] kullanılarak test edilmiş ve zahmetli bir hata düzeltme sürecinin ardından beklenen yerelleştirme ve eşleme sonuçlarına ulaşılmıştır. Bu durum, yerelleştirmenin diğer vSLAM uygulamalarıyla karşılaşırmasıyla tespit edilmiştir. Daha sonra özel veri kümemiz için veri toplanmaya başlandı. Sistemimizin sağlamlığını test etmek için her biri farklı özelliklere sahip çeşitli iç ve dış ortamlarda veri toplanmıştır. İç ortamlar, yurdumuzdaki çalışma odasını, mobilya ve duvarlar gibi belirgin özelliklere sahip iyi aydınlatılmış odaları ve çalışma odasının dışındaki iyi aydınlatılmamış ve her yerinde tutarlı donuk özelliklere sahip koridoru içermektedir. Yurdumuzun önündeki caddenin güneşli ve gölgeli bölgelerde dış ortamlardaki veri seti toplandı.. Ortamlardaki bu çeşitlilik, sistemin farklı aydınlatma koşulları ve özellik yoğunlukları altındaki performansını değerlendirmemize yardımcı oldu. Kamera hareketinin doğasını değiştirdik, hem yavaş ve kontrollü hareketleri hem de daha hızlı, daha düzensiz hareketleri denedik. Bu çeşitlilik, farklı hareket türlerinin sistemin kamerası takip etme ve konumlandırma becerisini nasıl etkilediğini değerlendirmemizi sağladı.

3.2.1 Veri Setler

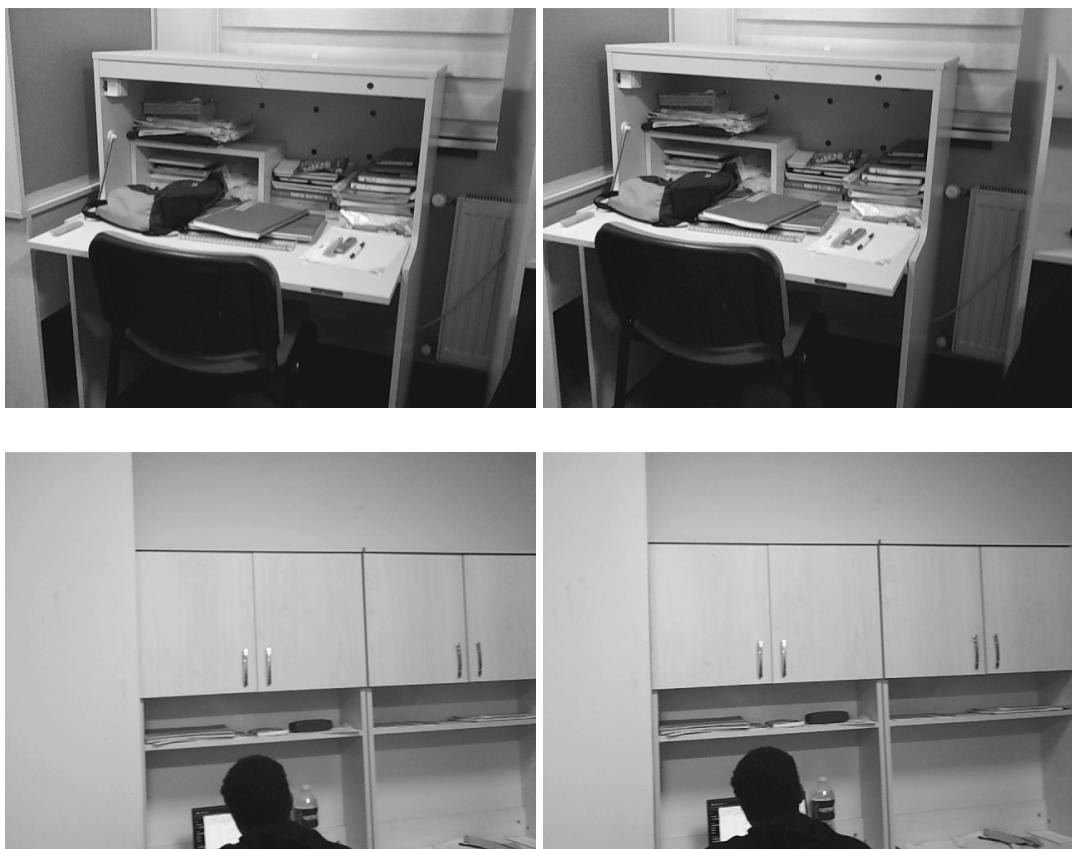
1. EuroC_MH_01: iç mekan, görüş alanında çok sayıda yakın mesafe nesnesi bulunan geniş ortam, parlak ve tutarlı ışık, çeşitli özellik aralığı. (harici veri kümesi) [16]





Şekil 13. EuroC Makine Odası veri kümelerinden kare çiftler.

2. study_room: iç mekan, orta büyüklükte ortam, parlak ve tutarlı aydınlatma koşulları, çeşitli özellik aralığı.



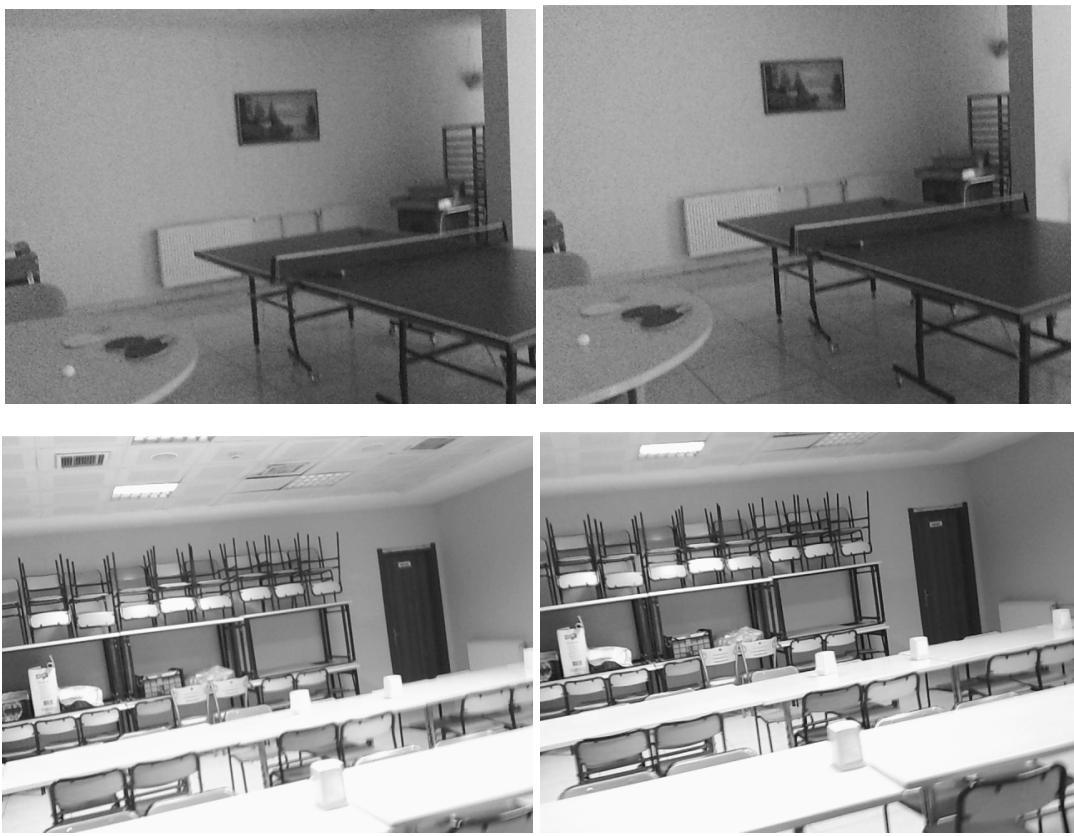
Şekil 14. Çektiğimiz Çalışma Odası veri kümelerinden kare çiftler.

3. corridor: iç mekan, uzun, dar ortam, düşük ve tutarlı aydınlatma koşulları, donuk özellik aralığı.



Şekil 15. Çektiğimiz Koridor veri kümesinden kare çiftler.

4. basement: iç mekan, geniş ortam, düşük ve tutarsız aydınlatma koşulları, çeşitli özellik aralığı.



Şekil 16. Bodrum Kat/Yemekhane veri kümesinden kare çiftler

5. ground: dış mekan, geniş ortam, yüksek ve tutarlı aydınlatma koşulları, çeşitli özellik aralığı.

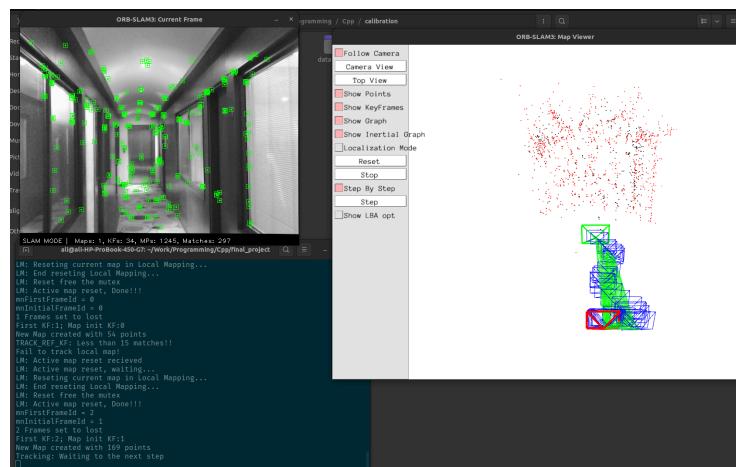


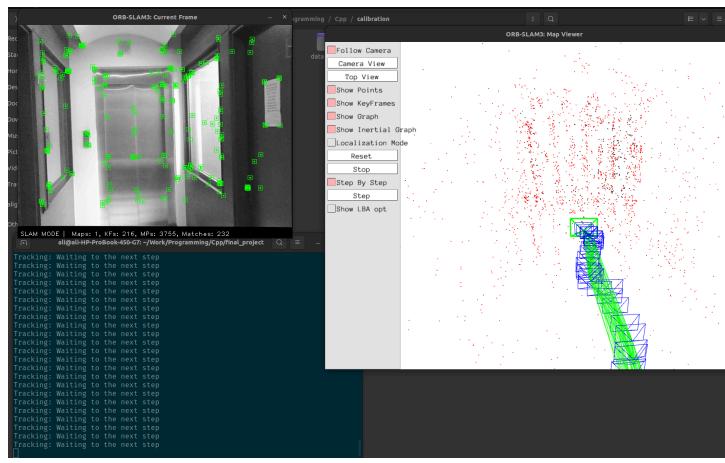


Sekil 17. Yurt önü veri kümelerinden kare çiftler

3.3. Performans Analizi

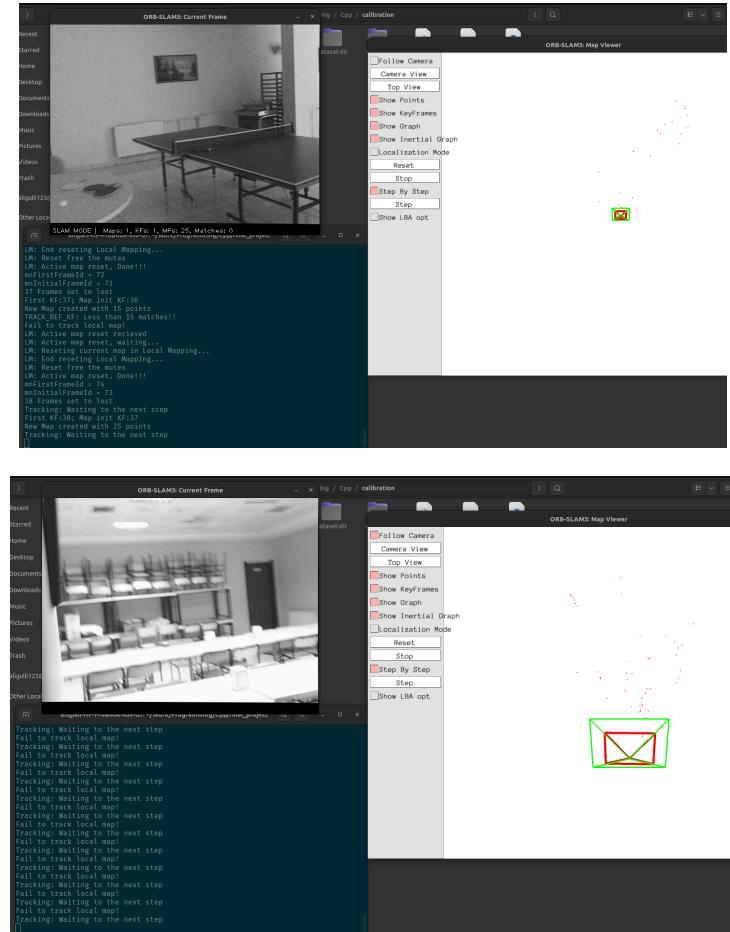
Kamera sisteminin performansı, çeşitli koşullar altında izleme ve yerelleştirme yeteneklerine dayalı olarak kapsamlı bir şekilde analiz edilmiştir. Sistemin en iyi performansı ayırt edici özelliklere sahip iyi aydınlatılmış ortamlarda gösterdiği tespit edilmiştir. Bununla birlikte, koridor veri seti gibi ideal olmayan aydınlatma ortamlarında bile sistemin aynı doğruluk seviyesinde olmasa da özellikleri izleyebildiğini görmek bizi şaşırttı. Genellikle 150-200 arasında feature tanımlama yapar bu veri set üzerinde. Gördüğünüz gibi Şekil 18. deki düşük ışık seviyeli ortamda bile başarılı bir şekilde konumlandırma yapılmaktadır.





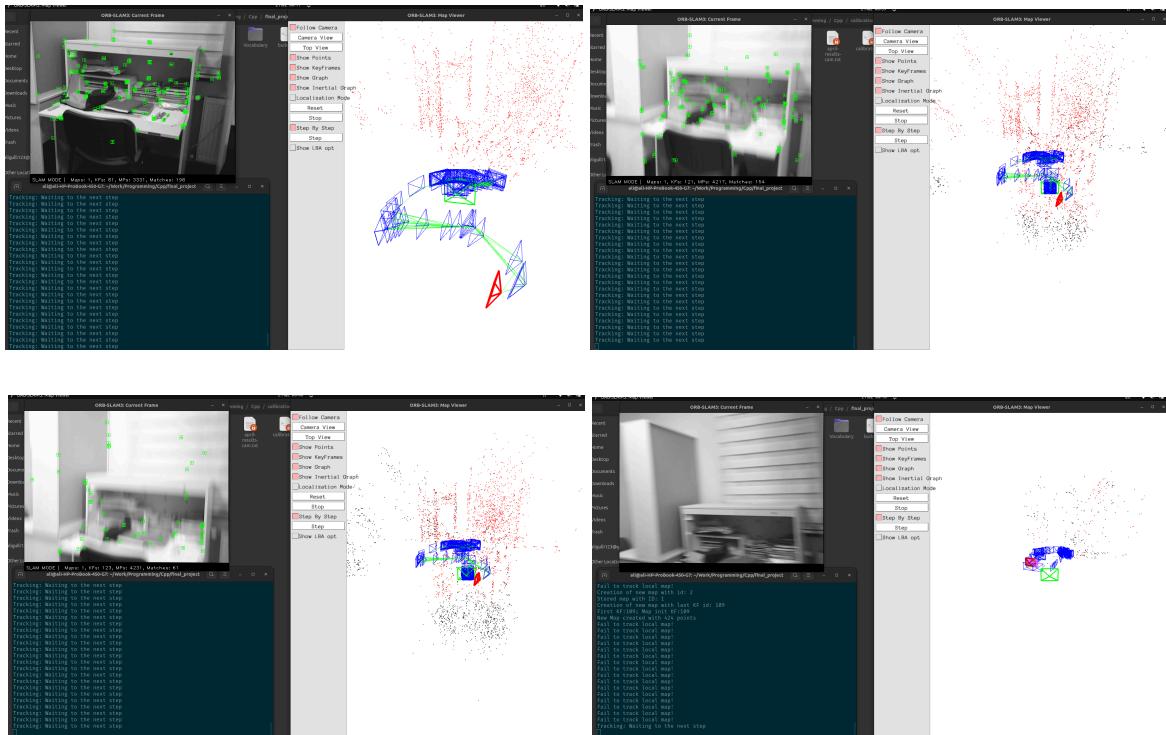
Şekil 18. Soldaki pencere Tespit edilmiş feature'leri göstermektedir. Sağdaki pencere 3D ortam çıktısı olarak kameranın konum, rotayı çizgiler olarak ve çevreyi noktalı bulut olarak göstermektedir.

Düşük ışık koşullarında doğruluk azalsa da sistem yine de çalışmayı başarmıştır. Bodrum veri setiyle karşılaştırıldığında (Şekil 19.), belirleyici farkın ışık seviyesi tutarlılığı olduğu görülmektedir.



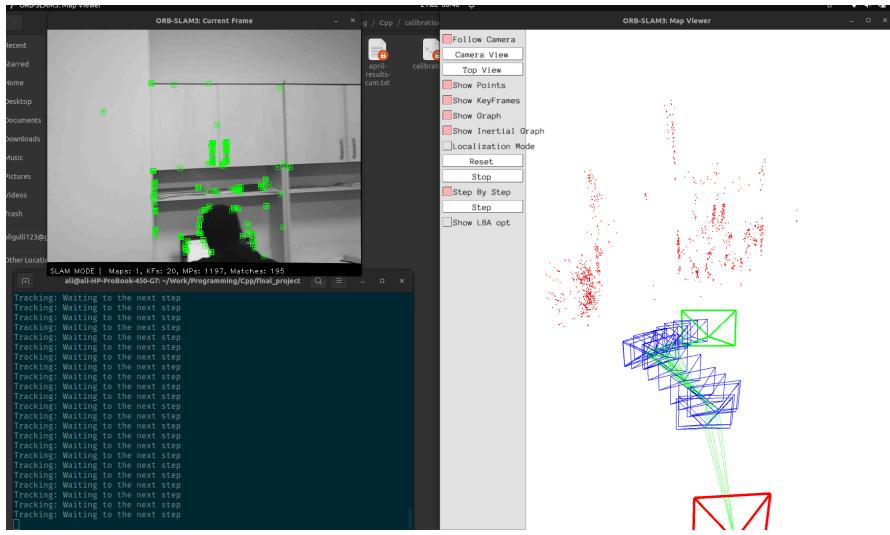
Şekil 19. Bodrum veri setinde ışık seviyesi tutarlı olmadığından sistem doğru bir feature algılama, izleme ve eşleşme süreç işlememektedir.

Her veri set için feature tanıma başarısında, kamera hareketinin niteliği, izleme başarısını etkileyen en önemli faktördü. Yumuşak, yavaş hareketler daha iyi takip ve lokalizasyonla sonuçlanırken, sarsıntılı veya hızlı hareketler hareket bulanıklığına ve özellik algılama başarısızlıklarına neden oldu. Sistem Şekil 20. de olduğu gibi hareket bulanıklığı ve fokus dışı olma durumunda da başarılı bir şekilde tracking ile devam edebilirdi. Ancak uzun süre bu devam ettiğinde feature başarı ile tanımlayamadığından tracking'i bırakıp yeni harita ile başlardı.



Şekil 20. (Yukarı soldan aşağı sağa) Fazla hızlı hareketinden kaynaklı bulanıklığı devam ettiği sürec sistem başarıyla konumlandırma yapamamaktadır.

Sistem, test ettiğimiz çoğu durumlarda başarılı bir konumlandırma yapabildi. Ancak yukarıda söz konusu olan durumlarda başarısız ya da tamamen doğru olmayan izleme sonuçları verilmektedir. Bu sorunlar kullandığımız kameradan kaynaklı. Sebepler ve olası çözümler ilerideki bölümlerde inceleneciktir.



Şekil 21. Başarılı izleme örneği.

3.4. Zorluklar ve Sınırlamalar

Proje sırasında çeşitli zorluklar ve sınırlamalarla karşılaşıldı. Kameramızın yavaş otomatik odaklılama, düşük deklanşör hızı ve ışiktaki hızlı değişimlere karşı hassasiyeti gibi bazı özellikleri hareket bulanıklığı gibi yapaylıklara yol açmıştır. Çok sayıda ardışık karenin bu artefaktları içerdiği veri kümelerinde, SLAM sistemi takibini sürdüremedi ve haritasını yeniden başlatmak zorunda kaldı. Kamerayı yalnızca elle hareket ettirebildiğimiz için bu durum veri toplamayı zorlaştırdı. Bu durumun bir örneği Şekil 22'de gösterilmiştir.





Sekil 22. corridor ve study_room veri setleri işlerken kameranın fazla hızlı dönmesi nedeniyle konumlandırma durdurup yeni harita açılmaktadır.

Kablosuz olmayan web kameralarının kullanılması, kameraların bir bilgisayara bağlanması gerektiğinden hareket alanını önemli ölçüde kısıtladı. Bu sınırlama, bilgisayarın kameralarla birlikte taşınmasını gerektirmekte ki bu da uygulama zorluğu arttırdı. Intel RealSense gibi özel kameralara kıyasla hazır web kameralarının kalibrasyonu önemli zorluklar içermektedir. Ayrıca, sistemimizde bir IMU'nun integre edilmemesi veri doğruluğunu daha da azalttı.

4. SONUÇLAR

4.1. Genel Açıklamalar

Sistemin poz tahminindeki doğruluğu deneyisel sonuçlarımızdan anlaşılmaktadır. Sistem, kameralanın yönelim ve hareket yönünü doğru bir şekilde tespit etme kabiliyetini göstermiştir. Bu doğruluk, kalibrasyon ve veri toplama sırasında ideal olmayan koşullara rağmen elde edilmiştir. Kullandığımız yöntem, farklı ortamlarda yüksek doğruluğun gerekli olduğu, ancak bütçe ve boyut kısıtlamalarının yüksek güçlü donanım kullanımını engellediği düşük ölçekli robotik uygulamalar için oldukça faydalı olduğunu kanıtlamaktadır.

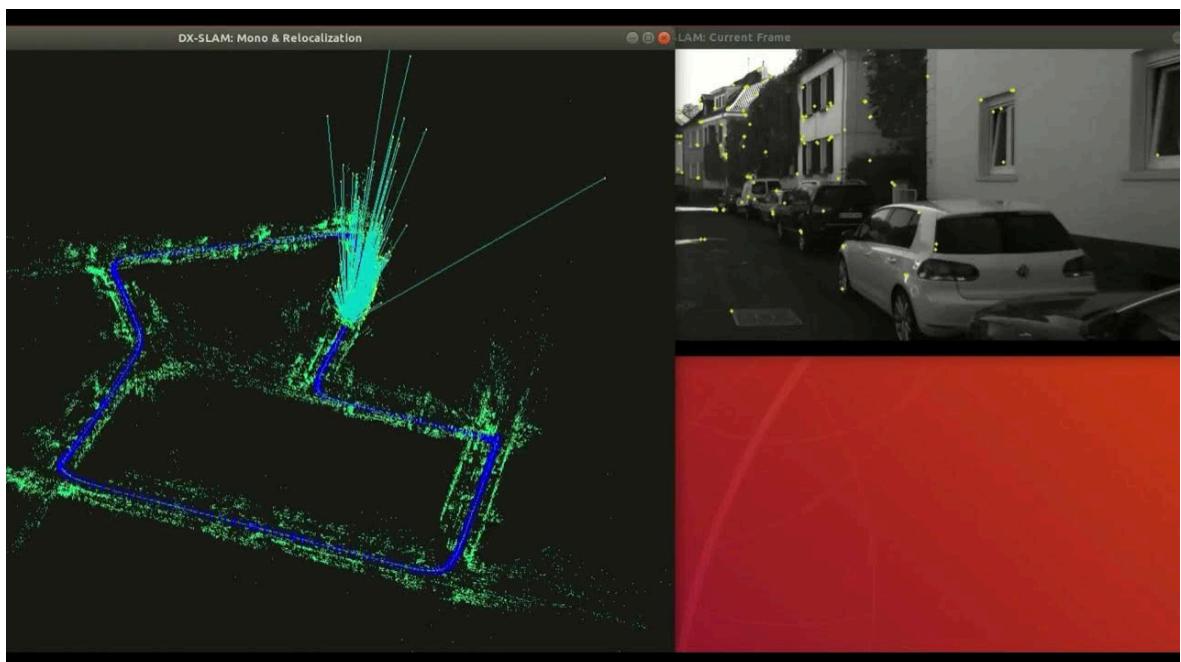
4.2. Gelecek İyileştirmeler

Bu projede kullanılan kamera sistemi ilk veri toplama ve algoritma geliştirmeye izin verirken, gelecekteki yinelemelerde ele alınması gereken birkaç sınırlama vardır. Mevcut kameralanın enstantane etkisi, otomatik odaklılama davranışları ve düşük enstantane hızı, onu yüksek doğruluk gerektiren genel amaçlı uygulamalar için uygunsuz hale getirmiştir.

Doğruluğu ve çok yönlülüğü artırmak için ileride global deklanşörlü, sabit odaklı ve yüksek deklanşör hızı özelliklerine sahip bir kamera kullanılmalıdır. Kamera ayrıca kablosuz olmalı veya SLAM algoritmasını yürüten ekli bir işlem birimine sahip olmalıdır. Bunun nedeni, mevcut kablolu sistemin kameranın hareketini büyük ölçüde kısıtlamasıdır.

Ayrıca, mevcut çevrimdışı uygulama gerçek zamanlı performans açısından sınırlamalara sahiptir. Bir sonraki büyük yükseltme, SLAM sisteminin gerçek zamanlı çalışmasını sağlamaya odaklanacaktır. Bu, yeni uygulama alanlarının ve kullanım durumlarının kilidini açacaktır.

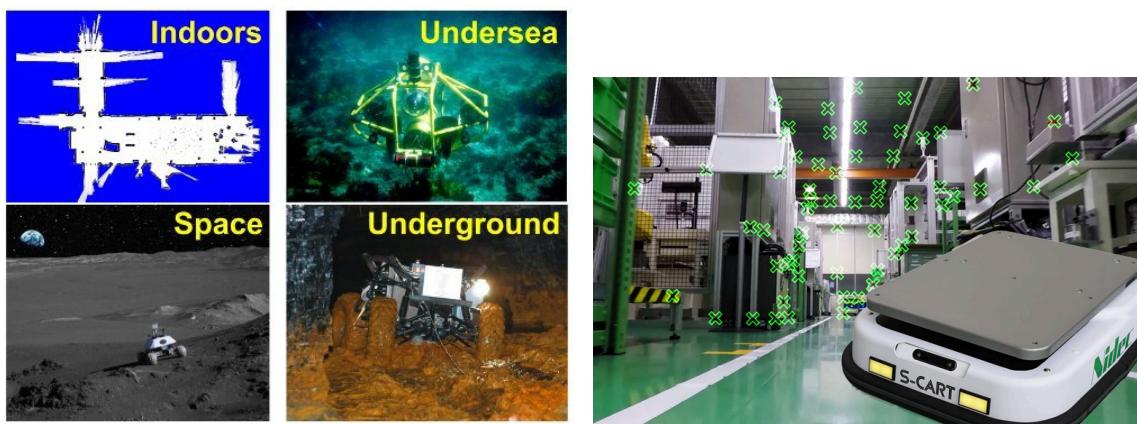
Belki de planlanan en önemli geliştirme, Li D. ve arkadaşlarının [10] çalışmalarını takiben, konvolüsyonel sinir ağları (CNN) tabanlı bir özellik algılama yaklaşımının entegrasyonudur. Önceden eğitilmiş CNN modellerinden yararlanmak, ideal olmayan çevresel koşullarda bile daha doğru ve sağlam özellik algılama sağlayacaktır. Şu anda kullanılan kelime torbası modeliyle karşılaştırıldığında, CNN'ler üstün doğruluk ve verimlilik sunmaktadır. Ayrıca, Li ve arkadaşlarının Intel OpenVINO uygulaması, modern CPU SIMD komut setlerinden tam olarak yararlanarak GPU hızlandırmaya güvenmeden gerçek zamanlı performans sağlar. Bu CNN tabanlı işlem hattının benimsenmesi, SLAM sisteminin doğruluğunu, verimliliğini ve çok yönlülüğünü önemli ölçüde artıracaktır.



Şekil 23. DxSLAM tanıtımından bir kare. vSLAM'in CNN destekli özellik algılama yaklaşımı, çok çeşitli ortamlarda önemli kareleri bulmada çok daha iyidir ve zayıf aydınlatma koşulları ve bulanık girdi gibi zorluklar karşısında daha sağlamdır [10]. Bu avantajlar sistemimizin etkinliğini ve çok yönlülüğünü büyük ölçüde artıracaktır.

4.3. Yaygın Etki

Sistemimiz ani hareket değişikliklerine karşı oldukça hassastır. Bu nedenle, AR veya VR gibi insanlar tarafından kullanılmayı içeren SLAM kullanım durumları için çok uygun olmayabilir. Ancak, yavaş veya sürekli hareketin gerekli olduğu robotik uygulamalarda oldukça etkili olabilir. Robotik ve tüketici uygulamaları için düşük maliyetli vSLAM, kaynak kısıtlı platformlarda doğru algılama, navigasyon ve haritalama sağlayabilir (Şekil 24.). Bu da ev işlerinde yardım için daha akıllı tüketici robotları, sürükleyici için uygun fiyatlı ticari dronlara olanak sağlar. Bilimsel araştırmalarda, çevresel izleme, tehlikeli veya uzak alanların otonom 3D haritalanması ve ekoloji ve jeoloji çalışmaları gibi uygulamalar için yerinde veri toplama için uygun fiyatlı mobil robotik platformlar sağlayacaktır. Genel olarak, bu teknolojinin demokratikleştirilmesi akademi, endüstri ve tüketici sektörlerinde inovasyon hızlandırır.



Şekil 24. Düşük maliyetli SLAM için olası uygulamalar.

5. KAYNAKLAR

- [1] GPS.gov, “GPS.gov: GPS Accuracy,” *Gps.gov*, 2019.
<https://www.gps.gov/systems/gps/performance/accuracy/> (accessed May 27, 2024).
- [2] Y. Cheng, M. W. Maimone, and L. Matthies, “Visual odometry on the Mars exploration rovers - a tool to ensure accurate driving and science imaging,” *IEEE Robotics & Automation Magazine*, vol. 13, no. 2, pp. 54–62, Jun. 2006, doi:
<https://doi.org/10.1109/mra.2006.1638016>.
- [3] R. Smith, M. Self, and P. Cheeseman, “A stochastic map for uncertain spatial relationships,” pp. 467–474, May 1988.
- [4] C. Kim, R. Sakthivel, and W. K. Chung, “Unscented FastSLAM: A Robust and Efficient Solution to the SLAM Problem,” *IEEE Transactions on Robotics*, vol. 24, no. 4, pp. 808–820, Aug. 2008, doi: <https://doi.org/10.1109/tro.2008.924946>.
- [5] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System,” *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, Oct. 2015, doi: <https://doi.org/10.1109/tro.2015.2463671>.
- [6] J. Engel, V. Koltun, and D. Cremers, “Direct Sparse Odometry,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 3, pp. 611–625, Mar. 2018, doi: <https://doi.org/10.1109/tpami.2017.2658577>.
- [7] M. Bloesch, J. Czarnowski, R. Clark, S. Leutenegger, and A. J. Davison, “CodeSLAM - Learning a Compact, Optimisable Representation for Dense Visual SLAM,” *Spiral (Imperial College London)*, Jun. 2018, doi:
<https://doi.org/10.1109/cvpr.2018.00271>.
- [8] Hriday Bavle, Paloma, J. P. How, and P. Campoy, “VPS-SLAM: Visual Planar Semantic SLAM for Aerial Robotic Systems,” *IEEE Access*, vol. 8, pp. 60704–60718, Jan. 2020, doi: <https://doi.org/10.1109/access.2020.2983121>.
- [9] R. Li, S. Wang, and D. Gu, “DeepSLAM: A Robust Monocular SLAM System With Unsupervised Deep Learning,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3577–3587, Apr. 2021, doi: <https://doi.org/10.1109/TIE.2020.2982096>.
- [10] D. Li *et al.*, “DXSLAM: A Robust and Efficient Visual SLAM System with Deep Features,” *IEEE Xplore*, Oct. 01, 2020.
https://ieeexplore.ieee.org/abstract/document/9340907?casa_token=XI_fyHjZV7oAAAAAA:T0XqJ8G1SzI09Gx7Hik5lm_LsGUhfTxPZXyNgQaZfwzl9i1znWjNySYlzuWqNVHpl5dVPM2eG8_W (accessed Dec. 26, 2022).

- [11] J. E. McCORMAC, A. Handa, A. J. Davison, and S. Leutenegger, “SemanticFusion: Dense 3D semantic mapping with convolutional neural networks,” *2017 IEEE International Conference on Robotics and Automation*, Jul. 2017, doi: <https://doi.org/10.1109/icra.2017.7989538>.
- [12] A. Tourani, H. Bavle, J. L. Sanchez-Lopez, and H. Voos, “Visual SLAM: What Are the Current Trends and What to Expect?,” *Sensors*, vol. 22, no. 23, p. 9297, Nov. 2022, doi: <https://doi.org/10.3390/s22239297>.
- [13] C. Campos, R. Elvira, J. J. G. Rodriguez, J. M. M. Montiel, and J. D. Tardos, “ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM,” *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1–17, 2021, doi: <https://doi.org/10.1109/tro.2021.3075644>.
- [14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “ORB: An efficient alternative to SIFT or SURF,” *IEEE Xplore*, Nov. 01, 2011.
<https://ieeexplore.ieee.org/document/6126544> (accessed Apr. 24, 2024).
- [15] L. Yu, C. Joly, G. Bresson, and Fabien Moutarde, “Monocular urban localization using street view,” *HAL (Le Centre pour la Communication Scientifique Directe)*, Nov. 2016, doi: <https://doi.org/10.1109/icarcv.2016.7838744>.
- [16] M. Burri *et al.*, “The EuRoC micro aerial vehicle datasets,” *The International Journal of Robotics Research*, vol. 35, no. 10, pp. 1157–1163, Jan. 2016, doi: <https://doi.org/10.1177/0278364915620033>.