

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DOMAIN ADAPTIVE VISUAL INERTIAL
GLOBAL LOCALIZATION IN GNSS-DENIED ENVIRONMENTS**



M.Sc. THESIS

Abdurbaki ŞANLAN

Department of Aeronautics and Astronautics Engineering

Aeronautics and Astronautics Engineering Programme

SEPTEMBER 2025

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DOMAIN ADAPTIVE VISUAL INERTIAL
GLOBAL LOCALIZATION IN GNSS-DENIED ENVIRONMENTS**



M.Sc. THESIS

**Abdulkaki ŞANLAN
(511241124)**

Department of Aeronautics and Astronautics Engineering

Aeronautics and Astronautics Engineering Programme

Thesis Advisor: Assoc. Prof. Dr. Emre KOYUNCU

SEPTEMBER 2025

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**GNSS OLMAYAN ORTAMLARDA GÖRÜNTÜ UYARLAMALI
GÖRSEL-ATALETSAL KÜRESEL KONUM BELİRLEME**

YÜKSEK LİSANS TEZİ

**Abdulkaki ŞANLAN
(511241124)**

Uçak ve Uzay Mühendisliği Anabilim Dalı

Uçak ve Uzay Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Emre KOYUNCU

EYLÜL 2025

Abdulkaki ŞANLAN, a M.Sc. student of ITU Graduate School student ID 511241124 successfully defended the thesis entitled “DOMAIN ADAPTIVE VISUAL INERTIAL GLOBAL LOCALIZATION IN GNSS-DENIED ENVIRONMENTS”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Dr. Emre KOYUNCU**
Istanbul Technical University

Jury Members : **Prof. Dr. İbrahim Özkol**

Dr. Murad Abu-Khalaf

.....

Date of Submission : **26 August 2025**

Date of Defense : **2 September 2025**





to my family...



FOREWORD

I would like to express my deepest gratitude to my advisor, Assoc. Prof. Dr. Emre KOYUNCU, for his invaluable guidance, insightful feedback, and continuous encouragement throughout the course of this research. His expertise and mentorship have been instrumental in shaping the direction and quality of my work.

I would also like to extend my sincere thanks to Dr. Murad ABU-KHALAF for his support and guidance throughout my academic career. His advice and encouragement have greatly influenced both my research and professional development.

I am grateful to my colleagues and friends at work, Mustafa DEMİR, Ahmet Talha ÇETİN , Fatih EROL, and Hasan Taha BAĞCI, Mehmet Zeki PAŞAOĞLU, Akın ÇATAK, Serkan ŞENEL, for their support, collaboration, and the motivating environment they have provided during this journey.

My heartfelt thanks also go to my friends from my undergraduate years, Sabire Betül ÇINAR, Ömer Emin ÇİÇEK, Bilal TÜRK, and Murat KURNAZ, for their friendship, encouragement, and the memorable moments we have shared.

Finally, I would like to express my deepest appreciation to my mother, Gülay ŞANLAN, and Arif ŞANLAN, for their endless love, unwavering support, and sacrifices, which have been the foundation of all my achievements.

September 2025

Abdulkali ŞANLAN

TABLE OF CONTENTS

	Page
FOREWORD	ix
TABLE OF CONTENTS	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
ÖZET	xxiii
1. INTRODUCTION	1
1.1 Literature Review	2
1.1.1 Visual inertial odometry	3
1.1.2 Image matching	4
1.2 Bridging the Domain Gap with Generative Adversarial Networks	6
1.2.1 Hypothesis	8
2. SYSTEM	9
2.1 Multi-State Constraint Kalman Filter (MSCKF) for VIO	9
2.2 IMU Motion Model	11
2.3 Marginalized Particle Filter	13
2.4 Measurement Model and Likelihood Calculation	18
2.5 Image Domain Adaptation/Synchronization	21
2.6 Data Collection	23
3. RESULTS AND DISCUSSIONS	25
3.1 Evaluation of CycleGAN Domain Adaptation Performance	25
3.2 Flight Test Results	28
3.2.1 Scenario 1: relatively higher domain gap	30
3.2.2 Scenario 2: relatively higher altitude	32
3.2.3 Scenario 3: relatively low feature uniqueness	34
3.2.4 Scenario 4: decent flight setup for image matching	36
4. FUTURE WORKS	39
5. CONCLUSION	43
REFERENCES	45
CURRICULUM VITAE	49



ABBREVIATIONS

AI	: Artificial Intelligence
ANN	: Artificial Neural Network
AWGN	: Additive White Gaussian Noise
BRIEF	: Binary Robust Independent Elementary Features
CNN	: Convolutional Neural Network
CPU	: Central Processing Unit
CUT	: Cycle-consistent Unpaired Translation
DKM	: Deep Kernel Map
EKF	: Extended Kalman Filter
FAST	: Features from Accelerated Segment Test
GAN	: Generative Adversarial Network
GNSS	: Global Navigation Satellite System
GPS	: Global Positioning System
GPU	: Graphics Processing Unit
IMN	: Image Matching Network
IMU	: Inertial Measurement Unit
INS	: Inertial Navigation System
MB	: Megabyte
MPF	: Marginalized Particle Filter
MSCKF	: Multi-State Constraint Kalman Filter
OKVIS	: Open Keyframe-based Visual-Inertial SLAM
ORB	: Oriented FAST and Rotated BRIEF
PF	: Particle Filter
RAM	: Random Access Memory
RANSAC	: Random Sample Consensus
RMSE	: Root Mean Square Error
SIFT	: Scale-Invariant Feature Transform
SLAM	: Simultaneous Localization and Mapping
SLAM3	: Simultaneous Localization and Mapping 3
SURF	: Speeded-Up Robust Features
UAV	: Unmanned Aerial Vehicle
VINS	: Visual-Inertial Navigation System
VIO	: Visual-Inertial Odometry



SYMBOLS

H	: Domain adaptation function
s	: State vector
δs	: Error state vector
q	: Quaternion
p	: Position
v	: Velocity
b_w	: Gyroscope (angular velocity) bias
b_a	: Accelerometer (linear acceleration) bias
$\mathbf{P}$	: Covariance matrix
Φ	: State transition matrix
$\mathbf{Q}$	: Process noise covariance
z	: Measurement vector
$\mathbf{H}$	: Measurement matrix (upright, non-italic)
θ	: Orientation (Euler angles or rotation vector)
$\mathcal{L}$	: Loss function
λ	: Weight of cycle-consistency loss
X_{domain}	: Set of samples for domain X
fps	: Frames per second
Hz	: Hertz (frequency unit)
m	: Metre
s	: Second
mm	: Millimetre
$^\circ$	: Degree (angle unit)



LIST OF TABLES

	<u>Page</u>
Table 3.1: Comparison of GAN models for the domain adaptation task in UAV image translation. Accuracy values are expressed as the relative increase in matched feature correspondences compared to the baseline (unaltered images).....	27
Table 3.2: MPU-9250 IMU specifications.	30
Table 3.3: Camera specifications.	30



LIST OF FIGURES

	<u>Page</u>
Figure 2.1: Main diagram of proposed Vision Aided Navigation method. Green block shows an example of two consecutive domain adaptation with CycleGAN H_1 and H_2 operator of general form $H_{X \rightarrow Y}$, Orange block shows Visual Inertial Odometry(VIO) in closed loop manner, where y_{pos} is captured relative motion of camera, x_{nom} is the nominal states of INS, x_{err} estimated error states of EKF. Blue block shows Image Matching Navigation(IMN) in open loop manner, where $\hat{x}_{nom}$ is nominal states of INS after EKF measurement update as $\hat{x}_{nom} = x_{nom} \oplus \hat{x}_{err}$, $\hat{s}_{full}$ is estimated INS states.....	9
Figure 2.2: Custom Likelihood Function $F(x)$ for various hyperparameter values v	20
Figure 2.3: Sample images from different domains	24
Figure 3.1: Demonstration of a trajectory of UAV point A to B in area of 400m x 400m satellite view	26
Figure 3.2: Different GAN model image to image translation results with ground truth label.....	28
Figure 3.3: The small UAV prepared for the flight tests carrying down-facing onboard camera and IMU	29
Figure 3.4: Scenario 1 position error: Big domain gap (UAV flights in winter/satellite map are in summer), low altitude, high feature uniqueness, 1000 m trajectory	31
Figure 3.5: Image matching performance under high domain discrepancy	33
Figure 3.6: Scenario 2 position Error: low domain gap (UAV flights in summer/satellite map are in iummer), high altitude, high feature uniqueness, 2800 m trajectory	34
Figure 3.7: Scenario 3 position Error: low domain gap (UAV flights in summer/satellite map are in summer), low altitude, low feature uniqueness, 1200 m trajectory	35
Figure 3.8: Scenario 4 position error: low domain gap (UAV flights in summer/satellite map are in summer), low altitude, high feature uniqueness, 1500 m trajectory	37



DOMAIN ADAPTIVE VISUAL INERTIAL GLOBAL LOCALIZATION IN GNSS-DENIED ENVIRONMENTS

SUMMARY

This thesis addresses the problem of long-range global localization for small unmanned aerial vehicles (UAVs) operating in Global Navigation Satellite System (GNSS)–denied environments. The central premise is that visual information from an onboard nadir-looking camera can be aligned with geo-referenced satellite imagery to obtain absolute position fixes, provided that (i) short-term motion is estimated reliably, (ii) the global estimator can reason over multi-modal hypotheses without instability, and (iii) the appearance gap between aerial/satellite domains is reduced so that feature-level correspondence becomes feasible. To this end, we propose an integrated pipeline that combines a lightweight visual–inertial odometry (VIO) front-end, a marginalized particle filter (MPF) for state estimation, and a learned image-to-image domain adaptation module based on cycle-consistent generative adversarial networks (CycleGAN). A custom likelihood function is introduced to convert feature-matching evidence into robust particle weights, thereby improving convergence in ambiguous conditions.

For VIO, we employ a tightly coupled multi-state constraint formulation (MSCKF-like), in which visual feature residuals are projected into the left null space of per-landmark Jacobians, enforcing constraints across a sliding window of camera poses without explicitly augmenting landmark states. This maintains bounded state dimension and estimator consistency under First-Estimates Jacobians while providing accurate short-term motion and pose estimates at high rate. The VIO output supplies drift-limited motion priors to the global estimator.

Global localization is performed with a marginalized particle filter. The state is partitioned into a low-dimensional nonlinear subspace, represented by particles, and a linear–Gaussian subspace, represented by per-particle Kalman filters (Rao–Blackwellization). This decomposition substantially reduces the number of particles required to cover the hypothesis space while preserving linear covariance propagation for velocity, attitude-error, and bias terms. The discrete-time form of the error-state dynamics is adopted for propagation, and observations are produced by rendering a particle-conditioned view (a crop and rotation of the satellite map) using the particle’s horizontal position, yaw, and altitude-dependent scale.

A key challenge is the domain gap between onboard imagery and satellite references, arising from seasonal appearance, illumination, and acquisition geometry. To mitigate this gap, we train three unpaired image translation models: the original CycleGAN, CUT (contrastive unpaired translation), and Turbo-CycleGAN. CycleGAN learns bidirectional mappings regularized by cycle consistency; CUT replaces

cycle consistency with a patchwise contrastive objective to better preserve content; Turbo-CycleGAN enhances texture fidelity and artifact suppression with multi-scale refinements at the cost of higher training and inference complexity. We evaluate these models on UAV/satellite domain adaptation using a task-aligned metric: the mean number of geometrically verified feature correspondences across 100 query images, normalized by the matches obtained from raw (unadapted) UAV images.

The proposed framework has been validated through multiple flight tests conducted under varying environmental conditions, altitudes, and scene characteristics. These tests demonstrate that domain adaptation significantly mitigates the appearance discrepancies between UAV and satellite imagery, resulting in more reliable feature correspondences and improved localization accuracy. In challenging cases with large seasonal differences or repetitive textures, adaptation helps maintain estimator stability by preventing divergence and reducing drift. At higher altitudes, where the effective resolution diminishes and feature observability becomes limited, the approach still provides measurable improvements in position accuracy.

Across scenarios, the combination of VIO, MPF, and domain adaptation consistently outperforms a purely visual baseline without adaptation, with the largest gains observed when the raw domain gap or the texture-induced ambiguity is severe. The MPF’s partitioned formulation enables robust multi-hypothesis reasoning without the instability that can arise when error states are reset at each update, and the proposed likelihood reduces false convergence. At the same time, we identify practical trade-offs. Turbo-CycleGAN offers the highest domain-alignment quality but incurs substantial model size and runtime, which may be challenging for the most resource-constrained airframes.

The thesis concludes that accurate, computationally tractable global localization in GNSS-denied settings can be achieved by: employing a consistent, tightly coupled VIO to provide drift-limited motion priors, using a marginalized particle filter to reason over global hypotheses while keeping computation bounded, adapting image appearance across domains with unpaired translation to restore descriptor consistency, and shaping particle likelihoods with a discriminative mapping from verified match statistics. The overall system advances the state of practice toward reliable, real-time-capable UAV navigation in diverse environmental conditions. Future avenues include night-to-day domain adaptation for low-light operations, image rectification strategies for non-gimballed sensors, lightweight generative models suitable for edge deployment, enhanced likelihoods that exploit spatial/geometric structure beyond match counts, and multi-modal fusion (e.g., magnetic field maps) for scenarios that are intrinsically challenging for vision alone.

GNSS OLMAYAN ORTAMLARDA GÖRÜNTÜ UYARLAMALI GÖRSEL-ATALETSAL KÜRESEL KONUM BELİRLEME

ÖZET

Bu tez, Küresel Navigasyon Uydu Sistemi (GNSS) bulunmayan ortamlarda görev yapan küçük insansız hava araçlarının (İHA) uzun menzilli küresel konumlandırma problemini ele almaktadır. Temel varsayım, İHA üzerine monte edilmiş aşağı bakan bir kameradan elde edilen görsel bilgilerin, coğrafi referanslı uydu görüntüleri ile hizalanarak mutlak konum bilgisi elde edilmesidir. Bunun mümkün olabilmesi için kısa vadeli hareketin güvenilir şekilde tahmin edilmesi, küresel kestiricinin çoklu hipotezleri kararlı biçimde işleyebilmesi ve hava/uydu görüntüleri arasındaki görünüş farklılıklarının azaltılarak özellik seviyesinde eşleştirmenin mümkün hale gelmesi gereklidir. Bu amaçla, hafif bir görsel-atıl ivmeölçer odometri (VIO) öne ucu, durum kestirimi için marjinalleştirilmiş parçacık filtresi (MPF) ve CycleGAN tabanlı öğrenilmiş görüntü alanı uyarlama modülünü birleştiren entegre bir boru hattı önerilmektedir. Ayrıca, özellik eşleşmelerinden elde edilen kanıtları sağlam parçacık ağırlıklarına dönüştüren özel bir olabilirlik fonksiyonu tasarlanmış ve belirsiz koşullarda yakınsamayı iyileştirmesi sağlanmıştır.

VIO tarafında, çok durumlu kısıtlı sıkı bağlı (MSCKF-benzeri) bir formülasyon benimsenmiştir. Görsel özellik artıklarının, her bir arazi noktası için oluşturulan Jacobi matrislerinin sol null uzayına izdüşürülmesiyle, arazi noktası durumları açıkça artırılmadan kamera pozlarının bir kaydırma penceresi boyunca kısıtlanması sağlanmaktadır. Bu yaklaşım, durum boyutunun sınırlı kalmasını ve kestiricinin First-Estimates Jacobians altında tutarlılığını korumasını sağlamaktadır. Böylece kısa vadeli hareket ve poz tahminleri yüksek oranda ve doğrulukla elde edilmekte, global kestiriciye sapma sınırlandırılmış hareket öncelikleri sağlanmaktadır.

Küresel konumlandırma, marjinalleştirilmiş parçacık filtresi ile gerçekleştirilmiştir. Durum, parçacıklarla temsil edilen düşük boyutlu doğrusal olmayan bir altuzay ve parçacık başına Kalman filtreleriyle temsil edilen doğrusal–Gaussian bir altuzay olarak ayrıştırılmaktadır (Rao–Blackwellizasyon). Bu ayrıştırma, hipotez uzayını kapsamak için gereken parçacık sayısını önemli ölçüde azaltırken, hız, tutum hatası ve bias terimleri için doğrusal kovaryans yayılımını korumaktadır. Hata-durum dinamiklerinin ayrık zamanlı formu kullanılarak yayılım yapılmış, gözlemler ise parçacığa bağlı konum, yönelim ve irtifa ölçeği ile uydudan kırılmış/girdaplanmış görüntülerin işlenmesiyle üretilmiştir. Kamera roll ve pitch eksenlerinde gimballı olduğundan, uçuş testlerinde nadir görüntüleme varsayımı geçerli olmuş ve işlem kolaylaşmıştır.

Onboard (uçuş sırasında alınan) görüntüler ile uydu görüntüleri arasındaki temel zorluk, mevsimsel görünüm farklılıkları, ışıklandırma ve alım geometrisinden kaynaklanan “alan farkı”dır. Bu farkı azaltmak için üç farklı eşleştirilmemiş görüntü çevirim modeli

eğitilmiştir: CycleGAN, CUT (contrastive unpaired translation) ve Turbo-CycleGAN. CycleGAN çift yönlü dönüşüm ve döngü tutarlılığı üzerine kuruludur; CUT, döngü tutarlılığı yerine yama düzeyinde karşıtlık tabanlı bir amaç fonksiyonu kullanarak içerik korunmasını artırır; Turbo-CycleGAN ise çok ölçekli iyileştirmeler ile doku doğruluğunu artırır, ancak daha yüksek hesaplama maliyeti getirir. Bu modeller, UAV/uydu uyarlamasında, 100 sorgu görüntüsünde geometrik olarak doğrulanmış özellik eşleşmelerinin ortalamasına dayalı bir metrikle değerlendirilmiştir. Deneylerde CycleGAN ve CUT mütevazı kazanımlar sağlarken, Turbo-CycleGAN daha yüksek (yüzde onlar seviyesinde) eşleşme artışı sunmuş fakat hesaplama yükü oldukça yüksek olmuştur.

Önerilen çerçeve, farklı çevresel koşullar, irtifalar ve sahne özelliklerinde gerçekleştirilen uçuş testleriyle doğrulanmıştır. Testler, alan uyarlamasının UAV ve uydu görüntüleri arasındaki görünüm farklılıklarını önemli ölçüde azalttığını, daha güvenilir özellik eşleşmeleri sağladığını ve konum doğruluğunu artırdığını göstermektedir. Mevsimsel farklılıkların büyük olduğu veya tekrar eden dokuların bulunduğu zorlu durumlarda, uyarlama kestiricinin kararlılığını koruyarak sapma ve hatalı yakınsamayı azaltmıştır. Daha yüksek irtifalarda, etkin çözünürlük düşse ve gözlenebilirlik azalsa dahi konum doğruluğunda ölçülebilir iyileşmeler elde edilmiştir. Uygun koşullarda ise sistem en iyi performansını göstermiş, alan uyarlaması ek olarak sağlamlık ve tutarlılığı artırmıştır. Genel olarak, uçuş deneyleri, VIO, marjinalleştirilmiş parçacık filtresi ve öğrenilmiş alan uyarlamasının bir araya getirilmesinin GNSS olmayan ortamlarda İHA küresel konumlandırması için güvenilir ve doğru bir çözüm sunduğunu kanıtlamaktadır.

Senaryoların tamamında, VIO, MPF ve alan uyarlamasının birleşimi, yalnızca görsel tabanlı yaklaşımlara kıyasla belirgin bir üstünlük göstermiştir. Özellikle, görsel eşleştirme sürecinde alan farklarının büyük olduğu veya dokuya dayalı belirsizliklerin yoğun olduğu koşullarda elde edilen kazanımlar oldukça dikkat çekicidir. Bu durum, görsel tabanlı sistemlerin, özellikle karmaşık çevresel koşullar altında, konumlandırma doğruluğunu önemli ölçüde sınırladığı bilinen bir gerçektir. MPF'nin ayrıştırılmış formülasyonu ise, sistemin çoklu hipotezleri daha kararlı bir şekilde ele almasını sağlayarak, hem başlangıçta daha doğru tahminler yapılmasına yardımcı olmuş hem de önerilen olabilirlik fonksiyonu sayesinde yanlış yakınsama durumlarını minimize etmiştir. Bu fonksiyon, daha iyi bir optimizasyon süreci sunmuş ve böylece elde edilen çözümün güvenilirliğini artırmıştır.

Bununla birlikte, bu gelişmelerin bazı pratik ödünleri de gözlemlenmiştir. Örneğin, Turbo-CycleGAN modelinin sağladığı en yüksek uyarlama kalitesi, özellikle kaynak kısıtlı hava araçlarında uygulama açısından bir sınırlama teşkil etmektedir. Modelin boyutu ve işlem süresi, düşük performanslı platformlarda kullanılabilirliğini kısıtlamakta olup, bu tür sistemlerin daha verimli hale gelmesi için farklı optimizasyon stratejilerine ihtiyaç duyulmaktadır. Ayrıca, kullanılan gimballı kamera varsayımı, düşük maliyetli platformlarda genellikle geçerli olmayabilir. Bu durumda, eğik (off-nadir) görüntülerle karşılaşılması, ek geometrik bozulmalara yol açabilir ve bu da eşleştirme öncesinde bu bozulmaların düzeltilmesini gerektirir. Bu sorunlar, pratik uygulamalarda dikkatle ele alınmalıdır, çünkü görüntü bozulmalarının düzeltilmesi, işlem süresi üzerinde ek bir yük oluşturabilir ve sistemin genel verimliliğini etkileyebilir.

Bu tez, GNSS olmayan ortamlarda doğru ve hesaplama açısından uygulanabilir küresel konumlandırmanın, sapması sınırlandırılmış hareket öncelikleri sağlayan sıkı bağlı bir VIO, küresel hipotezleri etkin biçimde değerlendiren marjinalleştirilmiş parçacık filtresi, UAV/uydu görüntüleri arasındaki alan farklılıklarını azaltan eşleştirilmemiş dönüşüm tabanlı alan uyarlama ve doğrulanmış eşleşme istatistiklerinden türetilmiş ayrıştırıcı olabilirlik fonksiyonunun bir araya getirilmesiyle mümkün olduğunu ortaya koymaktadır. Bu bütünleşik sistem, özellikle GNSS sinyalinin zayıf olduğu veya hiç mevcut olmadığı bölgelerde, İHA'ların doğru konumlandırmasını sağlayacak önemli bir adım atılmasına olanak tanımaktadır. Konumlandırma doğruluğunun, her bir modülün katkılarıyla sürekli olarak iyileştirilmesi, farklı çevresel koşullara uyum sağlamak için gerekli esnekliği sunmakta ve mevcut yaklaşımları ileriye taşımaktadır. Bu sayede, daha güvenilir ve gerçek zamanlı bir konumlandırma sağlanmaktadır.

Gelecek çalışmalar arasında, düşük ışık koşullarında çalışabilen gece-gündüz uyarlama algoritmalarının geliştirilmesi önemli bir yer tutmaktadır. Bu algoritmalar, hem şehir içi hem de kırsal alanlarda ortaya çıkabilecek dramatik aydınlatma değişimlerine karşı sistemin dayanıklılığını artıracaktır. Özellikle, şehir ışıkları ve doğal ortamların farklı ışık profilleri, mevcut algoritmaların performansını düşürebilmektedir; bu nedenle derin öğrenme tabanlı ışık adaptasyonu veya görüntü ön işleme yöntemlerinin geliştirilmesi büyük önem taşımaktadır.

Ayrıca, gimballsız sensörler için görüntü düzeltme yöntemlerinin araştırılması, sensörlerin daha geniş bir yelpazede kullanılabilmesini sağlayacaktır. Bu çalışmalar, platform üzerindeki donanım maliyetini düşürürken, farklı uçuş profillerine sahip İHA'larda da güvenilir konumlandırma sağlamayı mümkün kılacaktır. Kaynak kısıtlı platformlarda çalışabilecek hafif üretici modellerin tasarımı, daha verimli veri işleme ve daha düşük enerji tüketimi sağlamak için kritik öneme sahiptir. Bu modeller, özellikle mini veya mikro İHA'lar için uzun uçuş süreleri ve hızlı işlem yetenekleri sunabilir.

Bunun yanı sıra, eşleşme sayılarının ötesinde, mekansal ve geometrik yapıların dikkate alındığı yeni olabilirlik fonksiyonlarının geliştirilmesi, konumlandırma doğruluğunu daha da artırabilir. Bu tür fonksiyonlar, yalnızca görsel eşleştirmeleri değil, aynı zamanda çevresel parametreleri de modelleyerek daha kapsamlı bir analiz yapmayı mümkün kılacaktır. Örneğin, arazi eğimi, yapı yoğunluğu, gölgelendirme ve nesne boyutlarının hesaba katılması, sistemin yanlış yakınsama riskini azaltabilir ve partikül ağırlıklarının daha doğru şekilde güncellenmesini sağlayabilir.

Ayrıca, manyetik alan haritalarının ve diğer çoklu verilerin entegrasyonu, sistemin sağlamlığını ve doğruluğunu daha da pekiştirebilir. Özellikle, manyetik alan, barometrik ve atmosferik verilerle birleştirilen çoklu sensör füzeleri, GNSS olmayan ortamlar için güvenilir bir konumlandırma çözümü sunabilir. Bu tür yenilikçi yaklaşımlar, GNSS olmayan ortamlar için son derece etkili ve pratik çözümler sunma potansiyeline sahiptir. Dahası, simülasyon tabanlı doğrulama ve sahada yapılan testler, geliştirilen algoritmaların gerçek uçuş senaryolarındaki performansını daha iyi değerlendirmeye yardımcı olacaktır. Bu sayede, sistemin kararlılığı, doğruluğu ve dayanıklılığı artırılabilir ve ileri seviye İHA uygulamalarına doğrudan entegrasyon sağlanabilir.



1. INTRODUCTION

Autonomous navigation in Global Navigation Satellite System (GNSS)-denied environments presents a critical challenge, particularly for unmanned aerial vehicles (UAVs) operating in complex or contested areas. Traditional navigation methods, such as inertial navigation systems (INS) and terrain-aided navigation, often suffer from long-term drift, computational limitations, or impractical hardware requirements, especially for lightweight unmanned platforms. While terrain-aided navigation using digital elevation models (DEMs) offers an alternative, its effectiveness is limited by sensor dependency, altitude variations, and stealth concerns. Similarly, radar- or lidar-based mapping methods introduce high power demands and cost constraints, reducing their suitability for small UAVs.

To overcome these limitations, visual feature matching has emerged as a promising alternative, leveraging the widespread availability and affordability of camera-based systems. While Visual Inertial Odometry (VIO) systems show promising results in many cases, they can be sensitive to rapid motion or abrupt environmental changes, leading to inaccuracies and accumulated errors over time. Their performance can also degrade in low-light conditions or in environments with sparse visual features. To correct VIO solution inaccuracies, global visual position fix methods, specifically satellite image matching algorithms, have been developed. Positional fixes are obtained by comparing UAV-captured images with satellite referenced images. However, when the domain discrepancy between UAV images and satellite images is considerably high, such as due to significant temporal variations (season and weather) or different lighting conditions, the performance of image matching degrades sharply. Hence, this work proposes a novel domain adaptation technique within an aerial image visual matching framework. Limited hardware capabilities (GPU, CPU, RAM) are an unpreventable nature of small UAV applications due to lightweight constraints. Therefore, reducing computational

power is also another aim of this work. Light and effective alternatives for VIO, image matching, and marginalized particle filter-based state estimation methods are employed.

The main objectives of this thesis can be summarized as below:

- To evaluate, compare, and select the best suitable VIO algorithm considering the limited power resources on small UAVs.
- To optimize the computationally costly image matching algorithm for particle filter application by using a Marginalized Particle Filter and exploiting AI-based model properties.
- To propose a CycleGAN-based image-to-image translation method for increasing image matching performance by decreasing the domain gap between aerial image domains.
- To validate that the proposed method is a promising alternative to GPS/GNSS in real UAV applications across diverse scenarios.

1.1 Literature Review

As autonomous systems are increasingly used in complex, GNSS-denied environments, navigating without satellite signals remains a core challenge. Traditional methods like inertial navigation systems (INS) and terrain-aided navigation struggle with drift over time, high computational loads, or hardware that's too bulky for lightweight unmanned vehicles. Terrain-aided navigation using digital elevation models (DEMs) offers an alternative but is hampered by reliance on specific sensors, altitude sensitivity, and issues with stealth. Likewise, radar and lidar-based mapping demand significant power and are often too costly or heavy for small UAVs. Visual feature matching has gained traction as a low-cost, camera-based option, but it faces difficulties in reliably matching features across different sensor types—especially when comparing onboard camera images to satellite imagery from a different domain.

Early terrain-aided navigation relied on single-point elevation matching [1]–[4], which later evolved into point-cloud-based approaches [5,6]. These methods, while effective, often depend on specialized sensors like radar altimeters or lidar, which not only increase

system complexity but also compromise stealth due to their active emissions—limiting their use in certain missions.

1.1.1 Visual inertial odometry

Visual-Inertial Odometry (VIO) constitutes a fundamental enabling technology within the domain of contemporary state estimation, wherein high-frequency, short-term motion information originating from an Inertial Measurement Unit (IMU) is synergistically integrated with the long-term drift-compensating capabilities of a vision sensor, such as a camera. The historical trajectory of VIO development reveals a systematic evolution, beginning with initial loosely-coupled filtering strategies, advancing toward more precise and computationally robust tightly-coupled optimization frameworks, and more recently, transitioning into hybridized paradigms enriched by the capabilities of modern machine learning methodologies.

The earliest generations of VIO frameworks were predominantly grounded in filtering theory—most notably the Extended Kalman Filter (EKF)—applied within a loosely-coupled architectural context. In such systems, visual odometry pipelines operated in an independent manner to estimate the camera pose, with these pose estimates subsequently incorporated as measurement updates to refine the IMU-propagated state estimates [7]. While this methodology offered notable computational efficiency, it inherently discarded significant statistical correlations and interdependencies between the visual and inertial modalities, thereby constraining the achievable estimation accuracy.

A pivotal advancement emerged with the advent of tightly-coupled formulations, exemplified by the Multi-State Constraint Kalman Filter (MSCKF) [8]. This approach integrated visual feature observations directly into the measurement model of the filter, thereby eliminating the necessity for explicit three-dimensional point reconstruction. Instead, the MSCKF retained the historical poses of camera states that had observed the same feature, a strategy which not only enhanced estimation accuracy but also preserved computational tractability.

Following this, the research community gravitated toward optimization-based VIO solutions, wherein the estimation task is cast as a nonlinear least-squares problem defined over a sliding temporal window of states and sensor observations. This formulation enables the re-linearization of past states, substantially improving the fidelity of the solution. OKVIS (Open Keyframe-based Visual-Inertial SLAM) [9] serves as a representative example, performing joint minimization of visual reprojection errors and inertial residuals across selected keyframes. Expanding upon this paradigm, VINS-Mono [10] delivered a comprehensive monocular VIO pipeline that incorporates visual-inertial initialization, local bundle adjustment, and a global pose graph optimization mechanism for robust loop closure detection.

Present-day state-of-the-art frameworks continue to refine and extend these optimization-based methodologies. For instance, ORB-SLAM3 [11] demonstrates resilient multi-map visual-inertial fusion suitable for prolonged operational scenarios, while OpenVINS [12] offers a high-performance, open-source research platform designed for rigorous experimentation and reproducibility.

The most recent research trend involves the incorporation of deep learning techniques into the VIO pipeline, aiming to improve robustness under challenging operational conditions through enhancements in feature detection, noise modeling, and data association. Hybrid architectures, such as Adaptive VIO [13], represent a convergence of classical geometric optimization and adaptive, online machine learning, enabling the estimation process to dynamically respond to sensor degradation and environmental variability in real time—thus unifying the deterministic precision of geometric VIO with the adaptive resilience characteristic of modern data-driven approaches.

1.1.2 Image matching

The detection of local features in images has, in its traditional form, predominantly relied on meticulously hand-crafted algorithms designed to extract salient interest points based on variations in image gradients or localized intensity patterns. A well-known example is the Scale-Invariant Feature Transform (SIFT), which identifies scale-space extrema by computing the Difference-of-Gaussians and subsequently constructs gradient-based descriptors, thereby achieving robustness to changes in

both scale and rotation [14]. Another influential approach, the Speeded Up Robust Features (SURF) detector, employs a Hessian matrix–based interest point detection strategy in combination with box filter approximations, significantly accelerating the scale-space convolution operations required for multi-scale analysis [15]. Similarly, the Oriented FAST and Rotated BRIEF (ORB) method integrates the FAST keypoint detector with the BRIEF binary descriptor, while additionally introducing an orientation estimation component to secure rotation invariance, all within a framework optimized for high computational efficiency [16]. Diverging from these conventionally engineered feature extractors, SuperPoint represents a data-driven alternative that leverages a self-supervised convolutional neural network to concurrently predict stable keypoint locations and their associated descriptors. Its training process commences with synthetic corner data and is subsequently refined through a homographic adaptation scheme, ultimately yielding highly repeatable keypoints that remain robust even in regions of low texture, and which frequently surpass the performance of classical methods in challenging visual scenarios [17].

Following the localization of keypoints, the subsequent step involves establishing correspondences between descriptors across different images. The most straightforward approach to this task is nearest neighbor matching, wherein each descriptor from the first image is paired with its most similar counterpart from the second image, as measured by an appropriate distance metric—commonly Euclidean distance for descriptors such as SIFT and SURF, or Hamming distance in the case of binary descriptors like ORB [16]. While data structures such as k-d trees can be employed to accelerate these nearest neighbor searches, image regions containing repetitive patterns or ambiguous textures frequently induce false correspondences. To address this, robust geometric estimation techniques such as Random Sample Consensus (RANSAC) are utilized; these operate by repeatedly sampling minimal subsets of matches, fitting a geometric model such as the fundamental matrix, and identifying the maximal set of inlier correspondences that are consistent with the estimated model, thereby effectively filtering out spurious matches [18].

Recent advancements in deep learning have facilitated the development of end-to-end feature matching frameworks that markedly improve resilience in the face of severe

viewpoint, scale, or illumination variations. For example, SuperGlue conceptualizes the correspondence problem as a graph-based assignment task and applies a Graph Neural Network augmented with both self-attention and cross-attention mechanisms to jointly reason over the two sets of extracted local features, enabling accurate matches even under extreme appearance changes [19]. Extending this paradigm, LightGlue introduces an adaptive inference strategy, dynamically modulating its computational expenditure in accordance with the difficulty of the matching task—executing more rapidly on easy image pairs while retaining accuracy on visually challenging pairs [20]. In parallel, research has progressed beyond sparse keypoint-based pipelines toward detector-free dense matching methodologies. LoFTR exemplifies this class by initially performing coarse-level matching through transformer-based attention modules, followed by a refinement stage at the full image resolution, thus enabling dependable correspondences in low-texture or repetitive regions [21]. More recently, DKM has formulated dense correspondence estimation as a kernelized regression problem, merging globally informed kernel-based matching with convolutional refinement stages to produce per-pixel correspondences enriched with confidence measures, achieving leading performance on standard two-view geometry evaluation benchmarks [22].

1.2 Bridging the Domain Gap with Generative Adversarial Networks

One of the most persistent and practically significant challenges encountered in the deployment of visual navigation systems within real-world environments is the so-called domain gap. This term denotes the substantial discrepancy in visual appearance between imagery captured by an onboard camera during actual operation and the corresponding reference data sources, which may include aerial photographs or satellite imagery. Such discrepancies can be attributed to a range of factors, including differences in viewpoint (e.g., ground-level perspectives versus nadir or overhead views), variations in sensor modality, changes in illumination conditions (such as day versus night scenarios), as well as environmental influences related to weather patterns and seasonal shifts. Collectively, these variations degrade the direct comparability of visual data across domains, thereby reducing the reliability of feature matching and subsequent localization processes.

Generative Adversarial Networks (GANs), and more specifically the Cycle-Consistent Generative Adversarial Network (CycleGAN), have emerged as a highly effective approach for addressing this issue through the process of unsupervised domain adaptation. Introduced by Zhu et al. [23], CycleGAN enables the translation of images between two distinct domains without the necessity for paired training data—meaning that it is not required to have perfectly corresponding examples of the same scene in both domains. Instead, CycleGAN achieves this through the concurrent training of two generator networks, each responsible for translating images from one domain to the other, and two discriminator networks, which aim to distinguish real images from generated ones within their respective domains.

The defining innovation of CycleGAN lies in its cycle consistency loss. This loss enforces the constraint that if an image is translated from domain A to domain B , and subsequently translated back from B to A , the resulting reconstruction should closely match the original input image. This bidirectional consistency constraint acts as a powerful regularization mechanism, preventing degenerate solutions in which content information is lost, and ensuring that the semantic structure of the source image is preserved while adopting the stylistic characteristics of the target domain.

Within the scope of visual navigation, CycleGAN has been utilized to transform satellite-based reference maps into a visual modality that more closely resembles the imagery captured by a UAV's onboard camera [24]. By producing such pseudo-UAV perspectives from satellite imagery, it becomes possible to perform visual feature matching with markedly improved reliability, as the domain-specific discrepancies are substantially reduced. This translation harmonizes the visual characteristics of the reference and live-captured imagery, thereby enhancing localization robustness with respect to the pre-existing map. A related application involves day-to-night translation, in which nighttime images are converted into their daytime-equivalent counterparts, thus revealing more discriminative visual features and enabling reliable navigation under low-illumination conditions [25].

Through these capabilities, CycleGAN and similar generative models provide a systematic and adaptable means of bridging the domain gap in visual navigation systems.

By aligning the appearance of reference and operational imagery across diverse and dynamic environmental conditions, they expand the effective operational envelope of such systems, significantly enhancing resilience, adaptability, and overall navigation performance in complex real-world scenarios.

1.2.1 Hypothesis

The central hypothesis of this thesis is that an integrated framework—comprising a lightweight visual–inertial odometry (VIO) solution, a marginalized particle filter for state estimation, and a deep learning–based domain adaptation module for aerial imagery—can enable small unmanned aerial vehicles (UAVs) to obtain reliable and accurate global position estimates in environments where Global Navigation Satellite System (GNSS) signals are unavailable or degraded, all while remaining within the strict computational and hardware limitations inherent to onboard processing. Specifically, by incorporating a CycleGAN-based image translation stage to reduce the visual appearance disparity between UAV-acquired imagery and preexisting satellite reference maps, it is anticipated that the system will maintain robust and repeatable feature correspondences even under severe temporal shifts, seasonal variations, and challenging illumination conditions.

In addition, it is hypothesized that the deliberate selection and fine-tuning of a computationally efficient VIO algorithm, when fused with a marginalized particle filtering scheme, will effectively suppress cumulative drift and mitigate the influence of spurious or outlier measurements, without incurring the substantial processing overhead typically associated with conventional terrain-aided navigation or LiDAR-based localization techniques. By uniting these complementary components into a cohesive and resource-aware pipeline, the proposed approach is expected to deliver global positioning performance that is comparable to that of GNSS under nominal conditions, while ensuring compatibility with the stringent size, weight, power, and real-time operational constraints characteristic of small UAV platforms.

2. SYSTEM

The system that is proposed on the thesis is shown Figure 2.1. Components of all main system will be explained in following subsections.

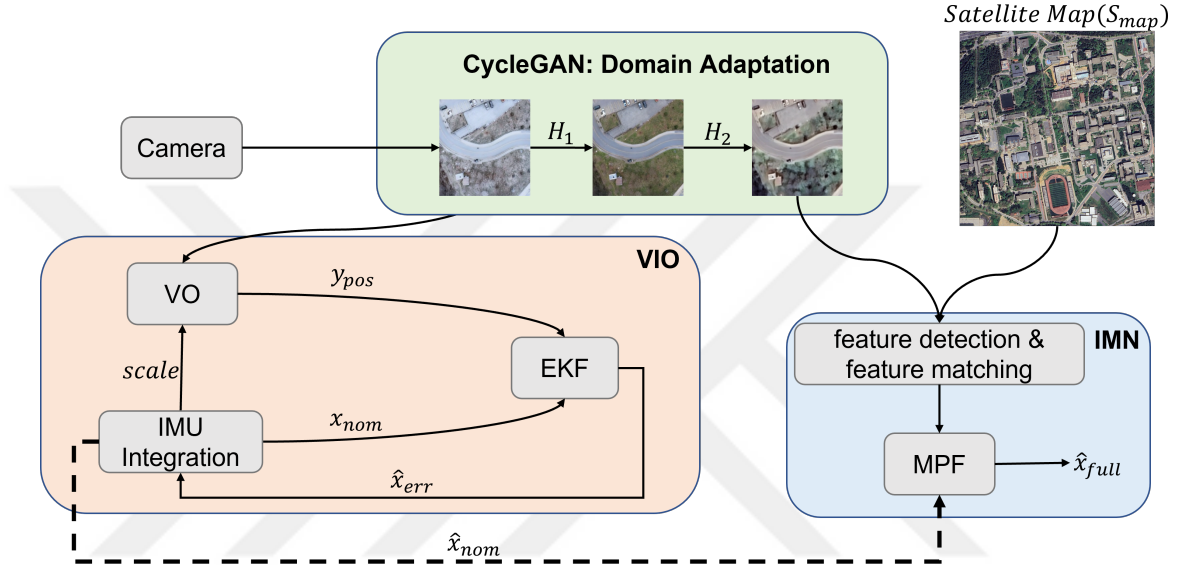


Figure 2.1: Main diagram of proposed Vision Aided Navigation method. Green block shows an example of two consecutive domain adaptation with CycleGAN H_1 and H_2 operator of general form $H_{X \rightarrow Y}$, Orange block shows Visual Inertial Odometry(VIO) in closed loop manner, where y_{pos} is captured relative motion of camera, x_{nom} is the nominal states of INS, x_{err} estimated error states of EKF. Blue block shows Image Matching Navigation(IMN) in open loop manner, where $\hat{x}_{nom}$ is nominal states of INS after EKF measurement update as $\hat{x}_{nom} = x_{nom} \oplus \hat{x}_{err}$, $\hat{s}_{full}$ is estimated INS states.

2.1 Multi-State Constraint Kalman Filter (MSCKF) for VIO

The Multi-State Constraint Kalman Filter (MSCKF) estimates the motion of a rigid body by performing a *tightly coupled* integration of inertial measurements with visual feature observations, without explicitly including the 3D positions of landmarks in the filter state. At a given time step k , the state vector comprises the current inertial state along with a sliding window of past camera poses:

$$\mathbf{s}_k = \left[{}^G_k \bar{\mathbf{q}}, {}^G \mathbf{p}_{I_k}, {}^G \mathbf{v}_{I_k}, \mathbf{b}_\omega, \mathbf{b}_a, \{ {}^G_{C_j} \bar{\mathbf{q}}, {}^G \mathbf{p}_{C_j} \}_{j=1}^M \right]^\top. \quad (2.1)$$

Here, ${}^G_k\bar{q}$ is the quaternion representing the rotation from the IMU frame at time k to the global frame G , ${}^G\mathbf{p}_{I_k}$ and ${}^G\mathbf{v}_{I_k}$ denote the IMU position and velocity in the global frame, $\mathbf{b}_\omega$ and $\mathbf{b}_a$ are the gyroscope and accelerometer bias vectors, and $\{{}_{C_j}^G\bar{q}, {}^G\mathbf{p}_{C_j}\}$ represent the orientations and positions of the past M camera poses stored in the sliding window. At each propagation step, the IMU's measured specific force $\mathbf{a}_m$ and angular velocity $\boldsymbol{\omega}_m$ are used to update all poses in the sliding window:

$$\hat{\mathbf{s}}_{k|k-1} = f(\hat{\mathbf{s}}_{k-1|k-1}, \mathbf{a}_m, \boldsymbol{\omega}_m), \quad (2.2)$$

$$\mathbf{P}_{k|k-1} = \Phi \mathbf{P}_{k-1|k-1} \Phi^\top + \mathbf{Q}, \quad (2.3)$$

where Φ is the state transition matrix from $k-1$ to k , and $\mathbf{Q}$ is the continuous-time process noise covariance mapped into the discrete domain.

When a feature f is tracked across n camera frames, the measurement residuals can be stacked as

$$\mathbf{r}_f = \begin{bmatrix} \mathbf{z}_1 - h(\hat{\mathbf{s}}_{k_1}) \\ \vdots \\ \mathbf{z}_n - h(\hat{\mathbf{s}}_{k_n}) \end{bmatrix} \approx \mathbf{H}_s \delta \mathbf{s} + \mathbf{H}_f \delta \mathbf{p}_f + \mathbf{n}, \quad (2.4)$$

where $\delta \mathbf{s}$ denotes the state error and $\delta \mathbf{p}_f$ is a small perturbation of the landmark position.

To avoid augmenting $\mathbf{p}_f$ into the state, the MSCKF projects $\mathbf{r}_f$ into the left null-space of $\mathbf{H}_f$. Let $\mathbf{A}$ satisfy $\mathbf{A}^\top \mathbf{H}_f = \mathbf{0}$; then:

$$\mathbf{r}'_f = \mathbf{A}^\top \mathbf{r}_f \approx (\mathbf{A}^\top \mathbf{H}_s) \delta \mathbf{s} + \mathbf{A}^\top \mathbf{n}. \quad (2.5)$$

A standard Kalman update is then applied:

$$\mathbf{K} = \mathbf{P}(\mathbf{A}^\top \mathbf{H}_s)^\top ((\mathbf{A}^\top \mathbf{H}_s) \mathbf{P}(\mathbf{A}^\top \mathbf{H}_s)^\top + \mathbf{R}')^{-1}, \quad (2.6)$$

$$\hat{\mathbf{s}}_{k|k} = \hat{\mathbf{s}}_{k|k-1} \boxplus \mathbf{K} \mathbf{r}'_f, \quad \mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}(\mathbf{A}^\top \mathbf{H}_s)) \mathbf{P}_{k|k-1}. \quad (2.7)$$

This null-space projection enables the efficient exploitation of visual constraints while keeping the filter state dimension bounded, and ensures estimator consistency when using First-Estimates Jacobians.

2.2 IMU Motion Model

In this work, the inertial motion modeling framework is based on the 15-dimensional state formulation presented in [26]. This formulation utilizes accelerometer and gyroscope measurements, each inherently corrupted by sensor noise, obtained from an Inertial Measurement Unit (IMU) rigidly mounted in the body frame. These measurements are employed to propagate the states of an Inertial Navigation System (INS) in continuous time. The model captures the complete navigation kinematics of the system, and its dynamic equations are expressed as follows:

$$\dot{p} = v, \quad (2.8a)$$

$$\begin{aligned} \dot{v} &= a \\ &= R_b^i(a_m - a_{\text{bias}} - n_a) + g, \end{aligned} \quad (2.8b)$$

$$\begin{aligned} \dot{q} &= \frac{1}{2}q \otimes w \\ &= \frac{1}{2}q \otimes (w_m - \omega_{\text{bias}} - n_\omega), \end{aligned} \quad (2.8c)$$

$$\dot{a}_{\text{bias}} = d_a, \quad (2.8d)$$

$$\dot{\omega}_{\text{bias}} = d_\omega, \quad (2.8e)$$

$$\dot{g} = 0, \quad (2.8f)$$

In the above equations, p , v , and a denote the true position, velocity, and linear acceleration vectors expressed in the inertial frame $\mathcal{F}_i$, which is considered the local navigation reference frame. The orientation of the body frame $\mathcal{F}_b$ relative to $\mathcal{F}_i$ is parameterized by the quaternion q , where the symbol $\otimes$ represents the Hamilton product between quaternions. The true angular velocity ω is defined in the body frame $\mathcal{F}_b$. The variables a_{bias} and ω_{bias} represent the slowly varying biases of the accelerometer and gyroscope sensors, respectively. The vector g denotes the local gravity vector in the inertial frame, assumed constant over time. The terms d_a and d_ω correspond to bias drift processes, which are modeled as zero-mean Gaussian white noise. The stochastic perturbations n_a and n_ω characterize the additive measurement noise of the accelerometer and gyroscope, respectively.

The outputs generated by the IMU are mathematically described by:

$$a_m = R_i^b(a - g) + a_{\text{bias}} + n_a, \quad (2.9a)$$

$$w_m = \omega + \omega_{\text{bias}} + n_\omega. \quad (2.9b)$$

Here, a_m and w_m denote the measured specific force (linear acceleration) and angular velocity, respectively. The terms n_a and n_ω represent the measurement noise associated with the accelerometer and gyroscope. The transformation matrix R_i^b is the rotation matrix that converts vector quantities from the inertial frame $\mathcal{F}_i$ to the body frame $\mathcal{F}_b$.

For estimation purposes, the complete state dynamics given in equation (2.8) are typically separated into nominal-state dynamics and error-state dynamics. This decomposition allows for a clear distinction between the deterministic state evolution derived from the IMU integration and the stochastic deviations arising from measurement and process noise.

The nominal-state dynamics, obtained through direct numerical integration of IMU measurements, yield the dead-reckoning solution as described in [26]:

$$\dot{\vec{p}} = \vec{v}, \quad (2.10a)$$

$$\dot{\vec{v}} = \bar{R}_b^i(a_m - \bar{a}_{\text{bias}}) + \bar{g}, \quad (2.10b)$$

$$\dot{\vec{q}} = \frac{1}{2}\bar{\vec{q}} \otimes (w_m - \bar{\omega}_{\text{bias}}), \quad (2.10c)$$

$$\dot{\vec{a}}_{\text{bias}} = 0, \quad (2.10d)$$

$$\dot{\vec{\omega}}_{\text{bias}} = 0, \quad (2.10e)$$

$$\dot{\vec{g}} = 0. \quad (2.10f)$$

The evolution of the error-state vector, which models small deviations from the nominal trajectory, is formulated as follows [26]:

$$\dot{\delta p} = \delta v, \quad (2.11a)$$

$$\dot{\delta v} = \bar{R}_b^i(-[a_m - \bar{a}_{\text{bias}}]_{\times} \delta \theta - \delta a_{\text{bias}} - n_a) + \delta g, \quad (2.11b)$$

$$\dot{\delta \theta} = -[\omega_m - \bar{\omega}_{\text{bias}}]_{\times} \delta \theta - \delta \omega_{\text{bias}} - n_\omega, \quad (2.11c)$$

$$\dot{\delta a}_{\text{bias}} = d_a, \quad (2.11d)$$

$$\dot{\delta \omega}_{\text{bias}} = d_\omega, \quad (2.11e)$$

$$\dot{\delta g} = 0. \quad (2.11f)$$

In this error-state representation, linear state variables such as position are expressed as the sum of the nominal value and a small perturbation, i.e., $p = \bar{p} + \delta p$. Rotational quantities are parameterized through the multiplicative form $R = \bar{R}\delta R$, where $\delta R = \exp([\delta\theta]_{\times})$ is the exponential map of the small-angle error vector $\delta\theta$ into a rotation matrix. The notation $[a]_{\times}b$ denotes the vector cross product between a and b , expressed via the multiplication of b with the skew-symmetric matrix associated with a .

2.3 Marginalized Particle Filter

Since the Extended Kalman Filter (EKF) framework fundamentally relies on the linearization of the measurement function, and such linearization is not practically feasible for measurement models derived from camera imagery, this work adopts the Marginalized Particle Filter (MPF) approach [27] to address the strong nonlinearities inherent in the system, particularly within the measurement process. In this context, particle filtering is employed to estimate the error state defined in equation (2.11) under an open-loop integration architecture, rather than utilizing a closed-loop Visual-Inertial Odometry (VIO) integration. The selection of an open-loop architecture is motivated by the observation that particle filter stability can be compromised if the error state is reset following each measurement update, as the posterior distribution in such cases often exhibits a multi-modal structure, in direct contrast to the unimodal Gaussian assumption underlying the EKF methodology.

The complete 15-dimensional state vector described in equation (2.11) introduces a considerable computational burden, because a larger number of particles is required to adequately sample the full high-dimensional state space. A key strategy to reduce this computational complexity is to decrease the dimensionality of the state space in which particles are sampled. This is achieved by exploiting the linear–nonlinear partitioning of the system’s dynamics: the MPF utilizes standard Kalman filtering techniques to process the linear substructure of the state-space model, while the nonlinear substructure is represented and propagated through particle filtering. This combined estimation strategy enables accurate handling of nonlinearities while maintaining computational tractability. Following the notation of [27], the discrete-time state-space model that separates the nonlinear and linear components is expressed as:

$$s_{k+1}^n = f_k^n(s_k^n) + A_k^n(s_k^n)s_k^l + G_k^n(s_k^n)w_k^n, \quad (2.12a)$$

$$s_{k+1}^l = f_k^l(s_k^n) + A_k^l(s_k^n)s_k^l + G_k^l(s_k^n)w_k^l, \quad (2.12b)$$

$$o_k = h_k(s_k^n) + C_k(s_k^n)s_k^l + e_k, \quad (2.12c)$$

where the nonlinear component of the state is given by $s_k^n = \delta p_k$, and the linear substructure is $s_k^l = [\delta v_k, \delta \theta_k, \delta a_{\text{bias}_k}, \delta \omega_{\text{bias}_k}]^\top$. The variable o_k denotes the observation or measurement vector.

A discrete-time formulation of equation (2.11) with sampling time T_s can be obtained as detailed in [26]. In the specific case considered in this work, the model functions and matrices take the following form:

$$f_k^n(s_k^n) = s_k^n, \quad (2.13a)$$

$$A_k^n(s_k^n) = [T_s I_{3 \times 3}, 0_{3 \times 9}], \quad (2.13b)$$

$$G_k^n(s_k^n) = I_{3 \times 3}, \quad (2.13c)$$

$$f_k^l(s_k^n) = 0_{12 \times 1}, \quad (2.13d)$$

$$G_k^l(s_k^n) = I_{12 \times 12}, \quad (2.13e)$$

$$C_k(s_k^n) = 0_{(3M) \times 12}, \quad (2.13f)$$

while the function $h_k(s_k^n)$ does not admit an analytical closed-form representation due to the highly nonlinear projection geometry and photometric characteristics associated with the camera sensing process. This aspect will be discussed in greater detail in Section 2.4.

The state transition matrix for the linear substructure is expressed as:

$$A_k^l(s_k^n) = \begin{bmatrix} I_{3 \times 3} & -\bar{R}_b^i[a_m - \bar{a}_{\text{bias}}] \times T_s & -\bar{R}_b^i T_s & 0_{3 \times 3} \\ 0_{3 \times 3} & \bar{R}_b^i \{(\omega_m - \bar{\omega}_{\text{bias}})T_s\}^\top & 0_{3 \times 3} & -T_s I_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} & 0_{3 \times 3} \\ 0_{3 \times 3} & 0_{3 \times 3} & 0_{3 \times 3} & I_{3 \times 3} \end{bmatrix}, \quad (2.14)$$

where the notation $R\{\omega T_s\}$ denotes the rotation matrix corresponding to a rotation about the axis defined by the angular velocity vector ω through a rotation angle of magnitude $\|\omega\|T_s$.

The Marginalized Particle Filter (MPF) exploits the existence of states within the system that evolve according to linear dynamics. Through the process of Rao–Blackwellization

[28], the linear substructures of a nonlinear state–space model are analytically marginalized, resulting in a partition of the overall state vector into two distinct components: a nonlinear portion and a linear portion, as expressed in equation (2.15)

$$\mathbf{s}_k = \begin{bmatrix} \mathbf{s}_k^n \\ \mathbf{s}_k^l \end{bmatrix}. \quad (2.15)$$

With the state vector so partitioned, the MPF implements a particle filter to propagate and update the nonlinear states, while concurrently running a bank of Kalman filters—one associated with each particle—to estimate the evolution of the linear states. This decomposition allows computational resources to be concentrated on the propagation of nonlinear states, which are typically more demanding to approximate, while still rigorously incorporating the influence of other system variables through the linear-state Kalman filtering. The subsequent equations follow the formulation presented by [27].

Consider the discrete–time, nonlinear stochastic system:

$$\mathbf{s}_{k+1} = f(\mathbf{s}_k) + \mathbf{w}_k, \quad (2.16)$$

$$\mathbf{z}_k = h(\mathbf{s}_k) + \mathbf{v}_k, \quad (2.17)$$

where $\mathbf{w}_k$ and $\mathbf{v}_k$ are additive white Gaussian noise (AWGN) processes, and $f(\cdot)$ and $h(\cdot)$ denote the nonlinear state transition and measurement functions, respectively. For clarity, control inputs $\mathbf{u}_k$ have been omitted from this example.

If the nonlinear system described by equation (2.8)–equation (2.9) can be reformulated in the following block-structured form, then the MPF framework can be applied:

$$\mathbf{s}_{k+1}^n = \mathbf{f}_n(\mathbf{s}_k^n) + \mathbf{A}_n(\mathbf{s}_k^n)\mathbf{s}_k^l + \mathbf{w}_k^n, \quad (2.18)$$

$$\mathbf{s}_{k+1}^l = \mathbf{f}_l(\mathbf{s}_k^n) + \mathbf{A}_l(\mathbf{s}_k^n)\mathbf{s}_k^l + \mathbf{w}_k^l, \quad (2.19)$$

$$\mathbf{z}_k = h(\mathbf{s}_k^n) + \mathbf{C}(\mathbf{s}_k^n)\mathbf{s}_k^l + \mathbf{v}_k. \quad (2.20)$$

Equations equation (2.18)–equation (2.20) explicitly describe the dynamic coupling between the nonlinear and linear partitions of the state vector in equation (2.8). The function $\mathbf{f}_n$ characterizes the evolution of the nonlinear states between discrete time steps, while $\mathbf{f}_l$, $\mathbf{A}_n$, and $\mathbf{A}_l$ describe, respectively, the linear–nonlinear cross–couplings and self–couplings of the linear states. In compact form, this can be expressed as:

$$\begin{bmatrix} \mathbf{s}_{k+1}^n \\ \mathbf{s}_{k+1}^l \end{bmatrix} = \begin{bmatrix} \mathbf{f}_n & \mathbf{A}_n \\ \mathbf{f}_l & \mathbf{A}_l \end{bmatrix} \begin{bmatrix} \mathbf{s}_k^n \\ \mathbf{s}_k^l \end{bmatrix} + \begin{bmatrix} \mathbf{w}_k^n \\ \mathbf{w}_k^l \end{bmatrix}, \quad (2.21)$$

with

$$\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad (2.22)$$

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_n & \mathbf{Q}_{nl} \\ \mathbf{Q}_{ln} & \mathbf{Q}_l \end{bmatrix}, \quad (2.23)$$

$$\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}). \quad (2.24)$$

When applying the MPF, it is convenient to describe the procedure in terms of prediction and correction phases, analogous to the Kalman filter. A practical MPF implementation proceeds according to the following six steps:

STEP 1 – Filter Initialization and Particle Weighting (Update): Initialize the set of particle states by drawing i independent samples from a selected prior distribution $p(\mathbf{s}_0^n)$ —commonly Gaussian for convenience. Initialize all associated Kalman filters of the linear states with an identical initial mean and covariance:

$$\mathbf{s}_0^{n,i} \sim p(\mathbf{s}_0^n), \quad (2.25)$$

$$\mathbf{s}_0^{l,i} = \bar{\mathbf{s}}_0^l, \quad (2.26)$$

$$\mathbf{P}_0^i = \bar{\mathbf{P}}_0. \quad (2.27)$$

Using the prior distribution as the proposal density and assuming $p(\mathbf{z}_k | \mathbf{s}_k^i)$ is Gaussian, the particle weights are updated via:

$$w_k^i = w_{k-1}^i \exp \left(-\frac{1}{2} (\mathbf{E}_k^i)^T (\mathbf{M}_k^i)^{-1} (\mathbf{E}_k^i) \right), \quad (2.28)$$

where the innovation $\mathbf{E}_k^i$ and its covariance $\mathbf{M}_k^i$ are given by:

$$\mathbf{E}_k^i = \mathbf{z}_k - h(\mathbf{s}_k^{n,i}) - \mathbf{C} \mathbf{s}_k^{l,i}, \quad (2.29)$$

$$\mathbf{M}_k^i = \mathbf{C} \mathbf{P}_k^i \mathbf{C}^T + \mathbf{R}_k. \quad (2.30)$$

STEP 2 – Kalman Filter Mean and Covariance Update: For the linear states, which are Gaussian by construction, the measurement update is computed for each particle's Kalman filter:

$$\hat{\mathbf{s}}_k^l = \hat{\mathbf{s}}_{k-1}^l + \mathbf{K}_k (\mathbf{z}_k - h(\mathbf{s}_k^n) - \mathbf{C}_k \hat{\mathbf{s}}_{k-1}^l), \quad (2.31)$$

$$\mathbf{P}_k = \mathbf{P}_{k-1} - \mathbf{K}_k \mathbf{M}_k \mathbf{K}_k^T, \quad (2.32)$$

where

$$\mathbf{M}_k = \mathbf{C}_k \mathbf{P}_{k-1} \mathbf{C}_k^T + \mathbf{R}_k, \quad (2.33)$$

$$\mathbf{K}_k = \mathbf{P}_{k-1} \mathbf{C}_k^T \mathbf{M}_k^{-1}. \quad (2.34)$$

STEP 3 – Calculation of Filter Output: The overall system state estimate and covariance are obtained by separately aggregating the nonlinear and linear estimates:

$$\hat{\mathbf{s}}_k^n = \sum_{i=1}^N w_k^i \mathbf{s}_k^{n,i}, \quad (2.35)$$

$$\mathbf{P}_k^n = \sum_{i=1}^N w_k^i (\mathbf{s}_k^{n,i} - \hat{\mathbf{s}}_k^n)(\mathbf{s}_k^{n,i} - \hat{\mathbf{s}}_k^n)^T, \quad (2.36)$$

$$\hat{\mathbf{s}}_k^l = \sum_{i=1}^N w_k^i \mathbf{s}_k^{l,i}, \quad (2.37)$$

$$\mathbf{P}_k^l = \sum_{i=1}^N w_k^i \left(\mathbf{P}_k^{l,i} + (\mathbf{s}_k^{l,i} - \hat{\mathbf{s}}_k^l)(\mathbf{s}_k^{l,i} - \hat{\mathbf{s}}_k^l)^T \right). \quad (2.38)$$

Finally:

$$\hat{\mathbf{s}}_k = \begin{bmatrix} \hat{\mathbf{s}}_k^n \\ \hat{\mathbf{s}}_k^l \end{bmatrix}, \quad \mathbf{P}_k = \begin{bmatrix} \mathbf{P}_k^n & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_k^l \end{bmatrix}. \quad (2.39)$$

STEP 4 – Resampling: Resampling is triggered when the effective sample size

$$N_{eff} = \frac{1}{\sum_{i=1}^N (w_k^i)^2} \quad (2.40)$$

drops below a pre-defined threshold. Because each particle is associated with its own Kalman filter, resampling must replicate both the state means and covariances of the linear filters.

STEP 5 – Particle Propagation: The particle states are advanced according to:

$$\mathbf{s}_{k+1}^{n,i} = \mathbf{f}_k^n(\mathbf{s}_k^{n,i}) + \mathbf{A}_k^n \hat{\mathbf{s}}_k^{l,i} + c \sqrt{\mathbf{A}_k^n \mathbf{P}_k^{l,i} (\mathbf{A}_k^n)^T + \mathbf{Q}_k^n} \mathbf{W}_n, \quad (2.41)$$

where $c\sqrt{\cdot}$ denotes the Cholesky decomposition, and $\mathbf{W}_n$ is an $n \times 1$ random vector drawn from the standard normal distribution.

STEP 6 – Kalman Filter Propagation: For each particle's linear state:

$$\hat{\mathbf{s}}_{k+1}^l = \bar{\mathbf{A}}_k^l \hat{\mathbf{s}}_k^l + (\mathbf{Q}_k^l)^T (\mathbf{Q}_k^n)^{-1} \mathbf{z}_k + \mathbf{f}_k^l(\mathbf{s}_k^n) + \mathbf{L}_k (\mathbf{Z}_k - \mathbf{A}_k^n \hat{\mathbf{s}}_k^l), \quad (2.42)$$

$$\mathbf{P}_{k+1}^l = \bar{\mathbf{A}}_k^l \mathbf{P}_k^l (\bar{\mathbf{A}}_k^l)^T + \bar{\mathbf{Q}}_k^l - \mathbf{L}_k \mathbf{N}_k \mathbf{L}_k^T, \quad (2.43)$$

where

$$\mathbf{L}_k = \bar{\mathbf{A}}_k^l \mathbf{P}_k (\mathbf{A}_k^n)^T \mathbf{N}_k^{-1}, \quad (2.44)$$

$$\mathbf{N}_k = \mathbf{A}_k^n \mathbf{P}_k (\mathbf{A}_k^n)^T + \mathbf{Q}_k^n, \quad (2.45)$$

and

$$\bar{\mathbf{A}}_k^l = \mathbf{A}_k^l - (\mathbf{Q}_k^{ln})^T (\mathbf{Q}_k^n)^{-1} \mathbf{A}_k^n, \quad (2.46)$$

$$\bar{\mathbf{Q}}_k^l = \mathbf{Q}_k^l - (\mathbf{Q}_k^{ln})^T (\mathbf{Q}_k^n)^{-1} \mathbf{Q}_k^{ln}. \quad (2.47)$$

Special Case – Shared Covariance Structure: In the general marginalized particle filter (MPF), each particle is associated with an individual Kalman filter characterized by its own mean and covariance. A useful simplification arises when the system matrices $\mathbf{A}_n$, $\mathbf{A}^l$, and $\mathbf{C}$ are independent of the nonlinear state $\mathbf{s}_n$, as is the case in our implementation. Under this assumption, a common covariance matrix $\mathbf{P}$ can be employed to represent the uncertainty of the linear states across all particles. A comprehensive treatment of this formulation, including theoretical foundations and algorithmic implications, is provided in [27].

2.4 Measurement Model and Likelihood Calculation

In order to incorporate the particle-specific measurement function h defined in equation (2.12), the methodology developed in this work utilizes each particle's hypothesized spatial position and orientation to project a corresponding view onto a pre-acquired satellite image of the operational region. This is achieved by emulating the UAV camera measurement model c introduced in equation (2.48). Considering that the UAV camera is mounted on a stabilized gimbal and captures nadir (orthogonally oriented) imagery, the pitch and roll components of the particle state can be neglected without significant loss of fidelity. In addition, barometric altitude measurements are incorporated to constrain and stabilize the vertical channel of the Inertial Navigation System (INS).

Under these operational assumptions, the image segment associated with each particle hypothesis can be extracted—commonly referred to as “cropping”—from the offline satellite map. This extraction process employs the particle's estimated horizontal coordinates (x, y) , its yaw angle (ψ) , and its altitude (z) to determine both the

appropriate crop location and scaling. The scaling is altitude-dependent and is governed by the measurement function h formulated in equation (2.49).

$$C_{image} = c(x, y, z, \psi, K, n_s), \quad (2.48)$$

where C_{image} denotes the actual image captured by the UAV camera, K is the 3×3 intrinsic camera calibration matrix, and n_s represents extrinsic sensor noise, which may originate from illumination variations, motion blur, perspective distortions, parallax effects, and similar imaging artifacts.

$$P_{view} = h(x, y, \psi, s(z), S_{map}, [w, h]), \quad (2.49)$$

In equation (2.49), P_{view} denotes the synthetic particle view extracted from the satellite image S_{map} , $s(\cdot)$ is a scaling function that adjusts S_{map} proportionally to the altitude z , and $[w, h]$ specifies the resolution of the cropped image, which is determined from the intrinsic matrix K .

To evaluate the measurement likelihood, the error signal between the UAV camera image and the particle-projected satellite image must be transformed into a probabilistic measure. Effective image comparison relies on the selection of robust similarity metrics, where higher similarity corresponds to lower error values. In this study, a feature descriptor-based image similarity approach is employed. Distinctive local keypoints, representing salient image structures, are first extracted. Traditional keypoint detectors such as ORB [16] and SURF [29] produce low-dimensional descriptors, whereas more recent convolutional neural network (CNN)-based approaches, e.g., SuperPoint [17] and R2D2 [30], yield high-dimensional descriptors. Given the computational demands associated with evaluating a large number of particles, the ORB detector is adopted in this work owing to its favorable balance between computational efficiency and matching performance.

Following keypoint and descriptor extraction from both images, descriptor matching is performed using efficient search strategies, with the Approximate Nearest Neighbors (ANN) method selected for its speed and scalability. Outlier matches are pruned via the RANSAC algorithm, ensuring geometric consistency. The cardinality of the remaining “good” matches serves as the primary similarity metric; a larger number of good matches

signifies higher similarity between the two images. An important property of this metric is its inherent rotation invariance, which supports robust matching even in the presence of image rotation. Within the particle filtering framework [31], particle weights are

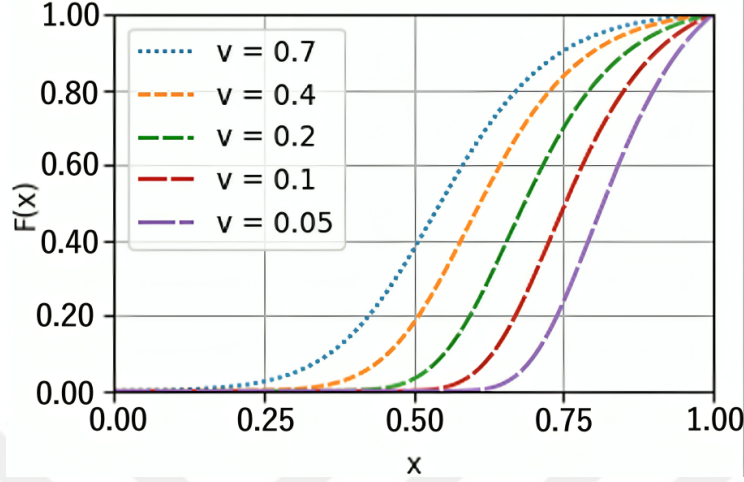


Figure 2.2: Custom Likelihood Function $F(x)$ for various hyperparameter values v .

computed from the likelihood values. Therefore, the design of the likelihood function must promote the survival of particles whose hypothesized states align closely with the observations. To this end, we construct a custom likelihood function inspired by [32], which maps the number of good matched keypoints into a normalized likelihood value in the range $[0, 1]$. Let M_t denote an $N \times 1$ vector containing the number of good matches for each of the N particles at time t . First, M_t is normalized by its maximum value $\text{argmax}(M_t)$, bounding the entries to $[0, 1]$. The normalized values are then transformed via the logistic mapping given in equation (2.50), which assigns low likelihoods to particles with few matches and higher likelihoods to those with more matches.

$$l(x) = \frac{(1 + e^{-5})^{\frac{1}{v}}}{(1 + e^{(-10x+5)})^{\frac{1}{v}}}, \quad (2.50)$$

where v is a hyperparameter controlling the curvature of the likelihood function. Lower values of v shift the curve such that particles with a small number of matches receive likelihood values close to zero, thereby enforcing a more selective weighting strategy. The effect of different v values on the likelihood curve is illustrated in Figure 2.2.

2.5 Image Domain Adaptation/Synchronization

In the context of this work, the fundamental objective is to enable a reliable comparison between an image captured in real time by the UAV's onboard camera and a corresponding view extracted from an offline satellite map. These two images inherently originate from different domains, and thus direct comparison using conventional similarity metrics presents a significant challenge. The primary difficulty arises from substantial discrepancies between the two domains caused by factors such as illumination changes, seasonal variations, temporal acquisition differences, and differences in imaging geometry, including viewing angles. These variations can cause extracted feature descriptors to diverge considerably, even when the two images depict the same physical scene, which in turn degrades matching performance and reduces the accuracy of similarity estimation.

To mitigate this domain-induced discrepancy, we introduce a domain adaptation framework based on the Cycle-Consistent Generative Adversarial Network (CycleGAN) architecture [33]. This approach facilitates the transformation of images from one domain to another while ensuring that their essential structural characteristics remain preserved. The adaptation is accomplished by learning a bidirectional mapping between the two domains, thereby aligning their visual representations and enhancing the compatibility of feature descriptors for similarity computation.

We define a domain adaptation operator $H_{X \rightarrow Y}(x)$, trained using CycleGAN, which transforms an image $x \in X$ from domain X into its corresponding representation $y \in Y$ in domain Y , where X and Y denote two domains with substantial visual or structural differences. Such a transformation is particularly advantageous when systematic inter-domain variations adversely affect feature extraction and matching. By learning both forward and backward mappings between the domains, the model ensures that feature consistency is maintained, even after undergoing domain transformations.

The CycleGAN architecture consists of two generator–discriminator pairs and employs a *cycle consistency loss* to enforce bidirectional transformations that preserve spatial and semantic structure.

Each generator G learns to map images between the two domains:

$$G_{X \rightarrow Y} : X \rightarrow Y, \quad G_{Y \rightarrow X} : Y \rightarrow X, \quad (2.51)$$

where $G_{X \rightarrow Y}$ translates an image from domain X to domain Y , and $G_{Y \rightarrow X}$ performs the reverse transformation.

Each discriminator D is trained to differentiate between real images and those synthesized by the generator:

$$D_Y(y) \approx 1, \quad D_Y(G_{X \rightarrow Y}(x)) \approx 0, \quad (2.52)$$

where D_Y classifies a real image $y \in Y$ as authentic (label 1), and a generated image $G_{X \rightarrow Y}(x)$ as fake (label 0).

The adversarial loss for the transformation $X \rightarrow Y$ encourages the generator $G_{X \rightarrow Y}$ to produce images that are indistinguishable from authentic images in domain Y :

$$\begin{aligned} \mathcal{L}_{\text{GAN}}(G_{X \rightarrow Y}, D_Y, X, Y) = & \mathbb{E}_{y \sim P_Y(y)} [\log D_Y(y)] \\ & + \mathbb{E}_{x \sim P_X(x)} [\log(1 - D_Y(G_{X \rightarrow Y}(x)))]. \end{aligned} \quad (2.53)$$

A symmetric adversarial loss is similarly defined for the reverse transformation $G_{Y \rightarrow X}$.

To ensure reversibility and structural preservation, the *cycle consistency loss* enforces the requirement that translating an image from one domain to the other and then back should yield a reconstruction close to the original:

$$\begin{aligned} \mathcal{L}_{\text{cyc}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}) = & \mathbb{E}_{x \sim P_X(x)} [\|G_{Y \rightarrow X}(G_{X \rightarrow Y}(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim P_Y(y)} [\|G_{X \rightarrow Y}(G_{Y \rightarrow X}(y)) - y\|_1]. \end{aligned} \quad (2.54)$$

This constraint discourages transformations that would otherwise result in excessive distortions or the loss of crucial image information.

The complete CycleGAN objective function is obtained by combining the adversarial losses for both transformation directions with the cycle consistency loss:

$$\begin{aligned} \mathcal{L}_{\text{CycleGAN}} = & \mathcal{L}_{\text{GAN}}(G_{X \rightarrow Y}, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(G_{Y \rightarrow X}, D_X, Y, X) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G_{X \rightarrow Y}, G_{Y \rightarrow X}), \end{aligned} \quad (2.55)$$

where λ is a tunable hyperparameter controlling the relative contribution of the cycle consistency term.

This formulation enables the CycleGAN to transform images between visually distinct domains in a way that harmonizes their representations for feature-based comparison, while preserving essential structural and semantic content necessary for robust localization and similarity evaluation.

2.6 Data Collection

Obtaining perfectly aligned and geometrically orthonormal photographs of the exact same geographical scene, captured by a UAV under identical conditions but at different temporal instances, is at present an exceedingly difficult—if not practically impossible—task. This inherent limitation in acquiring temporally diverse yet spatially consistent datasets significantly constrains the development of paired-domain learning strategies. One of the most notable advantages of unpaired image-to-image translation frameworks is the relative ease with which training data can be gathered, since such methods do not require strict one-to-one correspondences between source and target images. Leveraging this property, in the present work, we systematically sourced domain-specific UAV imagery from publicly available online photographic archives, thereby bypassing the need for labor-intensive paired data acquisition campaigns.

In the scope of this research, our study focuses on three distinct visual domains: *Summer*, *Winter*, and *Satellite*. The motivation for selecting these particular domains stems from the need to explicitly model seasonal variations in UAV orthogonal imagery, while also addressing the substantial visual discrepancy between UAV-captured scenes and satellite imagery. Due to the notable absence of sufficiently diverse and high-quality datasets in the literature for both Summer and Winter nadir-oriented UAV images, the datasets for these two domains were compiled manually by exhaustively searching and curating imagery from internet sources. This process required careful selection to ensure that the images met the necessary orthonormal viewpoint criteria and exhibited minimal perspective distortion, thereby maintaining suitability for training robust image translation models.

For the Satellite image domain, we employed Google Earth Pro’s data acquisition tools to systematically capture imagery from various geographic locations and across multiple altitude levels. However, it is important to emphasize that the satellite image domain lacks global visual consistency because of significant regional differences in capture conditions, illumination, and seasonal appearance. To reduce the variability introduced by such global heterogeneity and to achieve a more coherent domain representation, the satellite dataset was constrained exclusively to imagery from the geographical

boundaries of Turkey. This selection criterion was further motivated by the fact that the majority of high-resolution satellite images for this region are captured during the Summer season, thus ensuring a more homogeneous visual style within the dataset.

Representative example images from each of the three domains—Summer, Winter, and Satellite—are presented in Figure 2.3, illustrating the diversity and distinct visual characteristics inherent to each domain.

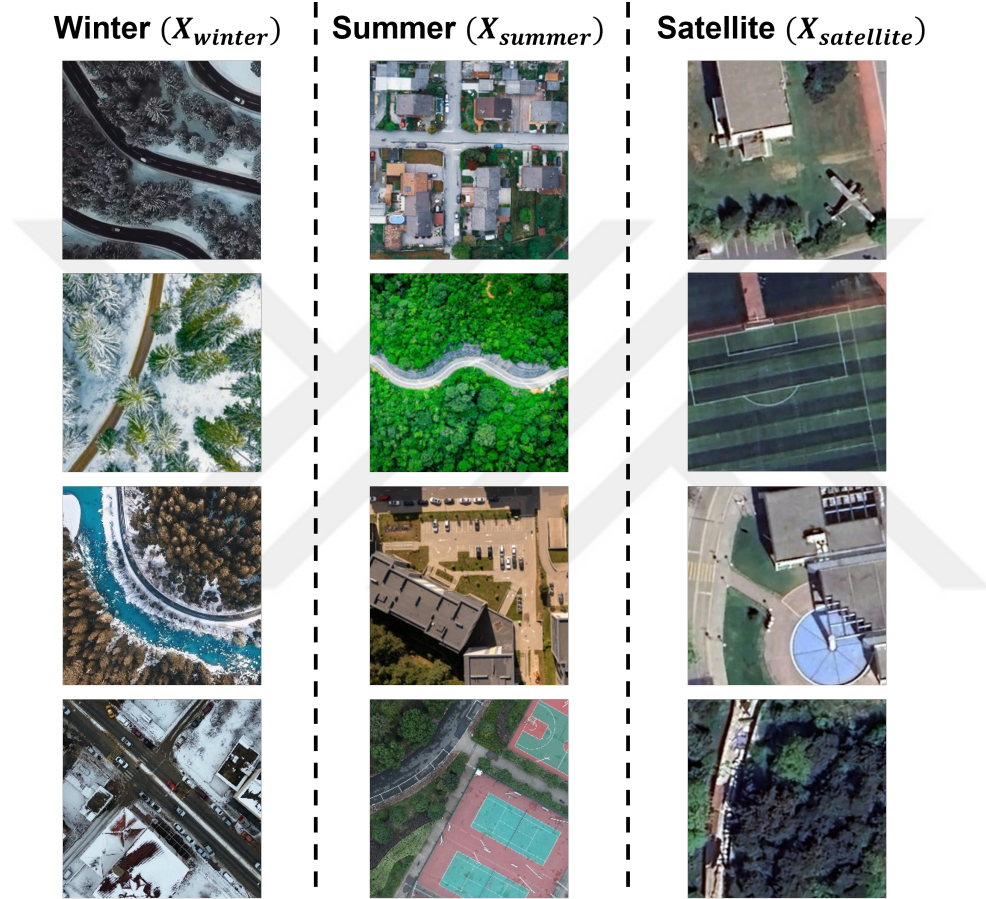


Figure 2.3: Sample images from different domains

3. RESULTS AND DISCUSSIONS

3.1 Evaluation of CycleGAN Domain Adaptation Performance

In the present study, three distinct generative adversarial network (GAN)–based architectures for unpaired image-to-image translation were trained and rigorously evaluated for their effectiveness in the domain adaptation task. These models include: the original Cycle-Consistent Generative Adversarial Network (CycleGAN) [23], the Contrastive Unpaired Translation (CUT) framework [34], and the recently proposed Turbo-CycleGAN architecture [33]. Each of these models is founded on the general principle of adversarial learning for unpaired image translation but differs considerably in terms of structural design, learning objective formulation, and optimization strategy, which in turn directly influence their domain adaptation capabilities.

The baseline CycleGAN model employs a two-way generator–discriminator setup combined with a cycle consistency loss, enabling the mapping between two domains without requiring paired datasets. This model has been widely used due to its simplicity and its capability to preserve structural information. However, the original CycleGAN architecture often suffers from the generation of repetitive patterns or local visual artifacts, particularly in high-frequency regions of the image, which may adversely affect subsequent feature matching.

CUT, in contrast, builds upon the CycleGAN philosophy but introduces a contrastive learning objective to enforce patch-wise correspondence between the source and translated images in the feature space of the generator. This approach, known as PatchNCE loss, aims to preserve semantic similarity across corresponding image patches without the need for a second generator or explicit cycle consistency. By maximizing mutual information between aligned features, CUT generally produces sharper and more detailed outputs with fewer structural distortions compared to standard CycleGAN. Nevertheless, like most patch-level optimization methods, CUT may still

introduce subtle high-frequency artifacts, especially in visually complex regions, which can reduce the repeatability of detected features.

Turbo-CycleGAN represents an advancement over both CycleGAN and CUT by integrating additional architectural refinements, improved normalization layers, and multi-scale feature integration strategies. Furthermore, Turbo-CycleGAN employs enhanced cycle consistency mechanisms and generator regularization techniques that aim to minimize perceptual artifacts while improving texture fidelity and style transfer quality. This architecture, however, achieves its superior performance at the expense of significantly higher computational requirements, both in terms of memory footprint and training time. Due to its multi-stage and feature-intensive design, Turbo-CycleGAN produces substantially larger model files and exhibits longer inference latency, which can be a limiting factor for deployment on computationally constrained platforms such as small UAVs.



Figure 3.1: Demonstration of a trajectory of UAV point A to B in area of 400m x 400m satellite view

Domain adaptation performance in this context is of critical importance for accurate global position estimation in GNSS-denied navigation scenarios. A thorough understanding of the capabilities and limitations of these image-to-image translation models is thus essential. Selecting an appropriate and reliable evaluation metric

poses a challenge because CycleGAN-based methods operate under an unsupervised learning paradigm, where paired ground truth images for quantitative comparison are not available. To address this, a custom evaluation methodology was devised. The method quantifies translation performance by calculating the mean number of matched feature correspondences across 100 query images that is collected from a trajectory shown in Figure 3.1 for each GAN model. The relative improvement in feature matching performance is then expressed as a percentage increase, with the number of matched keypoints in the original, unadapted images serving as the baseline reference.

In addition to measuring translation quality, the models were compared in terms of computational properties highly relevant to real-time UAV operation. These include total training time, model size, and inference time, which collectively determine the feasibility of deploying each architecture onboard a small UAV. All training and benchmarking experiments were conducted using an NVIDIA RTX 4090 GPU running on a Linux-based system, ensuring consistent computational conditions for fair comparison.

The quantitative results of this comparative study are presented in Table 3.1, which summarizes the trade-offs between translation accuracy, model complexity, and computational efficiency for the three models in the context of UAV-based domain adaptation.

Table 3.1: Comparison of GAN models for the domain adaptation task in UAV image translation. Accuracy values are expressed as the relative increase in matched feature correspondences compared to the baseline (unaltered images).

Model	Training Time	Model Size	Inference Time	Accuracy
CycleGAN	6 hours	50 MB	0.01 s	+10%
CUT	24 hours	100 MB	0.01 s	+15%
Turbo-CycleGAN	72 hours	1500 MB	0.15 s	+38%

A set of representative qualitative results comparing the visual outputs of the CUT and Turbo-CycleGAN models is shown in Figure 3.2. From the perspective of feature-based matching, one of the most critical aspects of evaluating image-to-image translation performance is the extent to which the generator introduces artifacts into the translated image. Both the original CycleGAN and the CUT framework tend to produce minor

visual inconsistencies or artifacts, which can negatively affect the robustness and repeatability of keypoint matching algorithms. Turbo-CycleGAN, on the other hand, demonstrates a substantial reduction in such artifacts, leading to a marked improvement in the preservation of distinctive features and thereby enhancing the overall feature matching performance.

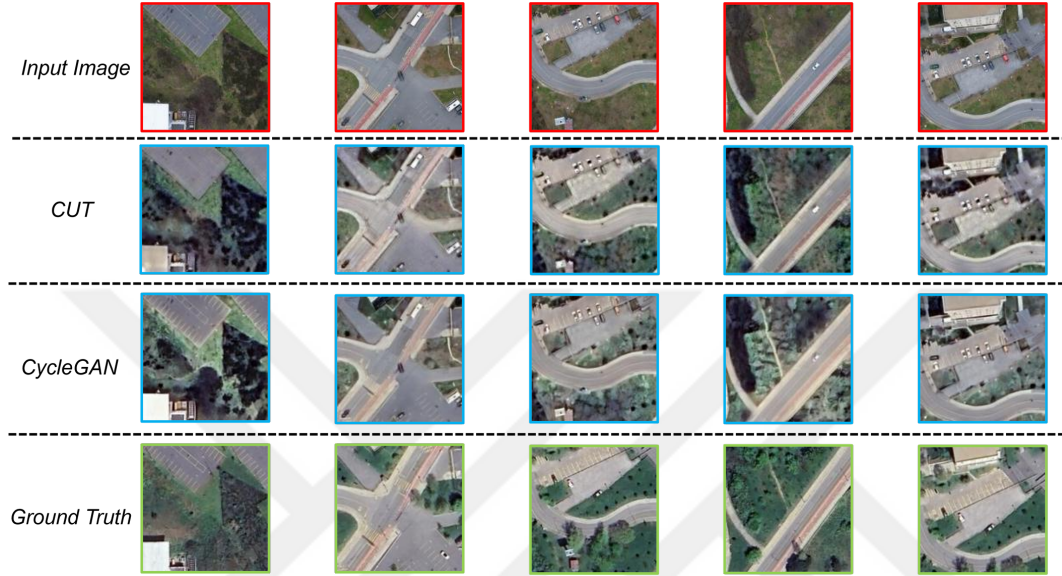


Figure 3.2: Different GAN model image to image translation results with ground truth label

3.2 Flight Test Results

In this section, we provide a detailed and comprehensive presentation of the performance evaluation results for the proposed domain adaptation–assisted navigation method, systematically comparing its performance against a baseline configuration in which no domain adaptation is applied. The purpose of this comparative analysis is to explicitly quantify the benefits of the proposed approach under a variety of operational conditions, with the goal of illustrating its robustness and adaptability in realistic UAV navigation scenarios. For that testing purpose, the UAV platform employed in the experiments is shown in Figure 3.3, while the specifications of the onboard camera system and the MPU-9250 IMU are summarized in Table 3.3 and Table 3.2, respectively.

To ensure that the evaluation encompasses a broad spectrum of environmental and mission-specific challenges, four distinct flight trajectory scenarios have been carefully



Figure 3.3: The small UAV prepared for the flight tests carrying down-facing onboard camera and IMU

selected. Each scenario has been designed to capture different combinations of domain gap magnitude, flight altitude, spatial feature distribution, and total trajectory length, thereby allowing for a nuanced examination of the proposed method’s strengths and limitations The scenarios under consideration are defined as follows:

- **Scenario 1:** Big domain gap (UAV flights in Winter/Satellite Map are in Summer), Low Altitude, High feature uniqueness, 1000 m trajectory
- **Scenario 2:** Low domain gap (UAV flights in Summer/Satellite Map are in Summer), High Altitude, High feature uniqueness, 2800 m trajectory
- **Scenario 3:** Low domain gap (UAV flights in Summer/Satellite Map are in Summer), Low Altitude, Low feature uniqueness, 1200 m trajectory
- **Scenario 4:** Low domain gap (UAV flights in Summer/Satellite Map are in Summer), Low Altitude, High feature uniqueness, 1500 m trajectory

3.2.1 Scenario 1: relatively higher domain gap

Table 3.2: MPU-9250 IMU specifications.

Specification	Value	Unit
Gyroscope		
Bias Instability	8.7300×10^{-2}	rad/s
Angular Random Walk	1.7453×10^{-4}	$rad/s/\sqrt{Hz}$
Accelerometer		
Bias Instability	6.8670×10^{-1}	m/s^2
Velocity Random Walk	2.9430×10^{-3}	$m/s^2/\sqrt{Hz}$

Table 3.3: Camera specifications.

Specification	Value
Resolution	4000×2250 pixels
Focal Length	4 mm
Sensor Size	1/1.3 inch
Field of View (FoV)	82.1°
Frame Rate	30 fps
Stabilization	3-axis Gimbal

In this particular scenario, the experimental setup is intentionally designed to represent a challenging condition by introducing a relatively large domain gap between the onboard UAV imagery and the reference satellite map. Specifically, the UAV flight data is collected during the Winter season, thereby creating substantial visual discrepancies with the satellite imagery, which corresponds to the Summer season. This setting provides a rigorous test case for assessing the performance of the CycleGAN-based domain adaptation framework under conditions of pronounced seasonal variation, including substantial changes in ground texture, vegetation cover, and illumination patterns. The evaluation process consists of comparing the localization results obtained using the raw UAV images against those obtained after applying the CycleGAN model to generate synthetic Summer-like UAV images, which are then employed for feature-based matching.

As can be observed in Figure 3.4, when the localization pipeline utilizes the raw UAV imagery captured in Winter, the position estimates exhibit noticeable divergence in both the North and East coordinates. This phenomenon arises because large inter-domain differences in the matched imagery lead to either premature convergence towards incorrect map regions or substantially prolonged convergence times before reaching

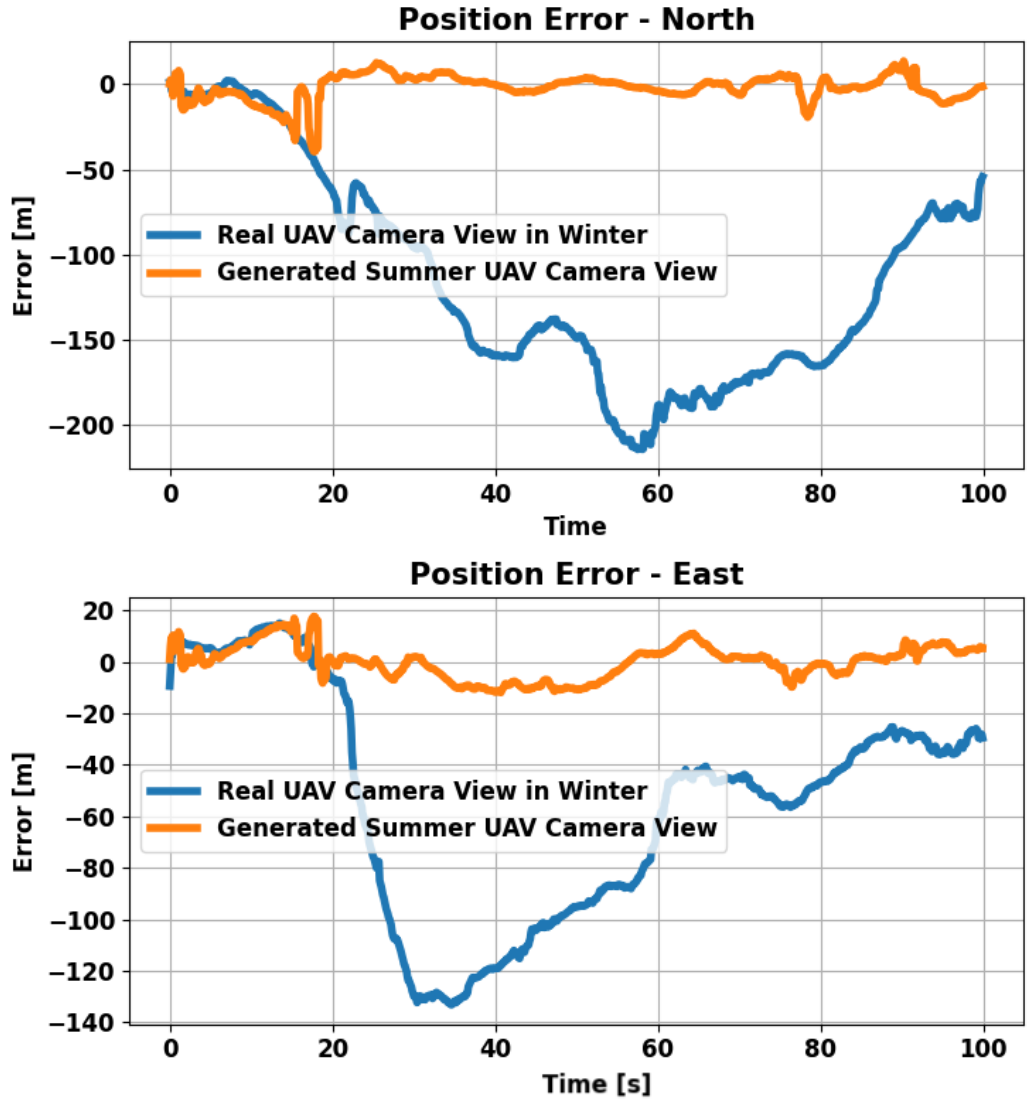


Figure 3.4: Scenario 1 position error: Big domain gap (UAV flights in winter/satellite map are in summer), low altitude, high feature uniqueness, 1000 m trajectory

the correct location. In situations where such significant domain discrepancies occur, the robustness of feature-based localization within a particle filter framework can be enhanced by jointly applying two strategies: firstly, increasing the number of particles to allow for broader hypothesis coverage, and secondly, tuning the hyperparameter ν in equation (2.50) to a lower value, thereby reducing the likelihood of premature and erroneous particle convergence towards incorrect matches.

In addition to the baseline results obtained using raw UAV imagery, the application of the proposed CycleGAN-based domain adaptation demonstrates highly promising improvements in scenarios with extreme domain mismatch, as illustrated in Figure 3.5.

For the purposes of visual clarity and to enable a direct qualitative comparison, the maximum number of keypoints extracted is limited to 50 in these experiments. This constraint makes it readily apparent that the LightGlue feature matcher consistently identifies a greater number of correspondences in the domain-adapted imagery than in the raw Winter images. Consequently, as demonstrated in Figure 3.4, the incorporation of domain synchronization via CycleGAN leads to significantly improved global localization performance in the presence of substantial seasonal differences.

3.2.2 Scenario 2: relatively higher altitude

Operating at significantly higher altitudes introduces notable challenges for both Visual-Inertial Odometry (VIO) and feature-based image matching, as the quality and observability of visual measurements degrade substantially under these conditions. The limited spatial resolution of the onboard UAV camera becomes a major limiting factor, as the projection of environmental features onto the image plane results in a greatly reduced apparent motion when the UAV is far from the ground surface. Consequently, the motion of distinct features becomes increasingly unobservable, leading to weaker visual constraints in the VIO estimation process. Furthermore, sensor noise—particularly noise in the camera image sensor that manifests at the pixel level—tends to have a more pronounced impact on distant objects, especially those located away from the principal point of the camera’s optical axis. This degradation in visual fidelity amplifies the difficulty of accurately extracting and matching features between successive frames or between UAV imagery and satellite reference images.

From the perspective of the feature matching process, high-altitude flights exacerbate the difficulty of capturing and resolving small pixel displacements that correspond to the UAV’s motion. At elevated altitudes, even a displacement of a single pixel in the image can correspond to a translation of several hundred meters in the physical world. This scaling effect significantly diminishes the sensitivity of pixel-level feature matching to UAV motion, making it difficult for the algorithm to detect meaningful correspondences. Moreover, the fixed number of particles used in the particle filter further limits the achievable resolution in the global localization output, as the search space covered by each particle becomes proportionally coarser at higher altitudes.

Real UAV vs. Real SAT
Inlier Matches: 8



Figure 3.5a: Raw UAV images captured in Winter (left) and satellite reference image (right).

GEN SAT vs. Real SAT
Inlier Matches: 15

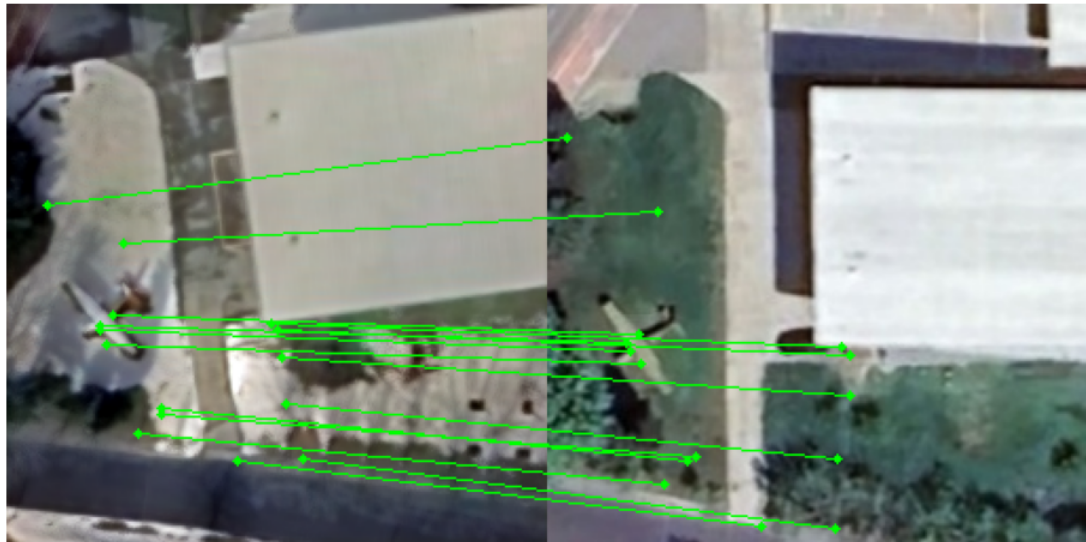


Figure 3.5b: CycleGAN-generated images (left) and satellite reference image (right).

Figure 3.5: Image matching performance under high domain discrepancy

As illustrated in Figure 3.6, the detrimental effects of operating at higher altitudes are evident in the elevated position errors observed for both the raw UAV images and their CycleGAN-based domain-adapted counterparts. Nevertheless, despite the inherent challenges of reduced feature observability and increased image noise, the application of

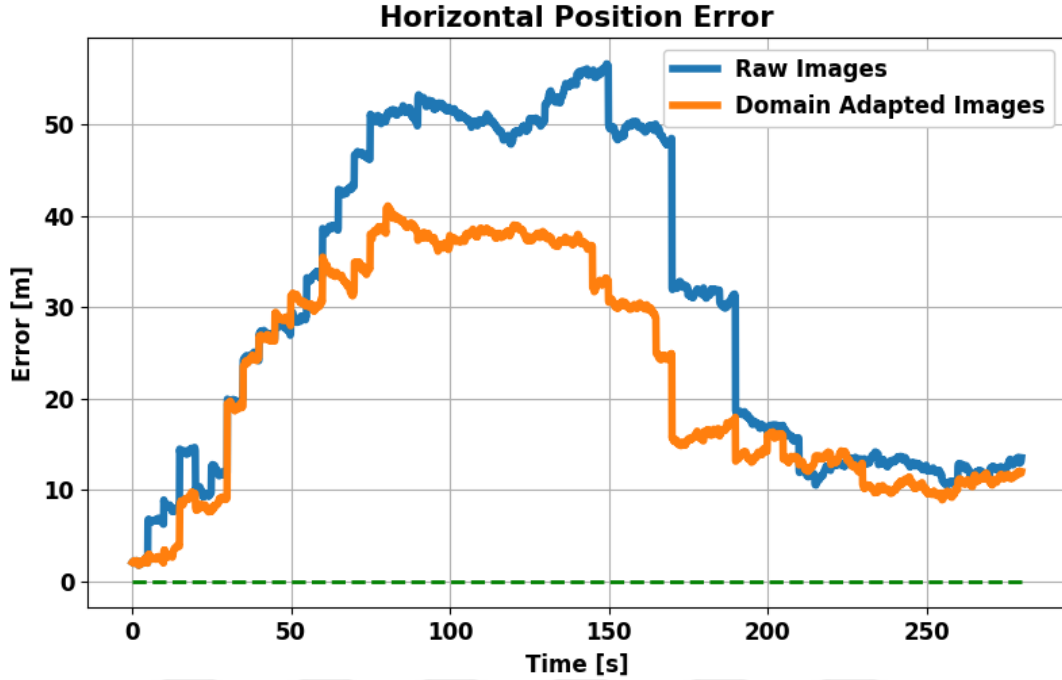


Figure 3.6: Scenario 2 position Error: low domain gap (UAV flights in summer/satellite map are in iummer), high altitude, high feature uniqueness, 2800 m trajectory

CycleGAN still yields a measurable improvement over the baseline performance using unprocessed imagery. When compared with low-altitude scenarios, the absolute number of matched features at high altitudes is reduced to near-zero levels, highlighting the extreme sparsity of viable correspondences. However, the synthetic images generated through CycleGAN retain a slight advantage, enabling the extraction of a small but meaningful increase in the number of matched keypoints. This modest improvement in available matches translates into a tangible enhancement in localization accuracy, with the CycleGAN-based approach achieving approximately 15 meters lower positioning error compared to the baseline in Scenario 2.

3.2.3 Scenario 3: relatively low feature uniqueness

Environments characterized by a low degree of feature uniqueness impose challenges that are, in many respects, analogous to those encountered at higher altitudes, in that both conditions lead to a degradation in the accuracy of Visual-Inertial Odometry (VIO) and terrain-based image matching. In scenarios where the environment contains repetitive, symmetric, or homogeneously textured patterns, the ability to extract and match distinctive visual features is substantially reduced. Representative examples of

such low-uniqueness environments include dense forest canopies where tree structures are visually similar over large areas, expansive marine environments with minimal surface texture variation, and mountainous terrains with repetitive ridge patterns and rock formations. In these contexts, sparse feature matching algorithms—particularly those relying on the detection of unique, high-contrast keypoints—are disproportionately affected, as the local visual neighborhoods around candidate features exhibit frequent and recurring similarities that lead to ambiguity in correspondence establishment.

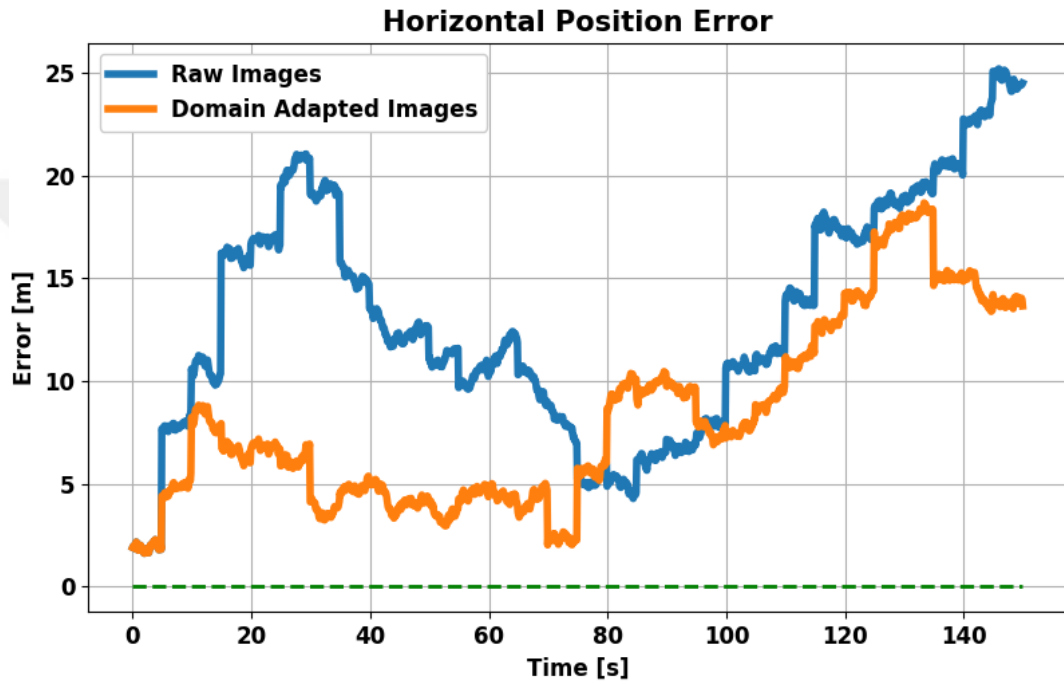


Figure 3.7: Scenario 3 position Error: low domain gap (UAV flights in summer/satellite map are in summer), low altitude, low feature uniqueness, 1200 m trajectory

For the evaluation of this scenario, a UAV flight trajectory was deliberately selected such that a significant portion of the mission was conducted over dense forested regions, yet at relatively low altitudes in order to partially mitigate altitude-related visual degradation effects. As anticipated, the localization errors in this setting are noticeably larger than those observed in more optimal conditions with high feature uniqueness. This degradation is primarily attributable to the scarcity of distinctive visual landmarks, which complicates both the VIO tracking process and the establishment of reliable correspondences in terrain matching.

Nevertheless, as depicted in Figure 3.7, the application of CycleGAN-based domain adaptation to transform raw UAV images into visually harmonized representations

results in a tangible improvement over the baseline performance obtained with unprocessed imagery. The robustness of the CycleGAN-generated images is particularly evident during the initial segment of the flight, specifically within the time interval between 0 and 60 seconds, when the UAV is traversing a region of high visual symmetry. In such challenging conditions, the domain-adapted images provide a modest but critical increase in the number of correctly matched features, thereby enabling the particle filter to maintain a more stable and accurate global position estimate compared to the raw image-based approach.

3.2.4 Scenario 4: decent flight setup for image matching

In this scenario, the environmental conditions and the parameters governing the UAV flight trajectory were deliberately configured so as to align closely with the optimal requirements for both Visual-Inertial Odometry (VIO) and feature-based image matching processes. Specifically, the mission was planned to ensure low-altitude flight over terrain exhibiting a high degree of visual distinctiveness and minimal repetitive patterns, thereby maximizing the potential for extracting a large number of stable and spatially well-distributed keypoints. Such conditions inherently enhance the robustness of both the motion estimation pipeline and the subsequent image-to-map correspondence stage, which is critical for accurate GPS-denied navigation.

The choice of flight path, seasonal alignment between UAV imagery and satellite maps, and the inherent richness of environmental features collectively result in conditions that are highly favorable for the performance of the proposed domain adaptation methodology. Under these circumstances, the inherent limitations posed by sensor noise, image resolution constraints, and domain discrepancies are significantly reduced. Consequently, the particle filter-based localization process operates with a more reliable and consistent set of measurement updates, minimizing the likelihood of particle divergence or false convergence.

As can be observed from Figure 3.8, the GPS-denied visual navigation framework exhibits highly promising results when operating under these optimal conditions. The alignment between the seasonal characteristics of the UAV imagery and the satellite dataset ensures that the domain gap is minimal, thereby reducing the need for extensive

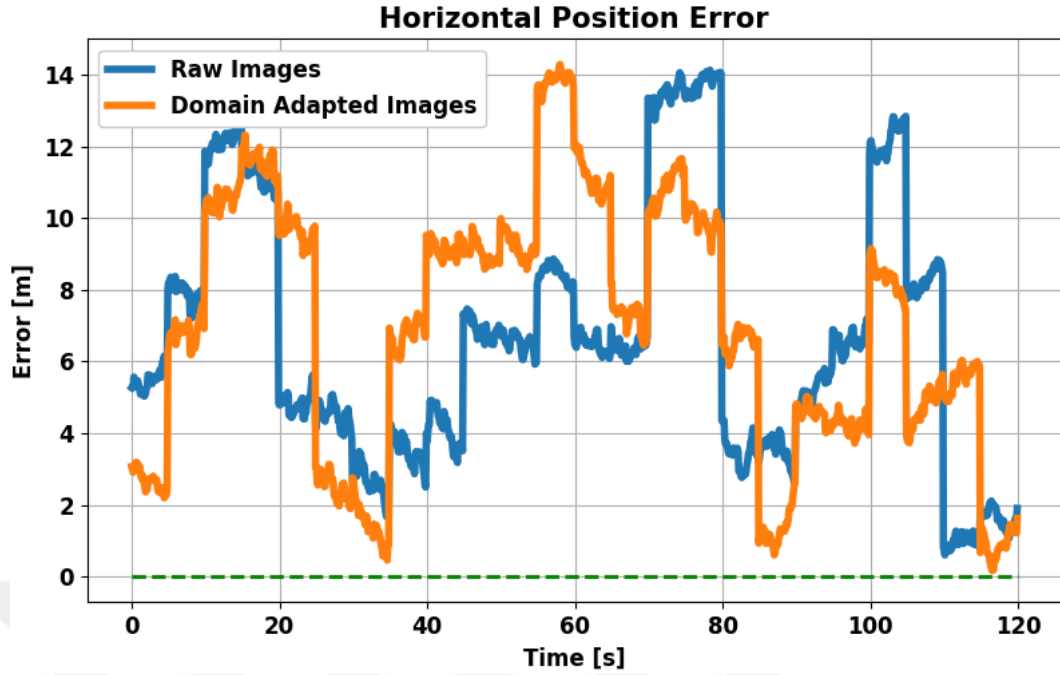


Figure 3.8: Scenario 4 position error: low domain gap (UAV flights in summer/satellite map are in summer), low altitude, high feature uniqueness, 1500 m trajectory

adaptation. Nevertheless, the proposed CycleGAN-based domain adaptation method further enhances performance by refining visual consistency and mitigating any residual differences in lighting, contrast, or texture distribution. This refinement enables the feature matching process to operate with higher precision, producing a greater number of correct correspondences and, in turn, leading to more accurate localization outputs. The combination of favorable environmental conditions and the application of the proposed domain synchronization framework underscores the system's capability to achieve near-optimal performance, suggesting that in practical deployment scenarios, when mission planning can ensure such conditions, the method could consistently deliver precise and stable navigation solutions.



4. FUTURE WORKS

The scope of the present study has primarily been focused on improving GPS-denied UAV localization performance through visual domain adaptation techniques, with emphasis on addressing seasonal and environmental discrepancies between UAV-captured imagery and satellite references. While the results obtained have demonstrated promising performance across various operational scenarios, there remain numerous opportunities to extend and refine the proposed approach in order to broaden its applicability, enhance robustness, and increase efficiency. One significant direction for future research is the investigation of worst-case operational scenarios for Visual-Inertial Odometry (VIO), particularly in low-light and night-time environments. Under such conditions, the reduction in illumination severely impacts the visibility and detectability of distinctive features, leading to degraded performance in both VIO and image-to-map matching. A promising line of work involves developing a specialized CycleGAN-based domain adaptation framework tailored specifically for transforming night-time UAV imagery into equivalent day-time representations. This would enable more effective feature extraction and matching against day-time satellite or aerial imagery, mitigating the adverse effects of low-light conditions. Achieving this will likely require incorporating physics-inspired constraints and training datasets that are explicitly designed to account for varying lighting conditions, sensor noise, and long exposure artifacts, in order to ensure that the generated synthetic day-time imagery retains geometric and photometric consistency for reliable feature correspondence.

Another area for potential enhancement lies in the consideration of camera mounting configurations. In the present work, the UAV camera system was mounted on a gimbal-stabilized platform, which actively compensates for roll and pitch motion of the airframe, thereby maintaining the nadir (orthonormal) orientation of captured images. This significantly simplifies the image matching process, as the majority of frames remain free from strong perspective distortions. However, for lower-cost UAV platforms

where a gimbal system may be impractical or unnecessary, image frames may be subject to noticeable off-nadir orientation, introducing geometric distortions and scale variations across the field of view. Instead of relying solely on expensive hardware stabilization, future research could focus on developing robust image rectification techniques capable of compensating for such off-nadir or oblique imagery prior to feature extraction and matching. This correction could be achieved through conventional photogrammetric transformations based on estimated attitude angles, or through learning-based methods capable of directly inferring rectified representations from raw inputs. The latter approach, if designed carefully, could operate in real time and adaptively handle variable degrees of orientation disturbance without significant loss of spatial resolution.

From a computational efficiency perspective, the use of Turbo-CycleGAN, while delivering superior domain adaptation quality, imposes substantial demands on both memory and processing power, which may be at or beyond the operational limits of many small or lightweight UAV platforms. In order to make the proposed method more compatible with a broader range of UAV applications, future work should explore the design or adaptation of lighter-weight generative architectures that maintain a balance between adaptation quality and computational feasibility. This could involve leveraging recent advances in model compression techniques such as knowledge distillation, pruning, or quantization, as well as exploring architectures that exploit multi-scale feature representations without significantly increasing the number of parameters. Furthermore, specialized hardware accelerators, such as low-power FPGAs or edge-optimized GPUs, could be integrated to support real-time domain adaptation while conserving power consumption.

In addition to improving domain adaptation and computational efficiency, refinements to the probabilistic localization framework itself represent another important research direction. The current implementation employs a likelihood function based on the number of successfully matched features for each particle in the filter. While this metric performs well under conditions of clear and distinct matches, it may exhibit limitations when a majority of particles achieve similar match counts that correspond to incorrect spatial locations. In such cases, the particle filter can converge prematurely towards an incorrect pose hypothesis, resulting in localization failure. A more discriminative

likelihood formulation could be developed that accounts not only for the number of matches but also for the spatial distribution, geometric consistency, and descriptor similarity scores of the matched keypoints. Such a likelihood function could be derived from a combination of geometric verification techniques, such as RANSAC-based homography validation, and statistical measures of descriptor confidence. Incorporating these additional cues could help prevent false convergence in ambiguous visual conditions and significantly improve the robustness of the localization process.

Furthermore, although the proposed approach focuses exclusively on vision-based navigation, there are operational environments in which visual information may be insufficient or unreliable, as highlighted in the flight test results. Examples include environments with low visual texture, such as open water or snow-covered landscapes, as well as adverse weather conditions that obscure camera visibility. For such challenging scenarios, integrating alternative navigation modalities into the framework could further enhance robustness. One promising complementary approach is magnetic field matching navigation, which exploits local anomalies in the Earth's magnetic field as a geospatial reference. By fusing magnetic navigation with the proposed visual domain adaptation and matching pipeline, a multi-modal localization framework could be developed that leverages the strengths of each modality while compensating for their respective weaknesses. This fusion could be implemented within a Bayesian filtering framework, allowing for adaptive weighting of visual and magnetic measurements based on their estimated reliability in real time.

In summary, future work could expand the proposed research along multiple dimensions, including addressing extreme visual degradation scenarios such as night-time operations through tailored domain adaptation networks, developing real-time image rectification strategies for UAV platforms lacking gimbal stabilization, designing lightweight generative models for computationally constrained UAVs, refining the likelihood function for more reliable convergence in ambiguous match scenarios, and incorporating alternative navigation modalities such as magnetic field matching to complement visual localization. Collectively, these advancements would not only improve the adaptability and robustness of the system under diverse and challenging conditions but also broaden its applicability to a wider range of UAV platforms and mission profiles.



5. CONCLUSION

In this thesis, a novel and computationally efficient visual navigation framework has been developed for global position estimation of small UAVs operating in GNSS-denied environments. The proposed approach synergistically combines a lightweight Visual-Inertial Odometry (VIO) pipeline, a Marginalized Particle Filter for robust state estimation, and a CycleGAN-based domain adaptation mechanism to address the persistent challenges posed by environmental appearance variations between UAV-captured imagery and satellite reference maps. By explicitly addressing the seasonal, temporal, and lighting-induced discrepancies between these two domains, the framework significantly enhances the reliability and accuracy of visual matching, even under adverse conditions.

The Marginalized Particle Filter, which is central to the state estimation process, effectively fuses the high-rate motion information from the VIO with intermittent yet globally informative position fixes derived from the image-to-image matching pipeline. By marginalizing out non-essential states and maintaining a reduced set of variables for direct particle representation, this filter configuration achieves a favorable balance between estimation accuracy and computational efficiency. This is particularly critical for small UAV platforms where processing resources and power budgets are inherently limited.

The visual front-end of the proposed framework incorporates a tailored CycleGAN-based domain adaptation module, designed to translate UAV imagery from its original domain into a visually consistent form that closely matches the appearance of satellite reference images. This transformation substantially mitigates the impact of domain gaps—such as seasonal foliage changes, snow cover, and varying lighting conditions—thereby improving the quality and number of matched features. The performance of the domain adaptation step has been quantitatively evaluated through a custom-designed metric, which uses the mean number of matched features over a

large set of queries, normalized against baseline matches from raw UAV images. This metric enables meaningful comparison between different generative models and directly correlates the domain adaptation performance with downstream localization accuracy.

A further contribution of this work is the development of a custom likelihood function for the particle filter, designed to address the inherent limitations of conventional match-count-based scoring. In scenarios where a large number of particles produce similar match counts at incorrect spatial locations, a naive likelihood formulation can lead to premature convergence towards erroneous solutions. The proposed likelihood function incorporates both the number of matches and additional measures of spatial and geometric consistency, thereby reducing the risk of false convergence and increasing the robustness of the localization output.

The integrated system has been evaluated across multiple flight scenarios. Results indicate that the proposed approach consistently outperforms baseline visual navigation methods that do not employ domain adaptation. Moreover, even in less challenging environments where the raw imagery already offers high match quality, the CycleGAN-enhanced imagery provides incremental improvements, demonstrating the generality and versatility of the proposed method.

Overall, the research presented in this thesis contributes a practical and adaptable solution for GNSS-denied UAV navigation, capable of achieving high localization accuracy without exceeding the computational constraints typical of small aerial platforms. By combining the strengths of VIO, Marginalized Particle Filtering, domain adaptation via generative modeling, and a more discriminative likelihood formulation, the proposed framework represents a step forward in enabling reliable and robust autonomous navigation in diverse and visually challenging operational environments. Future developments, as outlined in the Chapter 4, will further enhance the adaptability, scalability, and resilience of the system, paving the way for its integration into a wider range of UAV applications and mission profiles.

REFERENCES

- [1] **Golden, J.P.** (1980). Terrain Contour Matching (TERCOM): A Cruise Missile Guidance Aid, *T.F. Wiener, editor, Image Processing for Missile Guidance*, volume 0238, International Society for Optics and Photonics, SPIE, pp.10 – 18, <https://doi.org/10.1117/12.959127>.
- [2] **Turan, B. and Kutay, A.T.** (2016). Particle filter studies on terrain referenced navigation, *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp.949–954.
- [3] **Turan, B.** (2020). Comparison of Nonlinear Filtering Methods for Terrain Referenced Aircraft Navigation, *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp.144–149.
- [4] **Carroll, J.D. and Canciani, A.J.** (2021). Terrain-referenced navigation using a steerable-laser measurement sensor, *Navigation*, 68(1), 115–134.
- [5] **Vadlamani, A.K. and Uijt de Haag, M.** (2009). Dual Airborne Laser Scanners Aided Inertial for Improved Autonomous Navigation, *IEEE Transactions on Aerospace and Electronic Systems*, 45(4), 1483–1498.
- [6] **de Haag, M.U., Vadlamani, A., Campbell, J. and Dickman, J.** (2006). Application of laser range scanner based terrain referenced navigation systems for aircraft guidance, *Third IEEE International Workshop on Electronic Design, Test and Applications (DELTA'06)*, pp.6 pp.–274.
- [7] **Weiss, S. and Siegwart, R.** (2012). Real-time autonomous quadrotor navigation using a monocular camera and an IMU, *2012 IEEE International Conference on Robotics and Automation*, IEEE, pp.4571–4578.
- [8] **Mourikis, A.I. and Roumeliotis, S.I.** (2007). A multi-state constraint Kalman filter for vision-aided inertial navigation, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp.3565–3572.
- [9] **Leutenegger, S., Lynen, S., Bosse, M., Siegwart, R. and Furgale, P.** (2015). Keyframe-based visual–inertial odometry using nonlinear optimization, *The International Journal of Robotics Research*, 34(3), 314–334.
- [10] **Qin, T., Li, P. and Shen, S.** (2018). Vins-mono: A robust and versatile monocular visual-inertial state estimator, *IEEE Transactions on Robotics*, 34(4), 1004–1020.

- [11] **Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M. and Tardós, J.D.** (2021). Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam, *IEEE transactions on robotics*, 37(6), 1874–1890.
- [12] **Geneva, P., Eckenhoff, K., Lee, W., Yang, Y. and Huang, G.** (2020). OpenVINS: A research platform for visual-inertial estimation, *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, pp.4666–4672.
- [13] **Pan, C., Liu, Z., Zhang, Z. and Zhang, J.** (2024). Adaptive VIO: Online Continual Learning for Visual-Inertial Odometry, *arXiv preprint arXiv:2401.12630*.
- [14] **Lowe, D.G.** (2004). Distinctive image features from scale-invariant keypoints, *International journal of computer vision*, 60, 91–110.
- [15] **Bay, H., Tuytelaars, T. and Van Gool, L.** (2006). SURF: Speeded up robust features, *European conference on computer vision*, Springer, pp.404–417.
- [16] **Rublee, E., Rabaud, V., Konolige, K. and Bradski, G.** (2011). ORB: An efficient alternative to SIFT or SURF, *2011 International conference on computer vision*, Ieee, pp.2564–2571.
- [17] **DeTone, D., Malisiewicz, T. and Rabinovich, A.** (2018). Superpoint: Self-supervised interest point detection and description, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp.224–236.
- [18] **Fischler, M.A. and Bolles, R.C.** (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM*, 24(6), 381–395.
- [19] **Sarlin, P.E., DeTone, D., Malisiewicz, T. and Rabinovich, A.** (2020). Superglue: Learning feature matching with graph neural networks, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.4938–4947.
- [20] **Lindenberger, P., Sarlin, P.E. and Pollefeys, M.** (2023). Lightglue: Local feature matching at light speed, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp.17627–17638.
- [21] **Sun, J., Shen, Z., Wang, Y., Bao, H. and Zhou, X.** (2021). LoFTR: Detector-Free Local Feature Matching with Transformers, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.8922–8931.
- [22] **Edstedt, J., Athanasiadis, I., Wadenbäck, M. and Felsberg, M.** (2023). DKM: Dense kernelized feature matching for geometry estimation, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.17765–17775.
- [23] **Zhu, J.Y., Park, T., Isola, P. and Efros, A.A.** (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE international conference on computer vision*, pp.2223–2232.

- [24] **Author, A. and Author, B.** (2022). Domain Adaptation for UAV Visual Localization: A Satellite-to-Ground Image Translation Approach, *Journal of Field Robotics*, 39(1), 1–15.
- [25] **Porav, H., Bruls, T., Câmpeanu, R., Güneş, Z. and Newman, P.** (2019). Why did the self-driving car cross the road? A CycleGAN-based approach for day-to-night domain adaptation for navigation, *arXiv preprint arXiv:1910.05468*.
- [26] **Sola, J.** (2017). Quaternion kinematics for the error-state Kalman filter, *arXiv preprint arXiv:1711.02508*.
- [27] **Schon, T., Gustafsson, F. and Nordlund, P.J.** (2005). Marginalized particle filters for mixed linear/nonlinear state-space models, *IEEE Transactions on Signal Processing*, 53(7), 2279–2289.
- [28] **Murphy, K. and Russell, S.**, (2001). Rao-Blackwellised particle filtering for dynamic Bayesian networks, *Sequential Monte Carlo methods in practice*, Springer, pp.499–515.
- [29] **Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L.** (2008). Speeded-up robust features (SURF), *Computer vision and image understanding*, 110(3), 346–359.
- [30] **Revaud, J., De Souza, C., Humenberger, M. and Weinzaepfel, P.** (2019). R2d2: Reliable and repeatable detector and descriptor, *Advances in neural information processing systems*, 32.
- [31] **Gordon, N., Salmond, D. and Smith, A.** (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F (Radar and Signal Processing)*, 140, 107–113, <https://digital-library.theiet.org/doi/abs/10.1049/ip-f-2.1993.0015>, <https://digital-library.theiet.org/doi/pdf/10.1049/ip-f-2.1993.0015>.
- [32] **Jurevičius, R., Marcinkevičius, V. and Šeibokas, J.** (2019). Robust GNSS-denied localization for UAV using particle filter and visual odometry, *Machine Vision and Applications*, 30(7), 1181–1190.
- [33] **Parmar, G., Park, T., Narasimhan, S. and Zhu, J.Y.** (2024). One-step image translation with text-to-image models, *arXiv preprint arXiv:2403.12036*.
- [34] **Park, T., Efros, A.A., Zhang, R. and Zhu, J.Y.** (2020). Contrastive learning for unpaired image-to-image translation, *European conference on computer vision*, Springer, pp.319–345.



CURRICULUM VITAE

Abdlbaki ŐANLAN:

EDUCATION:

- **B.Sc.:** 2019-2024, Istanbul Technical University, Department of Aeronautical Engineering, Aeronautical Engineering Program
- **M.Sc.:** 2024-2025, Istanbul Technical University, Department of Aeronautical and Astronautics Engineering, Aeronautics and Astronautics Engineering Program

PROFESSIONAL EXPERIENCE AND REWARDS:

- **Researcher,** ITU Aerospace Center
- **Best Paper Award,** Best Paper in 5th AI/ML Session at 25th ICNS Conference

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Őanlan, A.,** Baęcı, H. T, Koyuncu, E (2025, April). Particle Filter-Based Localization Using Visual Feature Synchronization in GNSS-Denied Navigation. *In 2025 Integrated Communications, Navigation and Surveillance Conference (ICNS)* (pp. 1-8). IEEE.