GNSS-denied visual localization ameliorative method for UAVs in non-urban environments

Ngo Van Quan[1*], Duong Dinh Luyen[2], Phan Huy Anh[3], Nguyen Chi Thanh[1], Le Minh Ngoc[4], Pham Thi Hoai Thu[5]

[1]Institute of Information Technology, Academy of Military Science and Technology;
[2]Hanoi University of Science and Technology;
[3]Electronic Institute, Academy of Military Science and Technology;
[4]Academy of Military Science and Technology;
[5]Thach Ban High School, Hanoi, Vietnam.
[*]Corresponding author: ngoquanvnp@gmail.com.
Received 13 Oct. 2023; Revised 30 Nov. 2023; Accepted 12 Dec. 2023; Published 25 Dec. 2023.
DOI: https://doi.org/10.54939/1859-1043.j.mst.92.2023.130-136

ABSTRACT

In the context of Unmanned Aerial Vehicles (UAVs), localization is critical for both military and civilian applications. This is particularly true in environments without urban infrastructure, where Global Navigation Satellite System (GNSS) signals are unavailable. In these settings, vision-based methods have emerged as a promising solution. Despite their potential, current deep learning-based matching algorithms exhibit significant limitations in accurately localizing UAVs. To address this, our paper introduces enhanced algorithms that build upon existing methods. Specifically, we propose the use of the DC-ShadowNet shadow removal algorithm for UAV image preprocessing, a critical step in urban areas where shadows from large structures can obscure ground details, especially under sunny conditions. Additionally, we employ an improved matching algorithm based on the ASpanFormer model to increase accuracy in image matching. Our testing shows that these advancements lead to improved localization accuracy, both on a public dataset and on actual flight data. Furthermore, our method is well-suited for long-duration flights and offers considerable advantages in urban environments when compared to previous state-of-the-art Visual Odometry techniques.

Keywords: Visual localization; Unmanned aerial vehicles.

1. INTRODUCTION

Using unmanned aerial vehicles (UAVs) has become increasingly popular in various commercial settings. To ensure precise localization, different techniques are employed depending on the environment. The Global Visual-Inertial Simultaneous Localization and Mapping (VI-SLAM) [1] technique is typically utilized outdoors, where Navigation Satellite System (GNSS) [2] sensors can provide precise positional data. In recent years, aerial drones have advanced in autonomy and can fly over extended distances without human intervention [3]. However, providing accurate positioning data during long-distance flights without reliance on GNSS has not been adequately addressed, despite the necessity for a failsafe mechanism in the event of GNSS signal unavailability. This is particularly important as UAVs are integral to many missions and require the ability to activate a backup system automatically if GNSS signals are disrupted by factors such as jamming or spoofing [4]. Advancements in deep neural networks have enhanced image processing capabilities, enabling real-time recognition and matching of objects, even in satellite image data. The approach relies on the state-of-the-art algorithm ASpanFormer [7] to perform detector-based matching baselines in two-view pose estimation. Our experiment results on the GNSS dataset [7] and our test dataset outperform existing algorithms while having similar accuracy to traditional GNSS-based methods, which serve as the ground truth. In summary, the primary contributions of the proposed methods are as follows:

• A deep learning technique demonstrates a robust performance in delivering precise matching outcomes, even in scenarios involving variations in perspective and rotation between drone photographs and surfaces with limited salient features.

• Experimental results demonstrate that the proposed method outperforms the previous models significantly in terms of vision-based localization.

• Data preprocessing with shadow removal of large objects in sunny weather can obscure important locations.

The research paper under consideration relates to numerous other studies. These include path planning research based on semantic segmentation [9], using open-source Google Earth images for localization [10], pose estimation through georeferenced satellite photos via deep learning models [10], and an image segmentation model [11] originally designed for constructing virtual worlds. Over the past decades, Simultaneous Localization and Mapping (SLAM) [14] have become widely accepted for navigating unknown environments. These algorithms have achieved remarkable reliability, largely due to their reliance on Visual Odometry (VO) methods [15], which enable UAVs to accurately determine their locations in new and unfamiliar environments. Template matching [12] and SIFT [13] features have been proposed, but they lack competitiveness due to lower accuracy and longer computation times in complex cases. Learning-based techniques such as LIFT [18] and MagicPoint [19] have significantly improved local feature performance, especially in scenarios with substantial changes in viewpoint and illumination. SuperPoint [20] builds on MagicPoint [19], introducing a self-supervised training method through homographic adaptation. The SuperGlue [6] method, a novel learning-based regional feature matching approach, has been recently introduced. Wildnav [16] has adopted the SuperGlue [6] algorithm, setting a new standard in localization and mapping for outdoor images using UAVs. However, the model's accuracy is still inadequate for precision maneuvers or at low altitudes where obstacles are prevalent.

2. PROPOSED METHODS

2.1. Shadow removal processing

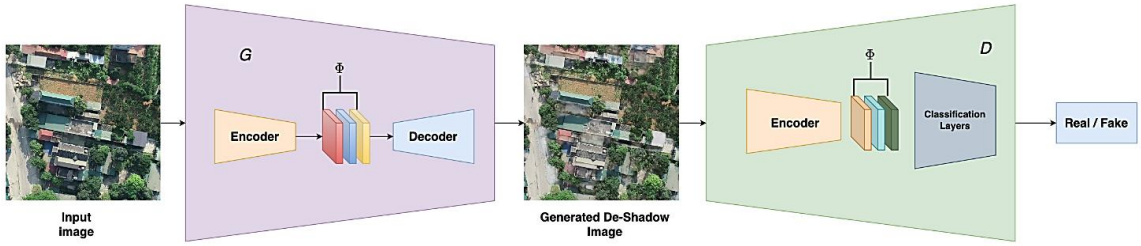

Figure 1. *DC-ShadowNet method to de-shadow the UAV image.*

For UAV remote sensing images with complex backgrounds and multiple shadows, compensating for shadows can lead to issues like color distortion or texture information loss. We adopt DC-ShadowNet, an unsupervised network guided by a shadow/shadow-free domain classifier. This model specifically integrates a domain classifier, which categorizes the input image as either in the shadow or shadow-free domain, into a Generative Adversarial Network (GAN) model, as illustrated in figure 1. The primary objective of the model is to concentrate on shadow regions and achieve more effective shadow removal. DC-ShadowNet introduces two novel types of loss: chromatic loss for entropy minimization in the log-chromaticity space, and perceptual features loss utilizing shadow-robust features derived from the pre-trained VGG-16 network. We employ this model as a preprocessing step to further enhance the performance of our method.

2.2. Georeferencing map sections

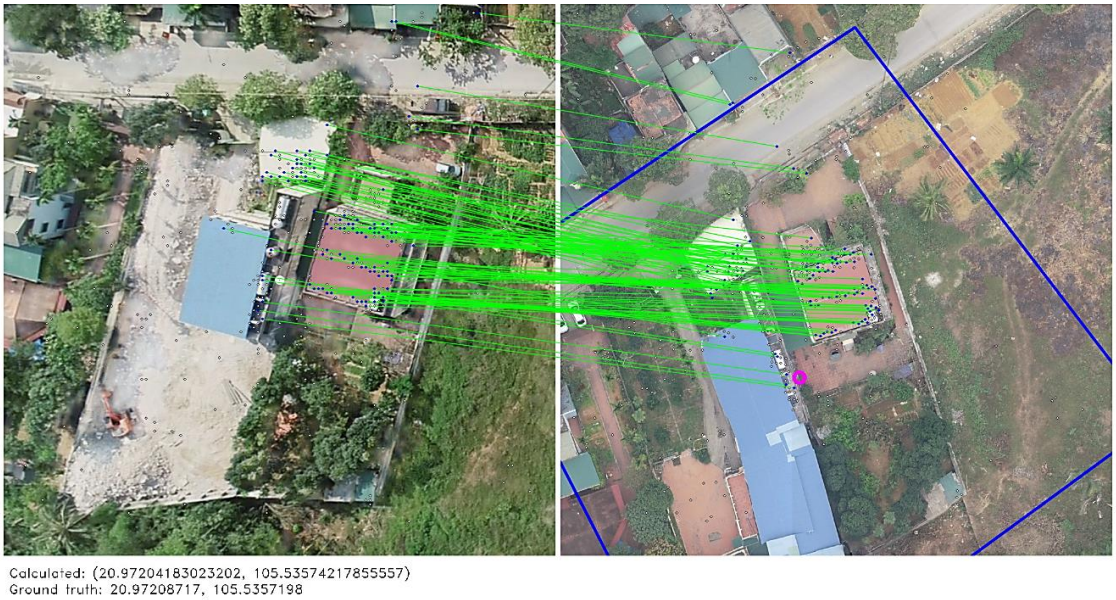

Figure 2. *The location of UAV is estimated based on the geography of the map image with the best matching result (based on the number of keypoints).*

Due to its enormous size, it is impossible to represent the entire flying region as a single picture file. Consequently, the map is divided into sections. Each section is defined by two sets of geographic coordinates: the top left corner and the bottom right corner, with each set comprising a pair of latitude and longitude values. This approach of using two sets of coordinates for the corners, rather than a single set for the center of the picture, allows for more accurate georeferencing of each segment. In this system, the georeferenced satellite image can encompass any set of matched features within a polygon. The center of this polygon, identified by the pixel coordinates $\boldsymbol{c_x}$ and $\boldsymbol{c_y}$, facilitates the direct computation of geographical coordinates, effectively transforming the coordinate system from local image file pixels to latitude and longitude. Figure 2 offers an aerial view of the designated region where drone photographs were captured to test the algorithms developed.

2.3. Detector-free image matching

In figure 3, we present an overview of our network structure, based on the ASpanFormer model [7].

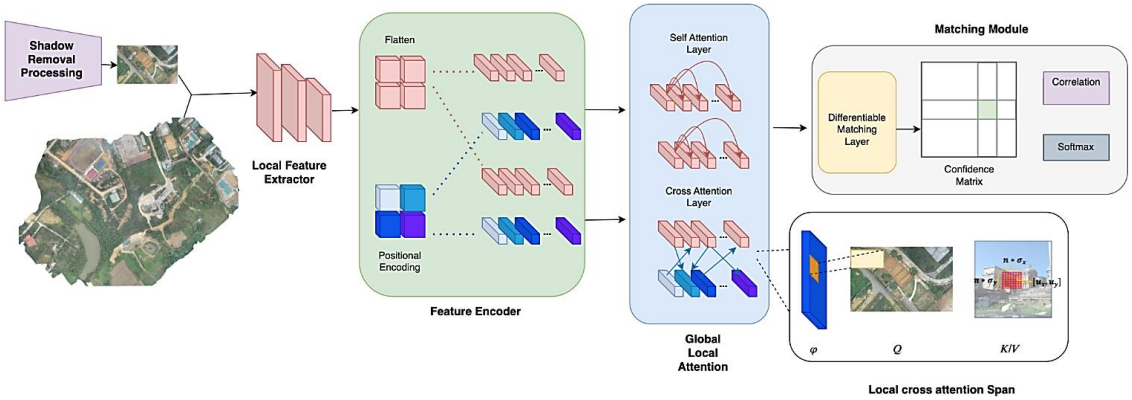

Figure 3. *Pipeline of Detector-Free Matching with Adaptive Span Transformer for UAV image.*

The network takes a pair of images $\boldsymbol{I_A, I_B}$ as input and produces reliable correspondences between them. The matching process begins with a CNN-based encoder that extracts initial features $\boldsymbol{F_A^0, F_B^0}$ for each image. These features are then transformed into $\boldsymbol{F_A^1, F_B^1}$ and fed into the Adaptive Span Transformer (ASpanFormer) module. This module updates the features through iterative global-local attention (GLA) blocks, which have a hierarchical structure. Each GLA block regresses auxiliary flow maps $\boldsymbol{\varphi_A, \varphi_B}$ which describe correspondence coordinates (flows) and their associated uncertainty. Rather than using these flow maps directly as our correspondence output, we employ them to guide local cross-attention. This approach enables an adaptive local attention span based on the matching uncertainty. After processing through N GLA blocks, the resultant features $\boldsymbol{F_A^{N+1}, F_B^{N+1}}$ are utilized to construct coarse-level matches. These matches are subsequently refined to establish the final correspondences.

3. EXPERIMENTS AND RESULTS

3.1. Dataset

The dataset employed in this study comprises two distinct subsets: the first, sourced from the Wildnav method, encompasses 124 photographs, and the second, a dataset autonomously assembled through UAV imagery. In an effort to rigorously assess the efficacy and precision of the proposed method, we conducted evaluations using a specially curated dataset, which provided insightful metrics. It was observed that, despite the Wildnav method's dataset encompassing a substantial volume of photographs, it fails to represent the full spectrum of potential scenarios and terrains. In order to demonstrate the robustness of our approach, particularly under conditions of intense sunlight and in environments where substantial shadows from large objects obscure ground features, we executed UAV flights over varied landscapes. These included a densely populated urban area (HL), a park area (CVHB), and a hilly terrain (Thach That). The specifics regarding the image count and area coverage for each zone are delineated in table 1.

Table 1. *Our datasets statistics include the area covered and the quantity of images acquired during flights in three separate places.*

Area name	**Number of images**	**Area covered** ($\boldsymbol{m^2}$)
ThachThat	2251	370000
CVHB	1630	380000
HL	2000	410000

The UAV flights, conducted at an altitude of 200 meters, facilitated the acquisition of images (each measuring 720x1080 pixels) that were subsequently aligned with the geographic coordinates recorded during flight. This data collection was strategically scheduled at two distinct times of the day—midday, when the sun is at its zenith, and in the afternoon under clear, sunny conditions—to optimize the capture of ground shadows. Furthermore, image labeling was automated through the utilization of the GNSS system integrated within the UAV.

Figure 4. *The map image of the dataset corresponds to ThachThat, CVHB, HL, respectively.*

3.2. Result evaluation

3.2.1. Shadow removal

The integration of anti-ghosting techniques significantly enhances the identification of obscured keypoints for image matching, a utility that proves particularly advantageous in sunny urban environments. In this context, we have advocated for the incorporation of DC-ShadowNet [23], an unsupervised learning-based shadow removal methodology steered by a domain classification network, as a preliminary processing step. The efficacy of this method is evident in figure 5, where it is observable that post-processing, the shadowed regions, particularly those cast by large, heavily pigmented objects, become more discernible. Beyond its applicability in UAV image localization, the employment of this shadow mitigation technique is deemed crucial in accurately identifying objects that are otherwise concealed or distorted by direct sunlight.

Figure 5. *The result of applying the shadow removal model.*

3.2.2. Matching accuracy

In this section, we rigorously test our approach to validate the accuracy of point-to-point correspondence and the precision of GPS coordinate positioning within the geographic maps of our datasets. We employ a two-pronged strategy, utilizing an outdoor pretrained ASpanFormer, trained on the MegaDepth [7] dataset for image matching, and a pretrained DC-ShadowNet [22] for consistent shadow removal across all datasets. Initially, images undergo processing through DC-ShadowNet [22] to mitigate shadow effects. Subsequent to these preparatory steps, we observe a notable enhancement in accuracy. Table 2 presents a comparative analysis of correct matches yielded by our proposed architecture against those obtained through classical key-point descriptors and other contemporary matching method-based approaches. For the dataset derived from the Wildnav [16] method, comprising 124 UAV images, we compare our results against those achieved using the SP+SuperGlue [6] matching algorithm. Additionally, we conduct experiments with the LoFTR [23] matching algorithm, recording accuracies of 77% and 86%, respectively.

To further substantiate the efficacy of our algorithm, particularly under conditions of bright sunlight obscuring ground objects, we extend our comparative analysis to a larger dataset, ThachThat, containing 2251 UAV images. Here, the combined application of ASpanFormer and DC-ShadowNet [22] demonstrates a significant improvement, achieving 67% accuracy compared to the 63% by LoFTR [23] and 55% by SP+SuperGlue [6]. Table 2 elucidates that the error margin for the ASpanFormer [7] combined with DC-ShadowNet [22] on the Wildnav set is reduced by 11.55 m, in contrast to the 15.28 m by SP+SuperGlue [6] and 13.32 m by LoFTR

[23]. In the ThachThat dataset, a densely populated area, the ASpanFormer [7] and DC-ShadowNet [22] combination achieved a reduced error of 8.23 m, showcasing greater efficiency than the 10.06 m by LoFTR and 11.38 m by SP+SuperGlue [17]. These results not only underscore the effectiveness of the DC-ShadowNet [22] preprocessing technique in shadow reduction but also highlight the variance in accuracy between our datasets and the Wildnav set, which can be attributed to the latter's limited representation of immutable objects like roads and buildings. This variance reinforces the suitability of our shadow mitigation technique and, by extension, our overall method in urban environments.

Table 2. *Comparison of localization error and matching methods via conventional keypoint descriptors.*

Localization datasets	Matching method	Total image	Localized accuracy(%)	Average error(m)
Wildnav	SuperPoint+SuperGlue	124	77%	15.82
Wildnav	LoFTR	124	86%	13.32
Wildnav	**ASpanFormer+DC-ShadowNet**	**124**	**96%**	**11.55**
ThachThat	SuperPoint+SuperGlue	2251	55%	11.38
ThachThat	LoFTR	2251	63%	10.06
ThachThat	**ASpanFormer+DC-ShadowNet**	**2251**	**67%**	**8.23**

4. CONCLUSIONS

To enhance GNSS-Free vision-based localization for UAVs in non-urban environments, the application of specific techniques appears promising. Recognizing the impact of sunny conditions, where ground objects are often overshadowed by larger structures, we have incorporated a shadow mitigation algorithm as a preprocessing step. This enhancement is part of a broader strategy involving a newly proposed algorithm for improved UAV operation. Addressing limitations in existing datasets, which were deficient in both size and variety, we curated a new dataset from three distinct suburban flight areas. Despite these advancements, challenges persist, particularly with the complexity of the new matching algorithm and its reduced efficacy at higher UAV altitudes. In conclusion, our proposed method yields encouraging results in non-urban settings. The integration of shadow removal techniques and advancements in matching algorithms heralds new possibilities for vision-based localization, extending its applicability across diverse environments, both urban and non-urban.

REFERENCES

[1]. Qin, Tong et al. *"VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator."* IEEE Transactions on Robotics 34, 1004-1020, (2017).

[2]. Alsalam, Bilal Hazim Younus et al. *"Autonomous UAV with vision-basedon-board decision making for remote sensing and precision agriculture."* IEEE Aerospace Conference, 1-12, (2017).

[3]. Nguyen, Thien-Minh et al. *"VIRAL-Fusion: A Visual-Inertial-Ranging-Lidar Sensor Fusion Approach."* IEEE Transactions on Robotics 38, 958-977, (2020).

[4]. Psiaki, M.L., Humphreys, T.E. *"GNSS Spoofing and Detection"*. Proceedings of the IEEE, 104, 1258-1270, (2016).

[5]. Shermeyer, Jacob et al. *"SpaceNet 6: Multi-Sensor All Weather MappingDataset."* 2020 IEEE/CVF Conference on Computer Vision and PatternRecognition Workshops (CVPRW), 768-777, (2020).

[6]. Sarlin, Paul-Edouard et al. *"SuperGlue: Learning Feature Matching WithGraph Neural Networks."* IEEE/CVF Conference on Computer Vi-sion and Pattern Recognition (CVPR), 4937-4946, (2019).

[7]. Chen, Hongkai et al. *"ASpanFormer: Detector-Free Image Matching withAdaptive Span Transformer."* European Conference on Computer Vision, (2022).

[8]. Yol, Aurelien et al. *"Vision-based absolute localization for unmannedaerial vehicles."* IEEE/RSJ International Conference on IntelligentRobots and Systems, 3429-3434, (2014).
[9]. Bartolomei, Luca et al. *"Perception-aware Path Planning for UAVs usingSemantic Segmentation."* IEEE/RSJ International Conference onIntelligent Robots and Systems (IROS), 5808-5815, (2020).
[10]. Shetty, Akshay, G. X. Gao. *"UAV Pose Estimation usingCross-view Geolocalization with Satellite Imagery."* InternationalConference on Robotics and Automation (ICRA), 1827-1833, (2018).
[11]. Gu'erin, Eric et al. *"Satellite Image Semantic Segmentation."* ArXivabs/2110.05812, (2021).
[12]. Briechle, Kai and Uwe D. Hanebeck. *"Template matching using fast nor-malized cross correlation."* SPIE Defense + Commercial Sensing, (2001).
[13]. LoweDavid, G. *"Distinctive Image Features from Scale-Invariant Key-points."* International Journal of Computer Vision, (2004)
[14]. A. Macario et al. "A Comprehensive Survey of Visual SLAMAlgorithms." Robotics 11, 24, (2022).
[15]. Delmerico, Jeffrey A. and Davide Scaramuzza. *"A Benchmark Com-parison of Monocular Visual-Inertial Odometry Algorithms for FlyingRobots."* IEEE International Conference on Robotics and Automa-tion (ICRA), 2502-2509, (2018).
[16]. Gurgu, Marius-Mihail et al. *"Vision-Based GNSS-Free Localization forUAVs in the Wild."* 7th International Conference on MechanicalEngineering and Robotics Research (ICMERR), 7-12, (2022).
[17]. Rublee, Ethan et al. *"ORB: An efficient alternative to SIFT or SURF."* International Conference on Computer Vision, 2564-2571, (2011).
[18]. Yi, Kwang Moo et al. *"LIFT: Learned Invariant Feature Transform."* ArXivabs/1603.09114, (2016).
[19]. DeTone, Daniel et al. *"Toward Geometric Deep SLAM."* ArXivabs/1707.07410, (2017).
[20]. DeTone, Daniel et al. *"SuperPoint: Self-Supervised Interest Point Detec-tion and Description."* IEEE/CVF Conference on Computer Visionand Pattern Recognition Workshops (CVPRW), 337-33712, (2017).
[21]. Yeying Jin et al. *"DC-ShadowNet: Single-Image Hard and Soft ShadowRemoval Using Unsupervised Domain-Classifier Guided Network"*. Inter-national Conference on Computer Vision (ICCV), 2207.10434, (2021).
[22]. Jiaming Sun et al. *"LoFTR: Detector-Free Local Feature Matchingwith Transformers"*. International Conference on Computer Vision(ICCV), 2104.00680, (2021).

TÓM TẮT

Phương pháp cải thiện định vị bằng hình ảnh cho UAV trong môi trường ngoài đô thị

Trong cả hoạt động quân sự và dân sự, vấn đề định vị của phương tiện bay không người lái (UAV) là một lĩnh vực nghiên cứu quan trọng. Đặc biệt, trong các môi trường tín hiệu GNSS không khả dụng, các phương pháp định vị dựa trên hình ảnh hứa hẹn cho hiệu quả tốt, tuy nhiên, các thuật toán đối sánh dựa trên học sâu có một số hạn chế dẫn đến không đáp ứng được độ chính xác theo yêu cầu. Vì vậy, trong bài báo này chúng tôi đề xuất các thuật toán cải tiến cho phương pháp đã sử dụng trước đó. Trong điều kiện thời tiết có ánh nắng mạnh, các vị trí trên mặt đất có thể bị bóng của các vật thể lớn che khuất, đặc biệt là ở môi trường ngoài đô thị, nên chúng tôi đề xuất áp dụng thuật toán loại bỏ bóng DC-ShadowNet để tiền xử lý ảnh chụp từ UAV. Thuật toán so khớp nâng cấp dựa trên mô hình ASpanFormer được sử dụng để tăng độ chính xác tại bước so khớp hình ảnh. Kết quả thử nghiệm cho thấy độ chính xác của phương pháp đề xuất có sự cải thiện so với tập dữ liệu của phương pháp trước, cũng như tập dữ liệu chuyến bay thực tế mà chúng tôi xây dựng. Ngoài ra, phương pháp của chúng tôi phù hợp cho các chuyến bay đường dài và rất phù hợp cho khu vực ngoài đô thị so với các phương pháp đo thị giác tiên tiến trước đây.

Từ khoá: Visual localization; Unmanned aerial vehicles.