

UAV LARGE OBLIQUE IMAGE GEO-LOCALIZATION USING SATELLITE IMAGES IN THE DENSE BUILDINGS AREA

Junqi Luo¹, Qin Ye^{1,2*}

¹ College of Surveying and Geo-Informatics, Tongji University, 200092, Shanghai, China

² State Key Laboratory of Geographic Information Engineering, 710054, Xi'an, Shaanxi, China
– 2310269@tongji.edu.cn, yeqin@tongji.edu.cn

KEY WORDS: Image-Based Geo-localization, UAV Large Oblique Image, Cross-view Image Matching, Dense Buildings Area, Image Retrieval.

ABSTRACT:

For UAV large oblique image geo-localization in the dense buildings area, there are still two main challenges. One is the presence of obvious occlusion and large viewpoint differences in UAV images, and the other arises from the fact that reference images, particularly orthographic satellite images, lack façade information of man-made structures (such as buildings and roads), which is crucial for UAV large oblique images. Most of existing image-based geo-localization methods only address the first challenge, neglecting the interference brought by the second challenge, especially for UAV large oblique image geo-localization in the dense buildings area. Motivated by both these two challenges, we have proposed a novel method for UAV large oblique image geo-localization in the dense buildings areas, with the segments direction statistics (SDS) features and their histogram descriptors designed. By considering both the local and global features of man-made structures, the proposed method effectively addresses the significant information difference encountered in cross-view image matching. We conducted experiments on both the public UAV images dataset University-1652 and our own collected dataset of UAV large oblique long focal whiskbroom (LO-LF-W) images. Comparative analysis with state-of-the-art (SOTA) methods demonstrated that the proposed method improves the geo-localization accuracy by approximately 10%. Furthermore, the proposed method exhibits greater robustness to noise and changing orientation of reference images, making it particularly well-suited for dense buildings areas that pose challenges for existing methods.

1. INTRODUCTION

Image-based geo-localization is to determine the approximate location of UAV images, which plays a crucial role in various UAV-based applications such as border management, building monitoring, and urban modelling. (Berton et al., 2022; Zhou et al., 2021) However, existing image-based geo-localization methods face challenges when applied to UAV large oblique images, especially in urban areas with dense buildings. There are several reasons for this limitation. Firstly, UAV oblique images significantly differ from orthographic images in terms of viewpoint, scale and resolution. Secondly, orthophotos lack detailed façade information of buildings that is present in UAV oblique images. This absence of façade information hampers the matching of cross-view images, leading to reduced accuracy in UAV image-based geo-localization.

The image-based geo-localization methods for UAV images can be categorized into two main groups: methods utilizing handcrafted features and methods leveraging deep learning. (Yao et al., 2019) Existing handcrafted features are often designed for specific types of scenes, making them less applicable to unknown or novel scenes. Consequently, there is a lack of suitable handcrafted features to address the challenges of image-based geo-localization in dense buildings areas. Moreover, gathering training datasets specifically for UAV large oblique images poses significant challenges. The scarcity of such datasets limits the accuracy of existing deep learning networks, as they are predominantly trained on a limited number of publicly accessible datasets.

The term “UAV large oblique image” in this paper refers to images captured at an oblique angle greater than 32 degrees. Specifically, in our own collected Dataset2, the oblique angle of UAV images is approximately 60 degrees.

We mainly made the following three contributions in this paper:

1. We have introduced feature filters to address the interference segments caused by façade information, which is crucial aspect of our method for solving geo-localization in dense buildings areas.
2. We have designed SDS features and histogram descriptors as the stable features for UAV large oblique images geo-localization with obvious viewpoint and scale gaps.
3. Instead of storing original reference images, we opted to store extracted feature descriptors, which significantly reduces the memory and computation costs on the online platform.

2. RELATED WORK

The airborne POS (Position and Orientation System) has emerged as a popular strategy for UAV positioning. However, the large oblique angle of UAV images amplifies the influence of POS data errors on image-based geo-localization. Additionally, the limited load capacity of UAVs often prevents them from carrying high-precision POS equipment. Therefore, UAV image geo-localization has gained significant attention, focusing on image matching between UAV images and

* Corresponding author

orthographic reference images with geo-tags, typically satellite-view images or Digital Ortho Maps (DOMs).

Early research on image-based geo-localization primarily focused on ground-view images collected by mobile platforms, and it allowed for image localization in environments where GNSS (Global Navigation Satellite System) signals are unavailable or unreliable. (Zamir et al., 2014) Compared to ground-view images, satellite-view images offer a closer view range that closely aligns with UAV images. Moreover, satellite-view images are less affected by occlusion and moving targets, making them valuable reference images for UAV image geo-localization. (Couturier et al., 2021)

Due to significant differences in flight height, acquisition time and shooting pose between UAV and satellite images, existing image-based geo-localization methods that rely on handcrafted features demonstrate poor performance. As a result, researchers have directed their efforts towards enhancing these traditional features or exploring approaches that combine multiple features for effective cross-view image matching. (Goel et al., 2022) combined extracted SURF (Speed-Up Robust Features) and FLANN (Fast Library for Approximate Nearest Neighbours) feature matcher to improve the accuracy of geo-localization. (Hasheminasab et al., 2021) improved SIFT (Scale-Invariant Feature Transform) with image consistency check to realize the UAV image geo-localization with DOM. However, the above improvements on existing handcrafted features still retain the limitations for images with large viewpoint differences. (Xu et al., 2021) designed a novel contour line feature descriptor, which can be used for image matching with deformed UAV images and reference images. While handcrafted features typically demonstrate high geo-localization accuracy and efficiency for small oblique images in non-building areas, they often encounter limitations when applied to UAV large oblique images due to the presence of façade information. Consequently, it becomes necessary to develop novel feature patterns, descriptors and matching strategies that are specifically tailored for UAV large oblique images.

Since deep learning networks are able to obtain more advanced features and descriptors, learning-based UAV image geo-localization has become a popular research topic. (Ding et al., 2021) transformed the image-based geo-localization into image classification task. (Zhang et al., 2020) extracted RCF (Richer Convolutional Features) generated feature maps and determined the similarity of UAV-view and satellite-view images. Meanwhile, to reduce the memory cost, many studies take pre-training or coding reference images into account. (Bianchi et al., 2021) encoded Google maps automatically, and compressed them into dimensional vectors and key features. Since CNN-based methods lead to information and features loss with down-sampling, feature extraction and matching network based on Transformer framework has been applied in image-based geo-localization. (Zhuang et al., 2022) proposed a Transformer-based network to extract more image context features, and matched them by the semantic guidance module (SGM). As another kind of strategy, (Tian et al., 2021) transformed UAV oblique images into nadir view images closer to satellite images by PPT (Perspective Projection Transformation), and used the local feature pattern network for image-based geo-localization. However, the application of the aforementioned deep learning networks is challenging due to the scarcity of datasets specifically tailored for UAV oblique image geo-localization. As a result, these networks face limitations in terms of generalizability and may struggle to perform effectively in scenarios beyond the scope of publicly available datasets.

In summary, for above handcrafted features based and learning-based geo-localization methods, they obey the development trend of low memory and computation costs. (Ma et al., 2021) Most existing methods mainly solve the problems caused by viewpoint differences. However, in densely built-up areas, the façade interference information can also easily lead to mis-localization, which is rarely considered in existing methods. To overcome these difficulties, we designed SDS features and their histogram descriptors, for accurate and efficient UAV large oblique images geo-localization.

3. METHODOLOGY

For UAV large oblique images, image-based geo-localization faces two primary challenges. Firstly, the large oblique viewpoints result in image deformation, leading to poor matching results. Extensive research has been conducted to address this challenge. Secondly, UAV large oblique images contain abundant façade information that is rarely found in orthographic reference images. This façade information poses a significant interference to the topological consistency of corresponding targets, especially in dense buildings areas. To overcome these challenges, particularly the latter, the proposed UAV large oblique images geo-localization method consists of three main components: contour segments extraction, SDS feature and descriptor, and image-based geo-localization.

3.1 Contour Segments Extraction

For extracting contour segments from images, the LSD (Line Segment Detection) algorithm is widely employed due to its favourable performance in accuracy and efficiency. In the proposed method, the target feature segments are the contour segments of building top surfaces and roads. As a result, the segments extracted by LSD serve as the initial extracted specifically vertical features. To enhance accuracy, filters are designed to eliminate interference segments, specifically vertical lines of buildings, from the initial extraction results.

3.1.1 Segments Length Filter: Due to the influence of noise, occlusion and shadow, there are plenty of scattered short segments in initial extracted results, which greatly interfere with statistics and description of features, and are necessary to be screened out. Segments length filter is usually used in this case, and it generally uses the average length or the maximum length to set the filter threshold. However, the contour segments extracted from UAV large oblique images contain both building and road contours with large length differences. Therefore, existing length filters could wrongly screen out many valuable contour segments due to long road contour segments. Since the interference segments still account for a small proportion in the initial extracted results, a length filter based on the median length is constructed, as shown in formula (1):

$$\frac{L_i}{L_m} > \varepsilon (i = 1, 2, \dots, n) \quad (1)$$

where L_i = the length of the i^{th} segment
 L_m = the median length of extracted segments
 ε = the set filtering threshold

3.1.2 Facades Segments Filter: In the extracted segments of UAV large oblique images, the interference features also include contour segments of architectural facades, which does not exist in the orthophotos. Architectural facades information is widely used in the field of buildings reconstruction. However,

this façade information on UAV large oblique images leads to great interference on image-based geo-localization using orthographic reference images. It is necessary to filter out these segments of façades in the initial extracted ones before cross-view image matching. Meanwhile, though low-precision POS directly assisted geo-localization shows poor accuracy with UAV large oblique images, it still can be used to provide constraints for architectural façades segments filter in geo-localization. The vertical segments are the significant feature of façade information, which should satisfy the constraints like $\Delta X/\Delta Z \rightarrow 0$ and $\Delta Y/\Delta Z \rightarrow 0$. In the proposed method, the ratio threshold is set for the façade segments filter, as formula (2):

$$\begin{aligned} \Delta X/\Delta Z < \eta_1 \\ \Delta Y/\Delta Z < \eta_2 \end{aligned} \quad (2)$$

where η_1, η_2 = the set filtering thresholds
 $\Delta X, \Delta Y, \Delta Z$ = the coordinate difference of endpoints

Since the façade constraint belongs to WCS (World Coordinate System) while the segments feature is located in IPCS (Image Plane Coordinate System), the transformation between these two coordinate systems is determined with the collinear condition equation, as shown in formula (3):

$$\begin{aligned} x &= -f \cdot \frac{a_1(X - X_s) + b_1(Y - Y_s) + c_1(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \\ y &= -f \cdot \frac{a_2(X - X_s) + b_2(Y - Y_s) + c_2(Z - Z_s)}{a_3(X - X_s) + b_3(Y - Y_s) + c_3(Z - Z_s)} \end{aligned} \quad (3)$$

where X, Y, Z = the ground coordinate of the feature point
 x, y = the image coordinate of the same feature point
 f = the focal length of images
 X_s, Y_s, Z_s = the position parameters of UAV
 a_i, b_i, c_i = the orientation parameters of camera

Based on the collinear condition equation and the homonymous points obtained by image matching algorithm, which we used here are SuperPoint (DeTone et al., 2018) and SuperGlue (Sarlin et al., 2020) for key points extraction and matching. The WCS can be determined with the IPCS of the homonymous points on two images using formula (4):

$$\begin{aligned} l_1X + l_2Y + l_3Z - l_x &= 0 \\ l_4X + l_5Y + l_6Z - l_y &= 0 \end{aligned} \quad (4)$$

where l_i = the coefficient values

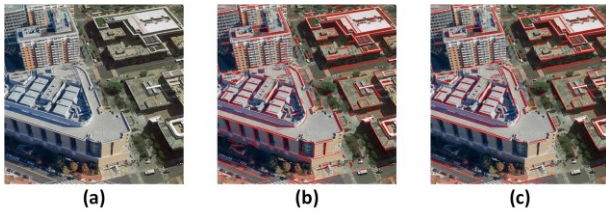


Figure 1. Contour segments extraction and filter results. (a) Original image; (b) Initial segments; (c) Filtered segments.

Due to the poor geometric intersection conditions of UAV-based large oblique images, the location directly solved are not accurate. However, we find that the spatial relationship between these locations is still stable, which can be used to filter out the façade information. With only the contour segments of building top surface retained, the façade segments are screened out. So far, based on the initial extraction segments by LSD, the designed filters can screen out interference like short segments and architectural façades segments. Contour segments of road and building top surface are extracted, as shown in Fig. 1.

3.2 SDS Feature and Descriptor

Due to the substantial disparities in viewpoint, resolution and scale between UAV large oblique images and orthographic images, the existing features of the same target exhibit noticeable differences, resulting in reduced accuracy in geo-localization. To address these limitations, this paper introduces the SDS (Segments Direction Statistics) feature and its corresponding descriptor. The SDS feature comprises both a local part and a global part, specifically designed to mitigate the impact of the impact of these significant differences on UAV large oblique image geo-localization.

3.2.1 Local SDS: With large viewpoint and scale gaps, there are significant differences in length and direction of extracted homonymous segments. It is difficult to apply segments features for cross-view image matching directly, but the statistics of segments direction shows stable for image translation, rotation and other deformations. Therefore, SDS is designed in this paper as features to be matched. For the extracted contour segments $\{L_1, L_2, \dots, L_n\}$, the IPCS is established with the lower left corner as the origin point. The slope of these segments $\{k_1, k_2, \dots, k_n\}$ and the angle with the horizontal axis $\{\theta_1, \theta_2, \dots, \theta_n\}$ are obtained, shown in formula (5):

$$\theta_i = \begin{cases} 0, y_1 - y_2 = 0 \\ \pi/2, x_1 - x_2 = 0 \\ \arctan k_i = \arctan \frac{y_1 - y_2}{x_1 - x_2}, \text{else} \end{cases} \quad (5)$$

where $(x_1, y_1), (x_2, y_2)$ = the endpoints of segments

Due to the large number of segments and uneven distribution density, segments in dense building areas are counted repeatedly. Therefore, the images are divided by set windows, and the major value of θ_i in the window is determined as the local SDS feature. The local SDS features are counted as 16 intervals (through lots of experiments, SDS with 16 intervals show the best geo-localization results), and each window provides the largest number of statistics. The local SDS features are counted to establish its histogram descriptor, as shown in formula (6):

$$Hist_L = \left\{ \theta_i^j, \theta_i^j \in \left[\frac{n\pi}{16}, \frac{(n+1)\pi}{16} \right], n = 0, 1, \dots, 15 \right\} \quad (6)$$

where $Hist_L$ = the histogram descriptor of local SDS

θ_i^j = the local SDS feature of the j^{th} image window

Since there is a large resolution difference between UAV images and orthographic images, the window size should be set inversely proportion to the resolution, which is able to ensure the scale invariance of the local SDS feature.

3.2.2 Global SDS: With the local SDS feature, it can ensure the robustness when there are only simple deformation and scale difference between images. But for UAV large oblique images, the deformation is more complex. (Wu et al., 2021) verified that the angle between feature lines can be used as a stable feature for matching the images with large deformation. Therefore, it can be applied to construct the global SDS feature in the proposed method. The angle between contour segments can be determined with formula (7):

$$\varphi_i = \arctan\left(\frac{|k_i - k_j|}{1 + k_i k_j}\right) \quad (7)$$

where φ_i = the angle between contour segments

Due to the specific segment slopes, there are several cases where φ_i needs to be determined separately, as shown in formula (8):

$$\varphi_i = \begin{cases} \pi/2, & k_i k_j = -1 \\ 0, & k_i = k_j = \text{nan} \\ \arctan k_i - \pi/2, & k_j = \text{nan} \\ \arctan k_j - \pi/2, & k_i = \text{nan} \end{cases} \quad (8)$$

where nan = this parameter cannot be obtained correctly

$$S = \sqrt{\sum_L^{16} P_L \cdot \left(\frac{\text{Hist}_L^{\text{UAV}}}{\text{Hist}_L^{\text{max}1}} - \frac{\text{Hist}_L^{\text{SAT}}}{\text{Hist}_L^{\text{max}2}}\right)^2} + \sqrt{\sum_G^{16} P_G \cdot \left(\frac{\text{Hist}_G^{\text{UAV}}}{\text{Hist}_G^{\text{max}1}} - \frac{\text{Hist}_G^{\text{SAT}}}{\text{Hist}_G^{\text{max}2}}\right)^2} \quad (10)$$

where S = the similarity of cross-view image matching

$\text{Hist}_L^{\text{UAV}}, \text{Hist}_G^{\text{UAV}}, \text{Hist}_L^{\text{SAT}}, \text{Hist}_G^{\text{SAT}}$ = the SDS feature histograms of UAV oblique images and satellite orthographic images

In order to highlight the effective features and suppress the interference information, two weight parameters p_L and p_G are introduced as similarity measure, as shown in formula (11):

$$p = 0.5 \cdot \left(\frac{\text{Hist}^{\text{UAV}}}{\text{Hist}^{\text{max}1}} + \frac{\text{Hist}^{\text{SAT}}}{\text{Hist}^{\text{max}2}}\right) \quad (11)$$

By determining the similarity between UAV images and orthographic reference images with formula (9), geo-tags with the highest similarity can be obtained as the location of UAV images, which means geo-localization has been completed. However, with some interference features in UAV large oblique images caused by viewpoint and scale gaps, mismatching is hard to be avoided in image-based geo-localization. With UAV sequence images, the bilateral matching strategy is designed to improve the robustness of the proposed method, as formula (12):

$$S = \max\left(S(\text{UAV}, \text{SAT})^i + \omega \cdot S(\text{SAT}, \text{UAV})^i\right) \quad (12)$$

where $S(\text{UAV}, \text{SAT})^i, S(\text{SAT}, \text{UAV})^i$ = the matching similarity
 ω = the set relaxation factor

Similarly, φ_i of contour segments is counted to form the global SDS feature, which is described by its histogram descriptor, as shown in formula (9):

$$\text{Hist}_G = \left\{ \varphi_i, \varphi_i \in \left[\frac{n\pi}{16}, \frac{(n+1)\pi}{16} \right), n = 0, 1, \dots, 15 \right\} \quad (9)$$

where Hist_G = the histogram descriptor of global SDS

For extracted contour segments, the SDS feature including the local and global parts is described with Hist_L and Hist_G , as shown in Fig. 2.

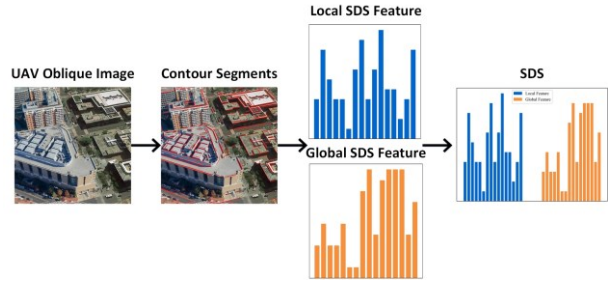


Figure 2. Segments direction statistical (SDS) feature and its descriptor.

3.3 Image-Based Geo-Localization

In this paper, UAV large oblique image geo-localization is to obtain the location of UAV images through matching them with the orthographic reference images. Based on SDS feature and its descriptor, the similarity measure is designed as formula (10):

Since $S(\text{UAV}, \text{SAT})$ is the main factor and $S(\text{SAT}, \text{UAV})$ plays a fine-tuning role, ω can usually be set within 0.5. Similarity with bilateral matching can be used as a better geo-localization evaluation criterion.

4. EXPERIMENTS AND ANALYSIS

4.1 Experimental Datasets

4.1.1 Dataset1: University-1652: University-1652 is a UAV oblique image dataset provided by (Zheng et al., 2020), which is a popular public dataset for image classification, matching and image-based geo-localization. It contains images of 1652 buildings from 72 universities around the world. Each target includes images from three viewpoints: UAV-view, satellite-view and street-view, as shown in Fig. 3. UAV-based image geo-localization experiments mainly use UAV-view and satellite-view images in this paper, usually with 54 UAV oblique images and a satellite-view image of each target.

4.1.2 Dataset2: UAV LO-LF-W Images: Different from UAV-view images in the University-1652, there is a significant

gap in the actually collected UAV large oblique image. In order to verify the reliability of the proposed method in real applications, UAV large oblique long focal whiskbroom images (Ye et al., 2022) are collected for geo-localization experiments. As shown in Fig. 4, the oblique angle of these images is about 66° , and the flight height is more than 4 km, with smaller scale and lower spatial resolution than Dataset1. Therefore, it is much more challenging for image-based geo-localization on this actual image dataset. This dataset is collected in Weinan city, Shaanxi province, China. Similarly, Google Map products are obtained as the orthographic images with geo-tags.



Figure 3. Dataset1: University-1652. Images of different views including UAV-view, Satellite-view and Street-view.

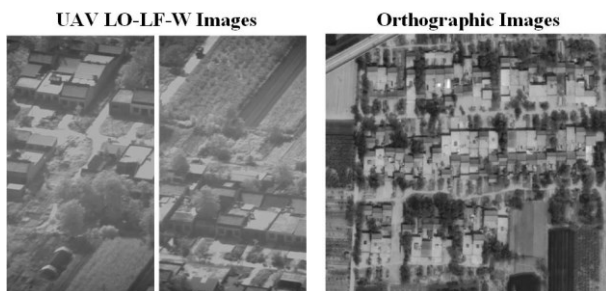


Figure 4. Dataset2: UAV LO-LF-W images and orthophotos.

4.2 Experimental Results

4.2.1 Evaluation Indicators: In most existing image-based geo-localization researches (Patel et al., 2022), $Recall@K$ is a popular accuracy evaluation indicator, which reflects the correct geo-localization result occurring in the top K results of matching score. K is often set to 1, 5 and 10. $K=1$ represents the ratio of automatically corrected geo-localization images, while $K=5,10$ can also reduce the labour cost in the man-machine interactive UAV-based image geo-localization task. $Recall@K$ can be determined with formula (13):

$$Recall@K = \frac{Recall_N}{N} \cdot 100\% \quad (13)$$

where N = the total number of geo-localization images
 $Recall_N$ = the number of correctly geo-located images

The average precision of image retrieval AP is another popular evaluation indicator for image-based geo-localization, which

shows the area under the precision-recall curve, and can be determined with formula (14):

$$AP = \frac{1}{N} \sum_I^N 0.5 \cdot (P^i + P^{i-1})$$

$$P^i = \begin{cases} 1, & i = 0 \\ \frac{TP + 1}{TP + FP}, & i > 0 \end{cases} \quad (14)$$

where P^i = the precision of the previous i^{th} images
 TP = the number of correctly geo-located images
 FP = the number of incorrectly geo-located images

4.2.2 Geo-Localization Accuracy Evaluation on Dataset1:

Three UAV-view images and the top five satellite-view images are selected from Dataset1, to show the geo-localization results with the proposed method, as shown in Fig. 5. The images with green border are the correctly matched satellite-view image, and images with the red border are the mismatched results.

On Dataset1, the geo-localization accuracy of the proposed method is evaluated and compared with Zheng's model (Zheng et al., 2020) and Ding's model (Ding et al., 2021). The experiments are carried out with 37854 UAV-view images, and the accuracy of geo-localization is counted in Tab. 1. Compared with Zheng's model, the proposed method shows the improvement of 16.66%, 8.80% and 16.89% on $Recall@1$, $Recall@5$ and AP ; Meanwhile, compared with Ding's model, the proposed method shows the improvement of 8.50%, 3.96%, 4.01% and 9.20% on $Recall@1$, $Recall@5$, $Recall@10$ and AP . On $Recall@1$ and AP indicators, the proposed method shows obvious advantages over the other two models.

Method	$Recall@1$	$Recall@5$	$Recall@10$	AP
Zheng's model	58.49	-	85.23	63.13
Ding's model	66.65	84.93	90.02	70.82
Our method	75.15	88.89	94.03	80.02

Table 1. Comparison of geo-localization accuracy among our method, Zheng's model and Ding's model.

To achieve automatic geo-localization of UAV images, $Recall@1$ needs to be close to 100%. However, due to the large differences between oblique images and orthographic images, the existing methods, including the proposed method, are still difficult to reach this ideal goal. $Recall@5$ and $Recall@10$ of the proposed method reach about 90% and 95%, which means that the vast majority of correctly geo-located results will occur in the top 10 or even the top 5 images in the matching score. For the man-machine interactive UAV image geo-localization, compared with selecting among a large number of UAV images, the efficiency and accuracy of geo-localization can be greatly improved with the proposed method.

4.2.3 Geo-Localization Accuracy Evaluation on Dataset2:

Similarly, three UAV LO-LF-W images and their top five orthographic images are selected from Dataset2, to show the geo-localization results with the proposed method, as shown in Fig. 6. With the proposed method, the geo-localization accuracy of 150 UAV image blocks in the UAV LO-LF-W images is

evaluated, and the results are shown in Tab. 2. Due to the larger viewpoint gap of images on Dataset2, the scene is more complex and the resolution is greatly different from that of orthographic images. The accuracy of geo-localization on this dataset is significantly lower than that of Dataset1, especially with an obvious decrease of about 20% on $Recall@1$. However, most methods (including Zheng’s model and Ding’s model) that perform well on public UAV large oblique images datasets even fail on Dataset2. Therefore, on the challenging Dataset2, the geo-localization accuracy of the proposed method is relatively high. Moreover, the $Recall@10$ of the proposed method on

Dataset2 can still reach 85%. Compared with the existing UAV-based image geo-localization methods used in engineering such as SIFT, ASIFT and HAPCG, the accuracy and efficiency of the proposed method have been improved.

Dataset	$Recall@1$	$Recall@5$	$Recall@10$	AP
Dataset1	75.15	88.89	94.03	80.02
Dataset2	52.67	74.67	86.00	69.11

Table 2. Comparison of geo-localization accuracy of our method on Dataset1 and Dataset2.

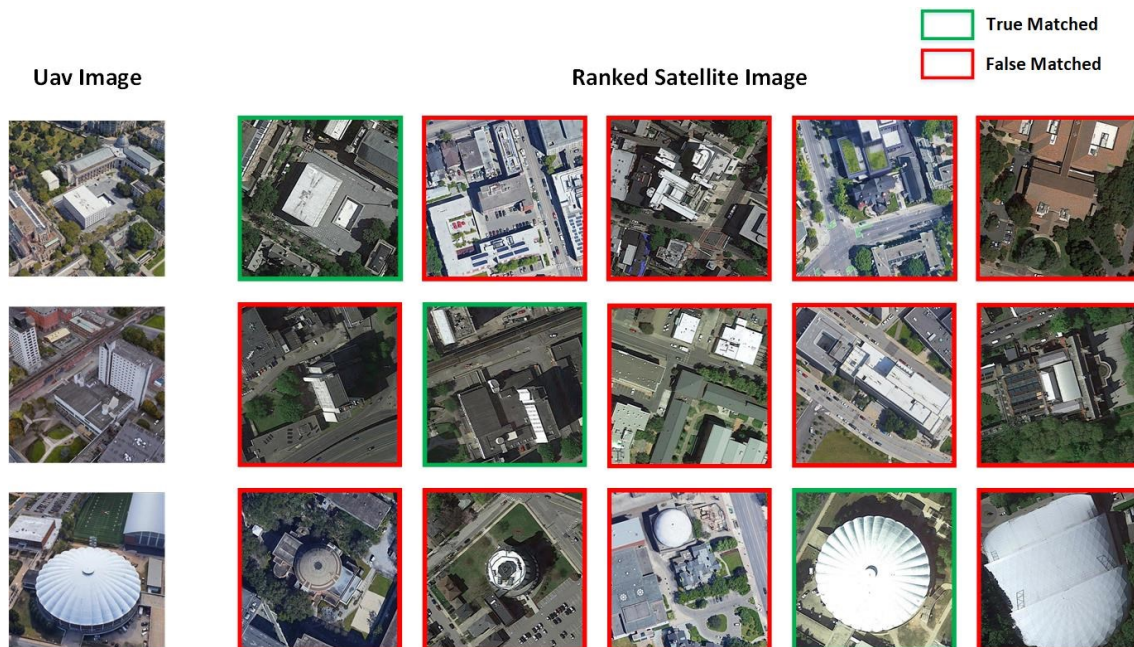


Figure 5. Top 5 matched satellite-view images of UAV images (1619th, 0839th and 1605th in Dataset1).

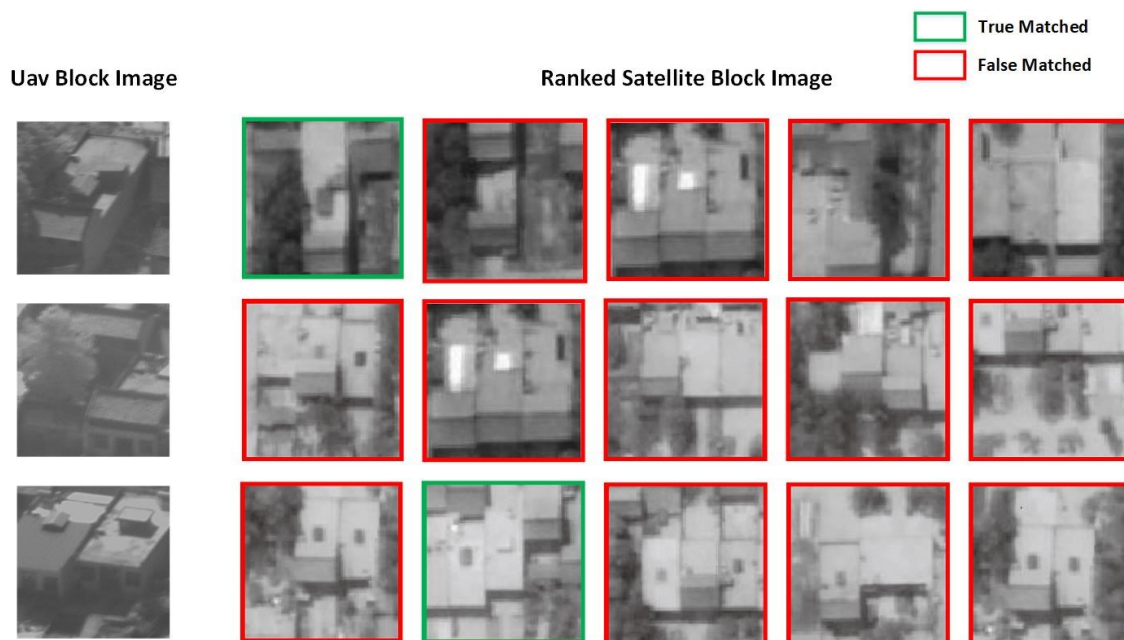


Figure 6. Top 5 matched orthographic images of UAV images (18th, 46th and 119th in Dataset2).

4.3 Influencing Factors Analysis

4.3.1 Orientation and noise of nonstandard reference images:

The two experimental datasets in this paper use the satellite-view images provided by Google Maps, which possess consistent orientation and quality. However, real-world scenarios often involve nonstandard reference images that exhibit varying orientations and significant noise. To assess the geo-localization accuracy of the proposed method under such conditions, the reference images in the two datasets have been rotated and subjected to noise augmentation. These modified reference images, featuring nonstandard orientations and Gaussian noise, are then used as input for the original dataset. Subsequently, the geo-localization accuracy of the proposed method has been evaluated as shown in Fig. 7.

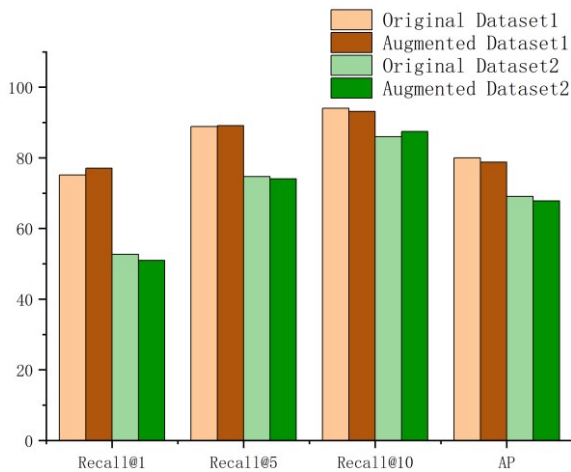


Figure 7. Geo-localization accuracy on original and augmented datasets.

The results indicate that the geo-localization accuracy of the proposed method experiences minimal fluctuations, about 1%, when reference images are subjected to rotation and Gaussian noise. Interestingly, the accuracy demonstrates both slight improvements and declines. This highlights the robustness of the proposed method in handling variations in reference image orientation and noise. Consequently, the method exhibits favourable suitability for practical applications involving multi-source reference images with diverse orientations and image qualities, facilitating UAV large oblique image geo-localization.

4.3.2 Artificial feature abundance: With designed SDS feature constructed by the contour segments of buildings and roads, the abundance of these artificial features should be one of the important influencing factors. Since the UAV images on Dataset1 are taken around single of a few buildings, the artificial feature abundance is much higher. However, for images of Dataset2, the artificial feature abundance is lower. Therefore, it is assumed that the artificial feature abundance is the main reason for lower accuracy of geo-localization on Dataset2. The artificial feature abundance is described by the proportion of artificial feature pixels on UAV images, determined with formula (15):

$$abd = \frac{PN_a}{PN} \cdot 100\% \quad (15)$$

where abd = the artificial feature abundance of images
 PN_a, PN = the number of artificial feature pixels and the whole image pixels

The artificial constructions on the UAV oblique image are segmented, and the number of pixels it contains is counted. According to the artificial feature abundance, the UAV images of Dataset2 can be divided into four intervals: {0~25%, 25%~50%, 50%~75%, 75%~1}. The average and standard deviation of the geo-localization accuracy in different features abundance intervals have been obtained respectively, as shown in Fig. 8. When the artificial feature abundance belongs to 0~25%, the geo-localization accuracy of the image is below 10%, that is, it is difficult for the proposed method to obtain accurate geo-localization results of UAV large oblique images with rare buildings areas. However, when the feature abundance is higher than 50%, the geo-localization accuracy of the proposed method is higher than that of the entire Dataset2. Moreover, when the feature abundance reaches 75%, the geo-localization accuracy evaluation index $Recall@10$ can reach 90%, indicating that when the feature abundance is high enough, the proposed method can well meet the geo-localization requirements of challenging UAV LO-LF-W images. It shows that the abundance of artificial features is the key influencing factor of the proposed geo-localization method.

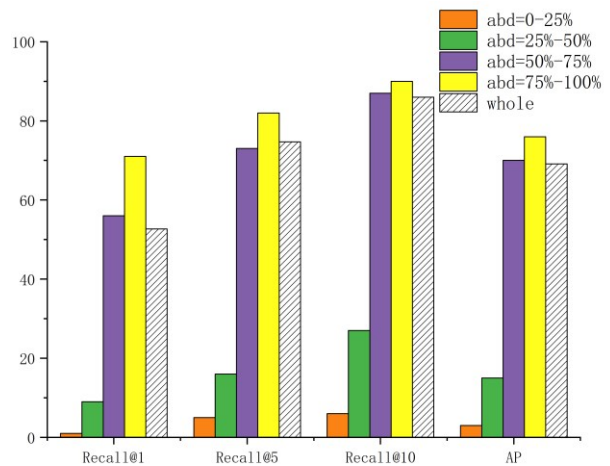


Figure 8. Geo-localization accuracy on images with different artificial feature abundance (abd).

In summary, for UAV large oblique images, the proposed method shows high accuracy and robustness, especially in the challenging scenes with dense buildings. On the public dataset University-1652, the geo-localization accuracy of the proposed method is much higher than existing methods; On the actually collected UAV LO-LF-W images, the proposed method can still obtain reliable geo-localization results, while many other methods that perform well on the public dataset are hardly to solve the image-based geo-localization on these actually collected images. Moreover, for reference satellite-view images with different orientation and noise, geo-localization results obtained by the proposed method are very stable, which proves that it can use challenging multi-source nonstandard reference images to realize UAV large oblique images geo-localization. And the accuracy of geo-localization depends on the abundance of artificial features.

5. CONCLUSIONS

Accurately geo-locating UAV large oblique images remains an immensely challenging task due to the substantial viewpoint differences and the presence of considerable façade interference information, particularly in dense buildings areas. To address these challenges, this paper presents a novel method for UAV

large oblique image geo-localization. The proposed method incorporates feature filters to eliminate interference features, introduces SDS feature patterns for extraction and description, and employs a bilateral similarity measure strategy for matching and geo-localization. The proposed method fully addresses the robustness of the designed features in handling image matching across varying viewpoints and scales while considering both local and global features. In addition to focusing on the significant viewpoint differences addressed by existing methods, the proposed method also accounts for the interference caused by architectural façade information, resulting in improved geo-localization accuracy. Experimental results show that the proposed method can better serve the geo-localization for UAV large oblique images, especially in the dense building areas.

Meanwhile, since the proposed method is based on the segment features of artificial structures, it is hard to obtain accurate geo-localization results for natural landforms or areas with sparse buildings and roads. Therefore, we will consider to expand the application scenes into scenes with more complex or sparse features as our future work.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Shanghai (Grant No. 22ZR1465700) and the Open Fund Project of State Key Laboratory of Geo-Information Engineering (Grant No. SKLGIE2018-K-3-1).

REFERENCES

- Berton, G., Masone, C., Caputo, B., 2022. Rethinking visual geo-localization for large-scale applications. *Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4868-4878.
- Bianchi, M., Barfoot, T., 2021. UAV localization using autoencoded satellite images. *IEEE Robotics and Automation Letters*, 6(2), 1761-1768.
- Couturier, A., Akhloufi, M., 2021. A review on absolute visual localization for UAV. *Robotics and Autonomous Systems*, 135, 103666.
- DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: self-supervised interest point detection and description. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 337-349.
- Ding, L., Zhou, J., Meng, L., Long, Z., 2021. A practical cross-view image matching method between UAV and satellite for UAV-based geo-localization. *Remote Sensing*, 13(1), 47.
- Goel, P., Dayal, P., Patra, T. Gupta, C., 2022. Automatic co-registration of aerial and satellite imagery using optimized minimum distance thresholds for matching. *Proceedings of 2022 International Mobile and Embedded Technology Conference (MECON)*, 25-28.
- Hasheminasab, S., Zhou, T., Laforest, L., Habib, A., 2021. Multiscale image matching for automated calibration of UAV-based frame and line camera systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14,

3133-3150.

- Ma, J., Jiang, X., Fan, A., Jiang, J., Yan, J., 2021. Image matching from handcrafted to deep features: a survey. *International Journal of Computer Vision*, 129(1), 23-79.
- Patel, Y., Tolia, G., Matas, J., 2022. Recall@k surrogate loss with large batches and similarity mixup. *Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7492-7501.
- Sarlin, P.-E., DeTone, D., Malisiewicz, T., Rabinovich, A., 2020: SuperGlue: learning feature matching with graph neural networks. *Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4937-4946.
- Tian, X., Shao, J., Ouyang, D., Shen, H., 2021. UAV-satellite view synthesis for cross-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(7), 4804-4815.
- Wu, Q., Xu, G., Cheng, Y., Wang, Z., Li, Z., 2021. Deformed contour segment matching for multi-source images. *Pattern Recognition*, 117, 107968.
- Xu, G., Wu, Q., Cheng, Y., Yan, F., Li, Z., Yu, Q., 2021. A robust deformed image matching method for multi-source image matching. *Infrared Physics and Technology*, 115, 103691.
- Yao, H., Qin, R., Chen, X., 2019. Unmanned aerial vehicle for remote sensing applications - a review. *Remote Sensing*, 11(12), 1443. doi.org/10.3390/rs11121443.
- Ye, Q., Zhang, X., Lin, Y., Jian, J., 2022. Accuracy analysis of real-time object positioning without gcp for images from UAV oblique ultra-long focal small view field whiskbroom camera system. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII/B1, 325-332.
- Zamir, A., Shah, M., 2014. Image geo-localization based on multiple nearest neighbor feature matching using generalized graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8), 1546-1558.
- Zhang, R., Xu, F., Yu, H., Yang, W., Li, H., 2020. Edge-driven object matching for UAV images and satellite SAR images. *Proceedings of 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 1663-1666.
- Zheng, Z., Wei, Y., Yang, Y., 2020. University-1652: a multi-view multi-source benchmark for drone-based geo-localization. *Proceedings of the 28th ACM International Conference on Multimedia*, 1395-1403.
- Zhou, T., Lv, L., Liu, J., Wan, J., 2021. Application of UAV oblique photography in real scene 3D modelling. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII/B2, 413-418.
- Zhuang, J., Chen, X., Dai, M., Lan, W., Cai, Y., Zheng, E., 2022. A semantic guidance and transformer-based matching method for UAVs and satellite images for UAV geo-localization. *IEEE Access*, 10, 34277-34287.