GPS-ASSISTED FEATURE MATCHING IN AERIAL IMAGES WITH HIGHLY REPETITIVE PATTERNS

Gonzalo Luzardo[⋆†]*, Michiel Vlaminck*[⋆]*, Dionysios Lefkaditis*[‡]*, Wilfried Philips*[⋆]*, Hiep Luong*[⋆]

[⋆] imec-IPI-URC, Ghent University, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
[†]ESPOL Polytechnic University, Km. 30.5 Vía Perimetral, Guayaquil, Ecuador
[‡]SITEMARK, Gaston Geenslaan 11, 3000 Leuven, Belgium

ABSTRACT

Matching aerial images might be challenging when they contain a large number of repetitive patterns. In this paper, we propose a feature-matching method that exploits the use of Affine Oriented FAST and Rotated BRIEF (AORB) as keypoint detector and feature descriptor and not accurate GPS (Global Position System) data to achieve a reliable feature matching of nadir UAV images that contain a large number of repetitive patterns. The proposed method assumes that the set of correct matches between two images only differ in a 2D translation. Experimental results show that the proposed method is able to correctly match pairs of very challenging images containing a large number of repetitive patterns.

1. INTRODUCTION

Feature matching is an important step in Structure-from-Motion (SfM) photogrammetry applications for 3D and 2D reconstruction purposes. It consists of comparing two sets of points, known as keypoints, from two different but overlapping images. The matching process compares the descriptors of the keypoints in both images based on a certain distance. This process usually includes a filtration step, where wrong matches or outliers are removed from the original set of matches. In the SfM pipeline, a wrong or a low number of correct feature correspondences, also know as inliers, between images makes the camera pose estimation less reliable and can lead to a wrong or incomplete reconstruction [5]. Matching aerial images is a challenging task mainly because they often contain repetitive structures, such as trees, houses, buildings, crops, solar panels, etc [2].

Descriptors of feature points are commonly based on local image information. Therefore, descriptors from repetitive patterns may not be unique which contributes to a lack of distinction in those regions [4]. This lack of distinction results in a large number of outliers, due to local and global ambiguities [3]. In [10], a probabilistic method based on a Bayesian model to remove these outliers by using global, local, and manifold regularizations is proposed. Likewise, a feature matching method for almost nadir-directed UAV images is proposed in [2]. This method assumes that the set of correct matches between two images will only differ in a 2D translation that is estimated by computing so-called pixel-distance histograms on a set of candidate matches. Distinct peaks located in pixel-distance histograms for X and Y coordinates represent an unknown 2D translation in each coordinate, respectively. More recently, a feature matching method for UAV images that combines the geometric information with the feature similarity is proposed in [8]. Here, the feature matching is restricted in pairwise geometric grid cells to avoid unnecessary feature similarities computations. Grid cells are defined by the result of matching large-scale SIFT features and selecting the top 10% of them to build the neighborhood pairs that define the grid cells.

Despite the fact, these techniques have shown good performance in matching images with repetitive patterns, their performance decreases when the number of repetitive patterns is larger. Regularizations as a step for filtering outliers proposed in [10] is partly based on agricultural UAV images that contain local structures among neighboring feature points that can be used as local geometrical constraints. However, when the repetitive structures are present on the entire image, this assumption does not hold. Likewise, the repetitive patterns on images can create a large number of peaks in pixel-translation histograms when the method proposed in [2] is used. In fact, a higher peak does not always represent the actual translation. In addition, large-scale SIFT features proposed in [8] will not be able to successfully find unique correct correspondences.

In this paper, we propose a feature-matching method that overcomes the problems related to having a large number of repetitive patterns. As in [2], our method is based on the assumption that the UAV images were captured in such a way that the set of correct matches between two images will only differ in a 2D translation. Additionally, it exploits the use of Oriented FAST and Rotated BRIEF (ORB) extended with affine transformations as keypoint detector and feature descriptor, and not accurate GPS (Global Position System) data to achieve a reliable feature matching of nadir oriented UAV images.

2. PROPOSED FEATURE MATCHING

As in [2], our proposed method is based on estimating a pixel-translation vector between matched keypoints coordinates by using pixel-distance histograms. However, our method is specially tailored to deal with images that contain a large number of repetitive patterns and differs in: (i) it uses AORB (Affine Oriented FAST and rotated BRIEF) feature extractor and descriptor, (ii) matches are filtered before computing the pixel-distance histograms, and (iii) it uses GPS information to discriminate false peaks in histograms caused by local ambiguities when images contain a large number of repetitive patterns.

The pixel-distance histogram is created by computing the coordinates differences of a set of candidate matches between a pair of images. The pixel shifts differences between the matched keypoints of a pair of images (I_i, I_j) are computed as follows:

$$\Delta_r^k = (r_i^m - r_j^n)^k \qquad \Delta_c^k = (c_i^m - c_j^n)^k \tag{1}$$

where k=1,...,N; N is the number of candidate matches found in the image pair, and (r_i^m, c_i^m) and (r_j^n, c_j^n) are the row and column pixel-coordinates of a m-th and n-th matched keypoints in I_i and I_j, respectively.

Due to the fact that image pairs are not perfectly aligned, pixel-distance histograms for rows and columns are computed by using bins of size $d > 1$. The value of d depends on the scene depth and how well the images are aligned. Higher values allow dealing with images that are not well aligned, increasing the range so that correct matches with not exactly the same pixel-distance belong to the same bin in the histogram.

2.1. Feature extraction

Despite the fact SIFT and ASIFT have proven to have a good performance in feature matching in [2], in our experiments using aerial images with a large number of repetitive patterns SIFT did not show a good performance. We found that SIFT is not able to find the number of correspondences between images with a large number of repetitive patterns, necessary to create highly prominent peaks in the pixel-distance histograms that allow identifying the correct translation between images. Therefore, we carried out a study to find the feature extractor and descriptor methods more suitable to be used in this feature matching approach. We found that Oriented FAST and rotated BRIEF (ORB) extended with affine transformations [9], which we refer to as AORB, are the best suited not only with images with a large number of repetitive patterns but also with aerial images in general when the translation estimation approach is used. This is due to the fact that ORB has shown high and stable repeatability for matching images [7], which is reinforced when affine transformations are used. To minimize the noise present in pixel-distance histograms, the images are rectified using the intrinsic camera-parameters.

2.2. Feature matching

After computing features using AORB, feature matching between image pairs, a query image (I_q) and a reference image (I_r) with overlap. The feature matching is performed using Hamming distance, which is the most suitable to compare binary descriptors as AORB. To tackle the problem of local ambiguities, we compute the matching using multiple nearest neighbors as matching candidates, known as k-nearest neighbor matching. This process involves matching one feature from one image with k features from the other image in a pair.

Based on the premise that most correct correspondences are not the closest matches when a large number of repetitive patterns are present [6], this approach significantly increases the probability of finding correct matches. Of course, this approach also has a downside: in addition to correct matches, it will introduce many false correspondences due to local ambiguities.

2.3. Geometric verification assisted by GPS data

The geometric verification is carried out in three steps: i) camera orientation estimation, ii) translation estimation, and iii) selection of the correct matching correspondences. As in [2], the translation estimation is done by using a pixel-distance histograms. However, we found that using the ratio test as a prior filter step results in less noisy histograms when a large number of repetitive patterns are present compared to using the raw matches.

2.3.1. Camera orientation estimation

UAV images do not always have the same orientation. For example, in a zig-zag flight pattern where the UAV turns for capturing the subsequent row, images between consecutive rows are not aligned but rotated 180 degrees. This misalignment can be estimated during the matching process taking the first image as the reference and labeling their matched pairs as aligned in the same direction or not with the reference. Subsequent pairs are labeled based on the pairs that have been already labeled. As in [2], a pixel-rotation histogram is employed to estimate if an image pair is aligned or not. However, we take into consideration only two discrete rotation values: 0 degrees when the image pair is aligned and 180 degrees when not.

2.3.2. Translation estimation assisted by GPS data

Translation estimation is performed by computing the pixel-distance histograms from filtered matches and identifying the peaks. The x- and y-coordinates of the keypoints are rotated

with respect to the center of the image for those images that are not aligned with the reference image.

When images have a large number of repetitive patterns, several peaks are present in the pixel-distance histograms. Therefore, peaks are considered as putative translations that need to be processed to identify the actual translation. This identification is carried out by using the GPS data and camera parameters to discard wrong translations in the presence of many peaks in the pixel-distance histograms. This discarding process is done in four steps: i) the orientation of the cameras with respect to the reference image is estimated, ii) the position of the peaks in the row and column pixel-distance histograms are identified and considered as the putative pixel-translations, iii) GPS data and intrinsic camera-parameters are used to compute the approximate pixel translation in rows and column respectively, iv) the putative translations in rows and columns closer to approximate pixel translation are considered as the actual translation.

To calculate the approximate pixel-translation for columns (d_x) and rows (d_y) between two images, the longitude and latitude coordinates are converted to x- and y-coordinates using a simple equirectangular projection which is reasonably accurate over small distances. Likewise, because the flight trajectory is not always rotationally aligned to the x-y plane, x- and y-coordinates are rotated to be aligned with respect to the flight trajectory. Therefore, the pixel-translation estimation d_x and d_y for columns and rows in an image pair (i, j), can be calculated as follows:

$$d_x = G_x(y_r^j - y_r^i)q \qquad d_y = G_y(x_r^j - x_r^i)q \tag{2}$$

where (x_r^i, y_r^i) and (x_r^j, y_r^j), are the rotated coordinates of the image pair (i, j), $q = 1$ when the camera orientation of the reference image is pointing to the positive x-coordinate and $q = -1$ when is pointing to the negative, and G_x and G_y represent the ground sample distance (GSD) for the x- and y-coordinate, respectively.

Finally, the actual pixel-translation in rows (d_r^{ij}) and columns (d_c^{ij}) between an image pair (i, j) are identified as the closer localization of peaks in the pixel-distance histograms to d_x and d_y, respectively, which allows dealing with not accurate GPS data.

2.3.3. Selection of the correct matching correspondences

All matches between the image pair (i, j) obtained by k-nearest neighbor matching that have the same actual pixel-translation in rows (d_r^{ij}) and columns (d_c^{ij}) taken into account a certain threshold t are selected as the correct matches. That means, correct matches should satisfy the following equation:

$$|\Delta_r^k - d_r^{ij}| \leq t \wedge |\Delta_c^k - d_c^{ij}| \leq t \tag{3}$$

where t should be larger or equal than the size of the bins d used to compute the pixel-translation histograms ($t \geq d$).

A higher threshold t allows dealing with small camera misalignments to the flight-path. Increasing t will increase the number of matches. However, it also will allow accepting more outliers as correct ones. To filter outliers present when a larger value t is used we performed a post-filtering using RANSAC.

3. RESULTS

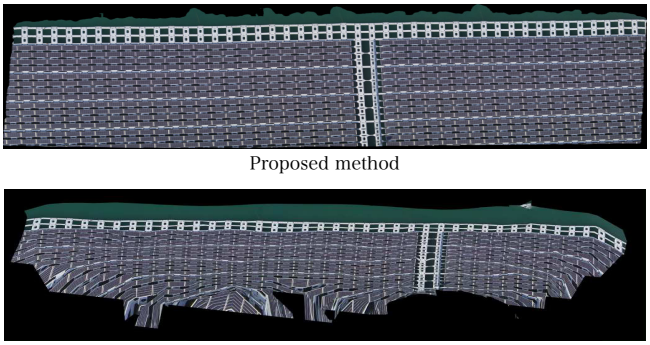
We evaluate our method using two datasets, one captured near a photo-voltaic (PV) power plant in Japan and the other one near a PV power plant in France. Both datasets contain nadir-oriented images with a moderate and a high number of repetitive patterns, respectively. The Japanese dataset contains 21 images with a resolution of 4000x3000 whereas the French dataset contains 42 images with a resolution of 4608x3456. In particular, the French dataset represents a very challenging scenario because all images contain repetitive patterns that cover almost the entire field of view. For both, the intrinsic camera parameters and the GPS data were known. A maximum of 43000 keypoints and features were extracted from each image using AORB. Matches were computed using 50 nearest neighbors and a bin size of 15 ($d = 15$) to create the pixel-distance histograms. Only images with at least 60% of overlap were matched.

We compare our results with those obtained by the method proposed in [2], labeled as KOCH, and the commercial software package Agisoft Metashape [1]. For this, we estimate the camera pose (location and rotation) using the matches obtained by our method and those obtained by Metashape and KOCH. Camera poses obtained were compared with the ground truth, which was obtained by manually matching the correspondences. If the difference between an estimated camera pose and the camera pose in the ground truth is less than a small threshold, the estimated camera pose is considered correct.

Table 1 shows the results obtained from this evaluation. As expected, when a large number of repetitive patterns are present as in the French dataset, local and global ambiguities cause the camera pose estimation in Metashape and KOCH to fail. As can be seen, matches obtained from the proposed method can be used to correctly estimate all the camera poses, even if images contain a large number of repetitive patterns such as solar panels.

Dataset	Number of correct camera poses		
	Agisoft Metashape	KOCH [2]	Proposed
Japan	42/42	41/42	42/42
France	3/21	0/21	21/21

Table 1. Results of the comparison between the number of correct camera-poses computed using matches obtained from our proposed method and those obtained by Agisoft Metashape and KOCH [2].



Proposed method

Agisoft Metashape

Fig. 1. Orthomosaics generated by our proposed method (top) and Agisoft Metashape (bottom). Our method is able to successfully reconstruct an accurate orthomosaic with highly repetitive patterns.

Additionally, for the French dataset we compute the orthomosaic generated using the camera position from both Metashape and our proposed method. As can be seen in Fig. 1, wrong camera position estimations in Metashape result in a wrong orthomosaic. Unlike Metashape, a complete and high-quality orthomosaic can be generated using the camera positions estimated from the matches generated by our proposed method.

4. CONCLUSIONS

The experiments show that our proposed method is capable of correctly estimate feature matches of UAV images containing a large number of repetitive that were captured in such a way that the set of correct matches only differ in a 2D translation.

The proposed method uses ORB extended with affine transformations (AORB), which is free to use and showed improved performance on matching images that contain highly repetitive patterns. Likewise, the use of AORB as a feature descriptor makes the matching process much faster and it becomes an efficient alternative to SIFT used in [2].

Our proposed method is able to remove outliers from local and global ambiguities created by the repetitive patterns using not accurate GPS data, which are equipped on most operational UAVs. This allows to generate a more precise and complete orthomosaic can be generated. Using the more expensive and more accurate differential GPS sensors can further improve to solve the ambiguities.

5. ACKNOWLEDGMENT

This work was financially supported by the following projects: ANALYST-PV, Flanders Innovation & Entrepreneurship project nr. HBC.2019.0050; COMP4DRONES ECSEL Joint Undertaking (JU) under grant agreement No 826610. The work of G. Luzardo is partially supported by Secretaría de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT) and Escuela Superior Politécnica del Litoral (ESPOL).

References

[1] Agisoft Metashape. https://www.agisoft.com/, 2020. Accessed: 2020-01-07.

[2] T. Koch, X. Zhuo, P. Reinartz, and F. Fraundorfer. A new paradigm for matching uav-and aerial images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2016*, 3:83–90, 2016.

[3] C. Le Brese, C. N. Young, and J. J. Zou. A robust match filtering algorithm for use with repetitive patterns. In *2013, 7th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–6. IEEE, 2013.

[4] T. Lin and X. Wang. Hierarchical clustering matching for features with repetitive patterns in visual odometry. *Journal of Intelligent & Robotic Systems*, 100(3):1139–1155, 2020.

[5] C. Stöcker, F. Nex, M. Koeva, and M. Gerke. High-quality uav-based orthophotos for cadastral mapping: Guidance for optimal flight configurations. *Remote Sensing*, 12(21):3625, 2020.

[6] F. Sur, N. Noury, and M.-O. Berger. Image point correspondences and repeated patterns. Research Report RR-7693, INRIA, July 2011.

[7] S. A. K. Tareen and Z. Saleem. A comparative analysis of sift, surf, kaze, akaze, orb, and brisk. In *2018 International conference on computing, mathematics and engineering technologies (iCoMET)*, pages 1–10. IEEE, 2018.

[8] C. Wei, H. Xia, and Y. Qiao. Fast unmanned aerial vehicle image matching combining geometric information and feature similarity. *IEEE Geoscience and Remote Sensing Letters*, 2020.

[9] G. Yu and J.-M. Morel. Asift: An algorithm for fully affine invariant comparison. *Image Processing On Line*, 1:11–38, 2011.

[10] Z. Yu, H. Zhou, and C. Li. Fast non-rigid image feature matching for agricultural uav via probabilistic inference with regularization techniques. *Computers and Electronics in Agriculture*, 143:79 – 89, 2017.