Proceedings of the 37th Chinese Control Conference
July 25-27, 2018, Wuhan, China

Automatic Panoramic UAV Image Mosaic Using ORB Features and Robust Transformation Estimation

Jun Chen[1], Linbo Luo [2], Siwei Wang[2], Hui Wu[1]

1. School of Automation, China University of Geosciences, Wuhan 430074, China

2. School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China
E-mail: wangsiwei1025@163.com

Abstract: Panoramic image mosaic, which aims to take many regular photographic or video images in order to cover the entire viewing space, plays an important role in many remote sensing tasks including map updating, change detection, environmental monitoring and surveillance. Typical mosaic methods involve four steps, namely, feature extraction, feature matching, transformation estimation, and blending. In this study, we introduce several novel strategies in these steps for the fast and automatic construction of unmanned aerial vehicle (UAV) panoramic image mosaic. First, we analyze and test several existing feature extraction techniques and propose to use oriented FAST and rotated BRIEF (ORB) because of its efficiency and capability to generate high-quality feature points. Second, we introduce a fast and robust feature matching strategy based on descriptor similarity along with a locality preserving geometric constraint. Third, we model the spatial transformation between a UAV image pair with an affine function and introduce a robust Bayesian framework to estimate this transformation from the ORB feature matches even if these matches are contaminated by false ones. Finally, we propose a gradual fading method to fuse and blend the matched images to create an attractive panorama. The qualitative and quantitative results of an image set demonstrate that our method exhibits superior performance over existing methods in terms of accuracy and efficiency.

Key Words: image mosaic, UAV images, ORB, LPM, feature matching.

1 Introduction

Unmanned aerial vehicles (UAVs) are regularly equipped with payloads that include high-resolution surveillance cameras. These systems have provided the military with the opportunity to monitor battlefields and remote terrains, conduct reconnaissance missions, and track targets from distant ground stations without endangering UAV operators. In addition to their military applications, the civilian applications of UAVs are also increasing widely with the improvement in the living standards of the people. However, similar to the development of other technologies, many technical challenges occur. The problem frequently arises because the human visual system has a field of view (FOV) of approximately 135° × 200°, whereas the FOV of a typical camera is only 35° × 50°. The decision-making process of detection and recognition algorithms can be rendered ineffective due to the limited visual FOV of individual image frames. An image mosaic technology is frequently adopted to address this issue. The objective of image mosaic is to combine multiple photographic images with overlapping FOVs to create a panoramic image [1].

Two essential approaches are used for image mosaic: direct approaches and feature-based approaches. The goal of the former is to directly minimize pixel-to-pixel mismatching [2]. By contrast, feature-based techniques rely on extracting and matching a series of features. Feature-based methods have become increasingly popular and widespread in image stitching, particularly because the superiority of robust feature algorithms has been proven in recent years[3,4]. In feature-based techniques, all the main feature points in an image pair are compared with all the features in another image by using one of the local descriptors.

In general, the feature-based image stitching process can be divided into four main steps or components: i) image feature extraction, ii) feature matching, iii) image transformation estimation, and iv) blending.

In recent years, many scholars have conducted considerable research on each of the aforementioned step. Vaghela et al. [5] provided a comprehensive comparison of new image stitching techniques. Many well-designed feature detectors have been developed in the past decades. Representative detectors include Harris [6], scale-invariant feature transform (SIFT) [7], speeded up robust features (SURF) [8], features from accelerated segment test (FAST) [9], and oriented FAST and rotated BRIEF (ORB) [10]. Feature matching and transformation estimation can be achieved separately. For example, a brute force [11] technique or its variant [12] is initially used to establish feature matches, and then a least squares technique is adopted to estimate image transformation. However, these two steps are typically performed simultaneously. A popular strategy for achieving this goal involves two steps: i) computing a set of putative correspondence and ii) removing outliers via geometric constraints. Representative strategies include random sample consensus (RANSAC) [13], least median of squares [14], and M-estimator [15] for parametric models, and locally linear transforming [3] and vector field consensus [4] for nonparametric models. For image blending in the final step, frequently used techniques include optimal seam blending, feathering blending, and pyramidal blending [16].

Although existing algorithms can solve most problems, they still cannot simultaneously consider real-time and

*This work is supported by the National Natural Science Foundation of China under Grant 61603354. Corresponding author: Siwei Wang

splicing effects, particularly in the UAV panoramic image mosaic problem. To address the aforementioned challenges, we propose an effective UAV image mosaic method in this study. We then verify the validity of the proposed method on a set of UAV images. In particular, we extract ORB feature points as control point candidates and then match them according to their descriptor similarity with a locality preserving geometric constraint. Subsequently, we propose a Bayesian framework to estimate the spatial transformation between two UAV images based on the ORB feature matches. This framework is robust to potential false matches. Finally, a panorama is created based on the transformed images by using a gradual fading method for blending.

The remainder of this paper is organized as follows. The proposed image stitching method is discussed in Section 2. In Section 3, the estimation of the transformation matrix of UAV images is presented in detail. The experimental results are provided in Section 4. Finally, conclusions are drawn in Section 5.

2 Proposed Image Mosaic Method

Feature-based image stitching systems comprise four stages. First, the features of the input images are extracted. Then, they are matched with one another based on the correspondence similarity of their descriptors. Subsequently, the transform relationship between the images is estimated based on matched feature point pairs. Finally, a blending algorithm is used to eliminate image stitching traces.

2.1 Feature Extraction

The most popular feature extraction algorithms in the literature are SIFT [7], SURF [8], and ORB [10]. These algorithms are introduced briefly as follows.

SIFT has the most common known vector descriptor. SIFT consists of four essential stages: scale-space extrema detection, key point localization, orientation assignment, and key point descriptor generation. In the first stage, the key points are extracted based on their strength, which are are invariant to orientation and scale, by using Gaussian difference. In the second stage, the wrong points are removed via interpolation, Taylor series expansion, and other methods. In the following stage, one or more orientations are assigned to each key point. In the final stage, an *n*-dimensional vector descriptor is made for each key point.

The SURF algorithm is built upon SIFT and is divided into the same four stages. However, it operates differently when extracting features. SURF is based on multi-scale space theory and speeds up its computations by rapidly approximating the Hessian matrix and descriptor using "integral images." Haar wavelets are applied during the stage of key point descriptor generation.

The ORB technique combines features from FAST key point detection and the binary robust independent elementary features (BRIEF) [17] descriptor. The FAST technique is used during the initial stage to determine the key points. FAST does not compute orientation and is rotation variant. It computes the intensity of the patch centroid with the corner at the center. The orientation is the direction of the vector from the center point to the centroid. Moments are computed to improve rotation invariance. In ORB, the rotation matrix is computed using the patch orientation. Then, BRIEF descriptors are steered according to the orientation. ORB describes the features of the input image in a binary string instead of a vector.

In the description of the binary bit string, the matching speed is considerably faster than that in the existing description method because the similarity among descriptors is determined by calculating the Hamming distance, which can be completed only by XOR operation or bit operation.

Although ORB features do not truly solve the scale invariance problem, the scale changes among the images to be spliced by the UAV images are insignificant, and thus, the ORB features can effectively meet the splicing requirements. Therefore, this study uses the ORB feature points for the feature extraction of spliced images.

2.2 Feature Matching

In feature matching, all the features in an image must be stitched to find the best matches. In general, the similarity among descriptors is determined by calculating the Hamming distance. Then, we can use brute force or the best-fin-first algorithm to obtain initial matches. If a feature exhibits a good discrimination degree in the image feature space, then it requires a considerable difference between the nearest neighbor features and the next-nearest-neighbor features. Thus, we can further utilize this constraint to obtain the putative set $S = \{x_i, y_i\}_{i=1}^{N}$, where x_i and y_i are 2D column vectors that denote the spatial positions of feature points.

The use of only local descriptor information will inevitably lead to a number of false matches in the putative set. This problem is typically worse if the image pairs suffer from low-quality, repeated structures. Therefore, we consider adding the locality preserving matching (LPM) algorithm [18] during this stage.

The principle of LPM is to maintain the local neighborhood structures of potential true matches. LPM can accomplish mismatch removal from thousands of putative correspondences within only a few milliseconds.

First, LPM constructs neighborhood $\{\mathrm{N_{x_i}}, \mathrm{N_{y_i}}\}_{\mathrm{i=1}}^{\mathrm{N}}$, where $\mathrm{N_{x_i}}$ denotes the neighborhood of point x_i. LPM evaluation adopts a simple strategy that searches the K nearest neighbors for each point in the corresponding feature set under Euclidean distance. Second, LPM calculates cost $\{c_i\}_{i=1}^{N}$ using Eq.(1). c_i is a cost that measures if the *i*-th correspondence (x_i, y_i) meets the geometric constraint of preserving the local neighborhood structure.

$$c_i = \sum_{j|x_j \in N_{x_i}} d(y_i, y_j) + \sum_{j|y_j \in N_{y_i}} d(x_i, x_j), \quad (1)$$

where

$$d(x_i, x_j) = \begin{cases} 0 & x_j \in N_{x_i} \\ 1 & x_j \in N_{x_i} \end{cases} \quad (2)$$

and the same as $d(y_i, y_j)$.

Third, the true inlier set $\mathcal{L}$ cannot be known in advance but it has to be solved in the problem. Thus, using Eqs. (3) and (4) can determine an approximation solution $\mathcal{L}_0$.

$$p_i = \begin{cases} 1 & \mathrm{c}_i \le \lambda \\ 0 & \mathrm{c}_i > \lambda \end{cases}, i = 1,2,3,\dots,\mathrm{N} \quad (3)$$

Parameter λ is a threshold that judges the match correctness of a putative correspondence.

$$\mathcal{L}_0 = \{i | p_i = 1, i = 1, \dots, N\} \quad (4)$$

Subsequently, we use $\mathcal{L}_0$ to construct the neighborhood for each correspondence in S and solve the optimal $\mathcal{L}^*$ as

$$\mathcal{L}^* = \arg\min_{\mathcal{L}} C(\mathcal{L}; \mathcal{L}_0, S, \lambda), \quad (5)$$

where

$$C(\mathcal{L}; \mathrm{S}, \lambda) = \sum_{i\in\mathcal{L}}(\sum_{j|x_j\in N_{x_i}} \left(d(x_i, y_j) - d(y_i, y_j)\right)^2 + \sum_{j|y_j\in N_{y_i}} \left(d(x_i, y_j) - d(y_i, y_j)\right)^2) + \lambda(\mathrm{N} - |\mathcal{L}|). \quad (6)$$

2.3 Transformation Estimation

After we obtain a reliable pair of matched feature points, we must find the geometric transform model M(x, y) between adjacent images, where x and y are the 2D column vectors that denote the spatial positions of optimal feature points $\mathrm{S}^* = \{x_i, y_i\}_{i=1}^N$. We use homogeneous coordinates for the image point set and M(x, y)=0 for the inlier point correspondence. Researchers always use RANSAC to obtain a minimum outlier-free subset to estimate a given parametric mode. However, RANSAC is dependent on a parametric model, which will not work in the case of image pairs with non-rigid motions because an explicit parametric model is unavailable for such case. Thus, we introduce a method under the Bayesian framework. We formulate our approach as a maximum likelihood problem and solve the geometric transform relationship among the putative point correspondences using an expectation–maximization (EM) algorithm [19-21].

Without losing generality, we assume that the noise on the inliers is Gaussian on each component with a zero mean and a uniform standard deviation σ, and the outlier distribution is uniform $\frac{1}{a}$, where a is a constant. Let γ be the percentage of inliers that is not known in advance. Thus, the likelihood is a mixture model as

$$p(S|\theta)\prod_{i=1}^{N}\frac{\gamma}{2\pi\sigma^2}e^{-\frac{\|M(x_i,y_i)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a}, \quad (7)$$

where $\theta = \{M, \sigma^2, \gamma\}$ is a set of unknown parameters.

We must find a maximum likelihood estimation of θ, which is equivalent to seeking the minimal energy

$$E(\theta) = -\sum_{i=1}^{N}\ln\left(p(x_i, y_i|\theta)\right). \quad (8)$$

The well-known EM algorithm provides a natural framework for solving the problem of estimating the parameters of a mixture model.

The E-step estimates the responsibility that indicates the degree to which a sample belongs to the inlier set under the given geometric transform model *M*, whereas the M-step updates *M* based on the current estimate of the responsibility. When the negative log-likelihood function, i.e., Eq.(8), is considered, we associate sample *i* with a latent variable $\mathrm{z}_i\epsilon\{0,1\}$, where $\mathrm{z}_i = 1$ indicates a Gaussian distribution and $\mathrm{z}_i = 0$ denotes a uniform distribution. We follow standard notations and omit certain terms that are independent of θ. The complete-data log-likelihood is then given by

$$Q(\theta, \theta^{old}) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}P(\mathrm{z}_i = 1|x_i, y_i, \theta^{old})\,\|M(x_i, y_i)\|^2 - \ln\sigma^2\sum_{i=1}^{N}P(\mathrm{z}_i = 1|x_i, y_i, \theta^{old}) + \ln\gamma\sum_{i=1}^{N}P(\mathrm{z}_i = 1|x_i, y_i, \theta^{old}) + \ln(1-\gamma)\sum_{i=1}^{N}P(\mathrm{z}_i = 0|x_i, y_i, \theta^{old}), \quad (9)$$

which may be maximized by regarding z_i as the missing data from the mixture model in Eq.(7).

E-step: We denote $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_N)$, where the responsibility $p_i = P(\mathrm{z}_i = 1|x_i, y_i, \theta^{old})$ can be computed using the Bayes rule:

$$p_i = \frac{\gamma e^{-\frac{\|M(x_i,y_i)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|M(x_i,y_i)\|^2}{2\sigma^2}} + \frac{2\pi\sigma^2(1-\gamma)}{a}}. \quad (10)$$

The posterior probability p_i is a soft decision.

M-step: We determine the revised parameter estimate θ^{new} as follows: $\theta^{\mathrm{new}} = \arg\max_\theta Q(\theta, \theta^{old})$. When the derivative of $Q(\theta)$ is considered with respect to the variance σ^2 and the mixing coefficient γ, which are set to zero, then we obtain

$$\sigma^2 = \sum_{i=1}^{N}p_i\,\|M(x_i, y_i)\|^2 / \sum_{i=1}^{N}p_i, \quad (11)$$

$$\gamma = \frac{\sum_{i=1}^{N}p_i}{N}. \quad (12)$$

To complete the EM algorithm, the geometric transform model *M* should be estimated in the M-step, which is the key step in our method, and is discussed in the next section.

Once the EM algorithm converges, we obtain the geometric transform model *M*.

2.4 Image Blending

After the corresponding transformation of the spliced image, some areas will overlap between the images; that is, two images may have different pixel values at the same pixel position. Thus, an image blending process should be performed on the overlapping images.

Evidently, the pixel values in the overlapping area are determined by the pixel values of the left and right images, supposing that the proportion factor is $\boldsymbol{\alpha}$. On the basis of priori knowledge, the color near the left in the overlapping area should be closer to the color of the left image in this area, and same rule applies to the color near the right. This principle requires the proportion factor in different parts of the fusion area to have varying values, and that the values should be determined by position weight. Therefore, the pixel values in the overlapping area is expressed by Eq.(13):

$$\boldsymbol{I_{blend}} = \frac{\boldsymbol{d_1}}{\boldsymbol{d_1}+\boldsymbol{d_2}}\boldsymbol{I_{left}} + \frac{\boldsymbol{d_2}}{\boldsymbol{d_1}+\boldsymbol{d_2}}\boldsymbol{I_{right}}, \quad (\mathbf{13})$$

where $\boldsymbol{d_1}$ is the distance of the pixel in the overlapping area to the left border, and $\boldsymbol{d_2}$ is the distance of the pixel in the overlapping area to the right border.

Our algorithm can be summarized as shown in Fig. 1.

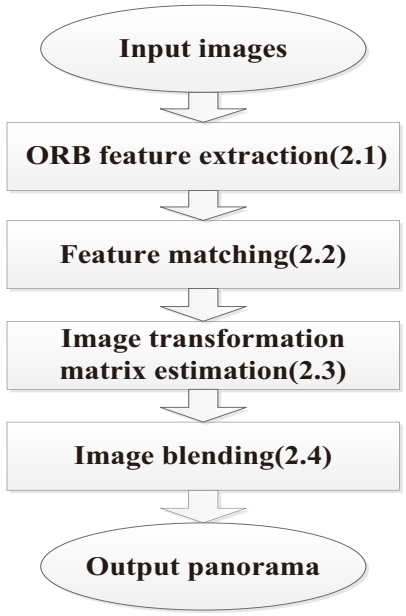


Fig. 1: Block diagram of the proposed panoramic image generation based on ORB features and the robust transformation estimation mosaic model

3 Estimation of the Transformation Matrix of UAV images

In accordance with a completely negative log-likelihood function, i.e., Eq. (9), the geometric transform model *M* is estimated by minimizing a weighted empirical error function as follows:

$$\boldsymbol{Q}(\mathbf{M}) = \sum_{i=1}^{N} \boldsymbol{p_i} \, \|\boldsymbol{M}(\boldsymbol{x_i}, \boldsymbol{y_i})\|^2. \qquad (14)$$

Unlike in other panoramic images, the ground is assumed to be generally planar when we handle UAV images because the flight altitude of UAVs is significantly higher than that of the ground. Moreover, proportions and viewing angles vary minimally, thereby making the UAV image geometric transform model considerably simpler than the traditional image mosaic transform model. All the correspondences can be related by an affine **A**: y=**A**x, which is a 3×3 matrix, and the last line is [0 0 1]. However, the geometric transform model *M* exhibits the form $\boldsymbol{M}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}$. Then, the weighted empirical error function, Eq.(14), is

$$\boldsymbol{Q}(\mathbf{A}) = \sum_{i=1}^{N} \boldsymbol{p_i} \, \|\boldsymbol{y_i} - \boldsymbol{A}\boldsymbol{x_i}\|^2. \qquad (15)$$

The three-element vectors $\boldsymbol{y_i}$ and $\boldsymbol{Ax_i}$ may differ in magnitude by a nonzero scale factor due to the affine vectors. Thus, we express the equation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ in terms of the vector cross product and then obtain **B***a* = 0, where *a* is a six-element vector that comprises the first two lines of **A** in a row-major order, and **B** uses the form:

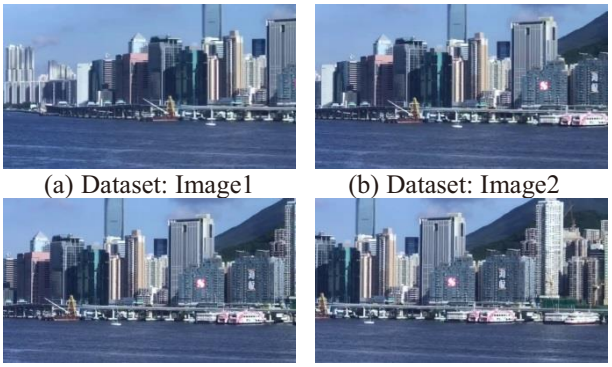
$$\mathbf{B} = \begin{bmatrix} x_1^x y_1^y & x_1^y y_1^y & y_1^y & -x_1^x y_1^x & -x_1^y y_1^x & -y_1^x \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_N^x y_N^y & x_N^y y_N^y & y_N^y & -x_N^x y_N^x & -x_N^y y_N^x & -y_N^x \end{bmatrix}. \qquad (16)$$

To estimate *a* in the M-step, the weighted error function, Eq. (15), uses the form: $\boldsymbol{Q}(\mathbf{A}(\mathbf{1}{:}\mathbf{2},:)) = \left\|\mathbf{P}^{\mathbf{1/2}}\boldsymbol{B}\boldsymbol{a}\right\|$. If $\mathbf{P}^{\mathbf{1/2}}\boldsymbol{B} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{Y}^{\boldsymbol{T}}$ with diagonal $\boldsymbol{D}$ and with positive diagonal entries arranged in descending order down the diagonal, then $\mathbf{A}(\mathbf{1}{:}\mathbf{2},:)$ is ranked 2nd. To enforce this constraint, we replace the calculated $\mathbf{A}(\mathbf{1}{:}\mathbf{2},:)$ in each iteration by its closest singular matrix under a Frobenius norm. In particular, if $\mathbf{A} = \boldsymbol{X}\boldsymbol{D}\boldsymbol{Y}^{\boldsymbol{T}}$ with $\boldsymbol{D} = \mathbf{diag}(\mathbf{d_1}, \mathbf{d_2}, \mathbf{d_3})$ arranged in descending order down the diagonal, then we replace **A** by

$$\widehat{\mathbf{A}} = \boldsymbol{X}\begin{bmatrix} \mathbf{d_1} & & \\ & \mathbf{d_2} & \\ & & \mathbf{0} \end{bmatrix}\boldsymbol{Y}^{\boldsymbol{T}}. \qquad (17)$$

4 Experimental Result

To test our method, we perform experiments on real images. The experiments are conducted on a laptop with a 2.0 GHz CPU, 4 GB RAM, and Windows 7 as the operating system. We implement a complete image stitching system in Microsoft Visual Studio 2013 and OpenCV version 2.4.13 library.

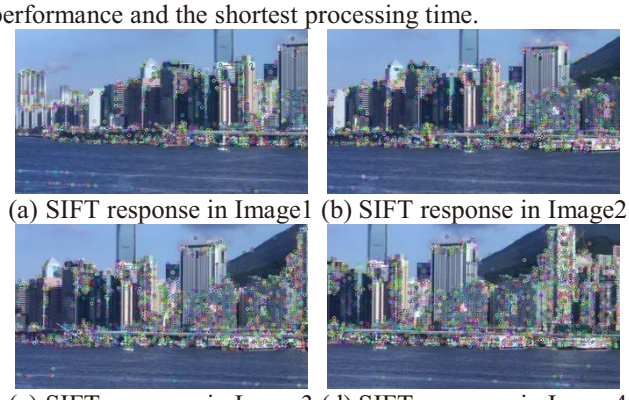


(a) Dataset: Image1 (b) Dataset: Image2

(b) Dataset: Image3 (d) Dataset: Image4

Fig. 2: Group of input images

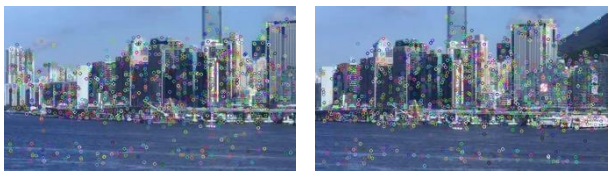
First, we evaluate SIFT, SURF, and ORB to demonstrate the superiority of ORB, which achieves the highest performance and the shortest processing time.



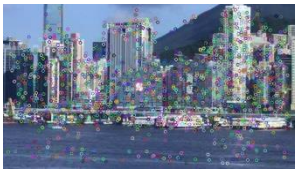
(a) SIFT response in Image1 (b) SIFT response in Image2

(c) SIFT response in Image3 (d) SIFT response in Image4

Fig. 3: SIFT response in the input images



(a) SURF response in Image1 (b) SURF response in Image2

(c) SURF response in Image3 (d) SURF response in Image4

Fig. 4: SURF response in the input images


(a) ORB response in Image1 (b) ORB response in Image2


(c) ORB response in Image3 (d) ORB response in Image4

Fig. 5: ORB response in the input images

Table 1 presents the number of detected features and the time required by the different detectors for the input images.

Table 1: Performance Analysis of the Feature-Extracting Algorithms for the Input Images

Feature-Extracting Algorithm	Test Image	Detected Features	Time (s)
SIFT	Image1	1402	2.78
	Image2	1910	3.31
	Image3	2466	3.61
	Image4	2814	3.76
SURF	Image1	1434	1.05
	Image2	1535	0.92
	Image3	1766	1.13
	Image4	1968	1.07
ORB	Image1	1925	0.10
	Image2	1934	0.10
	Image3	1918	0.11
	Image4	1929	0.12

The results of this step indicate that SIFT detects the highest number of feature points, but the requires the longest processing time among the three feature detectors. However, as shown in Figs. 3 to 5, ORB exhibits a performance similar to that of SIFT but consumes the least computation time.

After extracting the features from all the images, the next step is to match them with one another and discard incorrect points.

Fig. 6: Traditional matches between Image1 and Image2

Fig. 7: Matches applied using the proposed method between Image1 and Image2

Fig. 8: Traditional matches between Image3 and Image4

Fig. 9: Matches applied using the proposed method between Image3 and Image4

Figs. 6 to 9 show the comparison between the results of the traditional feature-matching method and the proposed method. In terms of speed and accuracy, the traditional method costs 1.45 s and 1.47 s, respectively, whereas the proposed method costs 1.53 s and 1.52 s, respectively. However, accuracy has improved from 91.04% and 85.36% to 98.64% and 98.40%, respectively.

Then, we compute the transformation matrix and stitch all the images. The result is presented in Fig. 10. Splicing seams appear as shown in the red box. Then, we apply an image blending algorithm and obtain the result shown in Fig. 11.

Fig. 10: Images stitching without blending

Fig. 11: Final panorama

The time required to complete all the steps between the two images is 2 s. However, the entire process generally takes 10 s or longer.

5 Conclusion

In this study, we proposed an automatic UAV image mosaic system that involves four major steps: ORB feature extraction, feature matching based on the LPM algorithm, transformation estimation under a Bayesian framework, and image blending via a gradual fading strategy. The experiments show that the proposed method can significantly improve efficiency while ensuring accuracy compared with state-of-the-art methods. Furthermore, it can accomplish quick and seamless splicing of UAV images.

References

[1] E. Adel, M. Elmogy, and H. Elbakry, Image stitching based on feature extraction techniques: a survey, *International Journal of Computer Applications*, 6 (99): 1-8, 2014.

[2] R. Szeliski, and S.B. Kang, Direct methods for visual scene reconstruction, in *Proceedings of IEEE workshop on Representation of Visual Scenes,* 1995: 26-33.

[3] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, Robust Feature Matching for Remote Sensing Image Registration via Locally Linear Transforming, *IEEE Transactions on Geoscience and Remote Sensing*, 53(12): 6469-6481, 2015.

[4] J. Ma, J. Zhao, J. Tian, A. Yuille, and Z. Tu, Robust Point Matching via Vector Field Consensus, *IEEE Transactions on Image Processing*, 23(4): 1706-1721, 2014.

[5] D. Vaghela, and K. Naina, A Review of Image Mosaicing Techniques, *Eprint Arxiv,* 2 (3), 2014.

[6] C. Harris, A combined corner and edge detector, in *Proceedings of the Alvey Vision Conference*, 1988: 147-151.

[7] G.L. David, Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision,* 60 (2): 91-110, 2004.

[8] H. Bay, T. Tuytelaars, and L. Van Gool, SURF: speeded up robust features, in *Proceedings of European conference on computer vision*, 2006: 404-417.

[9] E. Mair, G.D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, Adaptive and generic corner detection based on the accelerated segment test, in *Proceedings of European conference on Computer vision*, 2010: 183--196.

[10] E. Rublee, V. Rabaud, K. Konolige, and G.R. Bradski, ORB: An efficient alternative to SIFT or SURF, in *Proceedings of 2011 International Conference on Computer Vision*, 2011: 2564-2571.

[11] E. Bostanci, N. Kanwal, B. Bostanci, and M.S. Guzel, A Fuzzy Brute Force Matching Method for Binary Image Features. *arXiv preprint arXiv:1704.06018,* 2017

[12] J.S. Beis, and D.G. Lowe, Shape indexing using approximate nearest-neighbour search in high-dimensional spaces, in *Proceedings of Computer Vision and Pattern Recognition*, 1997: 1000—1006

[13] M.A. Fischler, and R.C. Bolles, Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Commun. ACM,* 24 (6): 381-395,1981.

[14] P.J. Rousseeuw, and A.M. Leroy, Robust regression and outlier detection, John wiley & sons, New York, 2005.

[15] P. Huber, Robust Statistical Procedures, in *Society for Industrial and Applied Mathematics*, Siam, 1996.

[16] K.S.V. Prathap, S. Jilani, and P.R. Reddy, A critical review on Image Mosaicing, in *Proceedings of International Conference on Computer Communication and Informatics*, 2016: 1-8.

[17] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, BRIEF: Binary Robust Independent Elementary Features, *in Proceedings of European conference on Computer vision*, 2010: 778-792.

[18] J. Ma, J. Zhao, H. Guo, J. Jiang, H. Zhou, and Y. Gao, Locality Preserving Matching, in 26th International Joint Conference on Artificial Intelligence, 2017: 4492-4498.

[19] J. Chen, J. Ma, C. Yang, and J. Tian, Mismatch removal via coherent spatial relations. Journal of Electronic Imaging, 23(4): 043012, 2014.

[20] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. Yuille, Robust Estimation of Nonrigid Transformation for Point Set Registration, in Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2013: 2147-2154.

[21] J. Ma, J. Zhao, J. Jiang, and H. Zhou, Non-Rigid Point Set Registration with Robust Transformation Estimation under Manifold Regularization, in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI), 2017: 4218-4224.