

# Robust Alignment for UAV Images Based on Adaptive Adjustment

Menghan Xia, Jian Yao<sup>†</sup>, Renping Xie, Xiaohu Lu, Li Li

*School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, China*

<sup>†</sup>E-Mail: [jian.yao@whu.edu.cn](mailto:jian.yao@whu.edu.cn) Web: <http://cvrs.whu.edu.cn/>

**Abstract**—In this paper, we investigate thoroughly the problem of aligning sequential images taken from low-altitude unmanned aerial vehicles (UAV). It is difficult because the photographic scene is no longer an approximate plane from the UAVs at low flight altitude. On this occasion, we assume the ground regions of the scene share the dominant plane, so sequential images are aligned based on the registration of the ground as a landmark, which precludes the obstruct of other objects of various altitudes. Specially, considering the occasional accumulating error and deformation degree, an adaptive bundle adjustment based on the updated reference image is creatively designed to guarantee the quality of incremental alignment in an efficient way. Finally, the global refinement on transformation models is performed under the constraint of no perspective distortion engendered, to improve the aligning accuracy further. Experimental results on several challenging datasets illustrate the superiority of our approach.

## 1. Introduction

A huge market is currently emerging from the wide applications and services offered by low-cost and low-flying UAVs. Equipped with imaging sensors, UAVs can easily provide aerial images taken from the target scene. Because of the limited field of view within a single UAV image, it's necessary to stitch sequential images together to create a wide-range mosaicked image with overall situational awareness. In the photogrammetry field, aerial images mosaicking usually requires extra information, such as the camera calibration parameters and poses or a reference map, to achieve accurate mosaicked results [1], [2]. However, the pose estimation inevitably introduces the additional computational cost, which is unworthy for just compositing a mosaicked image. Since its objective is not for geometrical measurement, many image mosaicking methods, requiring no extra pose parameters, have been proposed to obtain visually satisfactory mosaics in an efficient way.

As a strict aligning model, the homography is often used to describe the relationship between two images taken from a 3D scene plane [3], [4] or two images captured from the same camera center in different view directions [5]. Because of the limitation of motionless position, images of a fixed-camera-center viewpoint are mainly used to make a ground panorama with the omniscient view [6], [7]. On the contrary, mosaicking images of a 3D scene plane permits the camera moving freely for mosaicking, which is popular in robotic mapping and remote sensing applications. For aerial image mosaicking, feature-based approaches are usually applied to

recover the homographic model between images with assuming the ground as an approximate plane [8]–[10]. To improve the aligning precision, the bundle adjustment approach [11], [12] is often used as a typical global optimization method. To provide a good initial solution for global optimization, Xing et al. [13] proposed to first apply the Extended Kalman Filter onto the local area, and then refine all the parameters globally. However, because of the pseudo-planarity of the ground scene, mosaicking images from wide-range regions by some alignment methods based on minimizing registration error would result in a significant accumulation of perspective distortions [14]. To overcome this problem, the geo-referenced satellite images of the whole region can be used as a reference image [15], [16]. Another approach to avoid accumulated errors is to use extra sensors such as GPS and IMU to directly obtain camera poses for mosaicking [2]. In practice, the qualified satellite images or pose parameters are not always available. Therefore, optimizing aligning models for resisting perspective distortions have been studied recently [14], [17], [18]. Caballero et al. [17] proposed to use the hierarchical models according to the quality of data, where the model with less degree of freedom is for data with bigger parallax. The essence of this method is to make a trade-off between aligning precision and resisting perspective distortion. To allow the transition between aligning models, Xia et al. [14] proposed to initially align images by the robust affine model, followed by the model refinement under the anti-perspective constraint. By this way, both the global consistency and the aligning accuracy are able to be achieved at an optimal balance. However, the reference image fixed as the first image limited its application to large-scale scene, which presents more severe challenge in solving accumulation error.

In this paper, we concentrate on the problem of aligning low-flight UAV images, which can't be treated as an approximate plane any longer because the various altitudes of objects can't be ignored compared to the flight altitude. To this problem, we propose to make image alignment based on the registration of the ground as the landmark, which is under the basic premise that most ground regions share a approximately common plane. As to the occasional accumulation error and deformation degree, an adaptive bundle adjustment based on the updated reference image is specially designed to improve the robustness and precision of the incremental alignment, which conducts once a distortion is detected. At last, the global refinement of the model under the anti-perspective constraints will be performed to achieve a more accurate alignment. As for the misalignment of the objects above the ground, any seamline detection algorithm can be applied to remove these residual parallaxes visually, which is beyond the

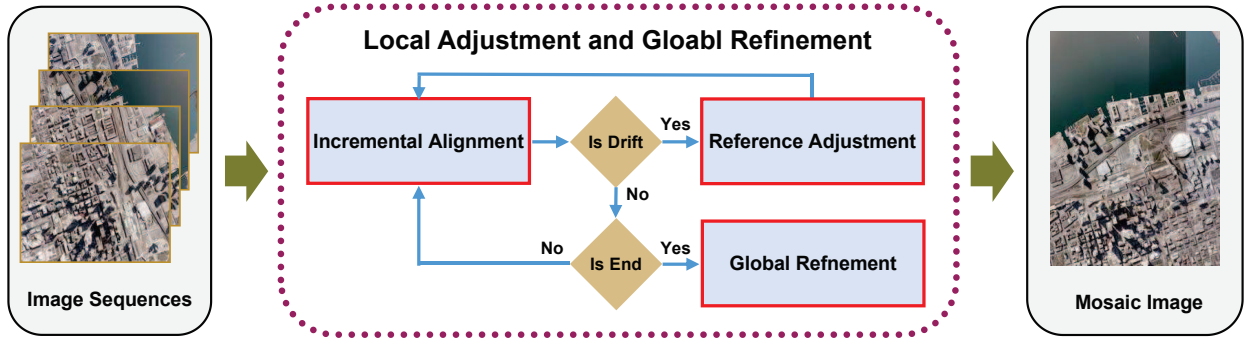


Figure 1. The flowchart of our approach for robust alignment of UAV images (Drift : existing accumulation error or perspective distortion).

research range of this paper. Our approach was tested through several groups of experiments on challenging datasets, which sufficiently illustrate its superiority.

## 2. Our Approach

As the registration landmark, the ground regions are used to make a global basis alignment, during which the bundle adjustment based on the updated reference image is performed adaptively once the accumulation error or perspective distortion is detected in real time. When all images are aligned, a global refinement under anti-perspective constraint is followed to achieve the optimal balance between global consistency and local alignment accuracy. The flowchart is depicted in Figure 1.

### 2.1. Dominant-Plane Based Alignment

To use the ground regions as the landmarks, we only utilize the matched features of the ground regions for alignment. In fact, it is not difficult to achieve this target only if the ground is the dominant plane of the scene. Fortunately, this requirement usually can be satisfied in most scenes (except for mountain areas), such as urban areas where the ground regions including parks, roads, squares form the dominant plane, even though many buildings exists.

Given a pair of images, the extracted SURF features [19] can be matched based on the similarity of descriptor vectors and the epipolar constraint, which distribute on various objects in the scene. Approximately, these matched features can be grouped into different planes, such as the roofs or facades of the buildings, the ground and so on. The dominant plane is defined as the plane with the most matched features, which would be presented by using the Random Sample Consensus (RANSAC) algorithm. The estimated homography by RANSAC can be used to remove the matched features not on the dominant plane. As shown in Figure 2, almost all the maintained features are located at the ground regions, and the features on the hill or buildings are efficiently removed. This is reasonable, on the one hand, there usually exists large differences of visual angles for those features on the tall objects between images, which increases difficulties in matching, on the other hand, the area of the planar ground is larger than any other existing planar objects, although maybe smaller than the total area of other planar regions. In fact, apart

from extracting the dominant plane, the coplanar constraint for feature matching is also an effective geometric constraint to remove outliers in the mosaicking for other planar scenes.

With the matched feature pairs consistent to the dominant plane, we make alignment of images incrementally. Although most of the tall objects above the ground are removed in registration period, the reserved ground is still a pseudo-plane which consists of different low-altitude objects. To efficiently deal with this problem, we employ the robust affine transformation model of 6 DoF as the initial aligning model, as its excellent property mentioned in [14]. To support real-time mosaicking, our approach temporarily sets the first image as the reference frame. While aligning a new image to the reference frame, all the previously aligned images having an overlap with the new image will be used as references. Let  $\mathcal{G} = \{\mathbf{I}_i\}_{i=1}^{m-1}$  be a set of previously aligned images and  $\mathbf{A}_i$  be the  $2 \times 3$  affine transformation matrix of the image  $\mathbf{I}_i$  with respective to the reference frame. The affine transformation  $\mathbf{A}_m$  of the newly introduced image  $\mathbf{I}_m$  for alignment will be optimized by minimizing the following cost function:

$$E(\mathbf{A}_m) = \sum_{i=1}^{m-1} \sum_{k=1}^{M_{i,m}} |\mathbf{A}_i \mathbf{x}_{i,m}^k - \mathbf{A}_m \mathbf{x}_{m,i}^k|^2, \quad (1)$$

where  $M_{i,m}$  denotes the total number of matches  $\{(\mathbf{x}_{i,m}^k, \mathbf{x}_{m,i}^k)\}_{k=1}^{M_{i,m}}$  between  $\mathbf{I}_i$  and  $\mathbf{I}_m$  ( $M_{i,m} = 0$  means no overlap existing between this image pair), and  $(\mathbf{x}_{i,m}^k, \mathbf{x}_{m,i}^k)$  denotes the homogeneous coordinates of two corresponding feature points extracted from  $\mathbf{I}_i$  and  $\mathbf{I}_m$ , respectively.

### 2.2. Adaptive Bundle Adjustment

The estimation process in Section 2.1 just recovers the locally optimal affine model of each image. Besides, the

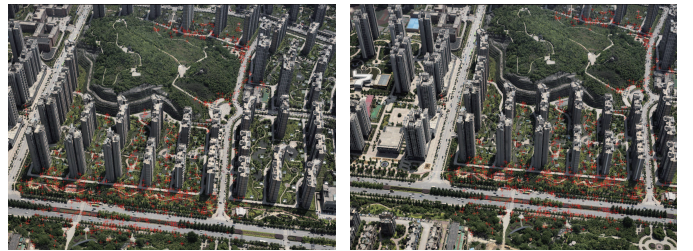


Figure 2. The distribution of matched features on the dominant plane.

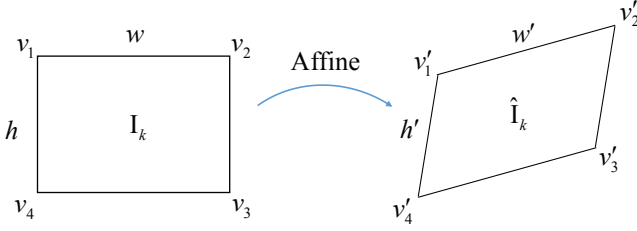


Figure 3. Configuration for computing the deformation  $D_e$  when projecting image  $\mathbf{I}_k$  to  $\hat{\mathbf{I}}_k$  through the affine transformation. Corresponding vertex pairs are  $(v_i, v'_i)$  for  $1 \leq i \leq 4$ .

reference image is the naively-selected first image. Thus, the accumulation error will appear inevitably. To solve this problem, the aligning precision of the newly added image is assessed in real time, which determines whether the bundle adjustment is needed for currently aligned images. Here, the aligning precision is evaluated in two aspects: the registration error and the geometric deformation. Specially, the latter describes the deformation degree from the original one to the projected one as in Fig. 3. Any projecting with  $|\alpha - 90^\circ| \geq C_1$  or  $\max(hw, h'w' \sin \alpha) / \min(hw, h'w' \sin \alpha) \geq C_2$  ( $C_1 = 5^\circ$ ,  $C_2 = 1.25$  in our experiments) is regarded as a deformation. Conclusively, when the projecting deformation exists or the registration error exceeds some predefined thresholds, it indicates that the accumulation error probably exists, and all the transformation models of currently aligned images will be adjusted jointly as described below.

Firstly, a more reasonable reference image will be selected by applying a graphic algorithm on the fast built graph. Based on the matching results obtained currently, we construct an undirected and weighted graph, where images are nodes, the overlap between an image pair is denoted by an edge or a link, and the weight of edge is defined as:

$$w_{ij} = \begin{cases} \inf, & \text{if } M_{i,j} = 0, \\ \frac{1}{\log(M_{i,j} + \varepsilon)}, & \text{if } M_{i,j} > 0, \end{cases} \quad (2)$$

where  $M_{i,j}$  denotes the total number of matches between  $\mathbf{I}_i$  and  $\mathbf{I}_j$  consistent to the dominant ground plane, and  $\varepsilon$  is a constant for regularization, which is set as 50 in our experiments. This weight setting formula is based on the idea that an image pair with more matched features usually provides a more accurate and reliable alignment, which is obvious as a common sense.

Based on the weighted graph, the optimal reference image selection problem is formulated as finding a node with the least sum weight of the shortest paths to all the other nodes, which can be solved by the Floyd Warshalls algorithm [20], [21]. With this algorithm, all-pairs shortest paths from a node to other nodes can be obtained. When there are  $m$  images aligned currently, we build a symmetric cost matrix  $\mathbf{W}$  of a size  $m \times m$  where each element represents the cost of the shortest path between two images, i.e.,  $\mathbf{W}(i, j)$  is the cost of the shortest path from  $\mathbf{I}_i$  to  $\mathbf{I}_j$ . Therefore, the  $i$ -th row or column of  $\mathbf{W}$  indicates the cost of every shortest path from other images to  $\mathbf{I}_i$ . Thus, the accumulated cost of each row in  $\mathbf{W}$  is calculated and the row with the minimum accumulated cost is selected as the reference image.

Secondly, with the updated reference image, the aligning models of currently aligned images are recalculated jointly in the optimization framework. Let  $\mathcal{G} = \{\mathbf{I}_i\}_{i=1}^m$  be the set of currently aligned images and  $\mathbf{I}_r$  be the newly selected reference image. Obviously, the affine transformation of  $\mathbf{I}_r$  is an identity matrix which is fixed during the optimization. The unknown affine transformation set  $\mathcal{A}$  of images in  $\mathcal{G}$  with respective to  $\mathbf{I}_r$  will be solved as follows:

$$E(\mathcal{A}) = E_1(\mathcal{A}) + E_2(\mathcal{A}), \quad (3)$$

where the first term stands for the energy cost between  $\mathcal{G}$  and the reference image as:

$$E_1(\mathcal{A}) = \sum_{\mathbf{I}_i \in \mathcal{G}} \sum_{k=1}^{M_{i,r}} \|\mathbf{A}_i \mathbf{x}_{i,r}^k - \mathbf{A}_r \mathbf{x}_{r,i}^k\|^2, \quad (4)$$

and the second term denotes the alignment error cost between images in  $\mathcal{G}$  as:

$$E_2(\mathcal{A}) = \sum_{i=1}^{m-1} \sum_{j=i+1}^m \sum_{k=1}^{M_{i,j}} \|\mathbf{A}_i \mathbf{x}_{i,j}^k - \mathbf{A}_j \mathbf{x}_{j,i}^k\|^2. \quad (5)$$

As a group of linear equations, the optimal solution of  $\mathcal{A}$  in Eq. (3) can be obtained by the Singular Value Decomposition (SVD) method. After this adjustment, the global consistency and aligning precision are bound to be improved, since the selected reference image makes effects in decreasing error accumulation and the bundle adjustment contributes on accurate alignment. Then, the next coming image is to be added as the procedure like before.

Note that the reference image adjustment is performed adaptively to the aligning accuracy, which won't be performed frequently unless the quality of the images are extremely inferior. However, such adjustment will be conducted compulsively when all the images are aligned, even if the registration precision is pretty high during the whole procedure.

### 2.3. Global Model Refinement

When all the images have been gradually aligned, the global refinement, allowing the affine model to transit adaptively to the homographic model with more freedom of degrees (DoF) under reasonable constraints, is conducted to achieve a more accurate alignment. The energy function is composed of two terms: the one minimizes the sum of the squared distances between matched feature pairs, and another one constrains the transformation parameters not to induce the strong perspective distortion. Let  $\mathcal{I} = \{\mathbf{I}_i\}_{i=1}^n$  be the image set composed of all the images, and the first energy term is defined as:

$$E_r(\mathcal{H}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=1}^{M_{i,j}} \|\varpi(\mathbf{H}_j \mathbf{x}_{j,i}^k) - \varpi(\mathbf{H}_i \mathbf{x}_{i,j}^k)\|^2, \quad (6)$$

where  $\mathcal{H} = \{\mathbf{H}_i\}_{i=1}^n$  represents the unknown homographic models with respective to the reference frame. Another optimization objective aims at keeping the global consistency by suppressing severe perspective distortions which may happen during the model refinement. In other words, the optimal homographic transformation should be close to the initially estimated affine transformation, which can be expressed as



the displacements of the warped features from their initial positions in the following cost function:

$$E_d(\mathcal{H}) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{k=1}^{M_{i,j}} (\|\varpi(\mathbf{H}_i \mathbf{x}_{i,j}^k) - \mathbf{A}_i \mathbf{x}_{i,j}^k\|^2 + \|\varpi(\mathbf{H}_j \mathbf{x}_{j,i}^k) - \mathbf{A}_j \mathbf{x}_{j,i}^k\|^2), \quad (7)$$

where  $\mathbf{A}_i$  and  $\mathbf{A}_j$  denote the previously estimated affine models for  $\mathbf{I}_i$  and  $\mathbf{I}_j$  with respect to the reference frame, respectively. So far, the cost functions defined in Eq. (6) and Eq. (7) can be linearly combined to define the final cost function as:

$$E(\mathcal{H}) = E_r(\mathcal{H}) + \lambda E_d(\mathcal{H}), \quad (8)$$

where  $\lambda$  is the balancing coefficient for the term  $E_d$ , which should be set to an appropriate small value since the constraint is not strict enough. Theoretically, a bigger value of  $\lambda$  strengthens the global consistency while decreases the accuracy of the local alignment. We set its value from 0.01 to 0.05 in all our experiments. As a typical non-linear least squares problem, the optimal solution of Eq. (7) can be solved by the LevenbergMarquardt (LM) algorithm. Specially, to save the memory and to speed up the computation, we would rather use the sparse LM algorithm [22].

### 3. Experimental Results

In this section, two groups of representative data sets acquired by UAVs were used to test the proposed approach. The first dataset, composed of 143 images belonging to 11 strips, was acquired by the professional UAV with aerial cameras. The average flying height is about 770 meters, and the forwarding overlapping rate is about 60% observing the urban district of Anyang City, Henan Province, China. The second dataset, composed of 62 images, was acquired in Yongzhou City, Hunan Province, China, of which the DOM at the precision of 0.2 meters is available. For the test efficiency, all the original images were down-sampled to the size with a width of 1500 in our experiments.

#### 3.1. Qualitative Evaluation

The experiment on the first dataset is designed to evaluate the visual effect of the proposed approach, where our mosaicking result is compared with that of a professional commercial software named PTGui<sup>1</sup>. Since aiming at comparing the alignment results only, the following seamline detection and tonal correction were skipped in PTGui and our image stacking order was made consistent with that of PTGui.

From the mosaics shown in Figure 5, the two mosaics have similar visual effects as a whole, both of which take on a good global consistency. However, when comparing carefully, the right bottom region on the mosaicking result of PTGui shows a slight tendency of down-scaling. On the contrary, the mosaic via our approach doesn't suffer from such a problem, thanks to the technology of topology analysis based reference image selection. To evaluate the local aligning precision, some typical regions marked with red/green boxes on the mosaic images, are enlarged for a detail observing, as listed at the

bottom of Figure 5. As a linear object in remote sensing images, the road is a kind of good symbol for distinguishing misalignment, which will be more noticeable when there is a parallax between them. Judged by the registration deviation on the roads, it's easy to find that our approach has an obvious superiority over PTGui on aligning accuracy, no matter in the central parts or the marginal parts. Besides, as the red curves shown in Figure 4, we can observe that the aligning precision increases a lot with the employment of the homography refinement, while the global consistency is not affected during the transition between models.

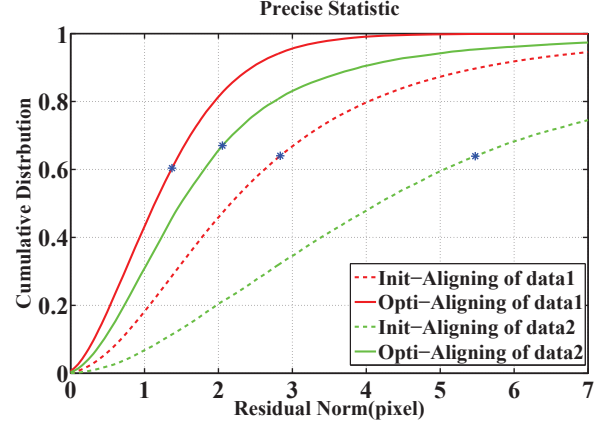


Figure 4. Cumulative probability distributions of the residual error norms with and without the global refinement performed in our approach, respectively. The red curves and green curves correspond to the first dataset and the second dataset respectively. Specially, the blue marks on curves indicate the RMS errors.

#### 3.2. Quantitative Evaluation

The second dataset is used to evaluate the mosaicking quality of our approach quantitatively, since the DOM of the corresponding area is available. The DOM is a kind of orthophoto map, which is produced by a series of strict procedures in photogrammetry. Here, the DOM is used as the ground-truth to assess the global consistency of the mosaic image, namely to analyze the magnitude of accumulation error. The mosaicking result and DOM of the corresponding area are depicted in Figure 6. For quantitative evaluation, 21 corresponding well-distributed check points were manually selected in both the DOM and the mosaic image. A four degrees-of-freedom 2D similarity transformation model was used to align the two sets of check points. The distribution of these check points after aligned into a common coordinate system is displayed in Figure 7. Obviously, the two groups of centroids have a similar distribution form although there are also some displacements between corresponding centroids. The mean aligning deviation is 2.508m when converted to metric units using the ground resolutions of the ground-truth images. In fact, as an image mosaicking approach based on the 2D feature registration, the recovered geometric positions are accurate enough to keep the global consistency of a mosaic, which emphasizes more on the visual effects than the geometric measurements.

As for the registration accuracy, the numerical analysis of the initial alignment and the final alignment with global

1. <http://www.ptgui.com/>



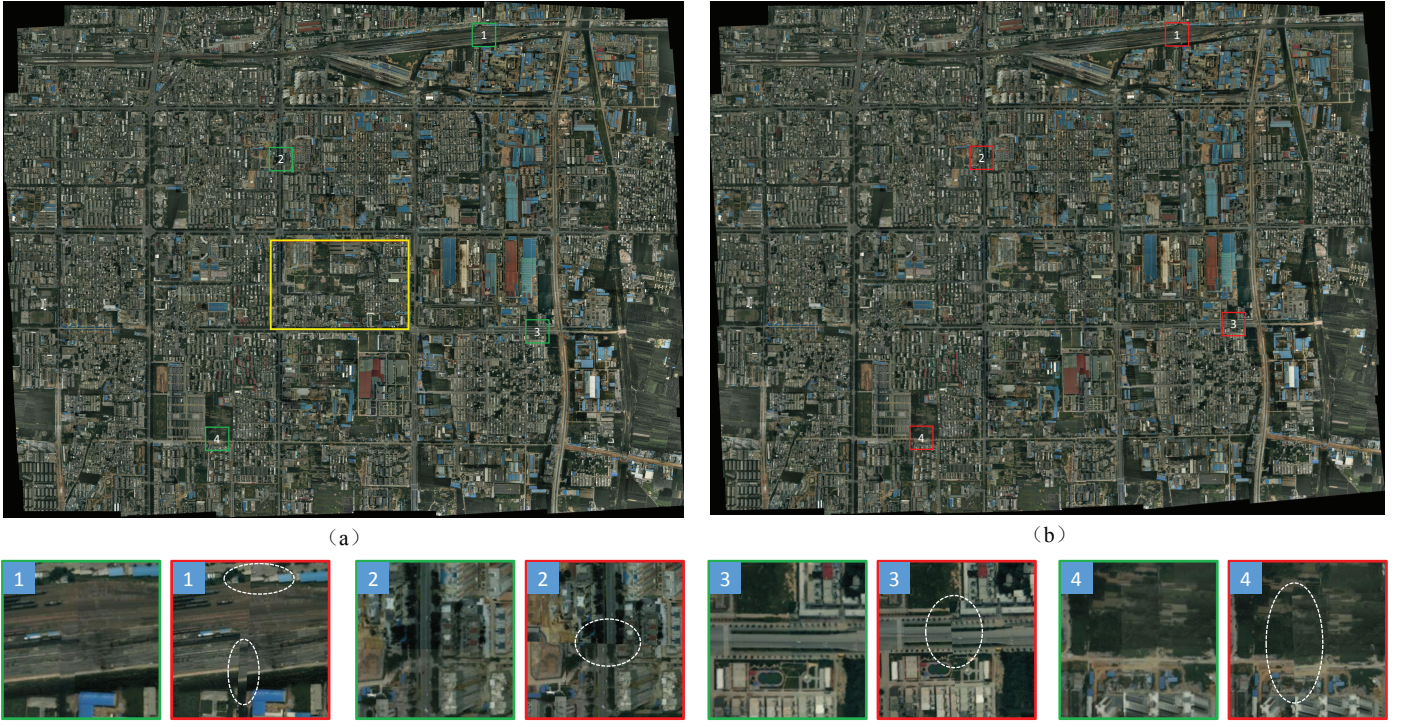


Figure 5. The mosaics composited from the first dataset (143 images) by: (a) our approach and (b) PTGui, respectively. Several typical regions grabbed from the mosaics are enlarged in pairs in the bottom row. The yellow box on the left mosaic image indicates the final reference image.

model refinement applied are demonstrated in Figure 4, corresponding to the green curves.

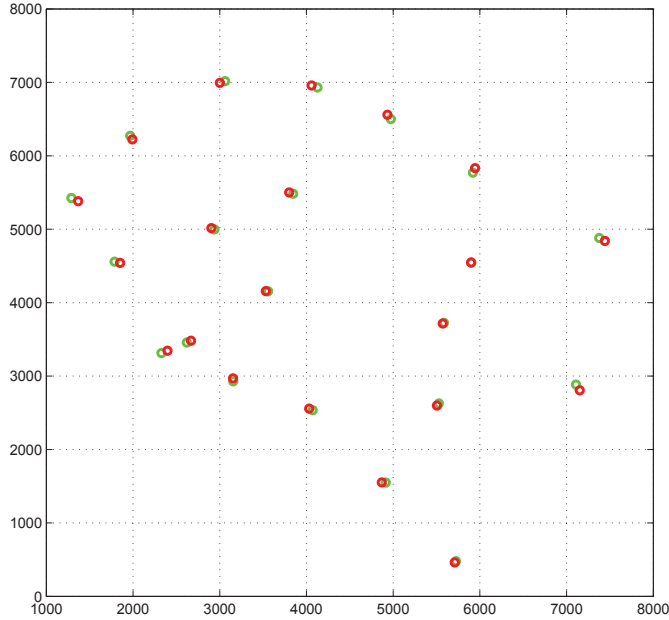


Figure 7. Distribution of the check point pairs after transformed to the coordinate system of the DOM (unit : pixel), where the red ones are from the DOM and the green ones are from the mosaic image.

## 4. Conclusion

This paper demonstrates an robust and effective approach for mosaicking UAV images taken at low-flight altitude. With the ground regions as the landmarks for geometrical alignment, all the images can be aligned under an uniform basis. The specially designed self-adaptive bundle adjustment based on the updated reference image makes effects in solving the accumulation error and perspective distortion online, and the final global model refinement manages to improve the alignment accuracy meanwhile preserves the global consistency. As the experimental results show, the proposed approach can generate quality-satisfied mosaics from sequential UAV images, which has been evaluated both qualitatively and quantitatively. In the future, considering the unstable flight posture of the micro UAV, the orientation of images with respect to the ground will be taken into account for selecting the optimal reference image, which determines the perspective effect of the final mosaicked image.

## Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Project No. 41571436), the Hubei Province Science and Technology Support Program, China (Project No. 2015BAA027), and the National Natural Science Foundation of China under Grant 91438203.

## References

- [1] M. Bryson, A. Reid, F. Ramos, and S. Sukkarieh, "Airborne vision-based mapping and classification of large farmland environments," *Journal of Field Robotics*, vol. 27, no. 5, pp. 632–655, 2010.

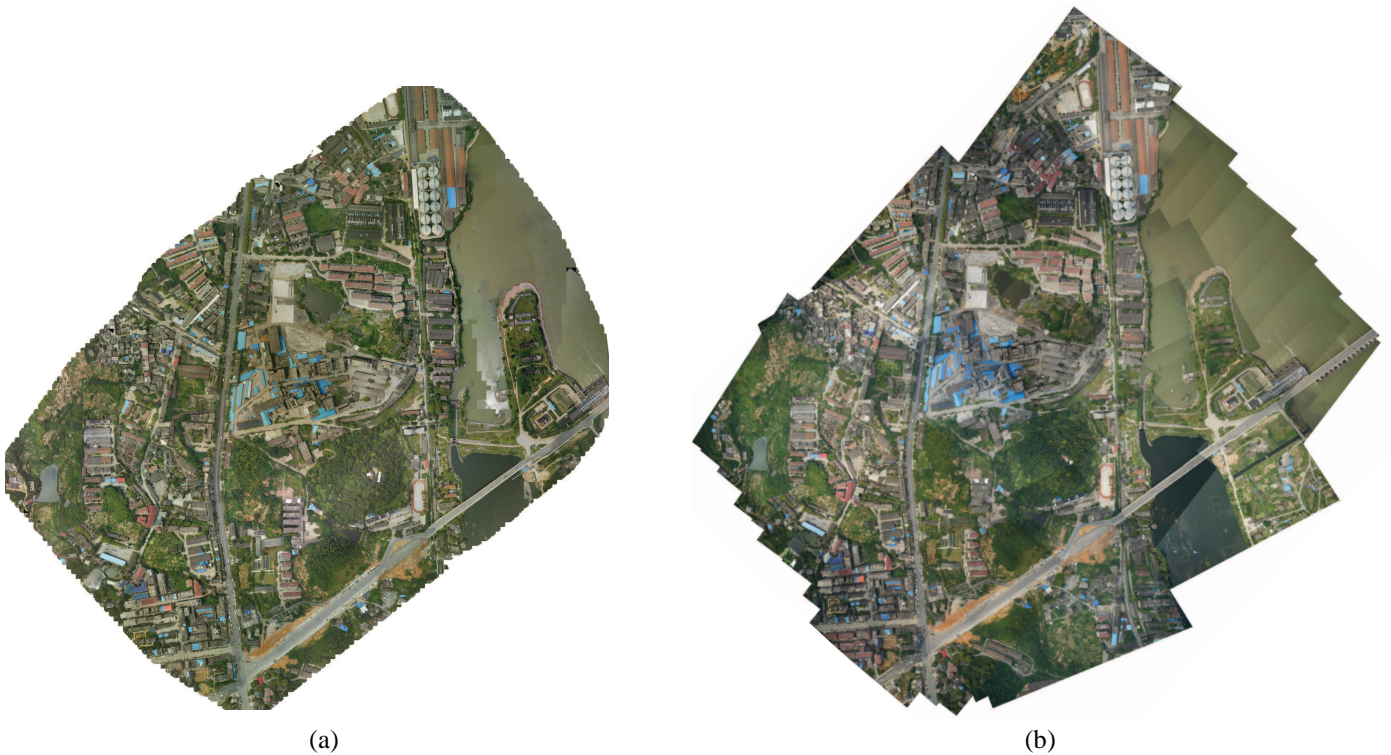


Figure 6. The DOM of the observing area (a) and the mosaicking result via the proposed approach (b). For the convenience of comparison, the orientation of the mosaic image is adjusted manually to be consistent with the DOM.

- [2] S. Yahyanejad, D. Wischounig-Strucl, M. Quaritsch, and B. Rin, "Incremental mosaicking of images from autonomous, small-scale UAVs," in *IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 329–336.
- [3] A. Elibol, R. Garcia, O. Delaunoy, and N. Gracias, *Efficient Topology Estimation for Large Scale Optical Mapping*. Springer, 2013, ch. A New Global Alignment Method for Feature Based Image Mosaicking, pp. 25–39.
- [4] F. Bellavia and C. Colombo, "Estimating the best reference homography for planar mosaics from videos," in *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2016.
- [5] L. Kang, L. Wu, Y. Wei, B. Yang, and H. Song, "A highly accurate dense approach for homography estimation using modified differential evolution," *Engineering Applications of Artificial Intelligence*, vol. 31, no. 4, pp. 68–77, 2014.
- [6] J. Zaragoza, T.-J. Chin, M. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1285–1298, 2014.
- [7] Z. Wang, Y. Chen, Z. Zhu, and W. Zhao, "An automatic panoramic image mosaic method based on graph model," *Multimedia Tools and Applications*, vol. 75, pp. 1–16, 2015.
- [8] Y. Chen, J. Sun, and G. Wang, "Minimizing geometric distance by iterative linear optimization," in *International Conference on Pattern Recognition (ICPR)*, 2010.
- [9] L. Zhang, Y. Li, J. Zhang, and Y. Hu, "Homography estimation in omnidirectional vision under the  $L^\infty$ -norm," in *IEEE International Conference on Robotics and Biomimetics*, 2010.
- [10] W. Mou, H. Wang, G. Seet, and L. Zhou, "Robust homography estimation based on non-linear least squares optimization," in *IEEE International Conference on Robotics and Biomimetics*, 2013.
- [11] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, *Vision Algorithms: Theory and Practice, Lecture Notes in Computer Science*. Springer, 2000, ch. Bundle adjustment: a modern synthesis, pp. 298–372.
- [12] K. Konolige and W. Garage, "Sparse sparse bundle adjustment," in *British Machine Vision Conference (BMVC)*, 2010, pp. 1–11.
- [13] C. Xing, J. Wang, and Y. Xu, "A robust method for mosaicking sequence images obtained from UAV," in *International Conference on Information Engineering and Computer Science*, 2010.
- [14] M. Xia, J. Yao, L. Li, and X. Lu, "Globally consistent alignment for mosaicking aerial images," in *IEEE International Conference on Image Processing (ICIP)*, 2015.
- [15] Y. Lin and G. Medioni, "Map-enhanced UAV image sequence registration and synchronization of multiple image sequences," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–7.
- [16] S. Se, P. Firoozfam, N. Goldstein, L. Wu, M. Dutkiewicz, P. Pace, and J. P. Naud, "Automated UAV-based mapping for airborne reconnaissance and video exploitation," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2009, pp. 73 070M–73 070M.
- [17] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Homography based Kalman filter for mosaic building. applications to UAV position estimation," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 2004–2009.
- [18] Y. Xu, J. Ou, H. He, X. Zhang, and J. Mills, "Mosaicking of unmanned aerial vehicle imagery in the absence of camera poses," *Remote Sensing*, vol. 8, no. 3, p. 204, 2016.
- [19] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: Speeded up robust features," *Computer Vision & Image Understanding*, vol. 110, no. 3, pp. 404–417, 2006.
- [20] R. W. Floyd, "Algorithm 97: shortest path," *Communications of the ACM*, vol. 5(6), p. 345, 1962.
- [21] B. Bollobás, *Modern graph theory*. Springer Science & Business Media, 2013.
- [22] M. I. A. Lourakis, "Sparse non-linear least squares optimization for geometric vision," in *Computer Vision—ECCV 2010*, vol. 6312, 2010, pp. 43–56.