



## A robust projection plane selection strategy for UAV image stitching

Renping Xie, Jingmin Tu, Jian Yao, Menghan Xia & Shiwei Li

To cite this article: Renping Xie, Jingmin Tu, Jian Yao, Menghan Xia & Shiwei Li (2019) A robust projection plane selection strategy for UAV image stitching, International Journal of Remote Sensing, 40:8, 3118-3138, DOI: [10.1080/01431161.2018.1539273](https://doi.org/10.1080/01431161.2018.1539273)

To link to this article: <https://doi.org/10.1080/01431161.2018.1539273>



Published online: 13 Nov 2018.



Submit your article to this journal [↗](#)



Article views: 35



View Crossmark data [↗](#)



# A robust projection plane selection strategy for UAV image stitching

Renping Xie<sup>a</sup>, Jingmin Tu<sup>a</sup>, Jian Yao<sup>a</sup>, Menghan Xia<sup>a</sup> and Shiwei Li<sup>b</sup>

<sup>a</sup>Computer Vision and Remote Sensing (CVRS) Lab, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, P.R. China; <sup>b</sup>Hubei Province in Western Hubei Geological Mapping Team, Yichang, P.R. China

## ABSTRACT

Unmanned Aerial Vehicles (UAVs) are the most popular way to collect ground data today, thanks to their low cost and matchless convenience. However, UAVs are prone to unstable flight poses because they are so light in weight, which has resulted in a new challenge for UAV image stitching. In this paper, we propose a robust approach to stitch UAV images captured from approximately planar scenes without pose parameters assistance. The key idea of the proposed framework lies in an effective projection plane selection strategy, which is capable of resisting the perspective distortion from existing pose-perturbed images. To select a reasonable reference image as the projection plane, we first divide all the images into two groups (stable group and unstable group) according to their registration error under the affine model. Then, a specifically designed approach is used to define a weighted topological graph, which guarantees that the reference image is selected from the stable group while maintaining a global minimum accumulated registration error. Based on our cost topological graph, each unstable group image is locally attached to a stable group image via a homography. Finally, alignment parameters of all the stable group images are solved using affine model, after which global optimization is performed on the model of both groups. Comparing our results to those of the conventional approaches indicates that our proposed approach produced superior results in several challenging experiments involving both qualitative and quantitative evaluation.

## ARTICLE HISTORY

Received 29 May 2018

Accepted 6 October 2018

## 1. Introduction

Owing to the rapid development of the procurement process of optical data from areas beyond human reach, there is a high demand from different science fields to stitch sequential images together to create a large mosaicked image. Image mosaicking is a procedure that merges two or more images covering overlapping areas into a single composite image as seamless as possible in both geometry and colour tone. The first step in image mosaicking is image stitching which accurately aligns images into a

unified coordinate system and directly influences the quality of mosaicking (Zagrouba, Barhoumi, and Amri 2009b; Chen et al. 2014b; Zitova and Flusser 2003b).

The first step in image stitching is to select an appropriate alignment model. Szeliski (2006) proposed different models to map pixel coordinates from one image to another, such as 2D plane transformation models, 3D models, cylindrical models, and spherical models, which were then used to construct a complete image alignment system. A homography is a strict alignment model to describe the relationship between two images captured from a 3D plane or from a fixed centre of projection (Szeliski and Shum 1997b; Kang et al. 2014b). For images with parallax, many stitching methods have been proposed in recent years (Zhang and Liu 2014b; Lin et al. 2015b). Chang, Sato, and Chuang (2014b) utilized a novel parametric warp which smoothly extrapolates the projective transformations of the overlapping regions into the non-overlapping regions, and the resultant warp gradually changes from projective to similar across the image. In order to deal with stitching images with parallax, Zaragoza et al. (2013b) used a local weighting method to create a local homography for each pixel or grid, which was highly accurate and could significantly reduce ghost effects. However, in order to fit in different complex environments, all of these existing methods mostly utilized local constraints to solve local models that were not suitable for stitching multiple images and achieved more degrees of freedom (DoFs) than global models, which can make the model's parameters more difficult to solve.

Unmanned Aerial Vehicle (UAV) image stitching mainly addresses stitching planar scenes for the purpose of recovering the homographies of image pairs (Xu et al. 2016b). Although many transformation models were proposed in the past decade (Chen, Sun, and Wang 2010b), their performances, considering both accurate alignment and global consistency, still need improvements.

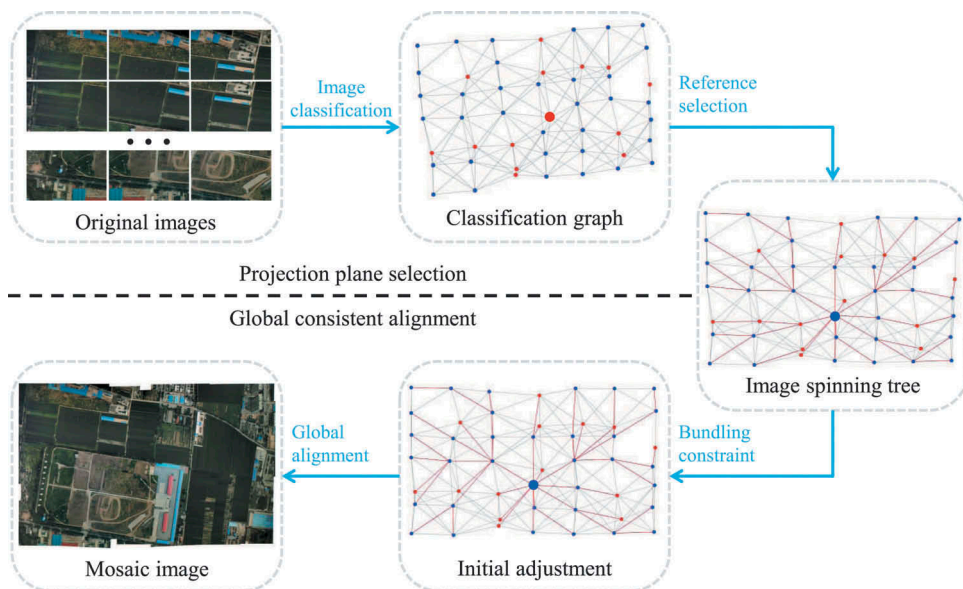
Most approaches to image alignment are feature-based (Mou et al. 2013b). The basic procedure is first to extract the features, including as many point, line, or edge features as possible. Then, the corresponding relationships are established between images of interest based on the extracted features. Finally, the transformation models are solved through optimization methods (Elibol, Gracias, and Gracias 2012b). Point features are most widely used because of their stability and universality. To improve the efficiency of their algorithm and obtain better alignment results, Li et al. (2015b) used the matches of sparse points and lines as constraints and proposed a dual-feature transformation model for image alignment to ensure the integrity of the line structures in the final mosaicked image. With an appropriate mathematical transformation model and matching features, many optimization algorithms also have been proposed to improve the precision of the stitching and to obtain global optimal results (Konolige and Garage 2010b). The bundle adjustment approach is often used as a typical global optimization method, which finds an optimal solution minimizing the total reprojection error (Liu et al. 2009b; Triggs, Zisserman, and Szeliski 2003b; Brown and Lowe 2007b). Xing, Wang, and Yaming's (2010b) proposed algorithm first applied the extended kalman filter onto the local area to obtain a good initial solution for global optimization and then refined all the parameters globally. Some methods utilized topological structure information of images to achieve global registration (Kang, Cohen, and Medioni 2000b; Marzotto, Fusiello, and Murino 2004b). To avoid the down-scaling effect on the image quality, Elibol et al.

(Elibol 2008) proposed optimizing the point positions in a mosaic frame and optimizing an alignment model in an alternate iteration scheme.

The above-mentioned approaches all concentrated on the optimization strategy or the registration scheme, seeking alignment with least registration errors, which usually produces a satisfying stitching result when there are several or dozens of images. However, because of the pseudo-planarity of observed scenes, many alignment methods based on minimizing registration errors would result in a significant accumulation of perspective distortions for hundreds of images acquired from wide-range regions. To overcome this problem, the geo-referenced satellite images of a whole region can be selected as the reference images (Lin and Medioni 2007b; Se et al. 2009b). Other approaches to avoid accumulated errors used extra sensors, such as GPS and IMU, to directly obtain camera poses for mosaicking (Yahyanejad et al. 2010b). In practice, pose parameters are not always available; therefore, optimizing the alignment models for resisting perspective distortions has been a popular research topic recently (Wang et al. 2016b; Xu et al. 2016b). Koo, Kim, and Cho (2009b) presented a new mathematical criterion to select an optimal warping function which minimizes the total number of perspective distortions of all the images. Multichannel bending (Burt and Adelson 1983) and stitching line cutting (Kwatra et al. 2003b) can alleviate this problem, but artefacts may still exist when images have large perspective distortions. Gao, Kim, and Brown (2011b) used two homographies to deal with two projection planes in an observed scene, through which better stitching result can be obtained. Many scholars used space variant distortion algorithms to address images with small perspective distortions (Lin et al. 2011b). Caballero et al. (2007b) proposed a hierarchical homography computation model to deal with quasi-planar scenes, where accumulated drifts can be compensated for and propagated to the rest of the mosaic when a loop is present in the sequence of images. To allow the transition between the alignment models, Xia et al. (2015b) proposed to initially align images by a robust affine model, followed by model refinement under an anti-perspective constraint. In this way, both the global consistency and the alignment precision could be achieved at an optimal balance. However, the reference image was fixed as the first image, which limited its application to large-scale scenes and brought more serious challenges to maintaining a global minimum accumulated error.

As for optimal reference image selection, Richard et al. (2006) stated that a reasonable selection is the most geometrical central image. This supposition is obvious because a central image usually has the shortest distance to all other images on average. However, the authors did not provide a way to find such an image. To solve this problem, Choe et al. (2006b) used the Floyd–Warshall's all-pairs shortest path algorithm to select a reference image with a minimal total registration error. To find the optimal reference image which minimizes the total propagated error, Xia et al. (2017b) proposed to organize all the images as a weighted topological graph and construct a symmetric cost matrix of the all-pairs shortest path to find a node with the least total weight of the shortest paths to all other nodes. However, the weights of the edges in the topological graph were set only according to the number of matched features and registration errors while the magnitude of the images' distortions was not considered. Huge distorted images usually should not be selected as reference images.

In this paper, we concentrate on projection plane selection for UAV images stitching captured from approximate planar scenes without pose parameters assistance. Because images are captured from approximate planar scenes and there are large pose-perturbed images at the same time, homography cannot describe the relationship between images perfectly and there are large unavoidable perspective distortion in stitching results. To address this problem, we propose an effective projection plane selection strategy, which can resist perspective distortion from possible existing pose-perturbed images. There are two general steps for the proposed approach: (1) projection plane selection and (2) global consistent alignment, as shown in Figure 1. First, the images are divided into two groups, the stable group and the unstable group, according to their registration errors under the affine model. Second, a specially designed approach is used to define a weighted topological graph, which guarantees that the reference image will be selected from the stable group while maintaining a global minimum accumulated registration error, and a spanning tree is built by the classical Floyd–Warshall algorithm, which determines the aligning order of the images in the following alignment. Third, a bundling strategy is creatively employed for the initial alignment, in which each unstable-group image is locally attached to a stable-group image via a homography, and the alignment parameters of all the stable-group images are solved with affine models. Finally, a global optimization is performed on the models of both groups to improve the alignment precision and avoid perspective distortion. Our approach is tested in several groups of experiments on both synthetic and challenging real datasets, which clearly illustrate the superiority of our approach.



**Figure 1.** The flowchart of our proposed approach for mosaicking UAV images captured from approximate planar scenes. The blue and red solid circles in the topological graphs denote the images in the stable group and the unstable group, respectively, and the large solid circles indicate the reference image.

The remainder of this paper is organized as follows. In [Section 2](#), a robust projection plane selection scheme is presented, and the globally consistent alignment strategy is detailed. In [Section 3](#), the qualitative and quantitative experimental results are presented. Finally, [Section 4](#) discusses our conclusions and our suggestions for future research directions.

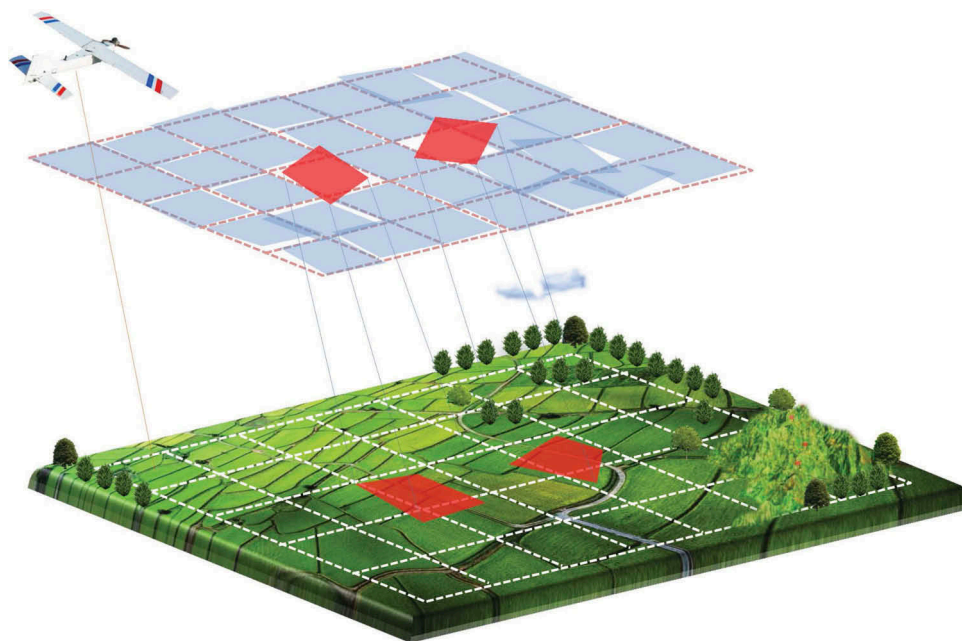
## 2. Our approach

### 2.1. Projection plane selection

The reference image that is generally treated as the projection plane of image mosaicking can be selected at the geometrical centre or through other simple methods. However, other approaches do not take the perspective distortion of images into consideration, which may result in considerable distortion of the mosaicked image. To select a reasonable reference image as the projection plane, we divide all the images into two groups: the stable group (e.g. the light blue rectangles as shown in [Figure 2](#)) and the unstable group (e.g. the red rectangles as shown in [Figure 2](#)), and search for an optimal reference image from the stable group, as demonstrated in [Figure 2](#).

#### 2.1.1. Image classification

The perspective distortion of images cannot be directly expressed based only on the image itself while the pose parameters are invalid, and it is also difficult to distinguish



**Figure 2.** An illustration of the robust projection plane selection strategy. The light blue rectangles represent the stable-group images during image classification, and the red rectangles represent the unstable-group images. The whole brown rectangle represents the projection plane we selected and the whole white rectangle represents the real plane of the observed scene.

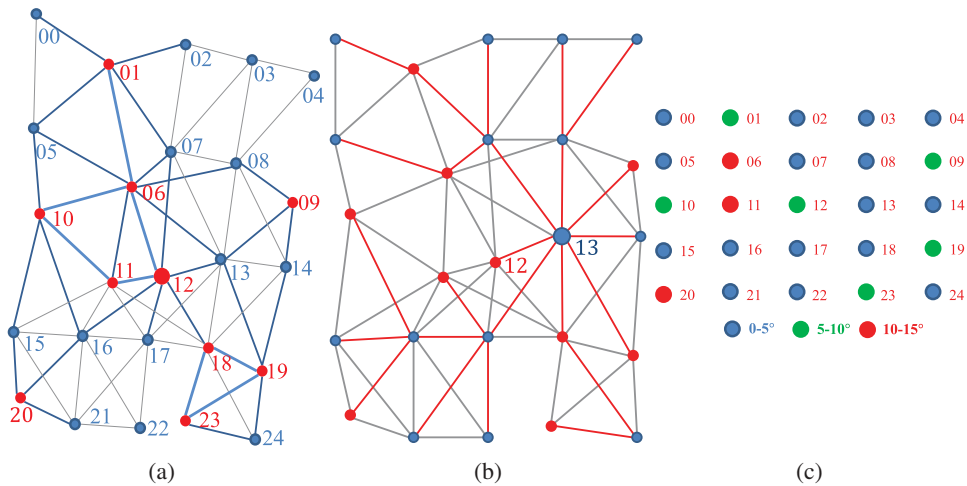


the magnitude of the perspective distortion of images. When all the images are warped onto a reference plane using affine model  $\mathbf{A}_i$  and there are a few images belonging to the image group of large perspective distortion, there should be larger registration errors between those unstable-group images and their surrounding stable-group images, but the registration errors between the images of small perspective distortion and their adjacent images are mostly very small. When the registration errors between each image and all the other images in a group are small enough, all the images in the group have similar perspective distortion. Based on these observations, we adopt a projection detection strategy to divide all the images into two groups: stable group (represented by  $\mathbb{S}$ ) and unstable group (represented by  $\mathbb{U}$ ), which includes three steps: error calculation, seed selection, and region growing.

The main idea of our strategy is to first divide all the images into small groups with similar perspective distortion and then merge these small groups through region growing. First, we warp all images onto reference image of geometrical centre via affine model and calculate registration error  $e_{i,j}$  between image  $I_i$  and adjacent images  $I_j$ , average registration error  $\bar{e}_i$  with all the adjacent images, average registration error  $\hat{e}$ , and mean square error  $\sigma$  for all the images. Second, the seed selection aims to divide all the images into small seed groups with similar perspective distortion or seed points having large perspective distortion differences with adjacent images. Considering that the perspective distortion of single image is hard to distinguish, we adopt a better strategy in which three images are chosen as a group in seed selection. In order to efficiently obtain image groups that have similar perspective distortion, two conditions are designed to choose the seed groups consisting of three adjacent images according to:

$$\begin{cases} \bar{e}_{v,v \in \{i,j\}} |\hat{e} - \theta_c \sigma|, \\ \sum_{I_v \cap \{t \neq \emptyset, v, t \in \{i,j,k\}, v \neq t\}} e_{v,t} \langle 3|\hat{e} - \theta_f \sigma|, \end{cases} \quad (1)$$

where  $I_i$  represents  $i$ th image in the sequential images  $I$ , and  $\theta_c$  and  $\theta_f$  are two given thresholds. The first condition indicates that the average registration errors of two images in this group are relatively small, which can avoid the perspective distortion of this group is great different with the most images. We set  $\theta_c = 3$  by default. The higher the value of  $\theta_c$  is, the smaller the allowable average registration error is and the fewer the number of large pose-perturbed images in seed groups is. The second condition is used to avoid that there are great registration errors between two images in this group. We set  $\theta_f = 0.5$  by default, which can constrain the similarity of the perspective distortion among three adjacent images. The smaller the value of  $\theta_f$  is, the higher the allowable registration error between the three images is. In order to find the images whose perspective distortion are quite different, images  $I_k$  are regarded as seed points when  $\bar{e}_k > |\hat{e} + \theta_c \sigma|$ . Third, the region-growing procedure aims to merge images of the same perspective distortion as much as possible, which is performed as follows: (1) Given two images  $I_i$  and  $I_j$  in a seed group, when  $e_{v,t,t \in \{i,j\}} < |\hat{e} - \theta_f \sigma|$ , isolated image  $I_v$  is added into this image group. (2) Classify seed groups into the stable group when there is no seed point among these groups. If  $I_i$  and  $I_j$  belong to the stable group, and  $e_{k,t,t \in \{i,j\}} > |\hat{e} + \theta_c \sigma|$ , seed points  $I_k$  are labelled as the unstable group, such as  $I_6$ ,  $I_{12}$  in Figure 3(a). (3) For the remaining unclassified seed points, seed groups, or isolated



**Figure 3.** An illustration of the weighted topological graph and the reference selection in our approach: (a) the weighted topological graph applying affine models with auto-selected reference image  $l_{12}$ ; (b) the spanning tree of the weighted graph with auto-updated reference image  $l_{13}$ ; and (c) the true classified graph referring to the perspective distortion. The grey edges imply the weights defined in Equation (2)(b), the blue edges indicate the weights defined in Equation (2)(c), and the bold blue edges imply the weights defined in Equation (2)(d). The red edges indicate the shortest way from each image to the reference image. The different of solid circles represent the different sizes of perspective distortion in Figure 3(c), and the corresponding degree represents the angle of the image with the observation plane.

images, such as  $l_{20}$ ,  $l_{24}$ , considering that there are only a few images of great perspective distortion, we classify them according to the principle of dichotomy. That is to say, when  $l_i$  and  $l_j$  belong to the unstable group, and  $e_{v,k,k \in \{i,j\}} > |\hat{e} + \theta_c \sigma|$ , image  $l_v$  would be divided into the stable group, such as  $l_{24}$  in Figure 3(a). The final classified topological graph of the synthetic dataset is shown in Figure 3(a), and the real classification of perspective distortion is shown in Figure 3(c). The perspective distortion of the images also can be determined by the images' poses (if any), machine learning, or manual intervention.

### 2.1.2. Reference image selection

To make stitching image as natural as possible, the stitching plane should be parallel to the real plane of the surveyed area. Since the flight plane is generally parallel to the ground and the images in this paper were captured from approximate planar scenes, the perspective distortion of the reference image should be small and the optimal reference image should be selected from the stable group. In general, the stitching procedure involves cascading a series of relative intermediate images that induce the accumulation of perspective distortion, especially the unstable-group images. In order to assure that as few as possible unstable-group images are selected as intermediate images, an undirected weighted topological graph is constructed, where the images are denoted by nodes and the overlaps between the image pairs are denoted by edges. The weight of an edge can be set according to the reciprocal of the number of matched features (Elibol, Gracias, and Garcia 2013b) or the registration error between the image pair (Choe



et al. 2006b). We set the weights of edges considering the association between the number of matched features and the registration error, and the weights are defined as

$$w_{ij} = \begin{cases} \inf, & \text{if } N_{ij} = 0, & (a) \\ \frac{1}{\ln(N_{ij} + \vartheta)}, & \text{if } N_{ij} > 0, l_i \text{ and } l_j \in \mathbb{S}, & (b) \\ \frac{1}{\ln(N_{ij} + \vartheta)} + \delta, & \text{if } N_{ij} > 0, l_i \text{ or } l_j \in \mathbb{U}, & (c) \\ \frac{1}{\ln(N_{ij} + \vartheta)} + 2\delta, & \text{if } N_{ij} > 0, l_i \text{ and } l_j \in \mathbb{U}, & (d) \end{cases} \quad (2)$$

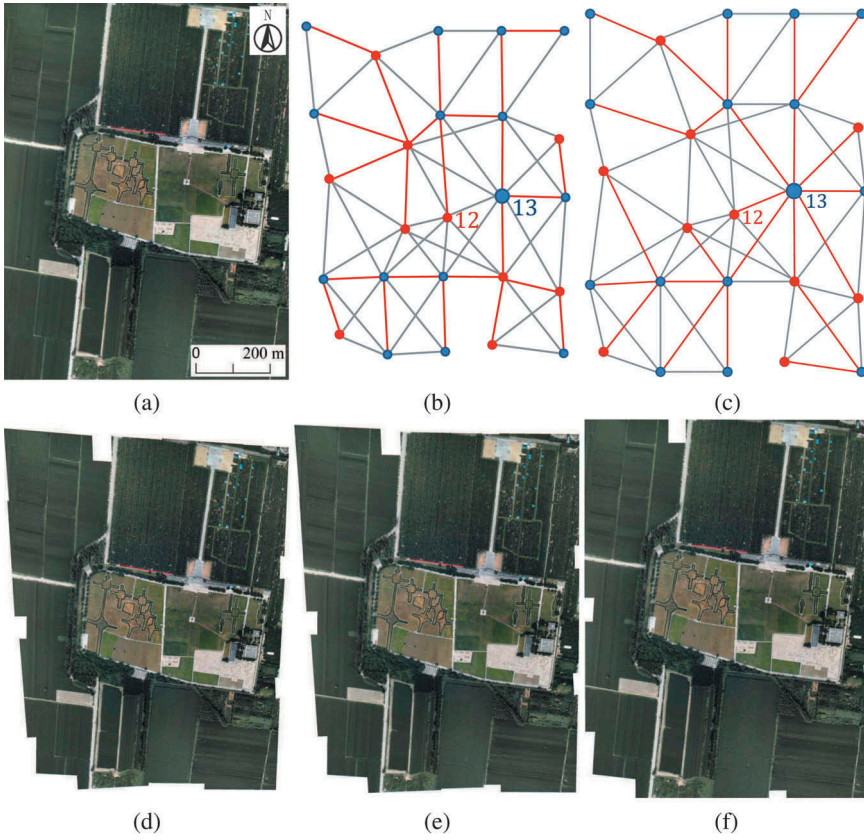
where  $N_{ij}$  denotes the total number of matches between  $l_i$  and  $l_j$ ,  $\delta$  is the maximum in  $\{w_{ij}\}$  got from Equation (2)(b) and  $\vartheta$  is a constant for regularization ( $\vartheta = 50$  by default). The final weighted topological graph is shown in Figure 3(a).

The optimal reference image should satisfy the minimal sum of accumulated errors from all the other images to the reference image (Xia et al. 2017b). In order to do this, we use the Floyd–Warshall's all-pairs shortest path algorithm (Floyd 1962b) to select the optimal reference image on the weighted graph. When the selected reference image belongs to the unstable group, such as  $l_{12}$  in Figure 3(a), its adjacent node is selected as the new reference image based on the idea of the shortest path algorithm, which belongs to the stable group and the weight of which with previous reference image is least, such as  $l_7$  in Figure 3(b). As soon as the optimal reference image is selected, the minimum spanning tree of the weighted graph can be obtained by choosing the optimal reference image as the root node and then applying the classical Floyd–Warshall algorithm.

To demonstrate the validity of the above procedure, a high-resolution image (centre coordinates are  $36^\circ 8' 42''$  N,  $114^\circ 18' 10''$  E), as shown in Figure 4(a), was posted onto the surface of an approximate planar model in a commercial software named 3D Studio Max. We then produced a variety of distortion images by setting the principle axis of the virtual camera with the ground at different angles ranging from  $70^\circ$  to  $90^\circ$ . The stitching result of Xia et al.'s algorithm (Xia et al. 2017b) is shown in Figure 4(d), which automatically selects  $l_{12}$  as reference image; and the result of our approach is shown in Figure 4(c,f). For effective comparison,  $l_{13}$  is set as reference image in Xia et al.'s algorithm and the stitching result are shown in Figure 4(b,e). In terms of this comparative experiment, our proposed approach is more efficient than Xia et al.'s approach and the spanning tree is also more reasonable.

## 2.2. Global consistent alignment

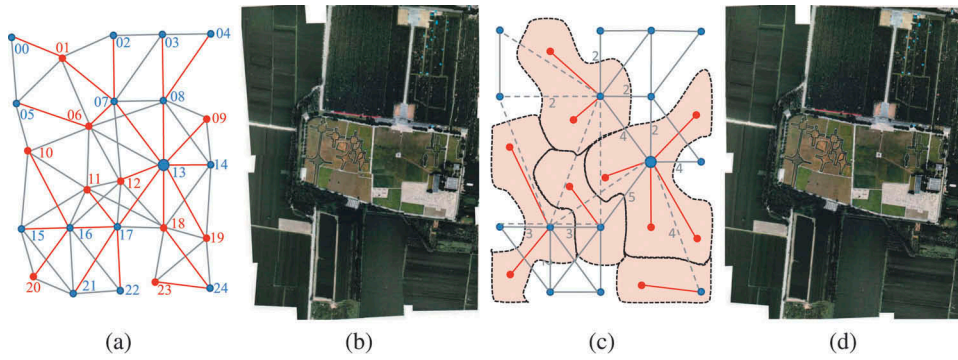
Since UAV images in this paper were captured from approximate planar scenes, homography model is adopted in global alignment, which transforms global alignment optimization to a non-linear problem in which the initial solution and the energy function become two essential components. However, during the process of obtaining the initial solution, a single transformation model is usually adopted in practical applications, which produces poor accurate alignment using only affine model or suffers severe perspective distortion using only homography. Therefore, we creatively adopt a bundling strategy in which different models are used for two groups of images during the initial alignment, and the models are refined with a global energy function considering an anti-perspective function.



**Figure 4.** The comparison results of the synthetic dataset (Syntheticdata-25) about the reference image selection: (a) the original high-resolution image; (b) the topological graph generated by Xia et al.'s approach with manually selected reference image  $I_{13}$ ; (c) the topological graph generated by our approach with auto-selected reference image  $I_{13}$ ; (d) the result aligned by Xia et al.'s approach with auto-selected reference image  $I_{12}$ ; (e) the alignment result corresponding to Figure 4(b,f) and the alignment result corresponding to Figure 4(c).

### 2.2.1. Initial alignment by bundling constraint

Based on the optimal reference image, a robust initial solution for the global refinement process is possible by warping other images to the reference plane group by group. But the unstable-group images would cause several accumulative error in the results due to the model's inaccuracy. In our approach, images  $I_u$  ( $I_u \in \mathbb{U}$ ) are not simply dropped, but instead are bundled to adjacent bundled images  $I_s$  ( $I_s \in \mathbb{S}$ ) which has the most overlap with itself to reconstruct new images  $\hat{I}_s$ , such as the reddish polygons in Figure 5(c). First, according to the spanning tree constructed in Section 2.1.2, its parent-node image is chosen as the bundled image, and the sibling node images belonging to the stable group are candidates when the parent-node image belongs to the unstable group. Then, homography  $\mathbf{H}_{s,u}$  from image  $I_u$  to image  $I_s$  is calculated, and new overlapping relations  $\{\hat{\mathbf{x}}_{s,j}^k\}$  with adjacent image  $I_j$  are reconstructed by adding matching points  $\{\mathbf{x}_{u,j}^k\}$  to adjacent relationship  $\{\mathbf{x}_{s,j}^k\}$  as



**Figure 5.** Comparison of initial solution using bundling strategy or not: (a,b) the result of initial alignment without using bundling strategy; (c,d) the result aligned by our approach using bundling strategy. The grey numbers in Figure 5(c) represent the number of merged overlaps and the dotted lines stand for newly generated overlaps. Considering simplicity, the reddish polygons indicate new images not including new overlaps.

$$\hat{\mathbf{x}}_{s,j}^k = \{\mathbf{x}_{s,j}^k\}_{l_s \in \mathbb{S}} \cup \left\{ \mathbf{x}_{u,j}^k \mathbf{H}_{s,u} \right\}_{l_u \in \mathbb{U}}, \quad (3)$$

where  $\mathbf{H}_i$  is the  $3 \times 3$  homography transformation matrix,  $\mathbf{x} = [x, y, w]^T$  denotes the homogeneous coordinate of a feature, and  $\mathbf{x}_{i,j}^k$  denotes 2D coordinate of  $k$ th matched feature in  $l_i$  corresponding to  $k$ th matched feature  $\mathbf{x}_{j,i}^k$  in  $l_j$ .

For example, image  $l_{01}$  is bound to image  $l_{07}$ , as shown in Figure 5(a,b); therefore, matching pairs  $\mathbf{x}_{01,02}$  between image  $l_{01}$  and image  $l_{02}$  are added to matching pairs  $\mathbf{x}_{07,02}$  and matching pairs  $\mathbf{x}_{01,05}$  are added to matching pairs  $\mathbf{x}_{07,05}$ . Of course, matching pairs  $\mathbf{x}_{01,05}$  would generate new matching relations  $\mathbf{x}_{07,05}$  that are shown as the dotted line in Figure 5(c). Finally, based on the reconstructed relations, affine models  $\mathbf{A}_s$  are solved by minimizing the cost function as follows:

$$E(\mathbf{A}) = \sum_{l_s, l_j \in \mathbb{S}} \sum_{k=1}^{N_{s,j}} \left\| \varpi(\mathbf{A}_s \hat{\mathbf{x}}_{s,j}^k) - \varpi(\mathbf{A}_j \hat{\mathbf{x}}_{j,s}^k) \right\|^2, \quad (4)$$

where  $N_{s,j}$  denotes the total number of matches between  $l_i$  and  $l_j$ ,  $\mathbf{A}_s$  is the  $3 \times 3$  affine transformation matrix, and  $\varpi(\mathbf{x}) = [x/w, y/w]^T$  represents the transformation of the homogeneous coordinate of a feature into the non-homogeneous coordinate. As linear equations, Equation (4) can be solved easily by the Singular Value Decomposition method.

### 2.2.2. Global optimization using anti-perspective function

Although the above initial alignment makes the stitching result robust as far as perspective distortion, several registration errors likely exist among the images since these deviations have never been processed in a global way. In this section, based on the primitive matching relationships, we align all the images in a unified energy framework which allows the transition from affine model to homography model under reasonable

constraints to reach an optimal balance between the alignment precision and the global consistency. The energy is composed of two terms: one minimizes the sum of the feature registration errors between images and the other suppresses the accumulation of perspective distortions. Let  $I = \{I_i\}_{i=1}^n$  be the primitive image set and the affine models  $A = \{A_i\}_{i=1}^n$  of all images as resolved above, which are adopted as the initial solution of global optimization, with the first energy term defined as

$$E_d(H) = \sum_{I_p, I_q \in I} \sum_{k=1}^{N_{p,q}} \|\varpi(H_p \mathbf{x}_{p,q}^k) - \varpi(H_q \mathbf{x}_{q,p}^k)\|^2, \quad (5)$$

where  $H = \{H_i\}_{i=1}^n$  represents the unknown homographies with respect to the reference plane.

Another optimization objective is to maintain global consistency by suppressing severe perspective distortion of stable-group images  $I_s$  and giving unstable-group images  $I_u$  more freedom to adjust during the homography ( $I_s$  and  $I_u$  defined in Section 2.2.1). First, the optimal homographic transformation should be close to the initially estimated models; therefore, the deviations between the optimal homographic transformations and the initially estimated transformations are set as a regularization term:

$$E_r^1(H) = \sum_{I_p, I_q \in I} \sum_{k=1}^{N_{p,q}} \left( \|\varpi(H_p \mathbf{x}_{p,q}^k) - \hat{A}_p \mathbf{x}_{p,q}^k\|^2 + \|\varpi(H_q \mathbf{x}_{q,p}^k) - \hat{A}_q \mathbf{x}_{q,p}^k\|^2 \right), \quad (6)$$

where  $\hat{A}_p$  (so does  $\hat{A}_q$ ) represents the image's warping model to the projection plane and is calculated as

$$\hat{A}_p = \begin{cases} A_p, & \text{if } I_p \in \mathbb{S}, \\ A_s \varpi(H_p), & \text{if } I_p \in \mathbb{U}, \end{cases} \quad (7)$$

where image  $I_s$  represents the bundled image of  $I_p$  (or  $I_q$ ).  $\varpi(\mathbf{x}) = [x/w, y/w]^T$  represents the transformation of the homogeneous coordinate of a feature into the non-homogeneous coordinate. Second, the plane of images  $I_s$  is nearly parallel to the real projection plane, and images  $I_s$  should not have severe perspective distortion during the global refinement. Therefore, we introduce the mathematical distortion coefficients for images  $I_s$  as a regularization term:

$$E_r^2(H) = \sum_{I_t \in I} \left\{ \frac{N_t}{4} \sum_{k=1}^{C_4} \log_{10} \left( (\delta_m^1 - 1)^2 + \left( \frac{1}{\delta_m^1} - 1 \right)^2 + (\delta_m^2 - 1)^2 + \left( \frac{1}{\delta_m^2} - 1 \right)^2 \right) \right\}, \quad (8)$$

where  $\delta_m^1$  and  $\delta_m^2$  represent the maximum and minimum local stretches of point  $\mathbf{x}_k$  defined by Koo, Kim, and Cho (2009b),  $N_t$  represents the total number of matching points with all the adjacent images, and  $C_4$  indicates a set containing image's four corners. If the two values  $\delta_m^1$  and  $\delta_m^2$  are both more close to 1, the transformed images look more like their source images (also in size). Specifically, the purpose of  $\log_{10}(\cdot)$  is to normalize the function value that represents the perspective distortion of a point, and the average distortions of the image's four corners are used to displace the image's

distortions for simplifying the calculation. The regularization terms defined in Equations (6) and (8) can be linearly combined to define the final regularization function as

$$E_r(H) = E_r^1(H) + \gamma E_r^2(H), \quad (9)$$

where parameter  $\gamma$  denotes the weight coefficient for balancing regularity terms  $E_r^1(H)$  and  $E_r^2(H)$ . Because the term  $E_r^2(H)$  represents the amount of average stretches of all images, the transformed images look like their source images in size when their value is approximate to zero. For unstable-group images  $I_u$ , we should give the models more DoFs to adjust to reduce the feature registration errors and set parameter  $\gamma$  as 0. In order to regard the stable-group images  $I_s$  as the main constraint plane, parameter  $\gamma$  is set as 0.15 for images  $I_s$  by default. So far, based on the energy function defined in Equations (5) and (9), the final energy function can be defined as follows:

$$E(H) = E_d(H) + \lambda E_r(H), \quad (10)$$

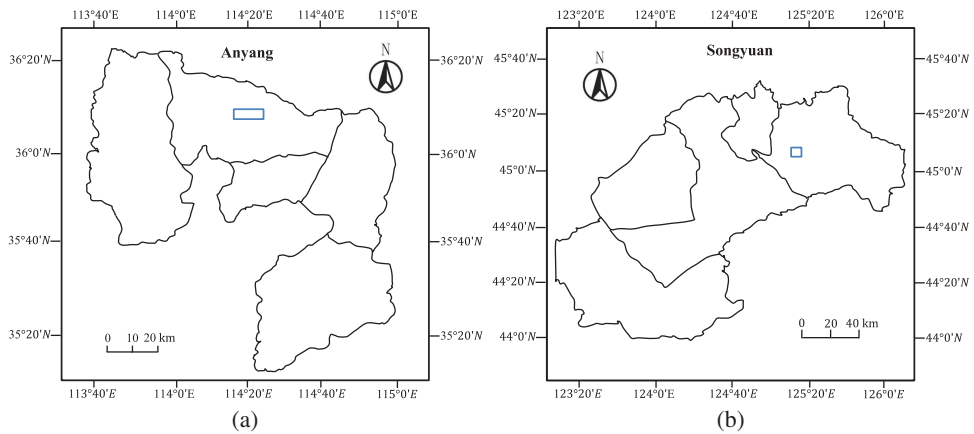
where parameter  $\lambda$  denotes the weight coefficient for balancing data term  $E_d(H)$  and regularity term  $E_r(H)$ . The two values of  $E_d(H)$  and  $E_r(H)$  reflect the alignment errors and the perspective distortion, respectively. The perspective distortion usually magnifies when the value of  $E_d(H)$  decrease, and the alignment precision often increase if the value of  $E_r(H)$  diminishes. Theoretically, a smaller value of  $\lambda$  enhances the precision of local alignment while reducing the global consistency. Considering that the constraint of the regularity term is not strict, we set  $\lambda$  as 0.05 to improve the global consistency in all the experiments. The above optimization can be solved by the Levenberg-Marquardt (LM) algorithm (Li, Deren, and Fan 2012b).

### 3. Experimental results

In this section, three typical datasets, including one synthetic dataset and two real datasets, are used to evaluate our proposed approach. The comparative experiments with the algorithm proposed by Xia et al. (2017b) dealing with approximate planar mosaicking and with the commercial software "PTGui" (PTGui is available at <http://www.ptgui.com/>.) are performed, which are evaluated both qualitatively and quantitatively. Because the aim of comparative experiments is comparing the alignment results only, the following seamline detection and tonal correction are skipped in PTGui. The qualitative evaluation is based on the global consistency, which is implemented by visual observation; and the quantitative evaluation is based on the alignment precision, which is evaluated based on the registration errors among the sequential images. The comprehensive results demonstrate the superiority of our approach.

#### 3.1. Dataset description

The first real aerial dataset was captured at an average flight height of 780 m over an urban area in Anyang city, Henan Province, China. The study area's average elevation is 76 m, and the area is 8 km in length and 1.8 km in width (centre coordinates are 36°8'10" N, 114°20'36" E), as shown in Figure 6(a). The average flight height of the



**Figure 6.** Location of the study area and the range of datasets marked with blue rectangles: (a) the aerial dataset, (b) the UAV dataset.

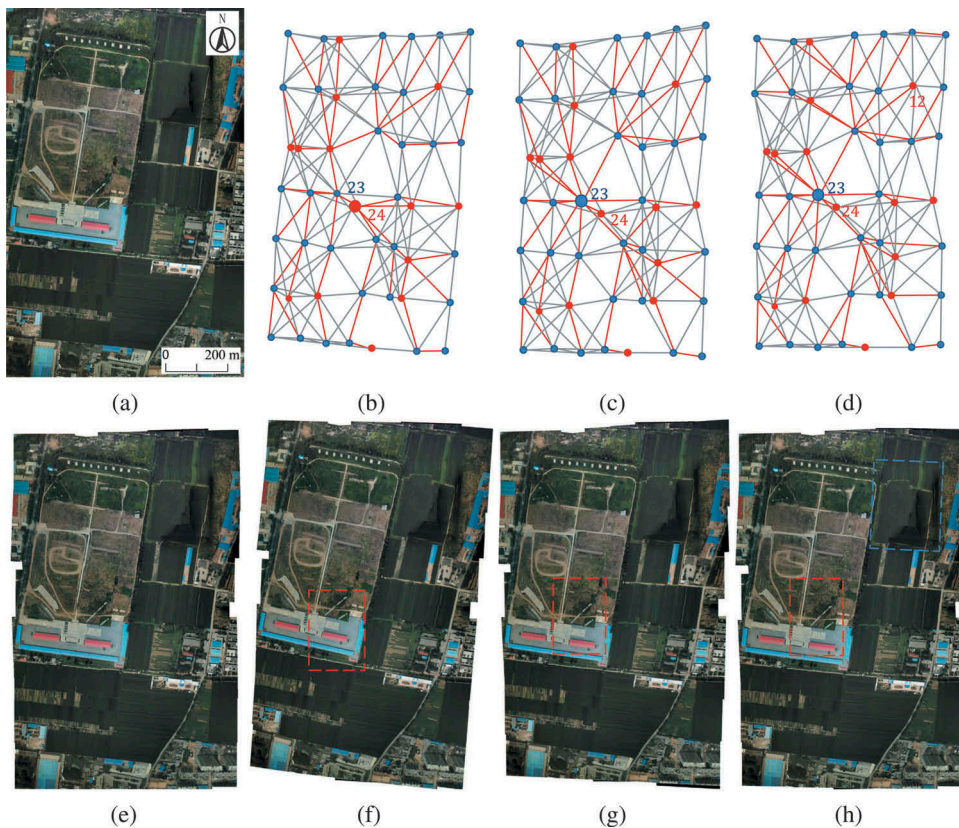
second real UAV dataset was 370 m while observing the Songnen Plain of Songyuan City, Jilin Province, China. The average elevation of the study area is 165 m, and it has an area of  $6.56 \text{ km}^2$  (4.44 km in length and 1.46 km in width) (centre coordinates are  $45^\circ 6' 37'' \text{ N}$ ,  $125^\circ 14' 9'' \text{ E}$ ), as shown in Figure 6(b).

### 3.2. Synthetic dataset

To test the robustness of our approach, a high-resolution aerial image with the size of  $9336 \times 6000$  pixels was selected from the first real aerial dataset, as shown in Figure 7(a) (centre coordinates are  $36^\circ 8' 23'' \text{ N}$ ,  $114^\circ 21' 39'' \text{ E}$ ). It was posted onto the surface of an approximate planar model in commercial software 3D Studio Max. Then, we produced diverse distortion images by setting the angle between the principle axis of the virtual camera and the model plane at different degrees ranging from  $75^\circ$  to  $90^\circ$ .

The stitching results using PTGui, Xia et al.'s approach, and our approach are shown in Figure 7(e–h), and the results of PTGui and our approach both show noticeable global consistency with original image about the visual effects. For the reference image selection, unstable-group image  $I_{24}$  is selected automatically in Xia et al.'s approach, which results in the deviation between projection plane with real ground, as shown in Figure 7(b,f). On the contrary, stable-group image  $I_{23}$  is selected as the optimal reference image in our approach, as shown in Figure 7(d,h), and it is also selected as the reference image in Xia et al.'s approach to make the comparison more convincing, as shown in Figure 7(c,g). In terms of visual effects, the stitching result applying our approach obtains better global consistency compared to the result from Xia et al.'s approach, whose top right-hand corner is overstretched severely and left bottom part is compressed. Although both grouping alignment and global optimization are applied, there are still severe perspective distortions in the result from Xia et al.'s approach because the size of the image's distortion is not considered when searching for an optimal projecting path for each image, which results in cascading many large distorted images during projection procedure. Besides, we should notice the image  $I_{12}$ , as shown in Figure 7(d,h),





**Figure 7.** The experimental results on the synthetic dataset (Syntheticdata-49): (a) the original high-resolution image; (b, f) the topological graph and the stitching result via Xia et al.'s approach with auto-selected reference image  $I_{24}$  marked with a red rectangle; (c, g) the topological graph and the stitching result applying Xia et al.'s approach with manually selected reference image  $I_{23}$ ; (d, h) the topological graph and the stitching result via our approach with auto-selected reference image  $I_{23}$ ; and (e) the stitching result with the commercial software PTGui. The reference image of each mosaic is marked with a red rectangle. The blue rectangle in Figure 7(h) stands for hill area.

and its adjacent images were captured from hill area, and the poses of these images are stable which range from  $0^\circ$  to  $5^\circ$ . However, the image  $I_{12}$  is divided into unstable-group because there are large registration errors with its surrounding stable-group images. Compared with the stitching result applying Xia et al.'s approach, the image  $I_{12}$  is not selected as the intermediate image during projection procedure in our approach which avoids the accumulation of perspective distortion, as shown in Figure 7(b–d, f–h). When UAV images captured from approximate planar scenes and when there are a few large pose-perturbed images, the perspective distortion caused by hill area is similar to the perspective distortion caused by large pose perturbation, and our approach can also recognize it.

As a quantitative evaluation, the numerical results shown in Table 1 (Syntheticdata-49) demonstrate that the alignment precision greatly improve after the application of both Xia et al.'s approach and ours, and the alignment precision of PTGui's result is the worst of the three approaches although the global consistency of PTGui's result is as



**Table 1.** Quantitative comparison on stitching results obtained through different methods by using the root-mean-square (RMS) errors. (IA: initial alignment; GR: global refinement; Unit: pixel.).

Method	Syntheticdata-25		Syntheticdata-49		Urbanarea		Songnenplain	
	RMS (IA)	RMS (GR)	RMS (IA)	RMS (GR)	RMS (IA)	RMS (GR)	RMS (IA)	RMS (GR)
Xia et al.'s method	3.7871	2.1344	3.3086	1.1376	2.8914	1.3044	6.3411	3.4303
PTGui	–	6.6009	–	6.0529	–	4.7356	–	6.9285
Our approach	1.5340	1.5390	4.1993	1.3052	3.1412	1.2323	6.8284	3.6116

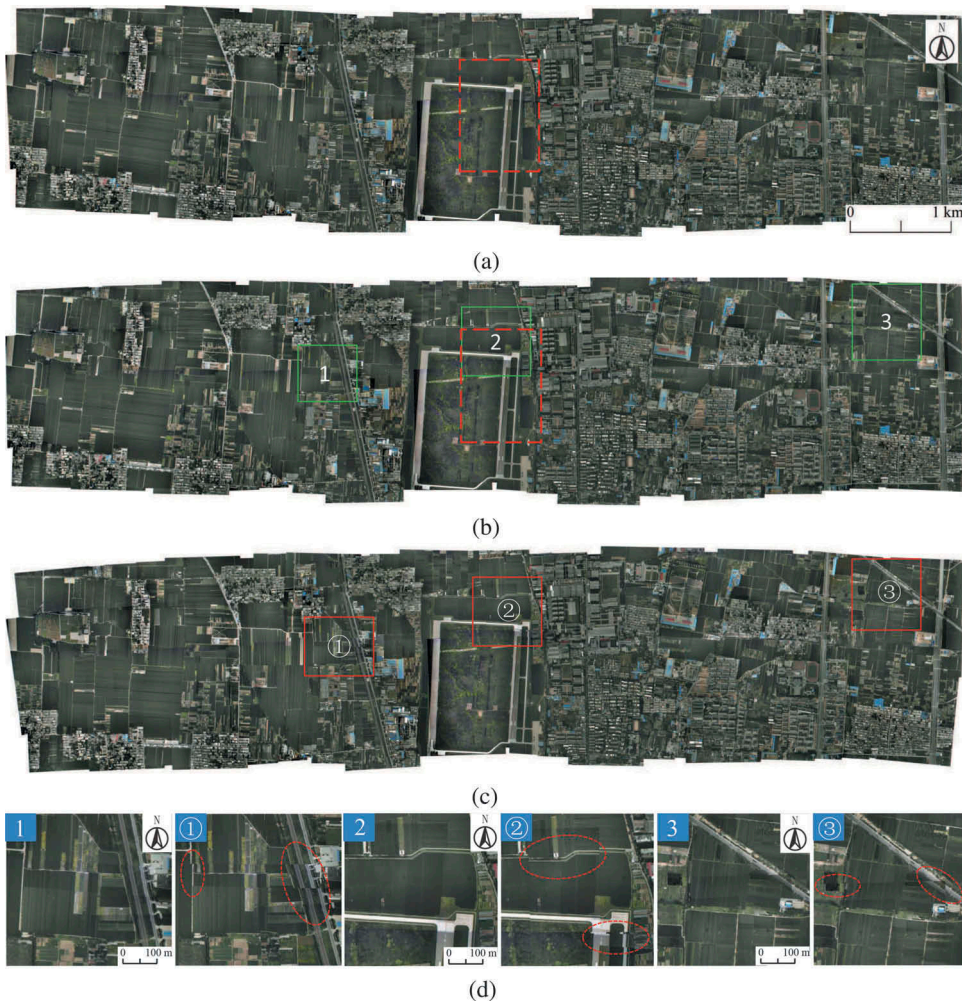
good as that of ours. Furthermore, our approach achieves worse result than Xia et al.'s approach after conducting the initial alignment. After applying global alignment, our approach also produces worse result compared with Xia et al.'s approach about alignment precision because the anti-perspective function in global optimization improves the global consistency of our result but simultaneously reduces its alignment precision. Therefore, considering the global consistency, our approach's result is approximate with PTGui's result, and the alignment precision is as good as Xia et al.'s approach simultaneously.

### 3.3. Real datasets

To evaluate our proposed approach on differently pose-perturbed datasets, we selected an aerial image set with small pose-perturbation, and a UAV image set with large pose-perturbation. The first real aerial dataset was composed of 93 images belonging to three strips, and the forwarding overlap rate was about 60%. The original images were downsampled to  $1000 \times 642$  pixels in our experiments. The stitching results corresponding to Xia et al.'s approach, ours, and PTGui are illustrated in Figure 8(a,b, and c), respectively. Because the principal axis of the aerial cameras is approximately perpendicular to the ground and the perspective distortion of this image dataset is small, the image classification automatically divides most of the images into the stable group, and only 12 images are divided into the unstable group. For this reason, affine model is used for most images during initial alignment in our approach. The initial alignment result and global optimization result of our approach are similar to Xia et al.'s results for visual effects and alignment precision, as shown in Figure 8 and in Table 1 (Urbanarea).

The perspective distortion of this image set is small, and the visual effects of these stitching results are similar. Only the perspective distortion of the PTGui's result is slightly obvious, in which slight flexion of the right part can be found, as shown in Figure 8(c). The numerical comparative analysis for this dataset is shown in Table 1 (Urbanarea), from which the Xia et al.'s approach and ours have similar aligning accuracy when the perspective distortions of all the images are small. Meanwhile, we can find that the aligning accuracy of Xia et al.'s approach and ours is superior than PTGui's, and some typical regions acquired from the result using our approach and PTGui's result are enlarged, as shown in Figure 8(d).

For more obvious evidence, the second real dataset, which was composed of 295 images, was captured by a UAV with aerial cameras with a downsampled size of  $900 \times 600$ . The forwarding overlapping rate was about 45%. In order to guarantee that the images captured by UAV are available, general practice is to obtain data



**Figure 8.** The experimental results on the first real dataset (Urbanarea): (a) the stitching result via Xia et al.'s approach XiaM2017 with auto-selected reference image  $I_{45}$  marked with a red rectangular box; (b) the stitching result using our approach with auto-selected reference image  $I_{45}$  marked with a red rectangular box; (c) the stitching result with PTGui; and (d) several enlarged regions grabbed from the result using our approach and PTGui's result shown in Figure 8(b,c).

using the UAV on a windless day as much as possible. However, the observed scene of this dataset is located in a wind power farm where there is wind all year round, which make the pose of the UAV change greatly and the original images present severe perspective distortion. The corresponding results of Xia et al.'s approach, PTGui, and ours are illustrated in Figure 9(a,b, and c), respectively. Some typical regions acquired from PTGui's result and our result are enlarged, as shown in Figure 9(d).

Compared to Xia et al.'s approach, PTGui and our approach achieve better global consistency. For example, the bottom left part and top right part of Xia et al.'s stitching result are severely compressed and the bottom right part is overstretched, as shown in



**Figure 9.** The stitching results of the second real dataset (Songnenplain, 295 images): (a) stitching result via Xia et al.'s approach with manually selected reference image  $I_{117}$  marked with a red rectangle; (b) the stitching result applying PTGui; (c) stitching result using our approach with auto-selected reference image  $I_{117}$ ; and (d) several enlarged local regions from two result in Figure 9(b,c). The green boxes indicate an image presenting a different perspective distortion after applying Xia et al.'s approach and ours.



Figure 9(a,b). Although Xia et al.'s approach manages to construct a spanning tree with less accumulated error with an optimal reference image, the edge weights of the topological graph are set based on only the number of matched features and the registration errors, which ignores the effect of the perspective distortion of the images themselves and causes the spanning tree cascading many distorted images in the middle of transferring procedure. On the other hand, the mosaicked image of our approach produces less perspective distortion than Xia et al.'s approach, as shown in Figure 9(c). The distorted images aren't selected as intermediate images through the image classification. Furthermore, during initial alignment, the unstable-group images are bundled to the stable-group images via homographies that have more DoFs, which can provide a more accurate initial solution.

As noted, the image marked with a green box presents severe perspective distortion in Figure 9 because there are only a few matching feature points in the low contrast and poor texture image with other images, and feature points are also unevenly distributed. However, the anti-perspective function in global optimization makes our approach more robust to Xia et al.'s approach. The anti-perspective function ensures that the transformed images are close to the original images after applying homographic transformation.

Compared to PTGui, the alignment precision of Xia et al.'s approach and our approach is better. From the whole alignment errors, especially the enlarged typical regions, as shown in Figure 9(c), we observed that the alignment precision of PTGui's result is worse than ours on visual effects. To demonstrate the superiority of our approach, the quantitative evaluation of the three approaches is performed based on the registration errors of the initial alignment and the final global alignment, as shown in Table 1 (Songnenplain). The results clearly show that the precision of our approach is very similar to that of Xia et al.'s approach and the global consistency of our approach outperforms Xia et al.'s approach. The global consistency of our approach is similar to PTGui's; meanwhile, the precision of our approach is superior to PTGui's.

#### 4. Conclusions

In this paper, a novel robust stitching framework for UAV image mosaicking without pose parameters is proposed. The key idea behind this framework lies in an effective projection plane selection strategy, which can resist the perspective distortion from possible existing pose-perturbed images. First, in order to construct the projection plane efficiently, we divide the sequential images into two groups, the stable group and the unstable group; and we select an optimal reference image through a weighted classification graph, which can prevent the reference image from being selected from the unstable group and maintain a global minimum accumulated registration error. Then, to achieve a better balance between global consistency and alignment precision, we propose an effective bundling strategy in which each unstable-group image is locally attached to a stable-group image via a homography, and the alignment parameters of all the stable-group images are solved with affine models. Finally, a global optimization using an anti-perspective function is employed to refine the model parameters of both groups. Our experimental results on several representative synthetic and real image datasets demonstrate that our proposed approach is superior to the state-of-the-art algorithm in terms of effectiveness and robustness. However, there are also some aspects of our framework would benefit from future work. The image classification of

the perspective distortion cannot be divided only according to the information of the images themselves, and the classified strategy is not robust because the parameters are manually set, which means that the global consistency cannot be improved much. We plan to further investigate all of these problems in our future work.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Natural Science Foundation of China [Project No. 41571436]; National Key Research and Development Program of China [Project No. 2017YFB1302400]; Hubei Province Science and Technology Support Program, China [Project No. 2015BAA027].

## References

- Brown, M., and D. G. Lowe. 2007b. "Automatic Panoramic Image Stitching Using Invariant Features." *International Journal of Computer Vision* 74 (1): 59–73. doi:10.1007/s11263-006-0002-3.
- Burt, P. J., and E. H. Adelson. 1983. "A Multiresolution Spline with Application to Image Mosaics." *ACM Transactions on Graphics* 2 (4): 217–236. doi:10.1145/245.247.
- Caballero, F., L. Merino, J. Ferruz, and A. Ollero. 2007b. "Homography Based Kalman Filter for Mosaic Building. Applications to UAV Position Estimation." In *IEEE International Conference on Robotics and Automation*, Roma, Italy, 2004–2009. doi:10.1109/ROBOT.2007.363616.
- Chang, C. H., Y. Sato, and Y. Y. Chuang. 2014b. "Shape-Preserving Half-Projective Warps for Image Stitching." In *The IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, 3254–3261. doi:10.1109/CVPR.2014.422.
- Chen, J., H. Feng, K. Pan, Z. Xu, and L. Qi. 2014b. "An Optimization Method for Registration and Mosaicking of Remote Sensing Images." *Optik-International Journal for Light and Electron Optics* 125 (2): 697–703. doi:10.1016/j.jileo.2013.07.034.
- Chen, Y., J. Sun, and G. Wang. 2010b. "Minimizing Geometric Distance by Iterative Linear Optimization." In *20th International Conference on Pattern Recognition*, Istanbul, Turkey. doi:10.1016/j.jileo.2013.07.034.
- Choe, T. E., I. Cohen, M. Lee, and G. Medioni. 2006b. "Optimal Global Mosaic Generation from Retinal Image." In *18th International Conference on Pattern Recognition*, Hong Kong, China, Vol. 3, 681–684. doi:10.1109/ICPR.2006.910.
- Elibol, A., N. Gracias, and R. Garcia. 2013b. "Fast Topology Estimation for Image Mosaicking Using Adaptive Information Thresholding." *Robotics and Autonomous Systems* 61 (2): 125–136. doi:10.1016/j.robot.2012.10.010.
- Elibol, A., N. Gracias, and R. Gracias. 2012b. *Efficient Topology Estimation for Large Scale Optical Mapping*. Vol. 82. Berlin: Springer.
- Elibol, A., R. Garca, O. Delaunoy, and N. Gracias. 2008. "A New Global Alignment Method for Feature Based Image Mosaicking." In *Advances in Visual Computing*. Berlin: Springer, 257–266. doi:10.1007/978-3-540-89646-3\_25.
- Floyd, R. W. 1962b. "Algorithm 97: Shortest Path." *Communications of the ACM* 5 (6): 345. doi:10.1145/367766.368168.
- Gao, J., S. J. Kim, and M. S. Brown. 2011b. "Constructing Image Panoramas Using Dual-Homography Warping." In *The IEEE Conference on Computer Vision and Pattern Recognition*, Colorado, USA, 49–56. doi:10.1109/CVPR.2011.5995433.

- Kang, E. Y., I. Cohen, and G. Medioni. 2000b. "A Graph-Based Global Registration for 2D Mosaics." In *15th International Conference on Pattern Recognition*, Barcelona, Spain, Vol. 1, 257–260. doi:[10.1109/ICPR.2000.905314](https://doi.org/10.1109/ICPR.2000.905314).
- Kang, L., W. Lingda, Y. Wei, B. Yang, and H. Song. 2014b. "A Highly Accurate Dense Approach for Homography Estimation Using Modified Differential Evolution." *Engineering Applications of Artificial Intelligence* 31: 68–77. doi:[10.1016/j.engappai.2013.11.015](https://doi.org/10.1016/j.engappai.2013.11.015).
- Konolige, K., and W. Garage. 2010b. "Sparse Sparse Bundle Adjustment." In *The British Machine Vision Conference*, Aberystwyth, UK, Vol. 10, 1–11. doi:[10.5244/C.24.102](https://doi.org/10.5244/C.24.102).
- Koo, H. I., B. S. Kim, and N. I. Cho. 2009b. "A New Method to Find an Optimal Warping Function in Image Stitching." In *IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, 1289–1292. doi:[10.1109/ICASSP.2009.4959827](https://doi.org/10.1109/ICASSP.2009.4959827).
- Kwatra, V., A. Schödl, I. Essa, G. Turk, and A. Bobick. 2003b. "Graphcut Textures: Image and Video Synthesis Using Graph Cuts." In *ACM Transactions on Graphics (Tog)*, ACM New York, USA, Vol. 22. doi:[10.1145/882262.882264](https://doi.org/10.1145/882262.882264).
- Li, M., L. Deren, and D. Fan. 2012b. "A Study on Automatic UAV Image Mosaic Method for Paroxysmal Disaster." In *The International Society of Photogrammetry and Remote Sensing Congress*, Melbourne, Australia, 123–128. Vol. 25. doi:[10.5194/isprsarchives-XXXIX-B6-123-2012](https://doi.org/10.5194/isprsarchives-XXXIX-B6-123-2012)
- Li, S., L. Yuan, J. Sun, and L. Quan. 2015b. "Dual-Feature Warping-Based Motion Model Estimation." In *The IEEE International Conference on Computer Vision*, Washington, USA, 4283–4291. doi:[10.1109/ICCV.2015.487](https://doi.org/10.1109/ICCV.2015.487).
- Lin, C. C., S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin. 2015b. "Adaptive As-Natural-As-Possible Image Stitching." In *The IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, 1155–1163. doi:[10.1109/CVPR.2015.7298719](https://doi.org/10.1109/CVPR.2015.7298719).
- Lin, W. Y., S. Liu, Y. Matsushita, T. T. Ng, and L. F. Cheong. 2011b. "Smoothly Varying Affine Stitching." In *Computer Vision and Pattern Recognition*, Colorado, USA, 345–352. doi:[10.1109/CVPR.2011.5995314](https://doi.org/10.1109/CVPR.2011.5995314).
- Lin, Y., and G. Medioni. 2007b. "Map-Enhanced UAV Image Sequence Registration and Synchronization of Multiple Image Sequences." In *Computer Vision and Pattern Recognition*, Minneapolis, USA, 1–7. doi:[10.1109/CVPR.2007.383428](https://doi.org/10.1109/CVPR.2007.383428).
- Liu, F., M. Gleicher, H. Jin, and A. Agarwala. 2009b. "Content-Preserving Warps for 3D Video Stabilization." *ACM Transactions on Graphics* 28 (3): 44. doi:[10.1145/1531326](https://doi.org/10.1145/1531326).
- Marzotto, R., A. Fusiello, and V. Murino. 2004b. "High Resolution Video Mosaicking with Global Alignment." In *The IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, USA, Vol. 1. doi:[10.1109/CVPR.2004.1315099](https://doi.org/10.1109/CVPR.2004.1315099).
- Mou, W., H. Wang, G. Seet, and L. Zhou. 2013b. "Robust Homography Estimation Based on Non-Linear Least Squares Optimization." In *IEEE International Conference on Robotics and Biomimetics*, Shenzhen, China, 372–377. doi:[10.1109/ROBIO.2013.6739487](https://doi.org/10.1109/ROBIO.2013.6739487).
- Se, S., P. Firoozfam, N. Goldstein, W. Linda, M. Dutkiewicz, P. Pace, and J. Pierre Naud. 2009b. "Automated UAV-based Mapping for Airborne Reconnaissance and Video Exploitation." In *International Society for Optics and Photonics on Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications*, Spie defense, security, and sensing, 73070M. Vol. 7307. doi:[10.1117/12.820225](https://doi.org/10.1117/12.820225).
- Szeliski, R. 2006. "Image Alignment and Stitching: A Tutorial." *Foundations and Trends in Computer Graphics and Vision* 2 (1): 1–104. doi:[10.1561/06000000009](https://doi.org/10.1561/06000000009).
- Szeliski, R., and H. Y. Shum. 1997b. "Creating Full View Panoramic Image Mosaics and Environment Maps." In *The 24th Conference on Computer Graphics and Interactive Techniques*, Los Angeles, USA, 251–258. doi:[10.1145/258734.258861](https://doi.org/10.1145/258734.258861).
- Triggs, B., A. Zisserman, and R. Szeliski, eds. 2003b. *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms*. Corfu, Greece: Springer. doi:[10.1007/3-540-44480-7](https://doi.org/10.1007/3-540-44480-7).
- Wang, Z., Y. Chen, Z. Zhu, and W. Zhao. 2016b. "An Automatic Panoramic Image Mosaic Method Based on Graph Model." *Multimedia Tools and Applications* 75 (5): 2725–2740. doi:[10.1007/s11042-015-2619-0](https://doi.org/10.1007/s11042-015-2619-0).

- Xia, M., J. Yao, L. Li, and L. Xiaohu. 2015b. "Globally Consistent Alignment for Mosaicking Aerial Images." In *IEEE International Conference on Image Processing*, Quebec, Canada, 3039–3043. doi:[10.1109/ICIP.2015.7351361](https://doi.org/10.1109/ICIP.2015.7351361).
- Xia, M., J. Yao, R. Xie, L. Li, and W. Zhang. 2017b. "Globally Consistent Alignment for Planar Mosaicking via Topology Analysis." In *Pattern Recognition*, Elsevier, 239–252. Vol. 66 (C). doi:[10.1016/j.patcog.2017.01.020](https://doi.org/10.1016/j.patcog.2017.01.020).
- Xing, C., J. Wang, and X. Yaming. 2010b. "A Robust Method for Mosaicking Sequence Images Obtained from UAV." In *Information Engineering and Computer Science*, Wuhan, China, 1–4. doi:[10.1109/ICIECS.2010.5678358](https://doi.org/10.1109/ICIECS.2010.5678358).
- Xu, Y., O. Jianliang, H. Hu, X. Zhang, and J. Mills. 2016b. "Mosaicking of Unmanned Aerial Vehicle Imagery in the Absence of Camera Poses." *Remote Sensing* 8 (3): 204. doi:[10.3390/rs8030204](https://doi.org/10.3390/rs8030204).
- Yahyanejad, S., D. Wischounig-Strucl, M. Quaritsch, and B. Rinner. 2010b. "Incremental Mosaicking of Images from Autonomous, Small-Scale UAVs." In *The 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, Boston, USA, 329–336. doi:[10.1109/AVSS.2010.14](https://doi.org/10.1109/AVSS.2010.14).
- Zagrouba, E., W. Barhoumi, and S. Amri. 2009b. "An Efficient Image-Mosaicing Method Based on Multifeature Matching." *Machine Vision and Applications* 20 (3): 139–162. doi:[10.1007/s00138-007-0114-y](https://doi.org/10.1007/s00138-007-0114-y).
- Zaragoza, J., T. J. Chin, M. S. Brown, and D. Suter. 2013b. "As-Projective-As-Possible Image Stitching with Moving DLT." In *The IEEE Conference on Computer Vision and Pattern Recognition*, Portland, Oregon, 2339–2346. doi:[10.1109/CVPR.2013.303](https://doi.org/10.1109/CVPR.2013.303).
- Zhang, F., and F. Liu. 2014b. "Parallax-Tolerant Image Stitching." In *The IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, 3262–3269. doi:[10.1109/CVPR.2014.423](https://doi.org/10.1109/CVPR.2014.423).
- Zitova, B., and J. Flusser. 2003b. "Image Registration Methods: A Survey." *Image and Vision Computing* 21 (11): 977–1000. doi:[10.1016/S0262-8856\(03\)00137-9](https://doi.org/10.1016/S0262-8856(03)00137-9).