UAV Image Stitching Based on Mesh-Guided Deformation and Ground Constraint

Quan Xu, Jun Chen, Linbo Luo, Wenping Gong, and Yong Wang

***Abstract*—This article introduces a drone image stitching based on mesh-guided deformation and ground constraint, which can closely match the characteristics of images and achieve precise registration and acquire ideal stitching effect. The traditional methods use the homography model to align the image, which causes artifacts in the result of stitching the images with parallax. To overcome this situation, the image is divided into meshes and the mesh vertices of the target image are used to guide the warping. A new energy function is designed to represent the deformation characteristics of the image. We propose a new alignment term by using local homography and a local scale term by using the edge information of the mesh. The established mesh-guided deformation model can overcome image parallax caused by some external factors and eliminate the ghostly parts of the result. Moreover, imaged scene is not effectively planar and some fluctuations exist in the scene of the images, which will distort the stitching result. We propose a ground constraint with the ground plane as the main plane to reduce projection distortions in non-overlapping areas between images. Finally, the method of creating groundtruth is proposed, which can evaluate the naturalness of results and make comparison more reasonable. Several sets of challenging drone images are tested, and the experimental results show that our stitching system has good results.**

***Index Terms*—Ground constraint, image matching, mesh-guided deformation, unmanned aerial vehicle (UAV) image stitching.**

I. Introduction

FROM the initial unmanned "balloon" to the sophisticated model, we have today, drone technology has made great progress. Because of its small size, low cost, convenient use, and low environmental requirements, unmanned aerial vehicle (UAV) is widely used and applied in various fields, such as environmental monitoring, military mapping, aerospace, and commercial fields [1], [2]. However, covering the complete target area is difficult due to the limitation of the single viewing angle [3]. To cover the complete target area, multiple images need to stitched together. Image stitching is an effective technique to achieve this function [4]–[8]. In addition, image stitching system is used in many tasks in our life, including remote sensing image processing [9]–[11], resource and environmental monitoring [12], [13], and so on.

There are often two main methods for producing image stitching with satisfactory visual effects [14]: 1) better alignment model, 2) using image synthesis techniques such as seam cutting [15], [16], Laplacian pyramid blending [17], [18], and Poisson blending [19]. Advanced image synthesis techniques can beautify the stitching result in the later stage. However, these methods cannot solve this situation where significant misalignment occurs in the image alignment phase. Here, we aim to design a robust alignment model to obtain a precise alignment of the image. Image matching [20], [21] is the key to this stitching system. All input source images need to be accurately matched to the same coordinate system, which directly affects the quality of the stitching.

Homography is often used to describe the relationship between two images in a three-dimensional (3-D) plane [22], as a strict alignment model. Autostitch [17] is a representative method of using homography model. The classic image stitching algorithm includes feature matching, image matching, bundle adjustment, automatic panorama straightening, gain compensation, and multiband blending [18]. The SIFT [23] feature points are detected to establish a point matching relationship between images. This algorithm uses a global homography to describe the model between the images in order to align the images. One image is automatically selected as the reference image when stitching multiple images, which has the most overlapping area with other images. The homography deformation is used to project other images onto the selected reference image. The error is inevitable during the stitching process. The effect is not obvious when stitching two images. However, these errors will be accumulated and the effect will appear when stitching multiple images. Brown and Lowe [17] took this into account and use bundle adjustment [24] to eliminate cumulative errors. This is a globally optimized way to get an optimal solution by minimizing the total projection error to get better results. This article also points out that only when there is no parallax in the image (the scene is close to the plane), the global method is applicable. Some significant ghostly parts occur in the stitching result when UAV images are aligned by homography model.

Some spatial transformation models [25]–[30] for image stitching have been proposed in recent years in order to

Manuscript received December 30, 2020; revised February 8, 2021; accepted February 19, 2021. Date of publication February 23, 2021; date of current version May 14, 2021. This work was supported by the National Natural Science Foundation of China under Grant 62073304, Grant 41977242, and Grant 61973283. *(Corresponding author: Jun Chen.)*

Quan Xu, Linbo Luo, and Yong Wang are with the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China (e-mail: xuquan950519@cug.edu.cn; luolb@ige-live.com; yongwang_cug@163.com).

Jun Chen is with the School of Automation, Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, China University of Geosciences, Wuhan 430074, China (e-mail: chenjun71983@163.com).

Wenping Gong is with the Faculty of Engineering, China University of Geosciences, Wuhan 430074, China (e-mail: wenpinggong@cug.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3061505

obtain good alignment results with parallax. As-projective-as-projective warp (APAP) is representative method of spatial transformation models. It meshes the image and computes some local warps for each small grid to solve the local mismatch in the image. The model can solve the problem that the deformation of the global homography model is insufficient, and the overlapping area in these images can be well aligned. However, projection distortion is caused in the nonoverlapping area, which affects the overall naturalness of the result.

For UAV images we studied, it has some features. UAV is small in size and is susceptible to external influences during flight. The camera cannot be adjusted in time when a slight jitter occurs in UAV, and the acquired image will definitely be affected. The most important problem is the image parallax. Moreover, different depth differences are present in the images due to the undulation of the ground. Some ghosting and distortion will be brought due to the influence of these characteristics when the UAV images are stitched by the previously introduced method.

Mesh-guided deformation is used to solve this situation, i.e., the image parallax. The local projection model has high degree of freedom and is flexible, which can handle local transformation and reduce artifacts. The image is meshed and the vertices of the mesh are used to guide the warping. A new alignment term is proposed by using local homography, which improves the initial matching ability. We use the edge information of the mesh to propose a local scale term, which further constrains the scene structure of the image. A new energy function is designed to specify the required deformation characteristics. The deformation of all images is solved to obtain an optimal solution and it can be solved efficiently in a sparse linear system. The form of the optimal solution is the result of mesh deformation, which can effectively eliminate local ghostly parts. On the basis of the UAV image characteristic that the ground area of the scene is the main plane, we propose a ground constraint by using a global matrix, which can alleviate the distortion of nonoverlapping area. These are crucial to the naturalness of the result. Finally, groundtruth is created by the method we proposed. For the ghostly parts in result, we can easily evaluate the results, but we cannot judge the distorted parts directly. Qualitative analysis is not reliable for this problem. Usually the scene structure in the source image is the natural standard, so we can artificially create a groundtruth to quantitatively compare with our results. The height of the UAV is adjusted to get a wider view of the area and take this large field of view image as our groundtruth. We do not judge the degree of distortion of the results based on feelings. A quantitative indicator is used to compare the difference between the results and groundtruth. This is also a contribution point for us.

There are several contributions to this article.

1) A mesh-guided deformation is used to warp UAV image, where a new alignment term and a local scale term are proposed and a new energy function is designed to specify the required deformation characteristics.
2) We propose a ground constraint with the ground plane as the main plane to reduce projection distortions of nonoverlapping area.
3) In order to better evaluate the distortion of the results, we propose a method to create groundtruth.

II. Related Work

Szeliski *et al.* [31] gave a detailed overview of the development of image stitching. Some parametric models are usually used to stitch images, such as affine deformation, similar deformation, and projection deformation. Autostitch [17] is a classic method of using the global projection model, which assumes that the camera's motion only contains 3-D rotation. A 2-D image is obtained by mapping a view sphere to an image plane by projection transformation. Its assumptions make it difficult for this method to meet various situations in reality.

A lot of work have been done to obtain a more robust stitching algorithm in the following three directions: 1) seam-driven stitching [15], [32]–[34]; 2) content-preserving warps [29], [32], [35]and; 3) local adaptive transformations [25]–[30]. Gao *et al.* [26] proposed a dual-homography warping method, which uses two homographies to match the two planes of an image scene, respectively. This method can seamlessly stitch some real scenes. But when the scenes are slightly more complex, this method fails. Lin *et al.* [27] proposed a smoothly varying affine model that uses multiple affine transformations to stitch images, which has stronger local deformation and alignment capabilities, and has a certain ability to handle parallax. But the degrees of freedom are not enough to represent a perspective transformation.

An as-possible-as-projection model was developed by Zaragoza [25], which divides the image into dense grids and each grid represents a local deformation. This local model is highly flexible and can handle image parallax. But when the number of detected feature points in the image decreases, this method will lose its superiority. In addition, due to the limitations of the model, it will also cause the projection distortion in the nonoverlapping area of the result. Some methods have been proposed to eliminate the projection distortion of non-overlapping area. Chang *et al.* [29] proposed shape-preserving half-projective. The method adds constraints and applies global similarity to change the appearance of the image. However, the similarity deformation method cannot solve the problem of image parallax. Lin *et al.* [30] improved the APAP algorithm and proposed adaptive as natural as possible (AANAP). This method uses t-distribution instead of Gaussian distribution in calculating the local homography, which reduces the difficulty of adjusting model parameters. In addition, it calculates the optimal similarity matrix of the image, and combines the local homography to smooth the image and eliminate the projection distortion of the overlapping area. Zhang and Liu. [32] combined the techniques of seam-selection and content-preserving warping and proposed a stitching method that can handle scenes with large parallax. Considering that when there are fewer feature points detected in low-texture images, it can not guarantee that the calculated local homography has a good alignment effect, Xia *et al.* added line features when calculating the local homography to make the local alignment ability stronger. Li *et al.* [36] proposed

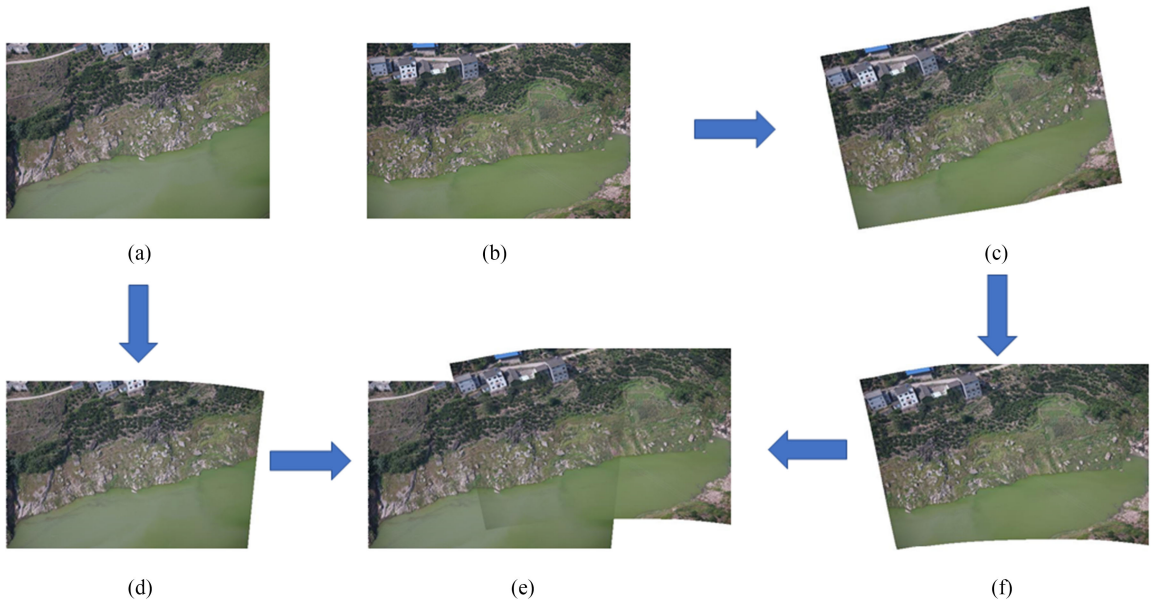

Fig. 1. Image stitching combined with mesh-guided deformation and ground constraint. (a) and (b) are the input images. (b) is warped using mesh-guided deformation to produce the warped image (c). (c) is then transformed using ground constraint to get (f). Meanwhile, (a) is transformed using ground constraint to get (d). The result image (e) is synthesized by (d) and (f) at last.).

a parallax-tolerant image stitching method based on robust elastic warping. It calculates the deformation function of point matching on the image plane and directly deforms the image. In addition, by analyzing the law of deformation model parameters, a feature refinement model based on Bayesian theory is proposed to remove mismatched points in image matching. This method can achieve accurate alignment and efficient processing simultaneously. In addition, some nonparametric deformation models [37]–[40] have also been used for image stitching, but they are limited to stitching two images. Because when stitching multiple images, the accumulated errors need to be dealt with by the bundle adjustment of the parameter model. In addition, some real-time UAV image mosaicing methods are also proposed [41].

III. Proposed Method

This section details the algorithm we propose, including local direct linear transformation (DLT) [42], mesh-guided deformation and ground constraint. Fig. 1 demonstrates the overall flow of the algorithm. Fig. 1(a) and (b) are reference image and target image. In the alignment phase, the mesh vertices are used for effective registration, and we optimize the designed energy function for mesh deformation. Fig. 1(b) is deformed by the mesh-guided deformation and becomes Fig. 1(c). Ground constraint are used to eliminate distortion in nonoverlapping area. Fig. 1(a) and (c) become (d) and (f) after ground constraint. Finally, Fig. 1(d) and (f) are merged into Fig. 1(e). In the input of two images, one is the reference image and the other is the target image. Only target image undergoes the first stage of warping with mesh-guided deformation. The proposed stitching algorithm is suitable for the characteristics of UAV images, and achieves accurate and natural image stitching. In addition, we also extend the algorithm to stitching multiple images.

A. Local DLT

In the feature registration phase, we first use the local DLT to get the matching points we need. Given the input images I and I' and the SIFT matching point $\mathbf{x}_i = (x_i, y_i)^T, \mathbf{y}_i = (x_i', y_i')^T, i = 1, \ldots, N$, the homographic transformation between two images $\mathbf{y} = \mathbf{h}(\mathbf{x})$ can be represented as

$$\mathbf{h}_x(\mathbf{x}) = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + h_9} \tag{1}$$

$$\mathbf{h}_y(\mathbf{x}) = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + h_9}. \tag{2}$$

It can be estimated by the relation

$$\widetilde{\mathbf{y}} \sim \mathbf{H}\widetilde{\mathbf{x}} \tag{3}$$

where $\widetilde{\mathbf{x}}$ denotes $\mathbf{x}$ in homogeneous coordinates, $\sim$ indicates the equality up to scale, and $\mathbf{H}$ is the global homography that is a 3×3 matrix. The rows of $\mathbf{H}$ are given by $\mathbf{h}_1 = [h_1\ h_2\ h_3]$, $\mathbf{h}_2 = [h_4\ h_5\ h_6]$, $\mathbf{h}_3 = [h_7\ h_8\ h_9]$. Taking a cross product on both sides of (3), we obtain

$$\mathbf{0}_{3\times 1} = \widetilde{\mathbf{y}} \times \mathbf{H}\widetilde{\mathbf{x}} \tag{4}$$

which can be rewritten as follows:

$$\mathbf{0}_{3\times 1} = \begin{bmatrix} \mathbf{0_{1\times 3}} & -\tilde{\mathbf{x}}^T & y'\tilde{\mathbf{x}}^T \\ \tilde{\mathbf{x}}^T & \mathbf{0}_{1\times 3} & -x'\tilde{\mathbf{x}}^T \\ -y'\tilde{\mathbf{x}}^T & x'\tilde{\mathbf{x}}^T & \mathbf{0}_{1\times 3} \end{bmatrix} \mathbf{h}, \mathbf{h} = \begin{bmatrix} \mathbf{h_1^T} \\ \mathbf{h}_2^T \\ \mathbf{h}_3^T \end{bmatrix}. \tag{5}$$

Only two of the rows are linearly independent and we will denote the 9×1 vector in (5) as $\mathbf{h}$.

Direct linear transformation is an effective way to calculate $\mathbf{H}$ from some sample data. We vectorize $\mathbf{H}$ into a vector $\mathbf{h}$ and $\mathbf{a}_i$ represent first-two rows of the LHS matrix in (5) computed for the ith point match $\{\mathbf{x}_i, \mathbf{y}_i\}$. Given an estimate $\mathbf{h}$, the quantity

$\|\mathbf{a}_i\mathbf{h}\|$ is the algebraic error. DLT minimizes the sum of squared algebraic errors

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \sum_{i=1}^{N} \|\mathbf{a}_i\mathbf{h}\|^2. \tag{6}$$

By stacking vertically $\mathbf{a}_i$ for all i into matrix $\mathbf{A}$ of size $2N \times 9$, the problem can be rewritten as

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \sum_{i=1}^{N} \|\mathbf{A}\mathbf{h}\|^2. \tag{7}$$

The solution is the least significant right singular vector of $\mathbf{A}$. After $\mathbf{H}$ is obtained, an arbitrary pixel $\mathbf{x}_*$ in the source image I can be projected to the position $\mathbf{y}_*$ in the target image I' by

$$\widetilde{\mathbf{y}}_* \sim \mathbf{H}\widetilde{\mathbf{x}}_*. \tag{8}$$

The UAV images have parallax and some nonrigid local distortion, and a global $\mathbf{H}$ cannot deform the image well. Therefore, a model based on local deformation is introduced by

$$\widetilde{\mathbf{y}}_* \sim \mathbf{H}_*\widetilde{\mathbf{x}}_*. \tag{9}$$

Each pixel $\mathbf{x}_*$ corresponds to a location-dependent homography $\mathbf{H}_*$, where $\mathbf{x}_*$ is estimated from the weighted problem

$$\mathbf{h}_* = \arg\min_{\mathbf{h}} \textstyle\sum_{i=1}^{N} \|w_*^i\mathbf{a}_i\mathbf{h}\|^2. \tag{10}$$

We can define the scalar weights $\{w_*^i\}_{i=1}^N$ as

$$w_*^i = \exp(-\|\mathbf{x}_* - \mathbf{x}_i\|^2/\delta^2) \tag{11}$$

where δ is a scale parameter and we set δ as 12. The exact solution steps are the same as before.

B. Mesh-Guided Deformation

In the previous step, we obtain the alignment relationship between I and I'. Note that local homography method is a mesh-based method and the image is meshed in the feature point matching stage. We establish a point set $M^{II'}$, which represents the mesh vertices of $I's$ in the overlapping part of I and I'. For each matching point in $M^{II'}$, we know its corresponding position in I', because the two images have been aligned in the previous step. We do not use feature points in the mesh-guided deformation phase, because the feature points of some scenes may not be evenly distributed, but only in specific parts. We use matching points to make full use of the image for more information.

We assume that I is the reference image and the matching points in the overlap of I and I' are $\mathbf{m}_i = (m_i, n_i)^T$, $i = 1, \ldots, k$. Matching points belong to $M^{II'}$ and the corresponding position on I' are $\mathbf{n}_i = (m_i', n_i')^T$, $i = 1, \ldots, k$. We have divided the target image I' into a series of squares in local DLT. The mesh vertices of the target image are used to guide the image deformation in this step. Let V_i denotes the set of vertices in the squares for the target image I'. $\mathbf{V}$ denotes the set of all vertices. Alignment refinement is expressed by minimizing the energy function to obtain a set of deformed vertices $\widetilde{\mathbf{V}}$. $\mathbf{n}_i$ can be represented by a linear combination of four mesh vertices $V = [V_1, V_2, V_3, V_4]^T$ in its locating quad: $\mathbf{n}_i = w^T V$, and $w = [w_1, w_2, w_3, w_4]^T$ are the corresponding bilinear weights [43]. We represent the deformation problem of the image as mesh-guided deformation. The purpose is to align the predeformed image with the reference image while making the overall image look more natural. Regarding the definition of nature, we assume that the original image is natural to the user. Therefore, we need to try to ensure the original view of each image as much as possible. For this reason, we design three items, alignment term, local scale term, and local rotation term in the energy function.

Alignment term E_a: It is used to make the matching points consistent with the corresponding points. This guarantees the quality of the alignment after deformation without ghosting. It is defined as follows:

$$E_a = \sum_{i=1}^{k} \|\widetilde{\mathbf{n}}_i - \widetilde{\mathbf{m}}_i\|^2 \tag{12}$$

where $\widetilde{\mathbf{m}}_i$ is the projection position that $\mathbf{m}_i$ of reference image I are projected to target image I' by local homography $\mathbf{H}_i$. $\widetilde{\mathbf{n}}_i = w^T\widetilde{V}$ is the corresponding point on the deformed target image. $\widetilde{V}$ represent the four vertices of a grid containing $\widetilde{\mathbf{n}}_i$ and w are the corresponding bilinear weights.

Local rotation term E_r: It is mainly to make the target image maintain the similarity of its mesh in the deformation, and avoid shape distortion as much as possible. We split each uniform square into two triangles and then apply the method of Igarashi *et al.* [44]. Given a triangle $\triangle V_1V_2V_3$, each vertex can be represented by the vector between the other two vertices. For example, V_1 can be defined using V_2 and V_3 as

$$V_1 = V_2 + \mu(V_3 - V_2) + \nu R(V_3 - V_2), R = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{13}$$

where μ and ν are the coordinates of V_1 in the local coordinated system defined by other two vertices. In the deformation, the correlation of the three coordinates is constrained by a similarity to avoid local rotation distortion. Its position in the local coordinate system is unchanged. Therefore, the local rotation term is

$$E_r(\widetilde{V}_1) = \tau\|\widetilde{V}_1 - (\widetilde{V}_2 + \mu(\widetilde{V}_3 - \widetilde{V}_2) + \nu R(\widetilde{V}_3 - \widetilde{V}_2))\|^2 \tag{14}$$

where τ is a weight used to measure the salience of the triangle as in [32] and [45]. We use this weighting strategy to distribute more distortion to less salient regions than those salient ones. This energy term E_r can be obtained by summation (10) over all the vertices.

Local scale term E_s: The previous one guarantees the similarity of the triangles, but the scale may become correspondingly larger or smaller. In this term, we use the edge information of the mesh to ensure that the length after deformation is not stretched or shrunk too much. Under this constraint, the scale is basically the same. The shape of the quadrilateral will not be distorted too much. It is defined as follows:

$$E_s(V) = \sum_{i=1}^{4}\sum_{j=1}^{4} \|(\widetilde{V}_i - \widetilde{V}_j) - (V_i - V_j)\|^2, i \neq j \tag{15}$$

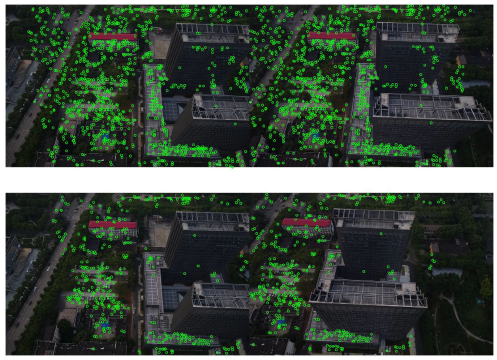

Fig. 2. Distribution of matched features. The first line is feature points detection of two images. The second line is the distribution of matched features.

where $E_s(V)$ is the energy term of a cell. This energy term E_s can be obtained by summation (11) over all the cells.

We combine these three terms into the following energy function and then minimize the energy function:

$$E = E_a + \alpha E_r + \beta E_s \tag{16}$$

where α and β are the weight assigned to local rotation term and local scale term, and their values are set to 0.001 and 0.01. The minimization problem can be solved using a standard sparse linear solver.

C. Ground Constraint

We introduce a mesh-guided deformation to eliminate ghosting in the overlapping area in the previous section. The projection distortions still occur in the nonoverlapping parts. In this section, we propose an approach to eliminate the projection distortions. The solution is to apply the matrix associated to the global similarity transform in the target image. The main area of the drone image scene is the ground plane, so we choose the deformation matrix of the ground plane to globally optimize the deformation of the entire image. We use this feature to design a ground constraint.

Let us give an example to illustrate our intention. Fig. 2 shows the feature points detection of two images to be stitched together. The scene can be divided into different planes by these matched feature points, such as the roofs or facades of the building, the ground and so on. As shown from the density of feature points in the figure, most of them are basically distributed on the ground, and the feature points on the high-rise objects appear sparse, and many feature points in the building area are removed as false match points. We explain the reasons for extracting the ground of the scene as a ground constraint from two aspects. One is that a large difference is present in the image of the object between the viewing angle characteristics, which increases the difficulty of matching. The other is that the area of the ground area is larger than any plane in the image.

The literature [29] shows that similarity transformation can effectively reduce distortions. This method uses all feature points to find global similarity transform, which can lead to nonoptimal solutions, especially when regions of image contain different planes. We combine this method with the homography to weight the image to solve the situation. First, a matrix $\mathbf{S}$ associated to the global similarity transform can be obtained by the principal ground plane and then $\mathbf{S}$ is applied to optimize the deformation of the target image to eliminate projection distortion. In addition, in order to smooth the entire image, a weighting strategy to combine the homography is used and the matrix $\mathbf{S}$ to make the image appear as a whole. We need to weight the entire image so that the overlapping and nonoverlapping parts are smoothly transitioned as follows:

$$\mathbf{H}_i' = \iota \mathbf{H}_i + \kappa \mathbf{S} \tag{17}$$

where $\mathbf{H}_i$ is ith local homography, $\mathbf{H}_i'$ is local transformation after the formula update. $\mathbf{S}$ is the global similarity transformation of the principal ground plane we obtained. ι and κ are weighting coefficients with $\iota + \kappa = 1$, where they are between 0 and 1.

The literature [46] points out that the geometric distortions of the homography occur on the axis of rotation on both images. If we calculate the weighting coefficients along this axis of rotation, we can provide a gradual change from projection transformation to similar transformation over the entire mosaic image to maintain the natural effects of nonoverlapping regions. We connect the centers of the two images as a rotation axis. The calculation of ι and κ can refer to literature [47].

A matrix $\mathbf{S}$ is applied to the target image, which results in misalignment of overlapping regions of the previously aligned reference image and the target image. Therefore, this corresponding deformation process should also be used for the reference image to compensate for the transformation. We can get a local transformation of the reference image

$$\mathbf{T}_i' = \mathbf{H}_i' \mathbf{H}_i^{-1} \tag{18}$$

where $\mathbf{T}_i'$ is ith updated local transformation in reference image.

We summarize our image stitching algorithm based on mesh-guided deformation and ground constraint in Algorithm 1.

Algorithm 1: Image Stitching Based on Mesh-Guided Deformation and Ground Constraint.

Input: Images I , $I^{'}$ and their matching points $\mathbf{n}_i$, $\mathbf{m}_i$;
Output: Updated location of matching points $\mathbf{n}_i$, $\mathbf{m}_i$;
for i = *1, ..., N* **do**
  Compute the local homography $\mathbf{H}_i$ by (9);
  Compute the projection of $\mathbf{m}_i$ in I to $I^{'}$ by $\mathbf{H}_i$, denote it as $\widetilde{\mathbf{m}}_i$;
end
Formulate (16) based on the matches $\widetilde{\mathbf{m}}_i$,$\mathbf{n}_i$;
Solve (16) to obtain the local alignment;
Update the location of matching points by ground constraint;
for i = *1, ..., N* **do**
  Recalculate the location of $\widetilde{\mathbf{n}}_i$ in I by (17);
  Recalculate the location of $\mathbf{m}_i$ in $I^{'}$ by (18);
end

Algorithm 2: Simultaneous Refinement of Multiple Image Based on Mesh-Guided Deformation and Ground Constraint

Input: Images $\{I^k\}_{k=1}^K$ with overlaps;
Output: Updated location of matching points;
Choose target image I^t from $\{I^k\}_{k=1}^K$;
Built the matching point sets P_i and $P_i^{'}$;
for k = *1, ..., K* **do**
  for i = *1, ..., N* **do**
    Compute the local homography $\mathbf{H}_i$ by (9);
    Compute the projection of $P_i^{'}$ in other images to I^t by $\mathbf{H}_i$;
  end
end
Formulate (16) based on the matches P_i and $P_i^{'}$;
Solve (16) to obtain the local alignment;
Recalculate the location of matching points by ground constraint;
for k = *1, ..., K* **do**
  for i = *1, ..., N* **do**
    Recalculate the location of P_i in I^t by (17);
    Recalculate the location of $P_i^{'}$ in other images by (18);
  end
end

D. Multiple Image Stitching

The proposed stitching model is carried out by deforming the image through the energy function, which is not related to the projection model. Therefore, the overall optimization technology bundle adjustment based on the matrix parameter model is not suitable for the proposed model. In fact, the incremental stitching scheme is powerful enough to provide satisfying results with the help of the mesh-guided deformation and ground constraint. Because the input image are incrementally aligned by mesh optimization and merged into a common image plane. What we need to notice is that the input image overlaps with multiple images in mesh-guided deformation. So the vertices of the input image may have a matching relationship with multiple images. While aligning the current image, these matches must be processed at the same time.

Enter some overlapping images $\{I^k\}_{k=1}^K$, the first is to find target image I^t that the image has the most overlap with other images. We use the techniques introduced in [14] to determine the target image I^t through the matching correspondence of key points, which has the most matching correspondence with other images. For image I^t, a pair of sets P_i and P_i' are allocated to collect the matching points. P_i are vertices of I^t. P_i' are the matching points related to I^t in other images. After all alignment images are traversed, the point sets P_i and P_i' are built, and these contain all point matches between I^t and the aligned images. Then, the matches are simultaneously refined by mesh-guided deformation. We summarize the entire procedure in Algorithm 2.

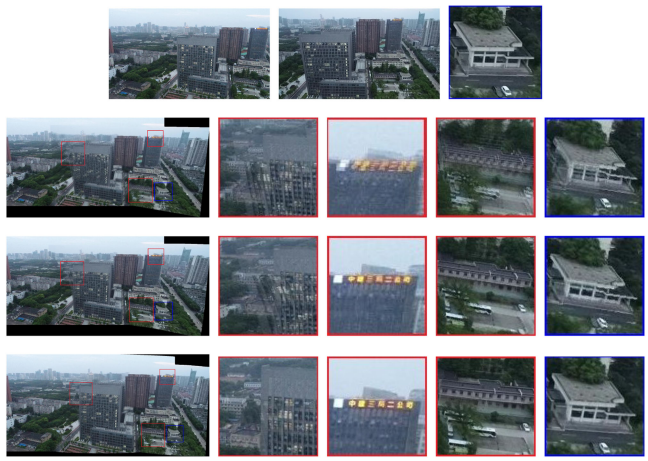

Fig. 3. Challenges on UAV images stitching. The first line is the input source image. Blue box of first line shows parts of the groundtruth. The second, third and fourth lines are the results of global homography [17], APAP [25] and our method. Red boxes highlight the ghostly parts. Blue boxes highlight distorted parts.

IV. Experimental Results

This part introduces the experiments performed by the proposed algorithm on some challenging UAV images and compares them with current algorithm. In our experiments, the test images are taken from outside by UAV and contain different scenes, which basically reflects their characteristics. We consider the experimental setup from the following aspects to illustrate that each of our system is indeed effective: 1) mesh-guided deformation and global homography and local homography comparison; 2) the effect of ground constraint; and 3) stitching multiple images.

In order to briefly illustrate the robustness of our method, We first conduct experiments on two images and finally extend the experiments to multiple images. In these experiments, we use the groundtruth we obtained to judge the effect of different methods of stitching.

A. Mesh-Guided Deformation

Some experiments are done to evaluate the effect of mesh-guided deformation. The goal of mesh-guided deformation is to eliminate ghosting. We can directly judge the quality of different methods from the results.

A comparative experiment on global homography [17], APAP [25] and our method is demonstrated in Fig. 3. In addition, several typical areas are highlighted in the box. The first line is the input source image. The second line shows the results of global homography. Ghosting occurs in many areas because a global homography cannot handle image parallax. In a high-building area, the parallax between the two source images is too large or enough feature points are not detected in this area, so that the fracture occurs. The result of APAP is displayed in the third line. This model introduces a local matrix, which has a stronger matching ability than a homography and can handle

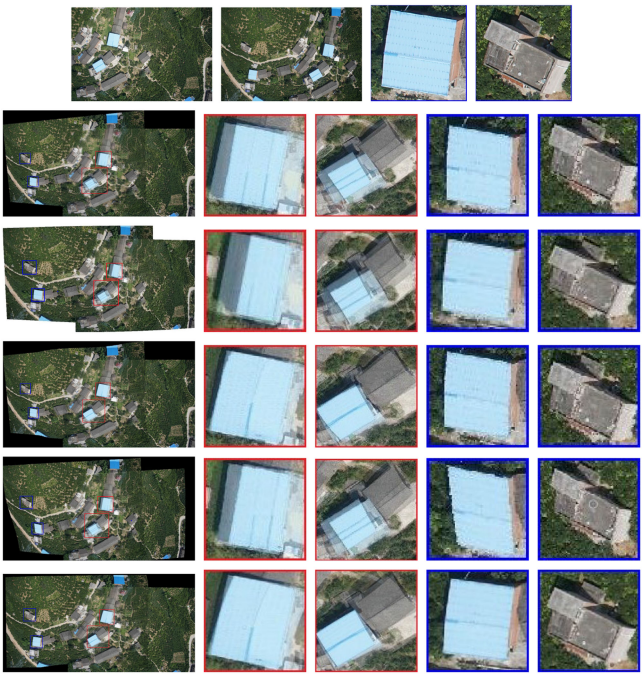

Fig. 4. Stitching results among various methods. The first line is the input source image. Blue boxes of first line show parts of the groundtruth. The second, third, fourth, fifth, and sixth lines are the results of global homography, SPHP, APAP, REW, and our method. Red boxes highlight the ghostly parts. Blue boxes highlight distorted parts.

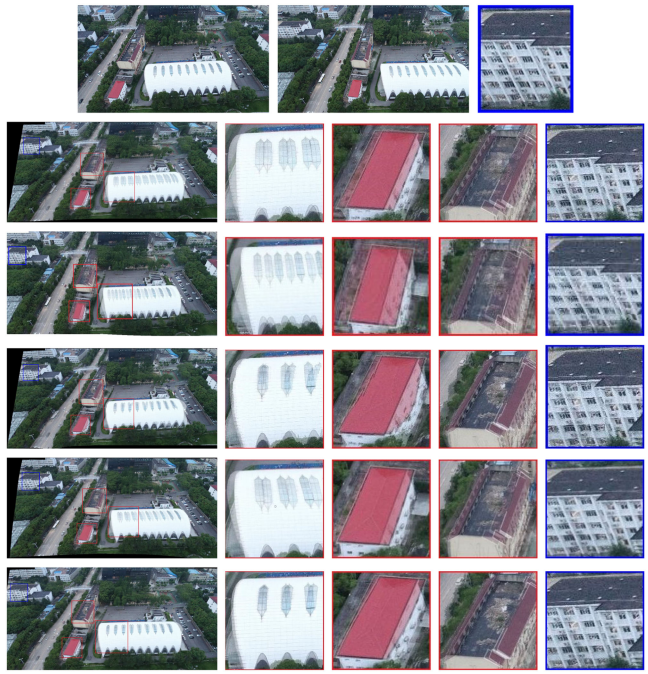

Fig. 5. Stitching results among various methods. The first line is the input source image. The second, third, fourth, fifth, and sixth lines are the results of global homography, SPHP, APAP, REW, and our method. Red boxes highlight the ghostly parts. Blue boxes highlight distorted parts.

areas with smaller disparity. In its results, some short buildings can be accurately matched, but there are still challenges for high-rise buildings (see first red box in the third row). Our model can solve these challenges and achieve a convincing effect.

Figs. 4 and 5 show the stitching results of global homography [17], SPHP [29], APAP [25], REW [36], and our algorithm. Several typical areas are highlighted in boxes. These two scenes are in the wild and in the city, and there are certain parallaxes in images. Ghosting is easy to appear on the roof of a building. Our method has the best results overall. Experimental results demonstrate that our results are obviously better than other methods, indicating that our algorithm is more suitable for UAV image stitching.

B. Ground Constraint

This part mainly analyzes the role of ground constraint. We design this constraint in order to prevent distortion in nonoverlapping regions. Some experimental results also prove the role of our constraint.

We evaluate the quality of different results in these methods through the two aspects of ghosting and distortion. The ghostly parts can be efficiently distinguished. But we often judge for the distorted parts by feeling, which is hard to convince. we propose a method to create groundtruth. To assess the degree of distortion of the results, root-mean-square error (RMSE) is adopted. We calculate the RMSE of the pixel of the distorted parts and the pixel of the relevant parts in the groundtruth, which is given as follows: $\mathrm{RMSE}(P, P') = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - p_i')}$, where P and P' are the distorted parts and the relevant parts in the groundtruth. N is the number of pixels of the distorted parts and p_i and p_i' are the pixel of the distorted parts and the relevant parts in the groundtruth. The larger the value of r, the worse the result. We can better evaluate the degree of distortion of the results and make it more reasonable.

Fig. 6. Comparative experiment of proving effectiveness of different parts of the proposed method.

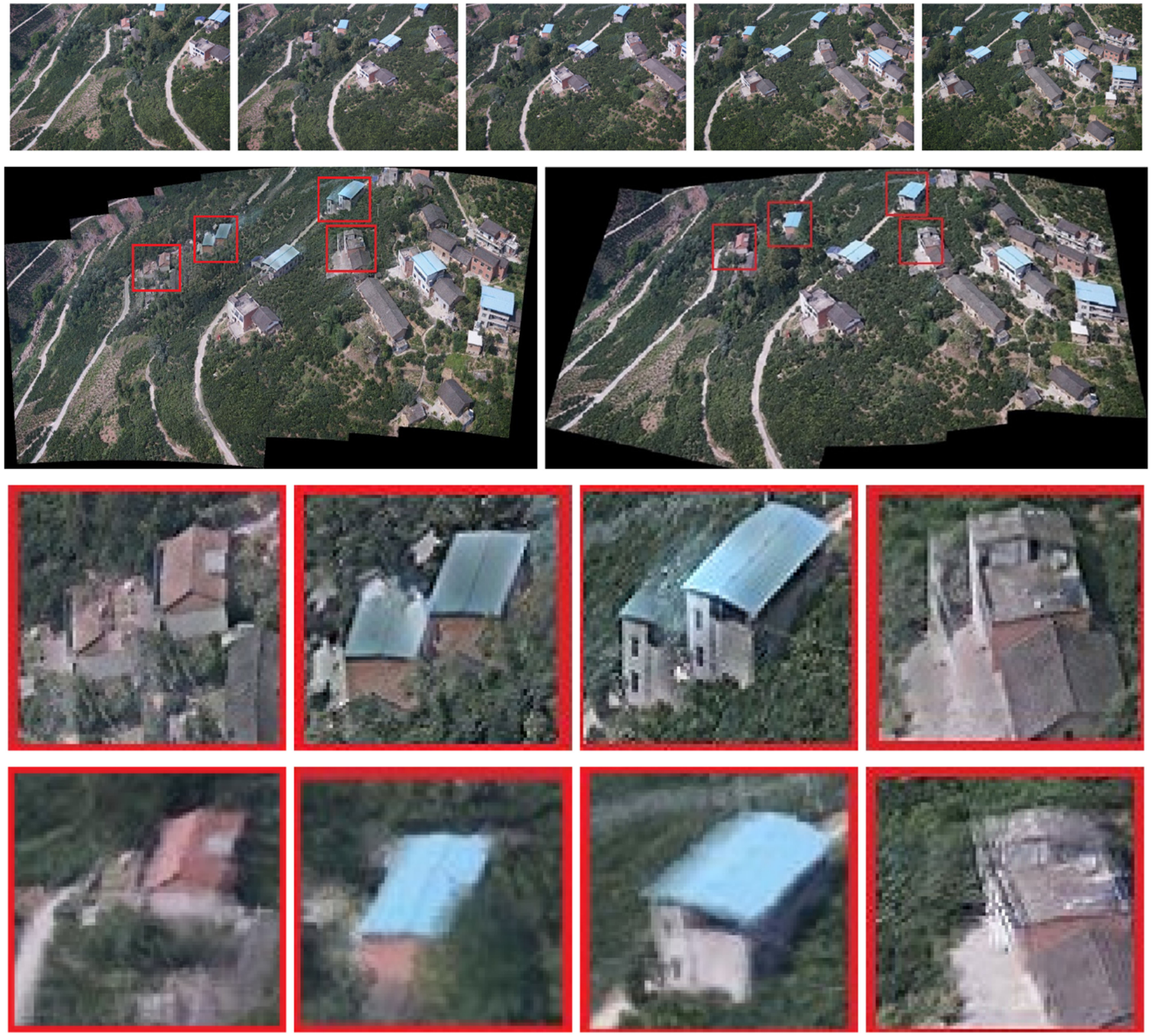

Fig. 7. Experimental results of stitching multiple images. The first line is the input source image. The second line is the results of Autostitch and our method. Red boxes highlight the ghostly parts. The last two line are the details of the red boxes.

TABLE I
COMPARISON OF DIFFERENT METHODS FOR DISTORTION IN BLUE BOXES OF FIG. 3

	Autostitch	APAP	Ours
Fig. 3	39.15	38.86	29.91

TABLE II
COMPARISON OF DIFFERENT METHODS FOR DISTORTION IN BLUE BOXES OF FIGS. 4 AND 5

	Autostitch	APAP	SPHP	REW	Ours
Fig. 4(1)	59.76	45.56	47.70	49.06	44.15
Fig. 4(2)	49.44	31.73	40.77	37.51	29.60
Fig. 5	28.58	22.13	26.42	23.67	19.97

In Fig. 3, blue box of first row shows parts of the groundtruth and blue boxes of other rows highlight the distorted parts. Table I shows the RMSE results of Fig. 3. Autostitch has the largest value because it uses a global homography to project the image. Our global similarity constraint reduce distortion, so the values are smaller than them. Table II shows the RMSE results of Figs. 4 and 5. In Fig. 4, two typical regions with severe distortion are selected and highlighted with blue boxes. The two blue boxes correspond to the first and second rows in Table II. In Fig. 5, one typical region is selected and highlighted with blue box. The blue box corresponds to the third row in Table II. We also conducted experiments on 5 UAV datasets we acquired, and these quantitative results are shown in Table III. From the

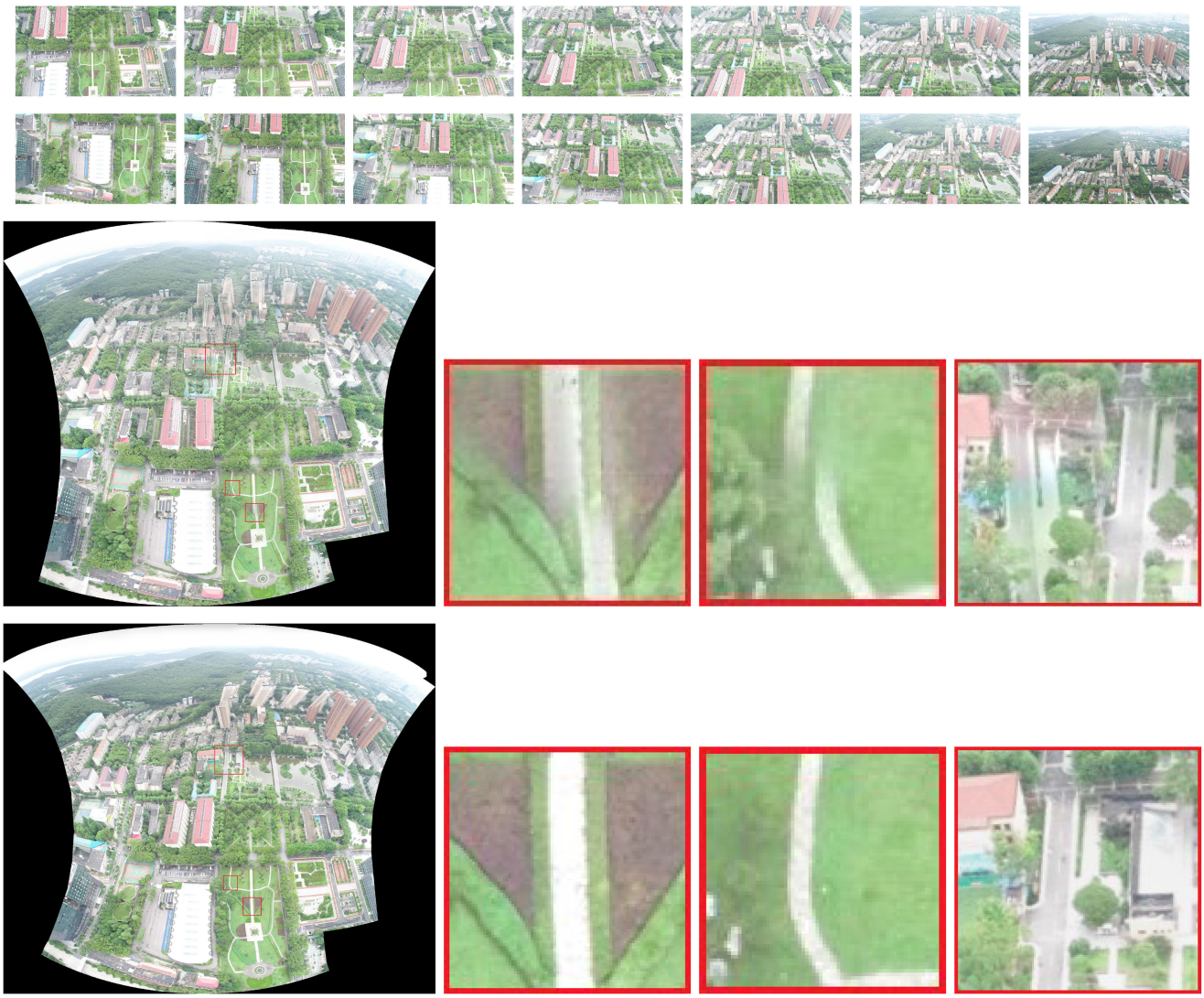

Fig. 8. Experimental results of stitching multiple images. The first and second lines are the image to be stitched together. The third line is the results of Autostitch. The fourth line is our method. Red boxes highlight the ghostly parts.

TABLE III
COMPARISON OF DIFFERENT METHODS FOR DISTORTION ON 5 UAV DATASETS

	Autostitch	APAP	SPHP	REW	Ours
1	38.55	29.56	34.72	28.83	25.22
2	44.36	35.98	40.41	36.86	30.35
3	40.56	30.95	37.81	32.93	27.73
4	23.57	17.98	21.88	18.21	16.45
5	33.67	26.76	30.08	24.72	22.86

perspective of images or indicators, our results are relatively better.

In addition, a comparative experiment is conducted to prove effectiveness of different parts of the proposed method. The experimental results are shown in Fig. 6. The first row is the input source image. The second row shows the result of global homography. Red boxes highlight the ghostly parts and blue boxes highlight distorted parts. There are many local ghostly parts and distortions in this result. The third row is the result of only mesh-guided deformation. There are only some distortions in this result. The fourth row is the result of mesh deformation and ground constraint. Distortions are reduced in this result. The experimental results show that mesh-guided deformation is used to eliminate local ghostly parts and ground constraint is used to reduce projection distortions of nonoverlapping area.

C. Stitching Multiple Images

The experiment of stitching multiple images is demonstrated in Figs. 7 and 8. Some parallax occur in these two sets of source images. Fig. 7 shows the result of stitching five images. The first line is the input source image. The second line is the result image of Autostitch (Baseline) and our method, respectively. Several areas with obvious ghosts are highlighted in red boxes, and the third line is the prominent detail part of Autostitch result. Some misalignments are evident in Autostitch result. Because

this method uses one homography when matching feature points, which leads to errors. In the process of stitching multiple images, the accumulated errors are not optimized by bundle adjustment. The fourth line is our detailed image. Basically no ghosts appear, and the shape of the building is not deformed. Fig. 8 shows a panoramic image of 14 images. We also highlight the obvious errors in the results. Autostitch produces unconvincing results. On the contrary, our model has better results because of mesh-guided deformation.

V. Conclusion

In this article, we propose a method for UAV image stitching based on mesh-guided deformation and ground constraint. The main goal of image stitching is to get a perfectly integrated, overall natural panoramic image. We use the edge information of the mesh in mesh-guided deformation phase. By minimizing the energy function designed, we get the optimal solution and make the overlapping regions accurately aligned. Then, the ground constraint is proposed on the basis of the characteristics of the UAV images. Weighting measures are taken to smooth the overall image from the overlapping area to the nonoverlapping area, and the overall effect is more natural. Finally, a method of creating groundtruth is proposed, which can better judge the overall natural effect of stitching results. Our method can match the characteristics of UAV images efficiently. However, for some particularly large parallax situations, providing perfect alignment is often not possible. Such as the high-rise buildings in the image, the angle of view changes too much and it cannot be precisely aligned. In the future work, we hope to explore ways to work in this large parallax situation.

References

[1] C. Liu, S. Zhang, and A. Akbar, "Ground feature oriented path planning for unmanned aerial vehicle mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1175–1187, Apr. 2019.

[2] X. Tian, Y. Chen, C. Yang, and J. Ma, "Variational pansharpening by exploiting cartoon-texture similarities," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3048257.

[3] X. Li, R. Feng, X. Guan, H. Shen, and L. Zhang, "Remote sensing image mosaicking: Achievements and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 4, pp. 8–22, Dec. 2019.

[4] E. Adel, M. Elmogy, and H. Elbakry, "Image stitching based on feature extraction techniques: A survey," *Int. J. Comput. Appl.*, vol. 99, no. 6, pp. 1–8, 2014.

[5] 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 Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015.

[6] B. Ye, Z. Cai, T. Lan, and W. Cao, "A novel stitching method for dust and rock analysis based on YUTU rover panoramic imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4457–4466, Nov. 2019.

[7] J. Chen, Q. Wan, L. Luo, Y. Wang, and D. Luo, "Drone image stitching based on compactly supported radial basis function," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4634–4643, Nov. 2019.

[8] X. Jiang, J. Ma, J. Jiang, and X. Guo, "Robust feature matching using spatial clustering with heavy outliers," *IEEE Trans. Image Process.*, vol. 29, pp. 736–746, 2020.

[9] X. Li, N. Hui, H. Shen, Y. Fu, and L. Zhang, "A robust mosaicking procedure for high spatial resolution remote sensing images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 109, pp. 108–125, 2015.

[10] D. V. Fedorov, L. M. G. Fonseca, C. Kenney, and B. S. Manjunath, "Automatic registration and mosaicking system for remotely sensed imagery," in *Proc. Image Signal Process. Remote Sens. VIII*, 2003, vol. 4885, pp. 444–451.

[11] X. Jiang *et al.*, "Robust feature matching for remote sensing image registration via linear adaptive filtering," *IEEE Trans. Geosci. Remote. Sens.*, vol. 59, no. 2, pp. 1577–1591, Feb. 2021.

[12] D. Turner, A. Lucieer, and C. Watson, "An automated technique for generating georectified mosaics from ultra-high resolution unmanned aerial vehicle (UAV) imagery, based on structure from motion (SFM) point clouds," *Remote Sens.*, vol. 4, no. 5, pp. 1392–1410, 2012.

[13] G. De Grandi, J.-P. Malingreau, and M. Leysen, "The ERS-1 central africa mosaic: A new perspective in radar remote sensing for the global monitoring of vegetation," *IEEE Trans. Geosci. Remote sensing*, vol. 37, no. 3, pp. 1730–1746, May 1999.

[14] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1285–1298, Jul. 2014.

[15] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, "Seam-driven image stitching," in *Proc. Eurographics (Short Papers)*, 2013, pp. 45–48.

[16] A. Eden, M. Uyttendaele, and R. Szeliski, "Seamless image stitching of scenes with large motions and exposure differences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2006, vol. 2, pp. 2498–2505.

[17] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *Int. J. Comput. Vis.*, vol. 74, no. 1, pp. 59–73, 2007.

[18] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, 1983.

[19] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, 2003.

[20] Z. Shao, C. Li, D. Li, O. Altan, L. Zhang, and L. Ding, "An accurate matching method for projecting vector data into surveillance video to monitor and protect cultivated land," *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 7, p. 448, 2020.

[21] Z. Shao *et al.*, "Exploring the relationship between urbanization and ecological environment using remote sensing images and statistical data: A case study in the yangtze river delta, china," *Sustainability*, vol. 12, no. 14, p. 5620, 2020.

[22] F. Bellavia and C. Colombo, "Dissecting and reassembling color correction algorithms for image stitching," *IEEE Trans. Image Process.*, vol. 27, no. 2, pp. 735–748, Feb. 2018.

[23] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[24] H.-Y. Shum and R. Szeliski, "Construction of panoramic image mosaics with global and local alignment," in *Panoramic Vision*. Berlin, Germany: Springer, 2001, pp. 227–268.

[25] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving DLT," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2339–2346.

[26] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 49–56.

[27] W.-Y. Lin, S. Liu, Y. Matsushita, T.-T. Ng, and L.-F. Cheong, "Smoothly varying affine stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 345–352.

[28] Z. Lou and T. Gevers, "Image alignment by piecewise planar region matching," *IEEE Trans. Multimedia*, vol. 16, no. 7, pp. 2052–2061, Nov. 2014.

[29] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3254–3261.

[30] C.-C. Lin, S. U. Pankanti, K. Natesan Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1155–1163.

[31] R. Szeliski *et al.*, "Image alignment and stitching: A tutorial," *Found. Trends Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1–104, 2007.

[32] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3262–3269.

[33] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, "Seagull: Seam-guided local alignment for parallax-tolerant image stitching," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 370–385.

[34] J. Jia and C.-K. Tang, "Image stitching using structure deformation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 617–631, Apr. 2008.

[35] K. Lin, N. Jiang, S. Liu, L.-F. Cheong, M. Do, and J. Lu, "Direct photometric alignment by mesh deformation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2405–2413.

[36] J. Li, Z. Wang, S. Lai, Y. Zhai, and M. Zhang, "Parallax-tolerant image stitching based on robust elastic warping," *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1672–1687, Jul. 2018.
[37] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, "Image matching from handcrafted to deep features: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021.
[38] J. Ma, J. Wu, J. Zhao, J. Jiang, H. Zhou, and Q. Z. Sheng, "Nonrigid point set registration with robust transformation learning under manifold regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3584–3597, Dec. 2019.
[39] J. Ma, W. Qiu, J. Zhao, Y. Ma, A. L. Yuille, and Z. Tu, "Robust l_2e estimation of transformation for non-rigid registration," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1115–1129, Mar. 2015.
[40] J. Ma, J. Zhao, J. Jiang, H. Zhou, and X. Guo, "Locality preserving matching," *Int. J. Comput. Vis.*, vol. 127, no. 5, pp. 512–531, 2019.
[41] S. Bu, Y. Zhao, G. Wan, and Z. Liu, "Map2DFusion: Real-time incremental uav image mosaicing based on monocular slam," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2016, pp. 4564–4571.
[42] Z. Zhang, "Parameter estimation techniques: A tutorial with application to conic fitting," *Image Vis. Comput.*, vol. 15, no. 1, pp. 59–76, 1997.
[43] P. S. Heckbert, "Fundamentals of texture mapping and image warping," 1989.
[44] T. Igarashi, T. Moscovich, and J. F. Hughes, "As-rigid-as-possible shape manipulation," in *Proc. ACM Trans. Graph.*, 2005, vol. 24, pp. 1134–1141.
[45] F. Liu, M. Gleicher, H. Jin, and A. Agarwala, "Content-preserving warps for 3 D video stabilization," in *Proc. ACM Trans. Graph.*, 2009, vol. 28, p. 44.
[46] O. Chum, T. Pajdla, and P. Sturm, "The geometric error for homographies," *Comput. Vis. Image Understanding*, vol. 97, no. 1, pp. 86–102, 2005.
[47] T.-Z. Xiang, G.-S. Xia, X. Bai, and L. Zhang, "Image stitching by line-guided local warping with global similarity constraint," *Pattern Recognit.*, vol. 83, pp. 481–497, 2018.

Quan Xu was born in Jingzhou, Hubei, China, in 1995. He received the B.S. degree from the School of Electronic Information, Yangtze University, Jingzhou, China, in 2018. He is currently working toward the M.S. degree with the School of Mechanical Electronic and Electronic Information, China University of Geosciences, Wuhan, China.

His research interests include computer vision and pattern recognition.

Jun Chen received the B.S. degree in electronic and information engineering and the M.S. degree in communication and information system from the China University of Geosciences, Wuhan, China, in 2002 and 2004, respectively, and the Ph.D. degree in communication and information system from the Huazhong University of technology, Wuhan, China, in 2014.

From 2004 to 2008, she was an Assistant Professor with the China University of Geosciences, where she is currently an Associate Professor with the School of Automation. Her research interests include computer vision and pattern recognition, geoscience and remote sensing.

Linbo Luo received the B.S. degree in electronics and information engineering from the China University of Geosciences, Wuhan, China, in 2002, the M.S. degree in engineering of traffic information and control from the Wuhan University of Technology, Wuhan, China, in 2005, and the Ph.D. degree from the Department of Electronic and Computer Science, Hanyang University, Seoul, South Korea, in 2012.

From 2014 to 2019, he was an Associate Professor with the China University of Geosciences. His research interests include image processing, image warping, and digital image stabilization.

Wenping Gong received the B.S. degree in civil engineering from Tongji University, Shanghai, China, in 2011, and the Ph.D. degree in civil engineering from Clemson University, Clemson, SC, USA, in 2014.

From 2015 to 2017, he was a Research Assistant Professor with the Glenn Department of Civil Engineering, Clemson University. He is currently a Professor with the Faculty of Engineering, China University of Geosciences, Wuhan, China. He has authored and coauthored more than 30 papers in various peer reviewed international journals, including Engineering Geology. His research interests include uncertainty, reliability, and risk in geological and geotechnical engineering; slope and landslide analysis and risk assessment; and tunneling and underground construction.

Prof. Gong was the recipient of the Excellent Paper Award from the Journal of GeoEngineering and the Best Paper Award from GeoShanghai 2014 International Conference. He is currently on the editorial boards of several international journals, in terms of Engineering Geology, Bulletin of Engineering Geology and the Environment, Marine Georesources and Geotechnology, and the International Journal of Geotechnical Engineering. He is also an Assistant Editor of Engineering Geology.

Yong Wang received the Ph.D. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009.

He is currently an Associate Professor with the China University of Geosciences, Wuhan, China. His current research interests include wireless sensor networks and image processing.