

Article

# A Robust Method for Automatic Panoramic UAV Image Mosaic

Jun Chen <sup>1,2,3</sup> , Quan Xu <sup>4</sup>, Linbo Luo <sup>4,\*</sup>, Yongtao Wang <sup>1,2</sup> and Shuchun Wang <sup>1,2</sup>

<sup>1</sup> School of Automation, China University of Geosciences, Wuhan 430074, China; chenjun71983@163.com (J.C.); ytwang@cug.edu.cn (Y.W.); wangscgt@126.com (S.W.)

<sup>2</sup> Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China

<sup>3</sup> Hubei Provincial Key Laboratory of Intelligent Robot, Wuhan Institute of Technology, Wuhan 430073, China

<sup>4</sup> School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan 430074, China; 1816312309@163.com

\* Correspondence: luolb@ige-live.com

Received: 1 February 2019; Accepted: 16 April 2019; Published: 22 April 2019



**Abstract:** This paper introduces a robust method for panoramic unmanned aerial vehicle (UAV) image mosaic. In the traditional automatic panoramic image stitching method (Autostitch), it assumes that the camera rotates about its optical centre and the group of transformations the source images may undergo is a special group of homographies. It is rare to get such ideal data in reality. In particular, remote sensing images obtained by UAV do not satisfy such an ideal situation, where the images may not be on a plane yet and even may suffer from nonrigid changes, leading to poor mosaic results. To overcome the above mentioned challenges, in this paper a nonrigid matching algorithm is introduced to the mosaic system to generate accurate feature matching on remote sensing images. We also propose a new strategy for bundle adjustment to make the mosaic system suitable for the UAV image panoramic mosaic effect. Experimental results show that our method outperforms the traditional method and some of the latest methods in terms of visual effect.

**Keywords:** image mosaic; nonrigid deformation; bundle adjustment; feature matching

## 1. Introduction

Unmanned aerial vehicle (UAV) remote sensing is a low-altitude technology, which has become an important means of information acquisition. UAV has been widely used in various ground survey applications due to its advantages of high acquisition speed, convenient operation, good safety, and low investment. However, a UAV image has a small view of scene due to the low flying altitude and limited focal length of the camera; thus, capturing a relatively complete target area becomes difficult. Multiple images of the same target should be combined by technical means to obtain a complete scene of the desired target. The purpose of the image mosaic is to combine multiple images with overlapping fields to form a panoramic image [1–3].

Traditional image stitching [4] includes feature matching, image matching, bundle adjustment, automatic panorama straightening, gain compensation, and multiband blending [5,6]. This system may appear perfect, but several errors emerge when images contain a large amount of nonideal data. For example, when UAV images are stitched using this system, ghosting and even incorrect matches may arise. Several efforts have been exerted to reduce these errors, for example, seam cutting methods [7,8] optimize pixel selection among the overlapping images to minimize visible seams, whereas Laplacian pyramid blending [4,5] and Poisson image blending [9] minimize blurring due to misalignments or exposure differences. Nevertheless, the effect remains unsatisfactory, and the problem may be even further severe when scenes are captured by UAVs involving nonrigid changes.

Two major challenges exist in the traditional stitching system when applying to UAV image mosaics. On the one hand, the geometrical relationship between UAV images is often complex due to variations in ground relief, changes in imaging viewpoints, and shooting at low altitudes, where image pairs cannot be accurately matched using a parametric transformation model (e.g., affine or homography) as in most existing methods. Image matching is a critical prerequisite of image mosaic, which aims to overlay two images of the same scene geometrically [10–13]. On the basis of the type of data given, image matching can be divided into rigid and nonrigid. In traditional mosaic matching, the implicit constraint is that the given image is rigid. However, for UAV images, nonrigid matching is crucial because these images often contain local deformation, which cannot be solved by rigid matching. On the other hand, remote sensing images captured by UAVs are from a low altitude, and the resulting images cannot be approximated as a plane. In the traditional stitching system, single planar perspective transform (homography) [14] is used with bundle adjustment for optimization, and this usage leads to ghosting errors. To overcome such limitation, Chin et al. proposed as-projective-as-possible (APAP) image stitching [15], which meshes images, and each grid is aligned with a homography, which is suitable for UAV image mosaic. However, the method has a prerequisite that the images have been accurately aligned.

To address the aforementioned challenges, we introduce a nonrigid matching algorithm on the basis of motion field interpolation, namely, vector field consensus (VFC) [16], to the mosaic system for generating accurate feature matching on remote sensing images. In bundle adjustment [17], we propose a new strategy that improves the original relationship of homography conversion and use the homography response to perform bundle adjustment, which is robust to our image mosaic. The experiments on different sets of UAV images reveal the superiority of our method.

## 2. Related Work

This section briefly reviews the background material on which our work is based, including nonrigid matching algorithms, local transformation descriptions, bundle adjustment, and aerial image mosaic.

### 2.1. Nonrigid Image Matching

The image matching phase is a key step in image stitching and many other computer vision tasks [18–21]. In the matching phase, the capability to minimize registration errors plays an important role in the subsequent steps of the system. Here, the matching problem aims to align the overlapping regions of two images pixel by pixel. Obtaining a good stitching effect requires a dense pixel-by-pixel matching of the overlapping regions. When the unknown complex relationship between images cannot be accurately modeled by a specific model, especially when a nonrigid change occurs in the image, achieving accurate pixel-by-pixel registration is difficult.

In general, the SIFT [22] feature points are extracted to register the relationship between images. Sparse feature point matching is used to guide dense pixel-by-pixel registration. Coping with a large proportion of feature points is possible with a general linear model, such as RANSAC [23], and its variants, including MLESAC [24], LO-RANSAC [25], and PROSAC [26], to estimate the spatial transformation between images. This approach seems to be a good solution. However, this series of algorithms relies on specific geometric parameter models. Images with nonrigid changes can no longer be applied because the transformation cannot be modeled parametrically.

Li and Hu [27] proposed an identifying correspondence function (ICF) algorithm based on nonparametric models to achieve nonrigid matching, but its matching precision drops sharply when many outliers are present. Another strategy for solving the matching problem is to estimate the corresponding matrix of the two point sets in combination with the parameter or nonparametric geometric constraints. The method based on estimating the corresponding matrix jointly estimates the correspondence between the point sets and transformation, compared with the previous independent estimated point correspondence and transformation. Representative methods for finding inter-image

correspondence matrices, such as iterated closest point (ICP) [28], coherence point drift (CPD) [29] and its variant [30], and locality preserving matching (LPM) [31,32], have also been proposed to match nonrigid cases, but such methods generally cannot tolerate excessive outliers. Moreover, for robustness purposes, such algorithms typically impose penalties on points that are not matched. The point set matching problem can also be solved by solving a graph matching problem [33]. Such methods construct an affinity matrix between point sets and then perform spectral analysis to obtain ordered features of the point set. These methods include dual decomposition (DD) [34], spectral matching (SM) [35], and graph shift (GS) [36]. However, the computational complexity of these methods is typically high, which cannot be applied to address large-scale real-time matching tasks. The VFC algorithm proposed by Ma et al. is also based on the nonparametric model [16]. This algorithm converts the point set matching problem into vector field interpolation and can efficiently generalize sparse matching to dense matching, which is suitable for our UAV image matching.

## 2.2. Local Transformation

The conversion relationship of the overlapping image alignment is estimated after acquiring the image matching relationship. Then, the aligned images are combined into a common plane. In the traditional stitching method, a homography is used to represent this conversion relationship. The goal is to minimize alignment errors between overlapping pixels by uniform variations. A homography is a commonly used transformation because it maintains the most flexible transformation of all lines, and the resulting panorama does not produce considerable distortion. The homography has excellent effect for all images in a plane. However, for the UAV images tackled in this study, this ideal condition typically does not hold.

As a homography cannot efficiently align the pixels of the overlapping part, Liu et al. [37] proposed a content-preserving warp that minimizes the registration error and maintains the stiffness of the scene. However, in image stitching, large differences are observed in the rotation and translation between different views. The interpolation of this method is not flexible enough due to the constraint of stiffness. Gao et al. [38] divided a scene into distant and ground planes that sweep out from the camera's location; they aligned the two planes with two homographies and then blended these to align the images. This method is more flexible than using a single homography, but it remains unsuitable for complex scenarios. Lin et al. [39] used a smoothly varying affine warp to align the image. The local deformability and alignment capability were strong and the method can handle parallax. However, the use of affine regularization is appropriate for interpolation, which may not be optimal for extrapolation, and the affinity is insufficient for completing the perspective transformation. Chin et al. [15] proposed a transformation that is as projected as possible by using an algorithm that is as transformable as possible. The purpose of the deformation is to perform a global perspective transformation and allow local non-projection transformations by sampling the local transformation model for image mosaic. The local model has high degree of freedom and is flexible, which can handle local transformation and reduce ghosting problems. Subsequent methods use global similarity to mitigate projection distortion in nonoverlapping regions, such as shape-preserving half-projective [40]. This approach adds constraints, combined with homography and similarity transformations, corrects the shape of the stitched image, and reduces projection distortion. Adaptive as natural as possible [41] also uses global similarity transformations to correct shapes, but it can adaptively determine angles to correct image shapes efficiently.

## 2.3. Bundle Adjustment

Given a set of overlapping images, we aim to project all images onto a common surface. This process will inevitably accumulate errors. To obtain an improved stitching effect, we must simultaneously refine these errors. The current method [4,42] optimizes the focal length of all views and the relative rotation of the camera pose through bundle adjustment and then aligns the series of images. The bundle adjustment can be based on the projection of all points in the image while

extracting the 3D point coordinates describing the scene structure, the relative motion parameters, and the optical parameters of the camera.

Bundle adjustment is often used in feature-based 3D scene reconstruction algorithms as the final step. This method is based on 3D structure and perspective parameters (e.g., camera position, orientation, inherent calibration, and radial distortion). Providing an initial estimate, the bundle adjustment simultaneously refines motion and structural parameters by minimizing projection errors between observed and predicted image points. The best reconstruction effect is obtained under the assumption that the obtained image features contain noises.

The ultimate goal of the bundle adjustment is to reduce the positional projection transformation (re-projection) error between the points of the observed and reference images (predicted image) by using a least squares algorithm. The most successful strategy is Levenberg-Marquardt algorithm [43], which is easy to implement and can quickly converge on a wide range of initial estimates.

#### 2.4. Aerial Image Mosaic

The theme of this study is the stitching of UAV images. In this part, we introduce the related work in aerial image mosaic.

Park et al. [44] proposed an algorithm for hierarchical multi-level image mosaicing for the autonomous navigation of UAV. The algorithm prevents the error accumulation propagated along the frames by incrementally building a long-duration mosaic on the fly, which is hierarchically composed of short-duration mosaics. The real-time processing requirements in autonomous navigation are fulfilled. In particular, the system can automatically adapt to the scene changes of the images to be spliced and can automatically select suitable methods for feature selection. This point is important in the autonomous navigation of UAV. The proposed system is a causal system and is suitable for real-time application. Ghosh et al. [45] proposed a spatial domain-based super-resolution mosaicing system and presented its evaluation results. This algorithm combines image mosaicing and super-resolution-reconstruction algorithms and has the characteristics of robustness and simple calculation. The authors used spatial domain-based super resolution and made it practical in real-time applications, such as surveillance and remote sensing. Ghosh et al. also introduced indicators for the quantitative evaluation of image mosaicing in multiple scene categories [46], which contributed to the defects in this aspect.

### 3. Methodology

Our task is to place a set of UAV images together in a system to obtain a mosaic panorama. The previous section indicates that the traditional stitching system cannot produce good results because of the particularities of UAV images. Hence, we perform modifications to the system to make it robust to UAV images.

#### 3.1. Nonrigid Feature Matching Using Vector Field Consensus

In the feature matching stage, the SIFT descriptor is used to conduct initial feature matching and obtain the corresponding relationship of feature points between images. UAV images may have nonrigid changes. Thus, we introduce a nonrigid matching algorithm based on VFC [16].

Given a set of putative SIFT matches  $S = \{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N$ , our purpose is to distinguish false matches (outliers) from the correct matches (inliers) and estimate the transformation  $\mathbf{f}$  to fit the inliers efficiently, where  $\mathbf{f} \in \mathcal{H}$ . We assume that  $\mathbf{f}$  is nonrigid and  $\mathcal{H}$  is a reproducing kernel Hilbert space (RKHS) [47].

We assume that the noise of inliers is Gaussian with zero mean and uniform standard deviation  $\sigma$ . The outliers are uniformly distributed in the image domain, and the distribution is assumed to be uniform  $1/a$ , where  $a$  is a constant. Let  $\gamma$  be the percentage of inliers that we do not know in advance. Thus, the likelihood is a mixture model of distributions for inliers and outliers:

$$\begin{aligned}
p(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) &= \prod_{n=1}^N p(\mathbf{y}_n, z_n | \mathbf{x}_n, \boldsymbol{\theta}) \\
&= \prod_{n=1}^N \left( \frac{\gamma}{(2\pi\sigma^2)^{D/2}} e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(x_n)\|^2}{2\sigma^2}} + \frac{1-\gamma}{a} \right),
\end{aligned} \tag{1}$$

where  $\boldsymbol{\theta} = \{\mathbf{f}, \sigma^2, \gamma\}$  is the set of unknown parameters. We associate sample  $n$  with a latent variable  $z_n \in \{0, 1\}$ , where  $z_n = 1$  represents Gaussian distribution and  $z_n = 0$  represents uniform distribution.

Considering the smoothness constraint, the prior of  $\mathbf{f}$  can be written as

$$p(\mathbf{f}) \propto e^{-\frac{\lambda}{2} \|\mathbf{f}\|_{\mathcal{H}}^2}. \tag{2}$$

By applying Bayes rule,  $p(\boldsymbol{\theta} | \mathbf{X}, \mathbf{Y}) \propto p(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) P(\boldsymbol{\theta})$ , we estimate  $\boldsymbol{\theta}$  as an MAP solution, i.e.,  $\boldsymbol{\theta}^*$ :

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta}} p(\mathbf{Y} | \mathbf{X}, \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{3}$$

where  $\boldsymbol{\theta}^*$  corresponds to the estimate of the true  $\boldsymbol{\theta}$ . Thus, the transformation  $\mathbf{f}$  can be obtained. We introduced and solved a latent variable  $z_n$  by using the EM algorithm. We can write the complete-date log posterior as follows [48]:

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) &= -\frac{1}{2\sigma^2} \sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(x_n)\|^2 + \ln \gamma \sum_{n=1}^N p_n \\
&+ \ln(1 - \gamma) \sum_{n=1}^N (1 - p_n) - \frac{D}{2} \ln \sigma^2 \sum_{n=1}^N p_n - \frac{\lambda}{2} \|\mathbf{f}\|^2,
\end{aligned} \tag{4}$$

where  $p_n = P(z_n = 1 | \mathbf{x}_n, \mathbf{y}_n, \boldsymbol{\theta}^{\text{old}})$ . The value may be maximized by regarding  $z_n$  as missing data from the mixture model.

**E-step:** Denote  $\mathbf{P} = \text{diag}(p_1, \dots, p_N)$ , where the responsibility  $p_n$  can be computed by applying the Bayes rule:

$$p_n = \frac{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(x_n)\|^2}{2\sigma^2}}}{\gamma e^{-\frac{\|\mathbf{y}_n - \mathbf{f}(x_n)\|^2}{2\sigma^2}} + (2\pi\sigma^2)^{\frac{D}{2}} (1 - \gamma) / a}, \tag{5}$$

The posterior probability  $p_n$  is a soft decision, which indicates that the degree of sample  $n$  agrees with the current estimated  $\mathbf{f}$ .

**M-step:** We determine the revised parameter estimate  $\boldsymbol{\theta}^{\text{new}}$  as follows:  $\boldsymbol{\theta}^{\text{new}} = \arg \max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ . By using the derivative of  $\mathcal{Q}(\boldsymbol{\theta})$  with respect to variance  $\sigma^2$  and the mixing coefficient  $\gamma$  and setting them to zero, we obtain

$$\sigma^2 = \frac{\sum_{n=1}^N p_n \|\mathbf{y}_n - \mathbf{f}(x_n)\|^2}{D \sum_{n=1}^N p_n}, \tag{6}$$

$$\gamma = \frac{\sum_{n=1}^N p_n}{N}. \tag{7}$$

To complete the EM algorithm, the vector field  $\mathbf{f}$  should be estimated in the M-step, which can be obtained by solving a regularization problem [16].

### 3.2. Local Wrap and Bundle Adjustment

After matching two images, the source image can be projected onto the target image through homography. However, the errors accumulate and magnify when stitching multiple images, especially in multiple overlapping regions [49]. We can simultaneously optimize projection functions to register panoramic images under the framework of minimum mean square error and reduce the cumulative error.

The first step when stitching multiple images is to find the reference plane [50,51] to which all images are projected through a basic homographic warp. We select an image from the input

image as the reference plane, as described in [4]. All overlapped images are identified through the feature matching of the panoramic image. We select the image with the most overlapped parts as the reference plane so that other images can be projected to the reference plane by a homographic warp. The traditional method uses direct linear transformation (DLT) [52] to calculate the single response between images (which requires *four* pairs of matched points). The registration of UAV images cannot be changed simply by a single response. Inspired by Chin et al. [15], the registration error caused by a single response was effectively reduced by differentiating the image, and each differential corresponds to a basic homographic matrix.

Given two overlapping images  $I$  and  $I'$  and their matched points  $\mathbf{x}_i = (x_i, y_i)^T$ ,  $\mathbf{y}_i = (x'_i, y'_i)^T$ ,  $i = 1, \dots, n$ , the global transformation between two images can be estimated by

$$\tilde{\mathbf{y}} \sim \mathbf{H}\tilde{\mathbf{x}}, \quad (8)$$

where  $\tilde{\mathbf{x}}$  denotes  $\mathbf{x}$  in homogeneous coordinates,  $\sim$  indicates the equality up to scale, and  $\mathbf{H}$  is the global homography that is a  $3 \times 3$  matrix.

In 2D projective warp, DLT is a basic method for estimating  $\mathbf{H}$  from a set of noisy point matches. We vectorize  $\mathbf{H}$  into a vector  $\mathbf{h}$ . Let  $\mathbf{a}_i$  be the two rows of the matrix for two matched points. Given an estimate  $\mathbf{h}$ , the quantity  $\|\mathbf{a}_i\mathbf{h}\|$  is the algebraic error. DLT minimizes the sum of squared algebraic errors:

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \sum_{i=1}^N \|\mathbf{a}_i\mathbf{h}\|^2. \quad (9)$$

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

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \sum_{i=1}^N \|\mathbf{A}\mathbf{h}\|^2. \quad (10)$$

The solution is the least significant right singular vector of  $\mathbf{A}$ . Given the estimated  $\mathbf{H}$  (reconstructed from  $\hat{\mathbf{h}}$ ), in the 2D projective warp, an arbitrary pixel  $\mathbf{x}_*$  in the source image  $I$  is warped to the position  $\mathbf{y}_*$  in the target image  $I'$  by

$$\tilde{\mathbf{y}}_* \sim \mathbf{H}\tilde{\mathbf{x}}_*, \quad (11)$$

For UAV images, global homography cannot efficiently match two images. In this case, we introduce a location-dependent homography.

$$\tilde{\mathbf{y}}_* \sim \mathbf{H}_*\tilde{\mathbf{x}}_*. \quad (12)$$

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

$$\mathbf{h}_* = \arg \min_{\mathbf{h}} \sum_{i=1}^N \|w_*^i \mathbf{a}_i \mathbf{h}\|^2. \quad (13)$$

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

$$w_*^i = \exp(-\|\mathbf{x}_* - \mathbf{x}_i\|^2 / \delta^2), \quad (14)$$

where  $\delta$  is a scale parameter. The exact solution steps are the same as before.

When we have a set of location-dependent homographies  $\{\mathbf{H}_*^k\}_{k=1}^K$ , where each  $\mathbf{H}_*^k$  corresponds to a pair of matching points in two images, we should consider the cumulative error during image mosaic. We minimize the cost to minimize the transfer error of all correspondences:

$$E_*(\Theta) = \sum_{i=1}^N \frac{w_*^i}{\sum_{k=1}^K \zeta_{ik}} \sum_{k=1}^K \zeta_{ik} \|\mathbf{x}_i^k - f(\mathbf{p}_i, \mathbf{H}_*^k)\|^2, \quad (15)$$

where  $\Theta = [\mathbf{H}_1^k, \dots, \mathbf{H}_*^k, \mathbf{p}_1, \dots, \mathbf{p}_N]$  and  $f(\mathbf{p}, \mathbf{H})$  is the project warp defined as

$$f(\mathbf{p}, \mathbf{H}) = \begin{bmatrix} \mathbf{r}_1[\mathbf{p}^T \mathbf{1}]^T & \mathbf{r}_2[\mathbf{p}^T \mathbf{1}]^T \\ \mathbf{r}_3[\mathbf{p}^T \mathbf{1}]^T & \mathbf{r}_3[\mathbf{p}^T \mathbf{1}]^T \end{bmatrix}^T, \quad (16)$$

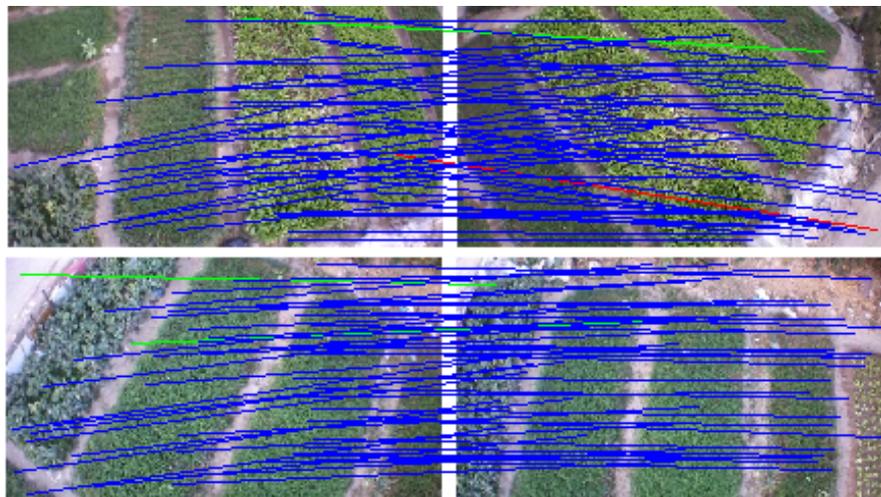
where  $\mathbf{r}_1$ ,  $\mathbf{r}_2$  and  $\mathbf{r}_3$  are the three rows of homography  $\mathbf{H}$ . We introduce a parameter  $\xi_{ik} \in \{0, 1\}$ , where  $\xi_{ik} = 1$  indicates that correspondence exists; otherwise,  $\xi_{ik} = 0$ .

#### 4. Experimental Results

We compare our method with the traditional image mosaic method. In the experiment, the UAV images have nonrigid changes, which cannot be treated as a plane. In the following, we initially report the feature matching results and then provide the image mosaic results.

##### 4.1. Remove Mismatches on Nonrigid Image Pairs

In stitching remote sensing images, a nonrigid matching algorithm must be added because of the high mismatching rate of remote sensing images. The false matching points of UAV images can be removed effectively by adding the VFC algorithm. To evaluate the feature matching performance of VFC, we use three other widely used methods for comparison, namely, RANSAC [23], ICF [27], and GS [36]. We select two representational image pairs [32] that have local nonrigid distortion for evaluation, where the ground truth feature matches are supplied by the dataset. Two common evaluation indicators, namely, precision and recall, are used to measure the performance of our algorithm [53]. The precision rate is defined as the ratio of the number of retained inlier points to the total number of initial feature points, and the recall rate is defined as the ratio of the number of retained inlier points to the number of initial inlier points. Experimental results show that this method is better than other methods when performing nonrigid registration. Figure 1 and Table 1 shows the results. VFC can achieve better matching performance on the UAV images than other state-of-the-art methods.



**Figure 1.** Experimental results on two typical image pairs of nonrigid objects. The lines indicate mismatch removal results (blue = true positive, green = false negative, red = false positive). For visibility, in the image pairs, only 50 randomly selected correspondences are presented, and the true negatives are not shown.

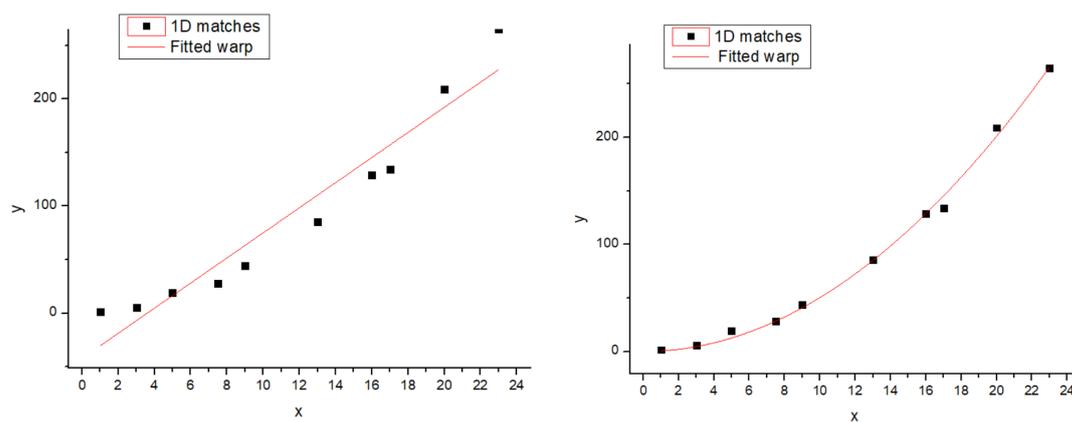
**Table 1.** The precision (P) and recall (R) results of different methods on the image pairs in Figure 1. The initial inlier percentage is 37.27% and 42.39% on the two image pairs, respectively.

|        | RANSAC [23]      | ICF [27]         | GS [36]          | VFC              |
|--------|------------------|------------------|------------------|------------------|
| (P, R) | (99.66%, 97.67%) | (98.98%, 91.76%) | (99.59%, 81.51%) | (99.67%, 98.02%) |
| (P, R) | (100.0%, 97.07%) | (99.32%, 92.86%) | (100.0%, 94.06%) | (100.0%, 97.62%) |

#### 4.2. Overall Differentiation and Local Warp

In projection transformations, the traditional method (Autostitch) [4] completes the transformation by using a homography matrix. Parallax changes in the image will bring negative effects. To solve these problems, we adopt the idea of differentiation. This method can tolerate the parallax in the image. We present a simple example in Figure 2 to illustrate the role of this idea in projection transformation. Local and global warps will have different effects in the stitching, especially when the two images are not on one plane (the two views differ by a rotation and translation). In Figure 2, a homography matrix can efficiently retain the rigid change of the image, but it is bound to leave matching points. Instead, our differentiation method can include most (even all) of the match points. Although the image stiffness may be sacrificed, the effect is acceptable. We also perform the experiment, and Figure 3 shows the results. In Figure 3, strictly aligning these images through the traditional method is difficult. Ghosts in the mosaic image are evident, because a homography matrix fails to match the objects in the two images efficiently.

Our method can greatly improve the situation. The image in the third row of Figure 3 is our mosaic result with evidently few ghosts. White boxes are used to highlight the ghostly parts. Moreover, we compare the difference in the effects of the two strategies. Our method is improved with local homography.



**Figure 2.** To generate a 1D analogy of image stitching, a set of 1D matches  $\{x_n, y_n\}_{n=1}^N$  are generated by projecting a 2D point cloud onto two 1D image “planes”. Here, the two views differ by a rotation and translation. The first indicates that the match points are warped by a global homography and it is unable to model the local deviations of the data. The second indicates that the match points are warped by some local homography and it interpolates the local deviations flexibly.

Differentiation is robust to local nonrigid transformation of images and can solve the problem that an image is not on a plane. In this way, the problem of local projection mismatch can be reduced (reducing the ghosting errors in image mosaic), and the performance of the subsequent bundle adjustment can be improved. We also conduct a series of experiments to verify the effect of our method. We compare our method not only with the traditional method (Autostitch) [4] but also with other state-of-the-art methods, including AANAP [41], SPHP [40], and parallax tolerant (PT) [54].



**Figure 3.** Mosaic results. The first row is the two images to be stitched together. The second row is two stitched images by global homography and local homography, respectively. White boxes highlight the ghostly parts.

The results are reported in Figures 4–6. Red boxes highlight the obvious parts of the ghost effect in the mosaic results. We enlarge the area in red boxes to make results clearer. Results show that our method is robust for stitching UAV images and can get better results than the competitors. The mosaic of panoramic image is based on the mosaic of two images. We can achieve improved results here, which alleviate the pressure of subsequent panorama stitching. However, our results still have shortcomings, and the ghosting of some parts cannot be completely eliminated, which can be further improved during bundle adjustment.



**Figure 4.** Qualitative comparison on UAV image mosaic. The first row is the two images to be stitched. The second and third rows are the stitching results of Autostitch [4], AANAP [41], SPHP [40], and PT [54]. Our result is shown in the last row. Red boxes highlight the ghostly parts.



**Figure 5.** Qualitative comparison on UAV image mosaic. The first row is the two images to be stitched. The second and third rows are the stitching results of Autostitch [4], AANAP [41], SPHP [40], and PT [54]. Our result is shown in the last row. Red boxes highlight the ghostly parts.



**Figure 6.** Qualitative comparison on UAV image mosaic. The first row is the two images to be stitched. The second and third rows are the stitching results of Autostitch [4], AANAP [41], SPHP [40], and PT [54]. Our result is shown in the last row. Red boxes highlight the ghostly parts.

#### 4.3. The Effect of Nonrigid Changes and Parallax

In this section, we will analyze how the mentioned challenges (such as nonrigid changes caused by ground relief, and parallax caused by changes in imaging viewpoints) affect the mosaic result.

Some experiments show the effect of those challenges to the mosaicing results. In Figure 4, nonrigid changes are observed in the ground. In the registration stage, the traditional method cannot efficiently align these nonrigid changes and it is easy to form many mismatches. The large number of small stones in the scene has formed a ghost because they are not efficiently matched. We enlarged the scene in the red boxes. Results are displayed next to the mosaicing results. Our method can improve the ghost effect. Figures 5 and 6 show the images obtained in two different scenes, but they have a certain parallax. In Autostitch, the overlapping areas will be misaligned when the parallax image is projected onto a common plane. There are some unsatisfactory areas in the mosaicing result. In Figure 5, the ground lines in the red boxes are clearly wrong in Autostitch. Such a mosaic result is not what we want to see. Other advanced methods can almost alleviate these ghosts. Our methods can acquire better results than these methods. Our results are clear and the line is still correct. In Figure 6, there are duplicate parts about cars and roads in the red boxes. The stitching results in the last row are better than the rest images. Our method is more suitable for such parallax scene.

At the same time, we choose some sequential images without parallax and nonrigid changes to carry out the experiment. Experimental results are shown in Figure 7. We can see that the results of Autostitch are almost identical to the results of our approach. In the Figure 8, these source images have some parallax and nonrigid changes. The result of Autostitch has errors, where the challenges mentioned above lead to these errors.



**Figure 7.** Experimental results of panoramic stitching. The source images are sequential images captured from left to right without nonrigid changes and parallax. The first row is the results of Autostitch. The second row is the results of our method.

#### 4.4. The Effect of Noise and Blurriness

When we acquire UAV images, introducing noise and blur into the images is inevitable. Therefore, a good system must resist the intrusion of noise and blur. We also conduct experiments on this aspect. Each of these images are convolved with a Gaussian blurring kernel of  $SD = 1.2$  (standard deviation). The images are also contaminated by additive white Gaussian noise with zero mean and variance 0.0004. We use these two strategies to evaluate the behavior of our mosaicing algorithm in the presence of blur or noise.

Figure 9 shows the results. Noise and blur in general have no effect on our system, because the ghost in the mosaic image is not enhanced or enlarged. Thus, the algorithm can still work efficiently and remove mismatched feature points further accurately in the stage of removing mismatches. In the registration stage, the image is unaffected by noise and blur. The algorithm can precisely align the pixels and then project the image to a common plane.

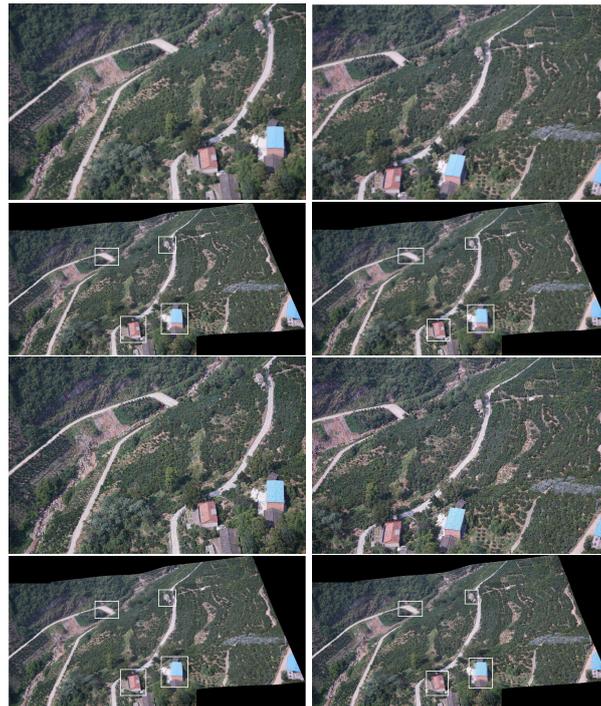


**Figure 8.** Experimental results of panoramic stitching. The first row is the image to be stitched together. The second row is the results of Autostitch and our method. Red boxes highlight the ghostly parts. The last two rows are the details of the red boxes.

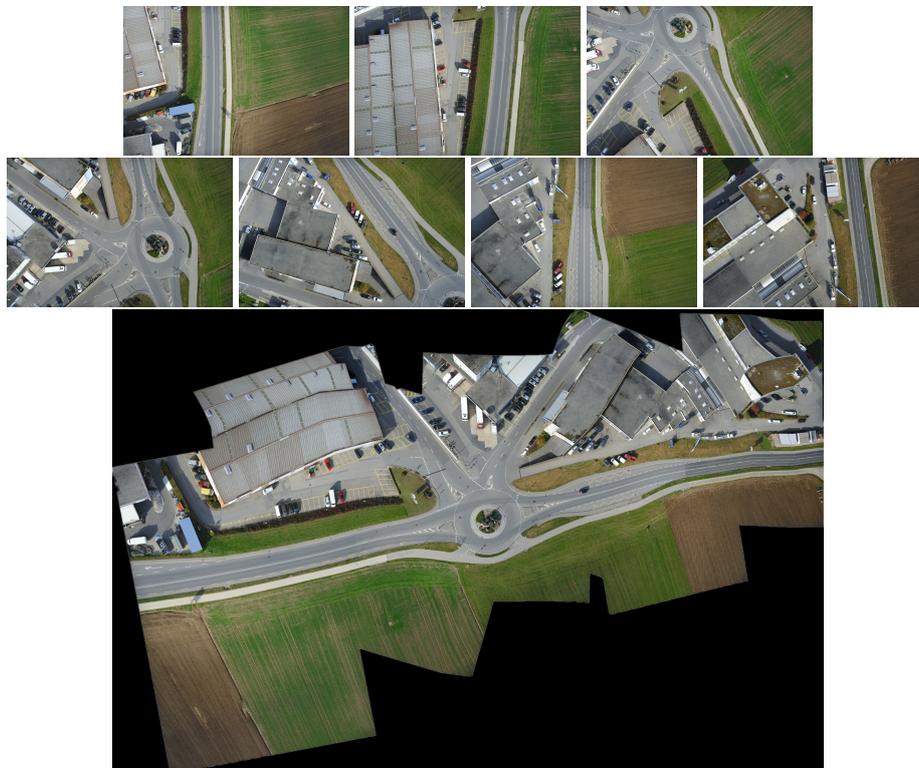
#### 4.5. Bundle Adjustment and Panoramic Stitching

At this stage, our method has the same capability as Autostitch and can achieve the mosaic in a group of arbitrary images. Hence, the source image can be any image of a scene. In Autostitch, only one global homography is used for stitching. Thus, the bundle adjustment optimizes the relative rotation and homographies between sets of overlapping images. This case only requires average intensity blending of the aligned image. Results show that our stitching system has a better effect than the traditional stitching system, and the ghosting effect is considerably reduced. Figure 8 presents the results.

As shown in Figures 3 and 8, when the traditional method (Autostitch) [4] mosaics the panorama image, the error is cumulatively enlarged. When Autostitch mosaics two images, the ghosting effect is not evident. When images are added to the system, the bundle adjustment cannot efficiently remove the accumulated error. By contrast, our method can efficiently remove the accumulated error during bundle adjustment. In Figures 10 and 11 [55], there are some images of arbitrary viewpoints and we also compare the splicing results. In Figure 10, errors are displayed when Autostitch stitches source images. This indicates that the method cannot process these images, so we do not report the result of Autostitch, and only give the splicing result of our method. In Figure 11, The Autostitch stitching results still have some misaligned areas. We have highlighted it in red boxes and zoomed in on the side.



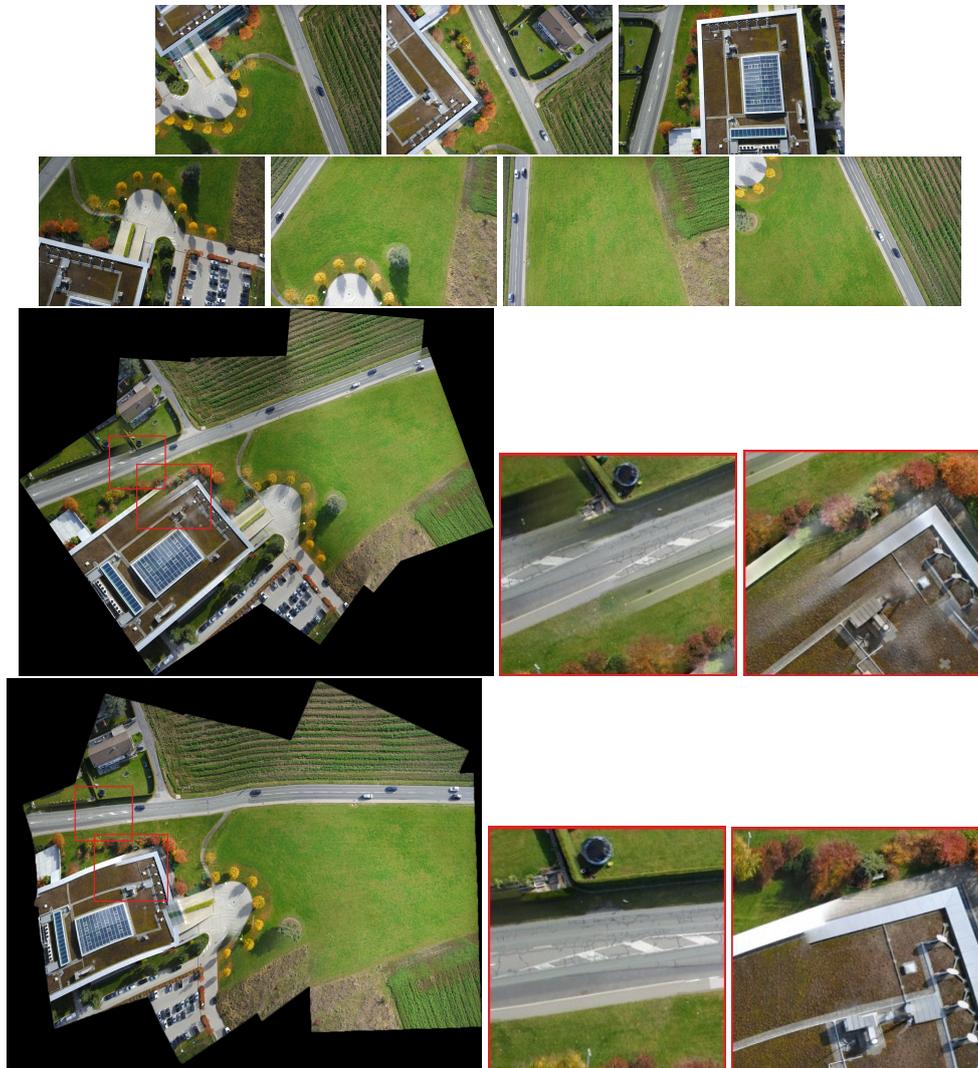
**Figure 9.** The first line is the blurred source image. The second line is the mosaic result of the unblurred source image and the mosaic result of the blurred source image. The third line is the noisy source image. The fourth line is the mosaic result of the source image without noise and the mosaic result of the source image with noise. White boxes highlight the ghostly parts.



**Figure 10.** Experimental results of panoramic stitching. The first and second rows are the images to be stitched together. The third row is the result of our method. Autostitch cannot process these images.

Our method sacrifices the stiffness of the stitching image. The effect is good when stitching two images. However, distortion is observed in the result when stitching multiple images. This

situation can be seen from our stitching results in Figures 7–11, but the overall effect is acceptable. For most UAV images, the mosaic results obtained by our system are acceptable. The distortion will affect the image effect only when the parallax of the source image is particularly large. This case can be improved later, i.e., how to obtain a good mosaic effect on a large-parallax image.



**Figure 11.** Experimental results on panoramic stitching. The first and second rows are the images to be stitched together. The third and fourth rows are the results of Autostitch and our method. Red boxes highlight the ghostly parts.

## 5. Conclusions

In this paper, we propose an image mosaic method based on robust feature matching and a new bundle adjustment strategy. The use of robust feature matching produces further accurate results and reduces the pressure on the subsequent steps of the system. When calculating the transformation relation between images, we differentiate the global homography matrix so that the new adjustment must only blend the aligned image with the average intensity. Results show that our approach provides a further natural panorama with no visible parallax in the overlapping regions and that the ghost effect produced by mosaicing is evidently reduced. Our system is more suitable for UAV images and has a wider application range than the traditional image mosaic method. Future research developments will improve the image effect in the case of a large parallax, which can be performed by integrating seam-cut methods into stitching framework.

**Author Contributions:** All authors have made great contributions to the work. L.L. and J.C. designed the research and analyzed the results. Q.X. performed the experiments and wrote the manuscript. J.C., Y.W. and S.W. gave insightful suggestions to the work and revised the manuscript.

**Funding:** This work was supported in part by the National Natural Science Foundation of China under grant no. 61603354, in part by the Fundamental Research Funds for the Central Universities grant nos. CUGL170221 and 2042018kf1009, in part by the 111 project under grant No. B17040, and in part by the open fund of Hubei Provincial Key Laboratory of Intelligent Robot.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Adel, E.; Elmogy, M.; Elbakry, H. Image stitching based on feature extraction techniques: A survey. *Int. J. Comput. Appl.* **2014**, *99*, 1–8. [[CrossRef](#)]
- Zhang, W.; Guo, B.; Li, M.; Liao, X.; Li, W. Improved Seam-Line Searching Algorithm for UAV Image Mosaic with Optical Flow. *Sensors* **2018**, *18*, 1214. [[CrossRef](#)] [[PubMed](#)]
- Ma, J.; Zhou, H.; Zhao, J.; Gao, Y.; Jiang, J.; Tian, J. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6469–6481. [[CrossRef](#)]
- Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73. [[CrossRef](#)]
- Burt, P.J.; Adelson, E.H. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.* **1983**, *2*, 217–236. [[CrossRef](#)]
- Li, M.; Chen, R.; Zhang, W.; Li, D.; Liao, X.; Wang, L.; Pan, Y.; Zhang, P. A Stereo Dual-Channel Dynamic Programming Algorithm for UAV Image Stitching. *Sensors* **2017**, *17*, 2060. [[CrossRef](#)]
- Agarwala, A.; Dontcheva, M.; Agrawala, M.; Drucker, S.; Colburn, A.; Curless, B.; Salesin, D.; Cohen, M. Interactive digital photomontage. *ACM Trans. Graph.* **2004**, *23*, 294–302. [[CrossRef](#)]
- Eden, A.; Uyttendaele, M.; Szeliski, R. Seamless image stitching of scenes with large motions and exposure differences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17–22 June 2006; Volume 2, pp. 2498–2505.
- Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. *ACM Trans. Graph.* **2003**, *22*, 313–318. [[CrossRef](#)]
- Ma, J.; Zhao, J.; Tian, J.; Bai, X.; Tu, Z. Regularized vector field learning with sparse approximation for mismatch removal. *Pattern Recognit.* **2013**, *46*, 3519–3532. [[CrossRef](#)]
- Wang, D.; Liu, H.; Cheng, X. A Miniature Binocular Endoscope with Local Feature Matching and Stereo Matching for 3D Measurement and 3D Reconstruction. *Sensors* **2018**, *18*, 2243. [[CrossRef](#)]
- Yu, M.; Deng, K.; Yang, H.; Qin, C. Improved WaSH Feature Matching Based on 2D-DWT for Stereo Remote Sensing Images. *Sensors* **2018**, *18*, 3494. [[CrossRef](#)]
- Ma, J.; Jiang, X.; Jiang, J.; Zhao, J.; Guo, X. LMR: Learning A Two-class Classifier for Mismatch Removal. *IEEE Trans. Image Process.* **2019**. [[CrossRef](#)]
- Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003.
- Zaragoza, J.; Chin, T.J.; Brown, M.S.; Suter, D. As-projective-as-possible image stitching with moving DLT. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 2339–2346.
- Ma, J.; Zhao, J.; Tian, J.; Yuille, A.L.; Tu, Z. Robust Point Matching via Vector Field Consensus. *IEEE Trans. Image Process.* **2014**, *23*, 1706–1721.
- Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—a modern synthesis. In Proceedings of the International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999; pp. 298–372.
- Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Inf. Fusion* **2019**, *45*, 153–178. [[CrossRef](#)]
- Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* **2019**, *48*, 11–26. [[CrossRef](#)]
- Wang, Z.; Yi, P.; Jiang, K.; Jiang, J.; Han, Z.; Lu, T.; Ma, J. Multi-memory convolutional neural network for video super-resolution. *IEEE Trans. Image Process.* **2019**, *28*, 2530–2544. [[CrossRef](#)] [[PubMed](#)]

21. Lu, T.; Xiong, Z.; Zhang, Y.; Wang, B.; Lu, T. Robust face super-resolution via locality-constrained low-rank representation. *IEEE Access* **2017**, *5*, 13103–13117. [[CrossRef](#)]
22. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
23. Fischler, M.A.; Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **1981**, *24*, 381–395. [[CrossRef](#)]
24. Torr, P.H.; Zisserman, A. MLESAC: A new robust estimator with application to estimating image geometry. *Comput. Vis. Image Underst.* **2000**, *78*, 138–156. [[CrossRef](#)]
25. Chum, O.; Matas, J.; Kittler, J. Locally optimized RANSAC. *Joint Pattern Recognition Symposium*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 236–243.
26. Chum, O.; Matas, J. Matching with PROSAC-progressive sample consensus. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; Volume 1; pp. 220–226.
27. Li, X.; Hu, Z. Rejecting mismatches by correspondence Li function. *Int. J. Comput. Vis.* **2010**, *89*, 1–17. [[CrossRef](#)]
28. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [[CrossRef](#)]
29. Myronenko, A.; Song, X. Point set registration: Coherent point drift. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2262–2275. [[CrossRef](#)] [[PubMed](#)]
30. Ma, J.; Zhao, J.; Yuille, A.L. Non-rigid point set registration by preserving global and local structures. *IEEE Trans. Image Process.* **2016**, *25*, 53–64.
31. Ma, J.; Zhao, J.; Jiang, J.; Zhou, H.; Guo, X. Locality preserving matching. *Int. J. Comput. Vis.* **2019**, *127*, 512–531. [[CrossRef](#)]
32. Ma, J.; Jiang, J.; Zhou, H.; Zhao, J.; Guo, X. Guided locality preserving feature matching for remote sensing image registration. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4435–4447. [[CrossRef](#)]
33. Yan, J.; Li, C.; Li, Y.; Cao, G. Adaptive discrete hypergraph matching. *IEEE Trans. Cybern.* **2018**, *48*, 765–779. [[CrossRef](#)]
34. Torresani, L.; Kolmogorov, V.; Rother, C. Feature correspondence via graph matching: Models and global optimization. In Proceedings of the European Conference on Computer Vision, Marseille, France, 12–18 October 2008; pp. 596–609.
35. Leordeanu, M.; Hebert, M. A spectral technique for correspondence problems using pairwise constraints. In Proceedings of the IEEE International Conference on Computer Vision, San Diego, CA, USA, 20–25 June 2005; Volume 2; pp. 1482–1489.
36. Liu, H.; Yan, S. Common visual pattern discovery via spatially coherent correspondences. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 1609–1616.
37. Liu, F.; Gleicher, M.; Jin, H.; Agarwala, A. Content-preserving warps for 3D video stabilization. *ACM Trans. Graph.* **2009**, *28*, 44. [[CrossRef](#)]
38. Gao, J.; Kim, S.J.; Brown, M.S. Constructing image panoramas using dual-homography warping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 49–56.
39. Lin, W.Y.; Liu, S.; Matsushita, Y.; Ng, T.T.; Cheong, L.F. Smoothly varying affine stitching. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 345–352.
40. Chang, C.H.; Sato, Y.; Chuang, Y.Y. Shape-preserving half-projective warps for image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3254–3261.
41. Lin, C.C.; Pankanti, S.U.; Natesan Ramamurthy, K.; Aravkin, A.Y. Adaptive as-natural-as-possible image stitching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1155–1163.
42. Shum, H.Y.; Szeliski, R. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *Int. J. Comput. Vis.* **2000**, *36*, 101–130. [[CrossRef](#)]
43. Lourakis, M.; Argyros, A.A. Is Levenberg-Marquardt the most efficient optimization algorithm for implementing bundle adjustment? In Proceedings of the IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2; pp. 1526–1531.

44. Park, S.; Ghosh, D.; Kaabouch, N.; Fevig, R.A.; Semke, W. Hierarchical multi-level image mosaicing for autonomous navigation of UAV. In Proceedings of the Intelligent Robots and Computer Vision XXIX: Algorithms and Techniques, International Society for Optics and Photonics, Burlingame, CA, USA, 23 January 2012; Volume 8301, p. 830116.
45. Ghosh, D.; Kaabouch, N.; Fevig, R.A. Robust spatial-domain based super-resolution mosaicing of CubeSat video frames: Algorithm and evaluation. *Comput. Inf. Sci.* **2014**, *7*, 68. [[CrossRef](#)]
46. Ghosh, D.; Park, S.; Kaabouch, N.; Semke, W. Quantitative evaluation of image mosaicing in multiple scene categories. In Proceedings of the 2012 IEEE International Conference on Electro/Information Technology, Indianapolis, IN, USA, 6–8 May 2012; pp. 1–6.
47. Aronszajn, N. Theory of reproducing kernels. *Trans. Am. Math. Soc.* **1950**, *68*, 337–404. [[CrossRef](#)]
48. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
49. Szeliski, R. Image alignment and stitching: A tutorial. *Found. Trends Comput. Graph. Vis.* **2007**, *2*, 1–104. [[CrossRef](#)]
50. Kang, E.Y.; Cohen, I.; Medioni, G. A graph-based global registration for 2d mosaics. In Proceedings of the International Conference on Pattern Recognition, Barcelona, Spain, 3–7 September 2000; pp. 257–260.
51. Marzotto, R.; Fusiello, A.; Murino, V. High resolution video mosaicing with global alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004.
52. Zhang, Z. Parameter estimation techniques: A tutorial with application to conic fitting. *Image Vis. Comput.* **1997**, *15*, 59–76. [[CrossRef](#)]
53. Ma, J.; Qiu, W.; Zhao, J.; Ma, Y.; Yuille, A.L.; Tu, Z. Robust L2E Estimation of Transformation for Non-Rigid Registration. *IEEE Trans. Signal Process.* **2015**, *63*, 1115–1129. [[CrossRef](#)]
54. Li, J.; Wang, Z.; Lai, S.; Zhai, Y.; Zhang, M. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Trans. Multimed.* **2018**, *20*, 1672–1687. [[CrossRef](#)]
55. Sensefly. Available online: <https://www.sensefly.com/drones/example-datasets.html> (accessed 28 March 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).