remote sensing

MDPI

Article

Mosaicking of Unmanned Aerial Vehicle Imagery in the Absence of Camera Poses

Yuhua Xu [2,†], Jianliang Ou [1,*,†], Hu He [3], Xiaohu Zhang [2] and Jon Mills [4]

1 College of Civil Engineering, Hunan University, Changsha 410082, Hunan, China
2 College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, Hunan, China; xyh_nudt@163.com (Y.X.); zxh1302@hotmail.com (X.Z.)
3 College of Mechanical and Electrical Engineering, Central South University, Changsha 410083, Hunan, China; hehu.mech@csu.edu.cn
4 School of Civil Engineering and Geosciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK; jon.mills@ncl.ac.uk
* Correspondence: oujl@hnu.edu.cn; Tel.: +86-13975153715; Fax: +86-731-88821491
† These authors contributed equally to this work.

Academic Editors: Pablo J. Zarco-Tejada, Magaly Koch and Prasad S. Thenkabail
Received: 12 December 2015; Accepted: 17 February 2016; Published: 2 March 2016

Abstract: The mosaicking of Unmanned Aerial Vehicle (UAV) imagery usually requires information from additional sensors, such as Global Position System (GPS) and Inertial Measurement Unit (IMU), to facilitate direct orientation, or 3D reconstruction approaches (e.g., structure-from-motion) to recover the camera poses. In this paper, we propose a novel mosaicking method for UAV imagery in which neither direct nor indirect orientation procedures are required. Inspired by the embedded deformation model, a widely used non-rigid mesh deformation model, we present a novel objective function for image mosaicking. Firstly, we construct a feature correspondence energy term that minimizes the sum of the squared distances between matched feature pairs to align the images geometrically. Secondly, we model a regularization term that constrains the image transformation parameters directly by keeping all transformations as rigid as possible to avoid global distortion in the final mosaic. Experimental results presented herein demonstrate that the accuracy of our method is twice as high as an existing (purely image-based) approach, with the associated benefits of significantly faster processing times and improved robustness with respect to reference image selection.

Keywords: UAV; sequential imagery; image mosaicking; homography energy model

1. Introduction

A huge market is currently emerging from the vast number of potential applications and services offered by small, low-cost, and low-flying unmanned aerial vehicles (UAVs). UAVs can carry payloads such as cameras, infrared cameras, and other sensors. Thus, they enable us to obtain a synoptic view of an area, which is helpful in applications such as surveillance and reconnaissance, environmental monitoring, disaster assessment and management (see, e.g., [1–3]). The aim of this paper is to present a novel method for mosaicking of UAV imagery to obtain visually satisfactory mosaics (not orthomosaics) that are not intended for precise measurement purposes.

A single image from a UAV mounted camera only covers limited area. In many digital earth applications, it is therefore necessary to stitch hundreds or even thousands of images together to create a larger image that can provide good overall situational awareness. Mosaicking of UAV imagery usually requires extra information, such as camera calibration parameters, position and orientation data from GPS/IMU, ground control points (GCPs) or a reference map, to achieve accurate mosaicking results [1–7]. When GPS/IMU data are not accurate enough for direct orientation determination, pose estimation using a 3D reconstruction method is usually employed to refine the camera

poses [1,8–10], which introduces additional computational cost. In this paper, we propose a novel image mosaicking method without the requirements of camera calibration parameters, camera poses, or any 3D reconstruction procedure. Our method can obtain visually satisfactory mosaicking results automatically using only the raw 2D imagery captured from a UAV.

Our approach assumes that the following two conditions are satisfied:

(1) The camera lens has no evident distortion (or, in cases of significant distortion, lens distortion coefficients are available).
(2) The ground is approximately planar.

The main contribution of this paper is the presentation of a novel objective function for mosaicking UAV imagery. The objective function consists of two terms. The first term minimizes the sum of squared distances of matched feature pairs to guarantee the geometric alignment of the aerial images. The second term directly constrains the image transformation parameters by keeping all transformations as rigid as possible to avoid global distortion in the mosaicking result. Based on this proposed objective function, visually plausible mosaicking results for aerial images from UAVs can be obtained without the requirement for camera poses and parameters. If the distortion of the camera lens is significant, the approach requires distortion coefficients to first rectify the images. However, camera parameters such as focal length and principle point are not needed.

The proposed method is inspired by the embedded mesh deformation model, which was proposed by Sumner *et al.* [11]. The primary challenge of embedded deformation is to find a deformation model that is general enough to apply to any object embedded in space, yet still provides intuitive direct manipulation and natural feature preservation. In this model, a space deformation is defined by a collection of affine transformations. Each affine transformation induces a localized deformation on the nearby space. In Sumner *et al.*, the objective function ensures detail preservation by specifying that the affine transformations should be rotations and translations. Consequently, local features deform in as rigid a manner as possible. Based on the method of Sumner *et al.*, we introduce the local rigid deformation constraint to the problem of UAV image mosaicking to largely preserve the original shape of the objects in image. Consequently, global distortions in the mosaicking results are effectively alleviated. Compared with existing image-based mosaicking methods (e.g., [12]), our method is found to be more accurate and efficient. It also demonstrates improved robustness with respect to the selection of the reference image.

The remainder of this paper is organized as follows: Section 2 reviews related works. The proposed methodology is presented in Section 3, and experimental results and comparisons with an existing mosaicking algorithm are provided in Section 4. Finally, Section 5 concludes the paper.

2. Related Work

Existing mosaicking methods can be roughly divided into two classes: single-viewpoint mosaicking [13] and multi-viewpoint mosaicking [3,7,12,14]. Single-viewpoint mosaicking can be achieved by rotating a camera around its optical center. An excellent review of the single-viewpoint mosaicking approach is given in [15], but further consideration is out of the scope of this paper. In multi-viewpoint mosaicking, the motion of the camera is not constrained purely to rotation only. For example, in UAV applications, the mounted camera undergoes not only rotations but also translations, making the mosaicking of the images more complicated. When the camera undergoes general 3D motion (including both rotation and translation), the eight degrees-of-freedom (DOF) homography is a general alignment model for the mosaicking of planar scenes. However, when the camera undergoes pure rotation, instead of the general 8-DOF homography relating a pair of images, we can get 3-, 4-, or 5-DOF 3D rotation motion models corresponding to cases in which the focal length is known, fixed, or variable. Estimating the 3D rotation matrix (and, optionally, focal length) associated with each image is intrinsically more stable than estimating a full 8-DOF homography [13,15]. Thus, the case of the camera undergoing general motion is more challenging.

A single image captured from a typical UAV covers only a limited area on the ground. To enlarge the field of view, feature-based pairwise image registration can be applied to generate precise short-term mosaicking with almost no visual artifacts. Compared to linear-pushbroom sensor mosaicking [4], this method has the advantage that non-linear flight-paths and varying view directions can be handled effectively. However, stitching hundreds, or even thousands, of images would result in a significant drift of parameters due to accumulated errors. Therefore, to obtain large mosaics, geo-referenced satellite images of the whole region may be used as a reference image [2,5] and the smaller mosaics are overlaid on top of the reference map using manually determined control-points. A framework where online mosaicking is used to compute and improve the localization of a UAV is presented in [16]. In instances where a region of the mosaic is revisited (namely loop-closure), an Extended Kalman Filter (EKF) can be adopted to minimize the accumulated drift.

Another approach to avoid accumulated errors is the poses-based method [6,7], which uses extra sensors such as GPS and IMU to directly obtain camera poses. Furthermore, if the intrinsic camera parameters are known, the images can be stitched together. However, this method necessitates sensors with a high level of known internal geometry, or a number of well distributed GCPs to determine the camera poses [3].

When sensors with sufficiently high known internal precision or GCPs are not available, image registration information and camera poses are combined to obtain seamless stitching while maintaining global consistence [1,6,7]. To combine image-based and pose-based methods together, an objective function to maximize correlation in overlapping areas constrained by camera poses calculated from GPS/IMU data is defined in [7]. The method described in [1] employed an EKF to integrate GPS and IMU data to estimate the initial camera poses. The initial poses are used to triangulate the 3D position of terrain features, and also to provide an initial estimate for the bundle adjustment procedure [17]. An overview image of the terrain is constructed by projecting each image onto a ground plane.

Compared with most of the methods above, the proposed method is purely image-based. It does not require satellite imagery as a reference, GPS/IMU data, or GCPs. Our work is closely related to the work by Capel [12], which also only employs image information, applying a bundle adjustment approach [17] to simultaneously optimize the image transformations and the 2D coordinates of the features in the reference image. It is a similar approach to the bundle adjustment method employed in the structure-from-motion (SFM) pipeline. The latest image-based approach in [18] aims to handle large parallax, while our work emphasizes the case of a large number of images of approximately planar scenes.

3. Methodology

3.1. Objective Function

We utilize a feature-based method for image registration. An 8-DOF homography matrix, $\mathbf{H}_i$, represents the transformation parameters of the i^{th} image.

$$\mathbf{H}_i = \begin{pmatrix} a_i & b_i & e_i \\ c_i & d_i & f_i \\ g_i & h_i & 1 \end{pmatrix} \tag{1}$$

When setting g_i and h_i to zero, then $\mathbf{H}_i$ becomes an affine homography. When further assuming that the submatrix $\begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$ is an orthogonal matrix, then $\mathbf{H}_i$ becomes a rigid homography.

For a point $\mathbf{x}$ in an image, the transformed coordinates of the point given the transformation represented by the homography $\mathbf{H}$ is:

$$\mathbf{x}' = \mathbf{H}\mathbf{x} \tag{2}$$

We assume that N feature pairs are found among the M images. The transformation parameter of the i^{th} image is represented by $\mathbf{X}_i$ (a column vector that consists of the eight independent parameters of the homography $\mathbf{H}_i$). Let $\mathbf{X} = \left(\mathbf{X}_1^T, \mathbf{X}_2^T, \cdots, \mathbf{X}_M^T\right)^T$.

We construct an objective function $E(\mathbf{X})$ as follows:

$$E(\mathbf{X}) = Ecor(\mathbf{X}) + \omega Ereg(\mathbf{X}) \tag{3}$$

where $Ecor(\mathbf{X})$ is the energy term based on feature correspondences, which is to minimize the sum of the squared distances of the feature pairs and to geometrically align the images. It is defined as:

$$Ecor(\mathbf{X}) = \sum_{i=1}^{N} \mathbf{e}_i^T \mathbf{e}_i + \sum_{i=1}^{\widetilde{N}} \widetilde{\mathbf{e}}_i^T \widetilde{\mathbf{e}}_i \tag{4}$$

where $\mathbf{e}_i = T_m\left(\mathbf{p}_{i,m}\right) - T_n\left(\mathbf{p}_{i,n}\right)$, $1 \leqslant m, n \leqslant M$. T_m and T_n denote the transformations of $\mathbf{X}_m$ and $\mathbf{X}_n$, respectively. $\widetilde{\mathbf{e}}_i = T_{n\_ref}\left(\mathbf{p}_{i,n\_ref}\right) - \mathbf{p}_{i,n\_ref}$. $\left(\mathbf{p}_{i,m}, \mathbf{p}_{i,n}\right)$ represents the i^{th} feature pairs. n_ref denotes the index of the reference image. $\widetilde{N}$ is the number of feature correspondences in the reference image.

The effect of the second term in Equation (4) is to constrain the reference image to preserve the original shape.

The regularization term $E_{reg}(\mathbf{X})$, inspired by Sumner *et al.* [11], enables $\mathbf{X}_i$ to be as rigid as possible in order to avoid global distortion in the mosaicking result. $E_{reg}(\mathbf{X})$ is defined as:

$$Ereg(\mathbf{X}) = \sum_{i=1}^{N} p_i E_{reg}(\mathbf{X}_i) \tag{5}$$

$$\begin{aligned} E_{reg}(\mathbf{X}_i) &= (a_i b_i + c_i d_i)^2 \\ &+ \left(a_i{}^2 + c_i{}^2 - 1\right)^2 \\ &+ \left(b_i{}^2 + d_i{}^2 - 1\right)^2 \\ &+ \left(g_i{}^2 + h_i{}^2\right)^2 \end{aligned} \tag{6}$$

where p_i is a weight factor, the value for which is determined by the number of the feature correspondences in the i^{th} image. Note that as the number of feature correspondences in an image increases, the number of the terms with respect to the image in $E_{cor}(\mathbf{X})$ increases accordingly. Therefore, the weight p_i should be larger in $E_{reg}(\mathbf{X})$.

ω is the constant weight factor. Currently, its value is determined heuristically. When ω is small, the mosaicking result appears seamless, but may contain global distortion. With an increase in the value of ω, global distortion is eliminated, but the alignment error of the feature correspondences may simultaneously increase. It is therefore necessary to choose an appropriate value for ω.

When the pitch and roll angles of the UAV vary during flight, the regularization term tries to maintain the original shape of the objects in the images. Meanwhile, the correspondence term tries to change the shape of the images to align them geometrically. The combined effect of the two terms results in an optimized mosaicking result.

There are $8M$ unknown parameters in Equation (3). Equation (3) is a typical non-linear least squares problem that can be solved by the Levenberg–Marquardt (LM) algorithm [19].

To solve Equation (3) using the LM algorithm, the Jacobian matrix has the size of $2N + 4M$ rows and $8M$ columns. When M and N are large, for example as in the experiments shown below where $M = 59$, $N = 220{,}977$, the Jacobian of 444,318 rows and 4728 columns becomes a huge matrix, which makes the execution of the standard LM algorithm impossible on a regular PC. To conserve

memory and to speed up the computation, we therefore employ the sparse LM algorithm [20] to solve Equation (3).

3.2. Mosaicking Work-Flow

The objective function proposed in Section 3.1 is embedded into a full mosaicking workflow as illustrated in Figure 1. It consists of four steps, as follows:

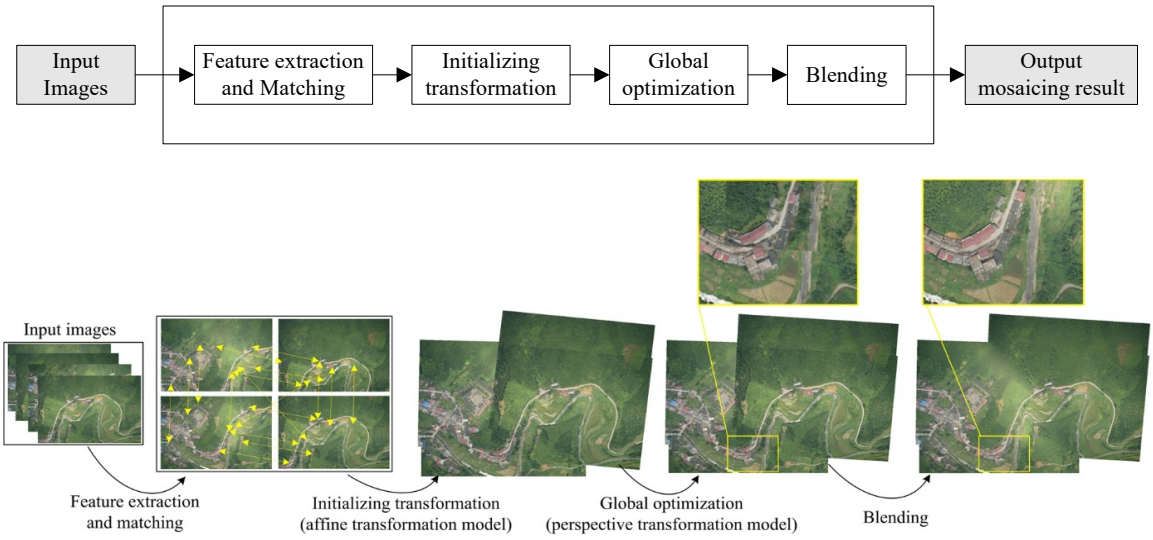


Figure 1. The pipeline for the proposed mosaicking algorithm.

3.2.1. Feature Extraction and Matching

Considering the varied height and pose of a UAV mounted camera during flight, we extract Scale-Invariant Feature Transform (SIFT) features [21], which are, to a certain extent, invariant to scale changes and affine distortions and use the Fast Library for Approximate Nearest Neighbors (FLANN) [22] to match the features. Subsequently, the RANdom SAmple Consensus (RANSAC) algorithm [23] is utilized to remove false matches.

3.2.2. Initialization of the Transformation Parameters

The image plane of the reference image is treated as the reference plane of the initial mosaicking result, with its homography set to an identity matrix. The reference image can be selected at random, but in practice we select an image captured by the camera for which the principal axis is approximately perpendicular to the ground as the reference. To obtain the initial transformations of other images, we simplify the transformation represented by an 8-DOF homography to the 6-DOF affine transformation and neglect the regularization term $E_{reg}(\mathbf{X})$ in Equation (3). Equation (3) then turns into a linear least-squares problem that is relatively simple to solve.

3.2.3. Global Optimization

After obtaining the feature correspondences in Step (1) and the initial transformations in Step (2), the sparse LM algorithm is deployed to optimize Equation (3).

3.2.4. Blending

Finally, based on the estimated geometrical transformation parameters for all images, we use the multi-band blending method [13] to process the mosaicking artifacts generated by alignment errors in geometry and differences in intensity among the images. Images with excessive overlap are not used to create the final results since their inclusion adds nothing to the final product and their exclusion saves computer memory and CPU time in the multi-band image blending step, as illustrated in Figure 2.

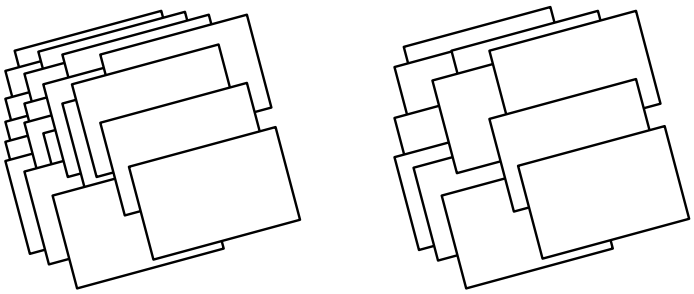


Figure 2. Images with excessive overlap (e.g., **left**) are not used to create the final result.

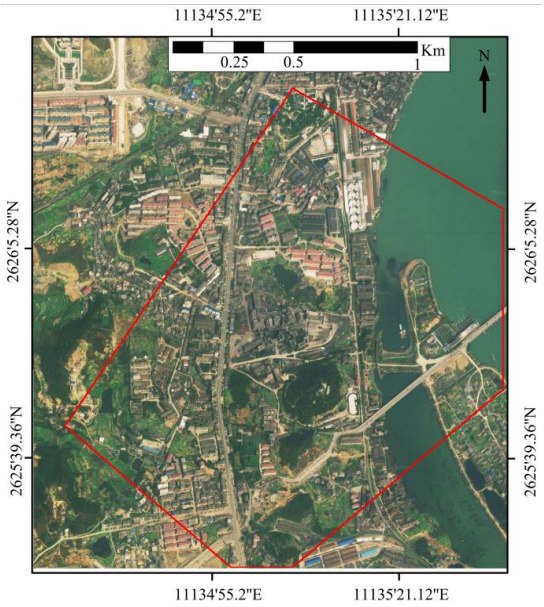
Note that no camera parameter or pose data were used in our experiments. ω in Equation (3) is set to 4000 for all tests.

4. Datasets

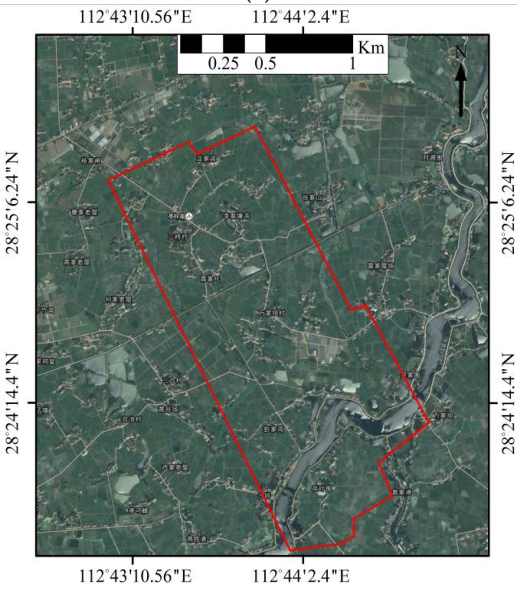
Four datasets (datasets can be obtained by sending an e-mail to the author) were used to test the proposed image mosaicking objective function and workflow. The images in Datasets 1, 2 and 3 were captured by a UAV (KC1600 [24]) over Yongzhou city, Hechi and HeJiangdong villages of Hunan province, China, respectively. Dataset 1, Dataset 2 and Dataset 3 contain 61, 182, and 51 images, respectively, each with resolution of 3680 by 2456 pixels. The average flying heights were 558 m, 405 m and 988 m, respectively. Although the UAV was guided by GPS, GPS data were not used in our mosaicking experiments. Digital orthophoto maps (DOMs) of the three datasets, each with ground resolution of 20 cm generated by DPGrid [25], were used as ground-truth to quantitatively evaluate the proposed algorithm.

The images in Dataset 4 were from Pix4D [26], and are openly available on the Internet. In total, 591 images were used for mosaicking purposes. The images (4000 $\times$ 3000 pixels) in the dataset were captured by a Panasonic DMC-GF1 camera with a 20 mm focal length lens. The average flying height was 742 m. In addition, the dataset provided by Pix4D contains a DOM with a ground resolution of 20 centimeters, which was used to quantitatively evaluate the proposed algorithm.

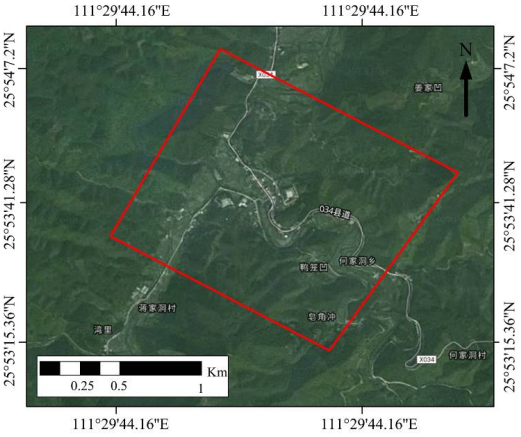
The sites of the four datasets are shown in Figure 3. The variations in terrain elevation for Dataset 1 and Dataset 3 are obvious, while for the other two datasets the variations are relatively small, as shown in Figure 4. The elevation of the terrain in Dataset 1 ranges from 62 m to 169 m. The differences in elevation for Dataset 2 are less than 25 m. The variations in elevation for Dataset 3 are the largest. It ranges from 354 m to 507 m. The Digital Elevation Model (DEM) of Dataset 4 is from Pix4D, as shown in Figure 4d. Although the exact elevation values are not shown, we can find out that the terrain is relatively flat.



(**a**)

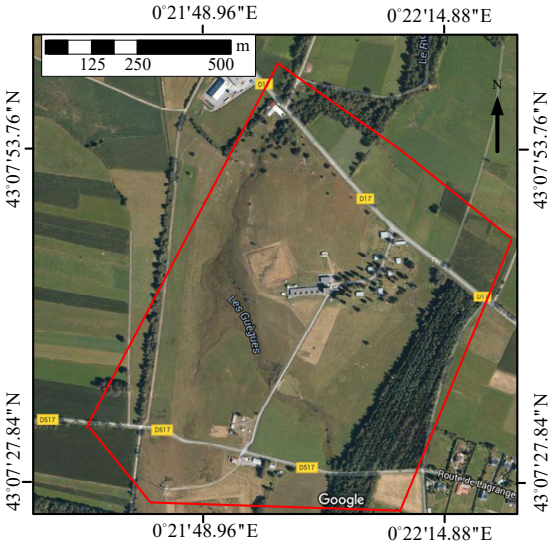


(**b**)



(**c**)

Figure 3. *Cont*.



(**d**)

Figure 3. Overview imagery of the three test sites: (**a**) Dataset 1 (BaiduMap [27]); (**b**) Dataset 2 (GoogleMap [28]); (**c**) Dataset 3 (GoogleMap [28]); and (**d**) Dataset 4 (GoogleMap [28]).

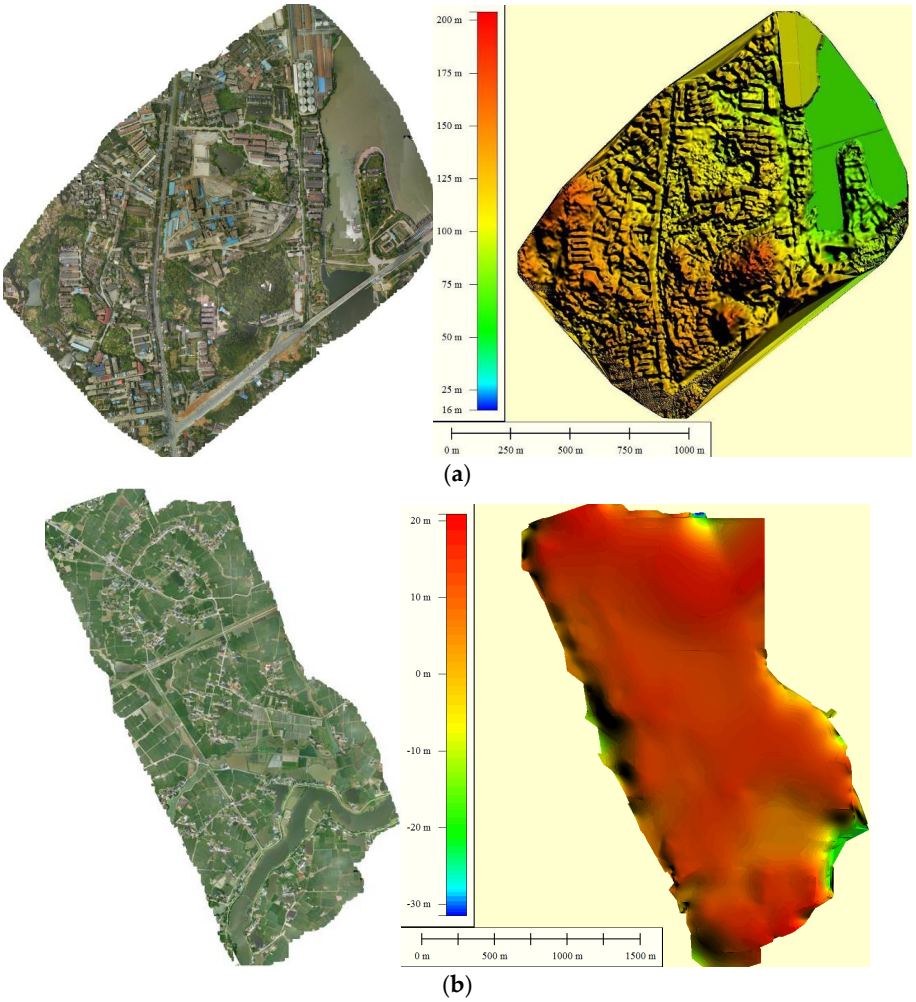
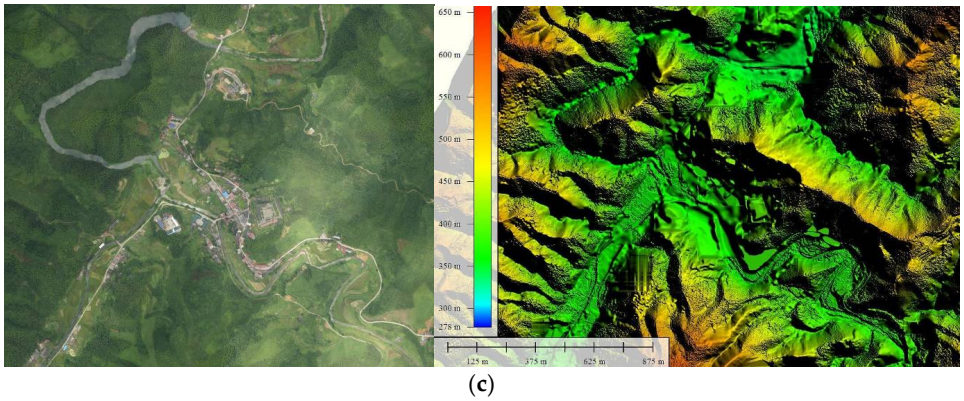
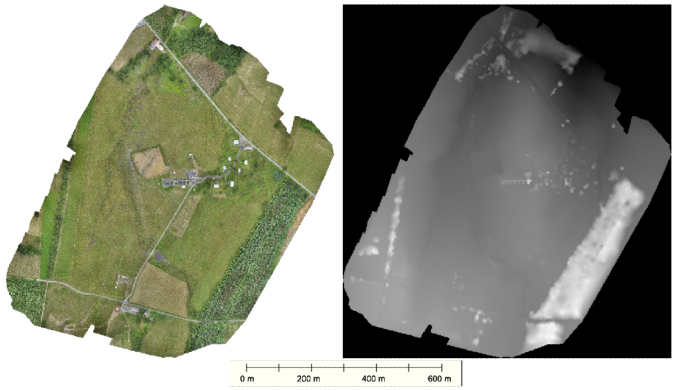


Figure 4. *Cont.*



(**c**)



(**d**)

Figure 4. Digital Orthophoto Maps (DOMs) and Digital Elevation Models (DEMs) of the four datasets: (**a**) Dataset 1; (**b**) Dataset 2; (**c**) Dataset 3; and (**d**) Dataset 4 (from Pix4D [26]).

5. Results and Discussion

Figures 5–8 illustrate the mosaicking results of the proposed algorithm for the four datasets. The results are found to be visually satisfactory for all of the experimental datasets. All of the mosaics produced by the proposed method are seamless and display no obvious global distortion.

(**a**) (**b**)

Figure 5. Mosaicking result of Dataset 1: (**a**) mosaicking result via Capel's method; and (**b**) mosaicking result using the proposed algorithm.

(**a**)

(**b**)

Figure 6. Mosaicking result of Dataset 2: (**a**) mosaicking result via Capel's method; and (**b**) mosaicking result using the proposed algorithm.

(**a**)

Figure 7. *Cont.*

(**b**)

Figure 7. Mosaicking result of Dataset 3: (**a**) mosaicking result via Capel's method; and (**b**) mosaicking result using the proposed algorithm.

(**a**)

(**b**)

Figure 8. Mosaicking result of Dataset 4: (**a**) mosaicking result via Capel's method; and (**b**) mosaicking result using the proposed algorithm.

Although image mosaicking has been studied for decades, many algorithms assume the camera rotates about the optical center [13,15]. For image mosaicking in UAV applications, the camera experiences not only rotations but also translations. Therefore, although they are purely image-based, these types of algorithms [13,15] cannot be directly compared to the proposed method. To obtain accurate mosaicking results, many algorithms for UAV imagery integrate the image with other information, such as data from GPS/IMU, or with GCPs [1,3,8,9]. By contrast, this additional information is not required in the presented method.

We compare the proposed method with that of Capel in aspects of accuracy and efficiency. We implement Capel's method incrementally, *i.e.*, when a new image is added, all the image transformation parameters and 2D feature locations are re-optimized simultaneously. At each iteration, the image with the largest overlap to the current mosaic is added into the optimization procedure. The steps of feature extraction and blending are the same for the two methods under comparison.

5.1. Accuracy

Thirty corresponding well-distributed check points were manually selected in both the ground-truth and the mosaicked image for each dataset. A four degrees-of-freedom 2D similarity transformation model was used to align the two sets of control points. The Root-Mean-Square (RMS) error, minimum error and maximum error of the alignment are analyzed. Note that values of various errors are converted to metric units using the ground resolutions of the ground-truth images. In each comparison, the same reference image is used for both methods.

Table 1 shows the mosaicking errors for both methods, including the RMS errors, minimum errors and maximum errors. For Dataset 1 and Dataset 4, the RMS error of the proposed method is twice as good as that resulting from Capel's approach. For Dataset 2, the improvement is fourfold.

Table 1. Mosaicking error for proposed algorithm and Capel's approach.

	Proposed Algorithm (m)			**Capel's Method (m)**		
	RMS	**MIN**	**MAX**	**RMS**	**MIN**	**MAX**
Dataset 1	10.5	5.6	16.5	25.8	3.7	39.8
Dataset 2	5.2	0.5	12.8	22.7	1.0	42.5
Dataset 3	13.4	3.9	24.6	22.5	10.8	57.2
Dataset 4	7.9	1.0	15.5	18.1	8.1	32.6

5.2. Efficiency

The run time for the two algorithms when processing three datasets is shown in Table 2. It shows that the proposed method is more than one order of magnitude faster for all three datasets. Capel's method needs to optimize $8M + 2n$ unknown parameters, where M is the number of images and n is the number of feature points. Generally, it is the case that $n >> M$. By contrast, our proposed method only needs to optimize $8M$ unknown parameters, and is therefore computationally faster. Moreover, our proposed algorithm can be initialized efficiently by solving a linear least-squares problem. Conversely, Capel's method has to optimize the transformation parameters of the images and the 2D feature locations incrementally in an iterative manner, rendering our proposed method theoretically more efficient.

Table 2. Run time for proposed algorithm and Capel's approach.

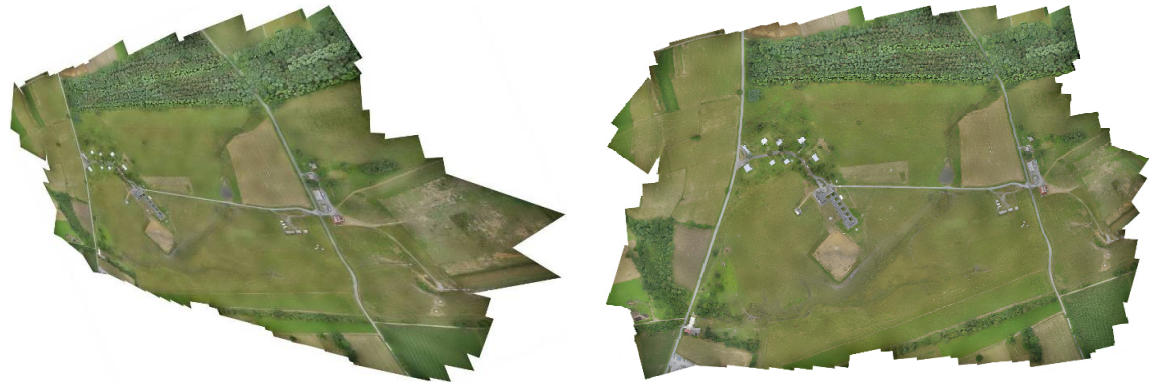
	Proposed Algorithm (s)	**Capel's Method (s)**
Dataset 1	2.3	43.1
Dataset 2	35.1	1145.5
Dataset 3	1.4	12.6
Dataset 4	36.4	1482.6

5.3. Impact of the Reference Image

The image illustrated in Figure 9a was selected as the reference image for the mosaic in Dataset 4 (see Figure 8). To assess the variability in the result based on the quality of the reference image, we repeated the mosaicking process using a sub-optimal reference image. The mosaicking results from the two methods when using the image illustrated in Figure 9b as the reference image, are shown in Figure 10. Comparing Figure 8 with Figure 10, we see that the results of Capel's method vary significantly, while the results of the proposed method are relatively stable. Moreover, the relative performances of the two methods for Dataset 4 are quantitatively shown in Table 3. We can see that Capel's method is highly affected by the selection of the reference image, displaying a six times degradation in error. By contrast, the proposed method is more robust to the selection of the reference image.

(**a**) (**b**)

Figure 9. The two different reference images in Dataset 4. (**a**) The reference image for the results in Figure 8; (**b**) The reference image for the results in Figure 10.



(**a**) (**b**)

Figure 10. Mosaicking result of Dataset 4 obtained using reference image shown in Figure 9b: (**a**) mosaicking result using Capel's method; and (**b**) mosaicking result using the proposed algorithm.

Table 3. Mosaicking errors for Dataset 4 obtained using different methods and reference images.

	Proposed Algorithm (m)	**Capel's Method (m)**
Figure 9a	7.9	18.1
Figure 9b	7.8	105.1

The results of Capel's method heavily rely upon the reference image. Once the image plane of the reference image deviates from the image plane of the ground, the final results of Capel's method

distort considerably. This will happen, for example, when the camera tilts significantly when capturing the reference image or when the ground elevation of the region covered by the reference image varies considerably. However, the impact of the reference image is minimal for the proposed method as a result of the regularization term in the objective function.

5.4. Impact of Variations in Terrain Elevation

The variations in terrain elevation for Dataset 1 and Dataset 3 are obvious, as shown in Figure 4, while the terrain for Dataset 2 and Dataset 4 are both relatively flat. The errors reported for Dataset 1 and Dataset 3 are significantly larger than Dataset 2 and Dataset 4, as shown in Table 1. The results show that the variations in terrain elevation adversely affect the mosaicking accuracy.

For the case of a camera undergoing purely rotation, two images of a 3D scene can be aligned seamlessly with the 8-DOF homography matrix [15]. However, for the case of a camera undergoing general 3D motion (including both rotation and translation), the parallax will induce artifacts in the mosaics (see Figure 1). By using the multi-band blending approach [13], the artifacts can be effectively eliminated, as shown in Figure 1.

6. Conclusions

This paper has introduced a novel mosaicking method for UAV imagery that was inspired by the non-rigid mesh deformation. The novel objective function consists of a feature correspondence energy term and a regularization term. The first term minimizes the sum of the squared distances between matched feature pairs to align the images geometrically. The second term constrains the image transformation parameters directly by keeping all transformations as rigid as possible to avoid global distortion in the final mosaic.

Experiments have demonstrated that the proposed method can effectively avoid global distortions and results in visually satisfactory mosaics for several different datasets. Compared with another popular existing image-based mosaicking algorithm [12], our approach was found to be more accurate (more than twice as high) and efficient (more than one order of magnitude faster). The proposed method is more efficient because it only needs to optimize $8M$ unknown parameters, *versus* $8M + 2n$ (generally, $n >> M$) unknown parameters in [12]. Moreover, the proposed objective function can be initialized efficiently by solving a linear least-squares problem, which provides the initial transformation parameters for the sparse LM algorithm. Conversely, the method in [12] has to optimize the transformation parameters of the images and the 2D feature locations incrementally in an iterative manner. Moreover, our approach is insensitive to selection of the reference image. Although we make the assumption that the ground is approximately planar, our method can adapt to non-planar terrain, at least to a certain degree. For featureless scenes (e.g., surfaces of large lakes or other water bodies), the image matching algorithm may fail since the proposed method is purely image-based.

When there is no demand for high accuracy mosaicking, especially when GPS data are either unavailable or not aligned to the imagery, the approach presented provides an effective and practical alternative method for mosaicking UAV imagery.

Future research will concentrate on the theoretical issues regarding the improvement of mosaicking, such as how to determine the optimal value of ω in Equation (3), this being a key factor in the mosaicking process. In addition, how best to develop the method to adapt to mountainous regions is also a very valuable consideration in engineering applications, and will be studied.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (No. 61402489), the Ph.D. Programs Foundation of the Ministry of Education of China (No. 20130161120028) and a China Scholarship Council Visiting Fellowship (No. 201308430177). The authors would also like to thank Pix4D for their public datasets, Baidu, and Google for providing satellite imagery. Finally, we would like to thank Qifeng Yu for fruitful discussions and Chenpeng Tong of VNUSK for providing the DOMs and DEMs of Dataset 1, Dataset 2, and Dataset 3.

Author Contributions: Yuhua Xu undertook the research that is reported in this paper as part of his post-doctoral studies at National University of Defense Technology of China. This included original development of the new mosaicking method and all other aspects including development and implementation of it. He also carried out the bulk of the manuscript preparation.

Xiaohu Zhang supervised Yuhua Xu and Jianliang Ou's post-doctoral research and guided the research development at National University of Defense Technology from 2010 to 2015. Their research is mainly concerned with image feature extraction, correspondence matching and 3D reconstruction.

Hu He took part in the discussion of the proposed method with Yuhua Xu. He helped analyze the experimental result, and made contributions to manuscript preparation.

Jon Mills supervised Jianliang Ou's academic visit at Newcastle University, UK, from 2014 to 2015, and supported his UAV image processing while in Newcastle. Jon Mills also contributed to the development and refinement of the manuscript.

Jianliang Ou joined the method development with Yuhua Xu, supported acquisition of UAV Dataset 1 and Dataset 2, and helped in the mosaicking accuracy appraisal. He undertook earlier UAV image matching research in his post-doctoral period at the National University of Defense Technology. Jianliang Ou also contributed to the development and refinement of the manuscript.

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

References

1. Bryson, M.; Reid, A.; Ramos, F.; Sukkarieh, S. Airborne vision-based mapping and classification of large farmland environments. *J. Field Robot.* **2010**, *5*, 632–655. [CrossRef]
2. Heinze, N.; Esswein, M.; Krüger, W.; Saur, G. Automatic image exploitation system for small UAVs. *Proc. SPIE 6946* **2008**. [CrossRef]
3. Zhou, G. Near real-time orthorectification and mosaic of small UAV video flow for time-critical event response. *IEEE Trans. Geosci. Remote Sens.* **2009**, *3*, 739–747. [CrossRef]
4. Chen, C.S.; Chen, Y.T.; Huang, F. Stitching and reconstruction of linear-pushbroom panoramic images for planar scenes. In Proceedings of the European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004.
5. Se, S.; Firoozfam, P.; Goldstein, N.; Wu, L.; Dutkiewicz, M.; Pace, P.; Naud, J.P. Automated UAV-based mapping for airborne reconnaissance and video exploitation. *Proc. SPIE 7307* **2009**. [CrossRef]
6. Stephen, H. Geocoded terrestrial mosaics using pose sensors and video registration. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, 8–14 December 2001.
7. Yahyanejad, S.; Wischounig-Strucl, D.; Quaritsch, M.; Rin, B. Incremental mosaicking of images from autonomous small UAVs. In Proceedings of the 7th International Conference on Advanced Video and Signal-Based Surveillance, Boston, MA, USA, 29 August–1 September 2010.
8. Turner, D.; Lucieer, A.; Watson, C. 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.* **2012**, *5*, 1392–1410. [CrossRef]
9. Turner, D.; Lucieer, A.; Wallace, L. Direct georeferencing of ultrahigh-resolution UAV imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *5*, 2738–2745. [CrossRef]
10. Yahyanejad, S.; Rinner, B. A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 189–202. [CrossRef]
11. Sumner, R.W.; Schmid, J.; Pauly, M. Embedded deformation for shape manipulation. *ACM Trans. Graph.* **2007**, *3*, 1–7.
12. Capel, D.P. Image mosaicing and super-resolution. Ph.D. Thesis, University of Oxford, Oxford, UK, 2001.
13. Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *1*, 59–73. [CrossRef]
14. Agarwala, A.; Agrawala, M.; Cohen, M.; Salesin, D.; Szeliski, R. Photographing long scenes with multi-viewpoint panoramas. *ACM Trans. Graph.* **2006**, *25*, 853–861. [CrossRef]
15. Szeliski, R. Image alignment and stitching: A tutorial. *Found. Trends Comput. Graph. Vis.* **2006**, *2*, 1–87. [CrossRef]
16. Caballero, F.; Merino, L.; Ferruz, J.; Ollero, A. Unmanned aerial vehicle localization based on monocular vision and online mosaicking. *J. Intell. Robot. Syst.* **2009**, *4*, 323–343. [CrossRef]

17. Triggs, B.; McLauchlan, P.; Hartley, R.; Fitzgibbon, A.W. Bundle adjustment—A modern synthesis. In Proceedings of the International Workshop on Vision Algorithms, Corfu, Greece, 21–22 September 1999.
18. Lin, C.C.; Pankanti, S.U.; Ramamurthy, K.N.; 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.
19. Madsen, K.; Nielsen, H.B.; Tingleff, O. *Methods for Non-Linear Least Squares Problems*, 2nd ed.; Informatics and Mathematical Modelling, Technical University of Denmark: Lyngby, Denmark, 2004; pp. 24–29.
20. Lourakis, M.I.A. Sparse non-linear least squares optimization for geometric vision. In Proceedings of the European Conference on Computer Vision, Hersonissos, Crete, Greece, 5–11 September 2010.
21. Lowe, D.G. Distinctive image features from scale-invariant key points. *Int. J. Comput. Vis.* **2004**, *2*, 91–110. [CrossRef]
22. Muja, M.; Lowe, D.G. Fast approximate nearest neighbors with automatic algorithm configuration. In Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP'09), Lisboa, Portugal, 5–8 February 2009.
23. Fischler, M.; Bolles, R. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Comm. ACM* **1981**, *6*, 381–395. [CrossRef]
24. Wuhan Intelligent Bird Co., Ltd. Available online: http://www.aibird.com/ (accessed on 18 August 2014).
25. DPGrid. Available online: http://www.supresoft.com.cn/ (accessed on 15 July 2015).
26. Sample Image Data of Pix4D. Available online: http://www.pix4d.com/ (accessed on 12 February 2015).
27. BaiduMap. Available online: http://map.baidu.com/ (accessed on 20 April 2015).
28. GoogleMap. Available online: https://maps.google.com/ (accessed on 25 May 2015).