

Article

A Simplified Method for UAV Multispectral Images Mosaicking

Xiang Ren ¹, Min Sun ^{1,*}, Xianfeng Zhang ¹ and Lei Liu ^{1,2}

¹ Institute of Remote Sensing and GIS, Peking University, Beijing 100871, China; xelmirage@pku.edu.cn (X.R.); xfzhang@pku.edu.cn (X.Z.); liuleiargis@pku.edu.cn (L.L.)

² The 61206 Troop, PLA, Beijing 100042, China

* Correspondence: sunmin@pku.edu.cn; Tel.: +86-010-6276-4484

Received: 31 July 2017; Accepted: 13 September 2017; Published: 17 September 2017

Abstract: This paper presents a method for mosaicking unmanned aerial vehicle (UAV) multispectral images. The main purpose of the proposed method is to reduce spatial distortion in the mosaicking process and increase robustness and the speed of the operation. Most UAV multispectral images have multiple bands, and in every band, ground targets have a variety of reflection characteristics that will result in diverse feature quality for feature matching. In this research, an information entropy-based evaluation method is used to select the optimal band for feature matching among the UAV images. To produce more robust matching results for the following alignment step, the evaluation method takes the contrast and spatial distribution of the feature points into consideration at the same time. In most common image mosaicking processes, the digital orthophoto map (DOM) is generated to achieve maximum spatial accuracy. During this process, the original image data will experience considerable irregular resampling, and the process is also unstable in some circumstances. The alignment step uses a simplified projection model that treats the ground as planar is provided, by which the alignment parameters are applied directly to the images instead of generating 3D points, to avoid irregular resampling and unstable 3D reconstruction. The proposed method is proved to be more efficient and accurate and has lower spectral distortion than state-of-the-art mosaicking software.

Keywords: Unmanned Aerial Vehicle; image mosaicking; multispectral; feature matching; optimal band

1. Introduction

Unmanned Aerial Vehicle (UAV) remote sensing is blooming for its low-cost, high resolution, and great flexibility. In such a situation, UAV multispectral remote sensing has become a new hot spot topic for researchers in the remote sensing community. UAV remote sensing can be used in fields such as photogrammetry, vegetation monitoring, and disaster management [1–3]. Agriculture is a main application area of vegetation monitoring based on UAV multispectral remote sensing. For example, Honkavaara used a Fabry-Perot Interferometer-based (FPI) spectral camera to monitor the growth of crops [4]; Gomez-Candon [5] used a quadcopter and a consumer-level digital camera to estimate the status of wheat growth; Sona [6] captured data of bare land and land with vegetation cover and evaluated the application potential of these data. However, most of these researches are based on consumer-level RGB cameras, and they do not have NIR bands, which greatly limits the vegetation characterization processes. Some consumer-level digital cameras can be modified to capture NIR data, but the spectral performance is limited. Thus, it is necessary to capture multispectral data using multispectral cameras that have more than three bands [7]. Moreover, studies on the corresponding data processing methods are required in UAV remote sensing applications.

UAVs mainly work in a relatively low altitude and, as a result, the area that a single image can cover is relatively small. Consequently, image mosaicking is an important technique for utilizing

UAV multispectral image data. In the process of UAV image mosaicking, accurate position and pose relations among images need to be determined. The process recovering positions and the poses of images is called the geometric correction. The geometric correction can be divided into image-to-map correction and image-to-image correction. Image-to-map correction registers the images in question to existing reference data, such as the Digital Surface Model (DSM) [8–11] or Ground Control Points (GCPs) [12,13]. Image-to-image correction recovers position and pose relations of images using tie points among the images [14–16], and is widely used in UAV data processing methods [17]. However, the image textures are usually weak or repetitive in most vegetated areas such as farmlands, which make it difficult to generate sufficient correct tie points. Generally speaking, ground objects have diverse reflective characteristics with various wavelengths of light and, thus, bands of one image may have various levels of feature richness. Therefore, to choose the right band for feature matching is significant for quality geometric correction and image mosaicking.

On the other hand, existing geometric correction methods such as SFM-based 3D reconstruction work on the regular RGB images and pay more attention to the geometric accuracy of the mosaicking results [2,3], thus, the correcting process is very computationally expensive. For vegetation monitoring tasks like agricultural applications, the adequate spectral performance needs to be considered seriously [18]. However, the characteristics of multispectral image data are not well considered in these methods. The SFM-based 3D reconstruction methods involve considerable irregular resampling [19], which may cause significant spectral distortion to the image data. Thus, it is significant and necessary for the UAV multispectral image mosaicking process to keep the original spectral information as much as possible and promote the efficiency of the geometric correction process.

In this paper, we will come up with a new method to determine the optimal band for feature matching so as to obtain sufficient correct tie points on the UAV multispectral images in agricultural areas. In order to avoid computationally expensive 3D reconstruction and irregular resampling, we put forward a simplified projection model, which can speed up and satisfy requirements of UAV multispectral image mosaicking in most agricultural areas.

2. Related Research

Image-to-image correction is widely used in UAV image mosaicking processes and can be divided into feature matching and transformation estimation [16].

Feature matching is a key step in image mosaicking. Scale invariant feature transform (SIFT) [20] is a feature descriptor that is invariant to image scaling and rotation in most cases and also has tolerance to illumination and 3D camera viewpoint change. It is widely used in various kinds of applications that need to match features between images, including image mosaicking. Other feature-matching algorithms such as PCA-SIFT [21], SURF [22], and ORB [23] are also applied to various fields depending on their characteristics. However, Mikolajczyk and Schmid [24] have proved that SIFT-based descriptors have a better performance than other descriptors in image feature-matching.

UAV multispectral images usually have multiple bands, and the selected band used in feature matching has an obvious influence on the matching result directly [25,26]. As the ground targets have various reflection characteristics, the environmental conditions, such as the atmosphere and the relative positions of the sun, the targets, and the sensor may also change. Thus, an optimal band for feature matching may vary among flights. Most current research simply uses the methods for regular RGB images, or directly designates a band without careful selection. For example, Navia [27] discussed the application of UAV multispectral image mosaicking techniques in the agriculture field and raised a mosaicking method based on the Lucas-Kanade-matching algorithm [28]. However, this research was based on a camera that was modified from a consumer-level RGB camera that had only three bands, and choosing an optimal band to conduct feature matching was not discussed. Du [29] directly used the NIR band for feature matching, and the variation of feature quality among images was not taken into consideration. For the mosaic data from the AISA Eagle hyperspectral sensor, YK Han [30] used the SIFT matching result to select the optimal band for feature matching. The weakness of this research

is that the optimal band is known after the feature points of all the bands are extracted and matched, and the spatial distribution and quality of the feature points are not taken into consideration.

The transformation estimation uses the matched point pairs to estimate the transformations among the images. In the process, the projection model is needed to describe the motion of the camera. For most regular frame cameras, the model is the perspective projection model. By using the perspective projection model, the camera motion can be restored from the image matching results, and the 3D positions of the matched points can also be estimated if necessary. In some studies, the affine transformation is used to simplify the estimation process [29]. In the method, affine transformation, the eight-parameter projective model, and the thin-plate spline model were applied to characterize the locations and poses of the images. However the affine transformation used in the research work was a general affine transformation that includes some unnecessary transformation for UAV multispectral images, and how to choose an optimal band to perform feature matching in the feature matching step was not well discussed. Li [31] combined viewport reselection and SFM to mosaic UAV multispectral images. In sparse farmland, this algorithm could improve spatial accuracy by distinguishing crop canopy and ground, but the loss of spectral accuracy was still not considered.

3. Methodology

3.1. Selecting the Optimal Band for Feature Matching

In our proposed method, SIFT is used to match feature points. Since feature matching algorithms such as SIFT mainly work on a single band grayscale image, it is necessary to select one optimal band for feature matching by evaluating all the bands of the image data. The way to evaluate the quality of a band for feature matching is based on the low-contrast point kick-out method in SIFT. The following equation is used in SIFT to describe the local contrast level of a keypoint.

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \quad (1)$$

where D is the first term of the Taylor expansion of the scale-space function $D(x, y, \sigma)$, and \hat{x} is the offset of the extremum from the sample point. By computing the average $D(\hat{x})$ of the keypoints, the feature richness and quality of an image can be evaluated. When comparing the average $D(\hat{x})$ of each band, the band that has the bigger average $D(\hat{x})$ is more suitable for feature matching.

However, the purpose of matching images is to reconstruct the accurate positions and poses of UAV images. If the spatial distribution of the feature points is non-homogeneous, the process of solving the positions and poses of the images may be unstable. The spatial distribution of the feature points also needs to be considered, besides the average quality value. The information entropy-based equilibrium degree [32] is often used to describe the distribution of a point set:

$$E = \frac{J}{J_{max}} = \frac{-\sum_{i=1}^N \frac{n_{points}^i}{\sum n_{points}^i} \ln \frac{n_{points}^i}{\sum n_{points}^i}}{\ln N} \quad (2)$$

where n_{points}^i is the number of points in the i -th sub area of the image, $\sum n_{points}^i$ is the total point number in the image, and N is the number of sub areas. $J = -\sum_{i=1}^N \frac{n_{points}^i}{\sum n_{points}^i} \ln \frac{n_{points}^i}{\sum n_{points}^i}$ is the information entropy of the point set when the image is divided into N sub areas, and $J_{max} = \ln N$ is the maximum of the information entropy. The unit of E is nat (natural unit) when using \ln .

As Figure 1 shows, to calculate the equilibrium degree, the image is divided into many sub areas. Since the aspect ratio of the image data is 5 to 4, the image can be split into 20 square sub areas, namely, in this case, $N = 20$. The product of the average $D(\hat{x})$ of the points and the information entropy-based equilibrium degree is used to describe the feature quality of a grayscale image in Equation (3).

$$Q = D(\hat{x})_{sum} \cdot E = \sum_{k=1}^M D(\hat{x})_k \cdot \frac{-\sum_{i=1}^N \frac{n_{points}^i}{\sum n_{points}^i} \ln \frac{n_{points}^i}{\sum n_{points}^i}}{\ln N} \quad (3)$$

where M is the number of the keypoints that can be detected, and $D(\hat{x})_{sum} = \sum_{k=1}^M D(\hat{x})_k$ is the sum of local contrast levels of the keypoints.

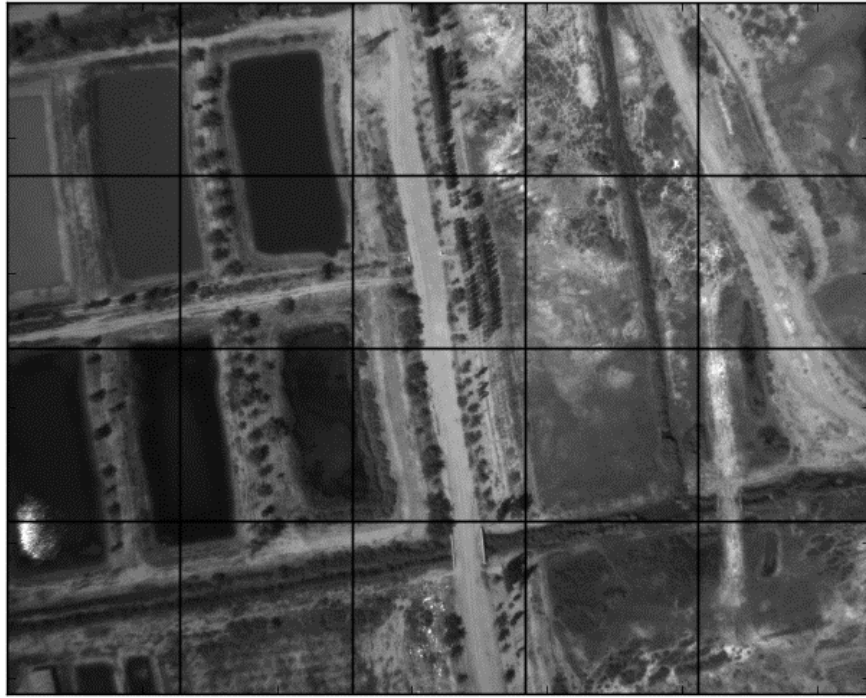


Figure 1. Image sub areas created from original image.

For a set of images, the average Q of the j -th band can be computed using Equation (4):

$$\overline{Q_{band-j}} = \frac{1}{m} \sum_{i_{im}=1}^m (Q_{band-j})_{i_{im}} \quad (4)$$

where $(Q_{band-j})_{i_{im}}$ is the feature quality of the j -th band in the i_{im} -th image, and m is the number of images that used to evaluate $\overline{Q_{band-j}}$. To reduce the time consumption of evaluating the $\overline{Q_{band-j}}$ of every band, the evaluation can be performed not on all the images. A subset of all the images can be used depending on the total number of images, for example, selecting one image out of every two to five images. The band that has the highest $\overline{Q_{band-j}}$ is considered to have better performance in the feature matching process.

In the following experiments, we discovered that the value $\overline{Q_{band-j}}$ could be used to pick up an optimal band to perform feature matching efficiently. However, it is highly possible that some individual images will have a different optimal band rather than the $\overline{Q_{band-j}}$ indicated. Taking this situation into consideration, a dynamic matching strategy is used. For a certain image pair, we denote the matched point pair set as P_0 ; if the size of the matched point pairs $size(P_0)$ are less than a threshold T_{pairs} , the optimal band only for these two images will be re-evaluated, and the two images will be re-matched using the new optimal band. The new matched point pair set is P_1 ; if $size(P_0 + P_1)$ exceeds T_{pairs} , the matching process of these two images is completed and the matching result is $P_0 + P_1$. If $size(P_0 + P_1)$ is still less than T_{pairs} , the matching process will be going on using the second band in

the re-evaluating result, until $size(P_0 + P_1 \dots)$ exceeds T_{pairs} , or all the bands of the two images have been matched.

3.2. The Simplified Projection Model

Limited by cost and weight, the inertial measurement unit (IMU) sensors on the UAVs have difficulty in providing accurate pose data for the image mosaicking. After feature matching, we have to restore the accurate positions and poses of the images using matching results. Fortunately, the GPS coordinates and pose data can be used for initializing the process. For most frame cameras, the projection model is [33]:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = KR \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = KR \begin{bmatrix} X_p - X_c \\ Y_p - Y_c \\ Z_p - Z_c \end{bmatrix}$$

where $[X_c, Y_c, Z_c]^T$ is the position of the projection center, $[X_p, Y_p, Z_p]^T$ is position of the ground point, and $[x, y, z]^T$ are the image space coordinates. R is the rotation matrix that represents the rotations around the roll, pitch, and yaw axes of the camera, and can be written as:

$$R = R_\varphi R_\omega R_\kappa = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & -\sin\omega \\ 0 & \sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and K is the interior parameters of the camera.

Using this model, the 3D positions of the ground points and the 6-DOF poses (3D positions and 3-Axis rotations) of the camera will be reconstructed at the same time. The time complexity of this process will be around $o(n^3)$ to $o(n^4)$ [34]. Considering that farmlands usually have a lower elevation variation, and the main purpose of capturing UAV multispectral data is to analyze the status of crop growth, the model can be simplified to reduce time consumption and avoid irregular resampling at the same time. First, in the rotation matrix, φ , ω , κ represent the roll, pitch, and yaw of the image, respectively. The image rotation around the roll and pitch axes will cause inconsistent scale variation around the image; this will lead to severe spectral distortion during resampling. Due to the fact that most UAV multispectral remote sensing systems can stabilize the camera in these two axes, we can treat φ and ω as zero and consider the two matrices R_φ and R_ω as the identity matrix, like:

$$\begin{aligned} R &= R_\varphi R_\omega R_\kappa = \begin{bmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & -\sin\omega \\ 0 & \sin\omega & \cos\omega \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= R_\kappa \end{aligned}$$

The projection model can be written as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = K \begin{bmatrix} \cos\kappa & -\sin\kappa & 0 \\ \sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_p - X_c \\ Y_p - Y_c \\ Z_p - Z_c \end{bmatrix}$$

For flat areas such as farmlands, $Z_p - Z_0$ can be approximated to the UAV flying height of H . At the same time, since our research is mainly focused on the image mosaic, not calibration, we treat

the images as being already calibrated, and then the K only has $\begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}$ left. The model can be further simplified as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_p - X_c \\ Y_p - Y_c \\ H \end{bmatrix}$$

The image captured by the camera is planar, so z is a fixed value, and the equation can be re-written as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{f}{H} \begin{bmatrix} \cos \kappa & -\sin \kappa \\ \sin \kappa & \cos \kappa \end{bmatrix} \begin{bmatrix} X_p - X_c \\ Y_p - Y_c \end{bmatrix} \quad (5)$$

According to (5), the coordinate of the ground point can be calculated using:

$$\begin{cases} X_p = \frac{H}{f} x \cos \kappa + \frac{H}{f} y \sin \kappa + X_c \\ Y_p = \frac{H}{f} x \sin \kappa - \frac{H}{f} y \cos \kappa + Y_c \end{cases} \quad (6)$$

After feature matching, for every matched point pairs, two ground coordinates (X_{pj_1}, Y_{pj_1}) and (X_{pj_2}, Y_{pj_2}) can be calculated using (6). Since the two points represent the same ground point, the correct exterior orientation elements should let the distance between (X_{pj_1}, Y_{pj_1}) and (X_{pj_2}, Y_{pj_2}) be minimum. For all the matched point pairs, the correct exterior orientation parameters should satisfy Equation (7):

$$\min_{C_j} \sum_{j_1=1}^m \sum_{j_2=1}^m \left((X_{pj_1} - X_{pj_2})^2 + (Y_{pj_1} - Y_{pj_2})^2 \right) \quad (7)$$

Replace $X_{pj_1}, X_{pj_2}, Y_{pj_1}, Y_{pj_2}$ using (6); the Equation (7) will only relate to the exterior orientation elements of the images and the coordinates on the images of the matched point pairs. By solving (7) with non-linear least square algorithms such as the Levenberg-Marquardt algorithm, the correct exterior orientation elements of images can be extracted.

The result resolved by the Levenberg-Marquardt algorithm corresponds to local minima. To increase the robustness and speed of the algorithm, the GPS coordinates and pose data are used to initialize the algorithm. In fact, what Equation (7) solves is the accurate relative positions and poses between the images; the actual georeference can be obtained using the method described in our previous work [35]. f can be found from the datasheet of the sensor. Since the UAV is able to maintain a steady flying height, H can be estimated by the average altitude of the whole working trajectory; for every image, the DOF (degree of freedom) of the exterior orientation parameters can be reduced from 6 to 3. This modification can promote the robustness of the model and reduce the time consumption remarkably. After estimating the locations of the ground points and the exterior orientation elements of the images, the images are transformed using the exterior orientation elements instead of the locations of the points. For one single image, such a transformation can be kept the same in every part of the image so that the irregular resample can be reduced.

3.3. Estimation of the Possible Overlapping Relations

To match a set of images, we may have to try every possible combination of image pairs to find possible overlapping relations if the spatial relations among the images are unknown. The time complexity of this process is $O(n^2)$, and the actual time consumption will increase quickly with the number of UAV images. In fact, every single image only has overlapping relations with surrounding images, and only these image pairs need to be matched. So when matching one image with others, it is necessary to restrict the match attempt only to the adjacent images.

As mentioned above, most UAV multispectral systems can stabilize the rotation of the camera around the pitch and roll axes, so the camera can be kept vertical to the ground. The footprint can be treated as a rectangle, and the long side is w and the short side is h . When having overlapping relations, the longest center distance l for such two rectangles is $\sqrt{w^2 + h^2}$. Obviously, setting $l < \sqrt{w^2 + h^2}$ as the threshold for trying to match two images may cause a lot of false positives, so we tightened the restriction to $l < w$. If the center distance l of the two images satisfies the following equation, it is assumed these two images are close enough to have an overlapping relation, and the matching processing of these two images will be tried in the feature matching step.

$$l < 2H \tan \frac{\theta}{2} \quad (8)$$

where H is the height of the camera from the ground and θ is the field of view (FOV) of the camera. Since we treat the flying trajectory as planar, H can be set with the average of all the heights of the images. As shown in Figure 2, using the distance threshold to restrict matching attempts can significantly improve the efficiency of the matching process.

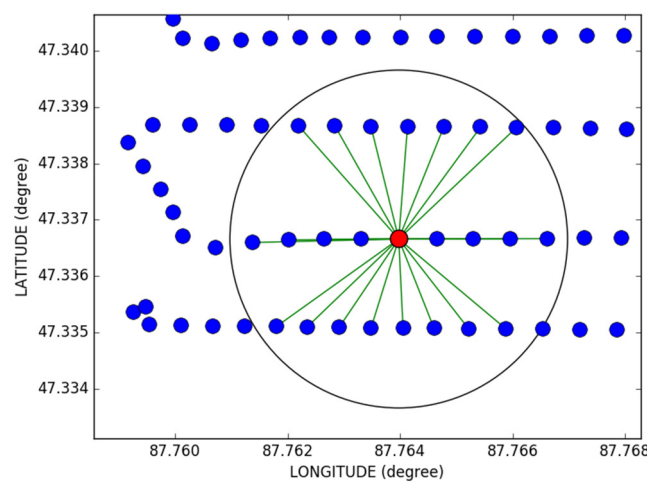


Figure 2. Estimating possible overlapping relations using distance threshold. One dot represents one image. For a certain image represented by the red dot, all the images in the circle of radius l are supposed to have possible overlapping relations with the image represented by the red dot.

3.4. The Workflow of the Proposed Method

The complete workflow of the proposed method is shown in Figure 3. The original image data and the GPS/pose data are captured by the UAV. The proposed image mosaicking process is composed of four steps:

Step 1. In the pre-processing stage the GPS/pose records, and original images are processed separately. The GPS/pose data are used to estimate potential overlapping relations using the distance-based method described above. The original images are used to find the optimal band for feature matching. The detailed steps are:

1. Pickup sample images according to a fixed interval, such as every three or five images for feature quality estimation.
2. Divide every band image into several sub areas as close to the square as possible.
3. Detect SIFT feature points on all the bands of the images.
4. For every band of each frame of the image, calculate the Q_{band-j} using Equation (3).
5. Calculate the $\overline{Q_{band-j}}$ of each band among all the sample images using Equation (4).
6. The band that has the highest $\overline{Q_{band-j}}$ will be used for feature matching.

Step 2. In the feature matching step, the optimal band that was chosen in **Step 1** is used to match the image pairs that have potential overlapping relations. For every pair of images, the matching steps are:

1. Try matching the two images using the optimal band selected in **Step 1**. The matched point pair set is P_0 .
2. If $\text{size}(P_0)$ is less than T_{pairs} (in this research T_{pairs} is set to 30), redo the optimal band evaluation according to **Step 1** on only these two images, and use the new optimal band to re-match the image pair. The new matched point pair set is P_1 .
3. If $\text{size}(P_0 + P_1)$ is still less than T_{pairs} , continue matching with the band that has the second highest $\overline{Q_{band-j}}$ in the re-evaluation result until $\text{size}(P_0 + P_1 \dots)$ exceeds T_{pairs} , or every band of the two images is matched.

Step 3. In the alignment step the matched results are used to compute the accurate position and pose of the images. Every pair of matched points corresponds to two images (j_1, j_2). The two points positions on the two images are $[x_1, y_1]^T$ and $[x_2, y_2]^T$. The exterior orientation parameters of the two images are $[X_{c1}, Y_{c1}, \kappa_1]^T$ and $[X_{c2}, Y_{c2}, \kappa_2]^T$. Substitute them into Equations (6), then get X_{pj_1}, Y_{pj_1} and X_{pj_2}, Y_{pj_2} . Then, the Levenberg-Marquardt algorithm is used to find the accurate exterior orientation parameters of all the images that satisfy Equations (7). In fact, this process is a global Bundle Adjustment, since the Bundle Adjustment can handle “almost any predictive parametric model [36]”. In order to make the process stable and fast, the GPS/pose data are used to initialize the exterior orientation parameters for the LM-algorithm.

Step 4. In the blending step, the accurate exterior orientation parameters are applied to all the bands of all the images at the same time. Then, the resampled images are blended together. In our research, the blending algorithm described in reference [37] is used to blend the resampled images.

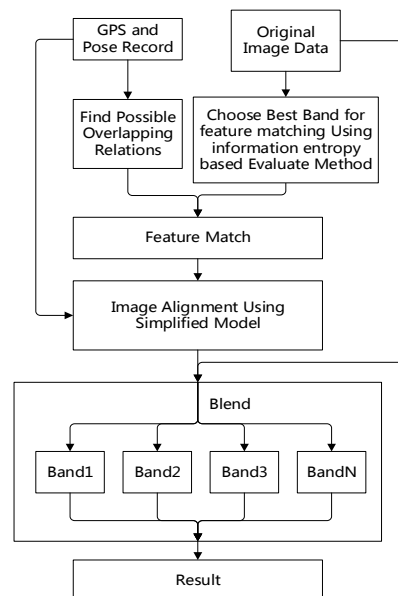


Figure 3. The workflow of the proposed method.

4. Datasets

A Tetracam Mini-MCA6 multispectral camera onboard the AF25b UAV platform is used to capture the UAV multispectral image in Beitun city, Xinjiang, China. The camera has six bands, including three visible and three NIR bands, and can capture 1280×1024 pixels images with a horizontal field of view (FOV) of 38.26 degrees. The sensor dimension of the MiniMCA6 is $6.66 \text{ mm} \times 5.32 \text{ mm}$, and the focal length of the lenses is 9.6mm. The example images captured by the MiniMCA6 are shown in Figure 4, and the specification of the MiniMCA6 is listed in Table 1. The UAV system was used to acquire two datasets in the study area, as shown in Figures 5 and 6. One dataset has 412 shots and contains urban land, bare land, farmland, and a waterbody, and the other dataset has 223 shots and contains farmland and roads. The main crop in the study area is popcorn. The first dataset will be called the dataset-urban, and the other one will be called the dataset-popcorn in the following sections.

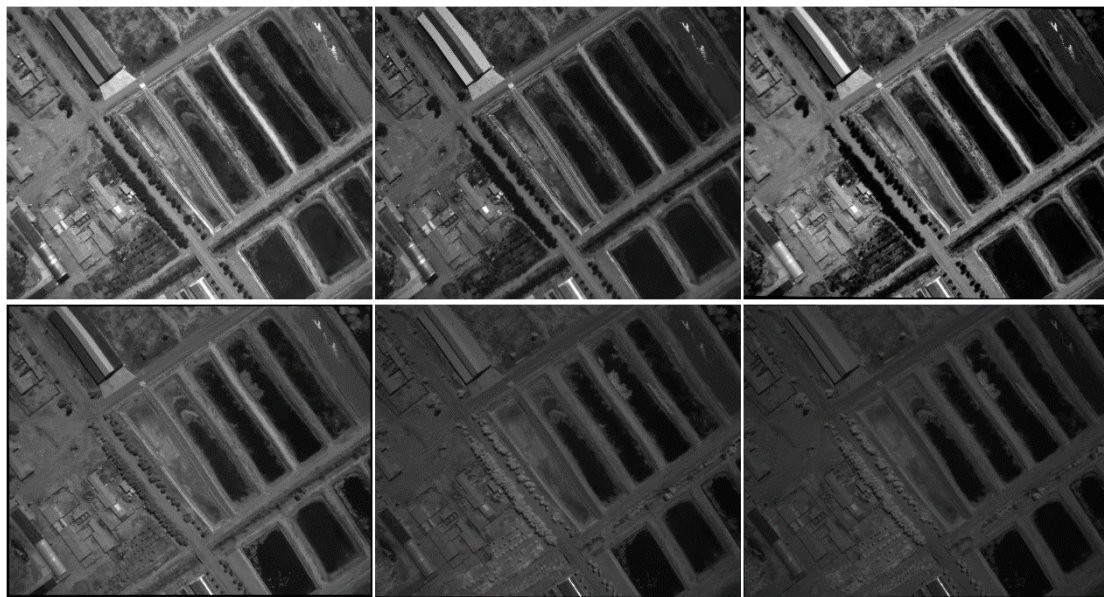


Figure 4. Six image bands of MiniMCA6 camera; the feature quality of each band has significant differences.

Table 1. Band Parameters of MiniMCA6.

Band	Wave Length-Band Width (nm)	Name	Relative Exposure (%)
Band1	550-10	Green	70
Band2	490-10	Blue	127
Band3	680-10	Red	100
Band4	720-10	NIR	100
Band5	800-10	NIR	80
Band6	900-20	NIR	100

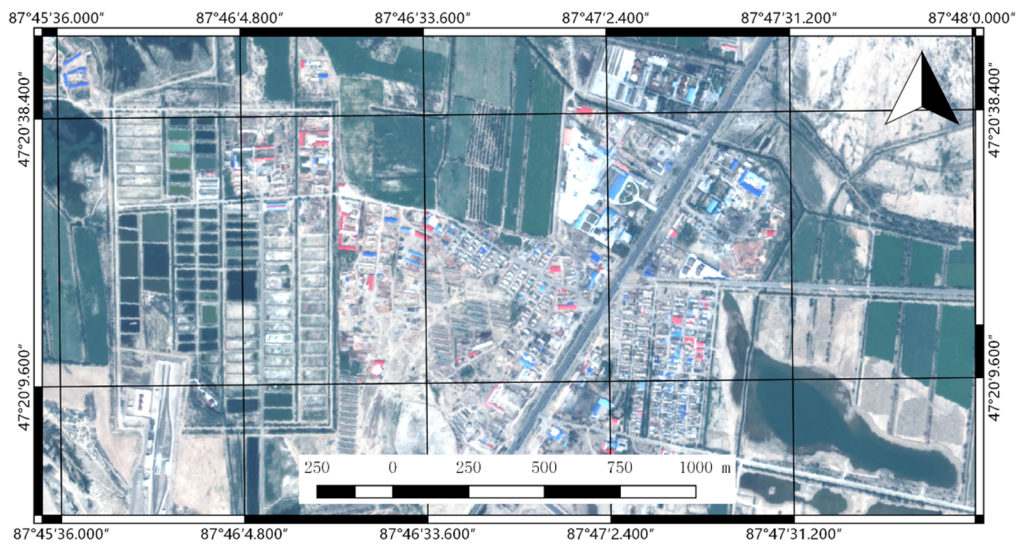


Figure 5. Overview of the dataset-urban.

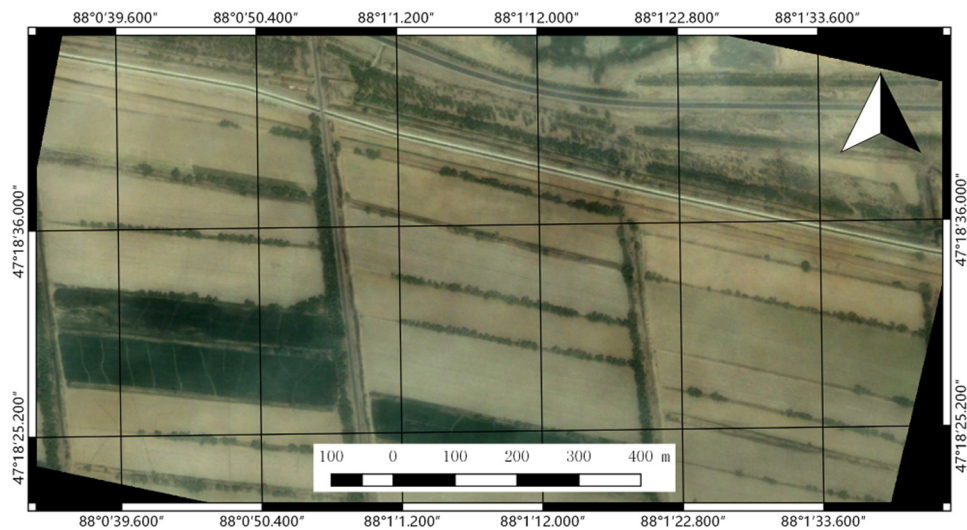


Figure 6. Overview of the dataset-popcorn.

5. Experiment Results

5.1. Selection of the Optimal Band for Feature Matching

Based on the evaluation method described in 3.1, the feature quality evaluation is conducted among the two datasets. In order to reduce the time consumption on evaluation, one out of every three images is used to perform the evaluation on the dataset-urban, and the evaluating results are illustrated in Figures 7 and 8; the average Q-Values of each band of the two datasets are presented in Table 2.

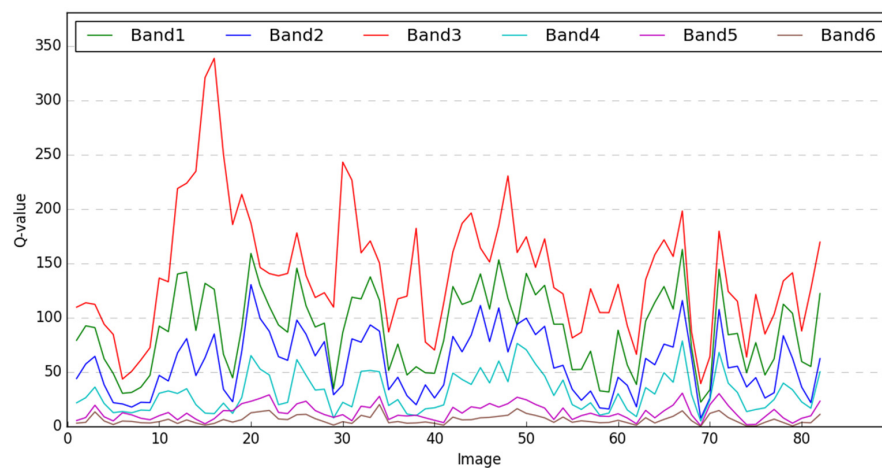


Figure 7. Q-value of dataset-urban; band3 has the best performance, and the other bands also have good performance.

For the dataset-urban, band3 has the highest Q-Values among all the images, which means band3 (the red band) can provide better performance in the matching process all the time. Considering that the dataset-urban covers a large area of urban land and bare land, the evaluation result is reasonable. For the dataset-popcorn, things are a little bit complicated. As we can see in Figure 6, the test site of the dataset-popcorn is mostly farming land, and the image texture is relatively weak. Although band3 still has higher global average Q-Values, band5 performs better, and band3 performs much worse in some images. The green band (band1) performs not as well as the red band (band3) in the farming areas because the red band can have a higher contrast level between vegetation and bare land and provide feature points that have a much higher quality. Using only one certain band for feature matching may be not suffice to handle such a complex situation well. The flexible matching strategy of the proposed method can get a steady matching result for further processes.

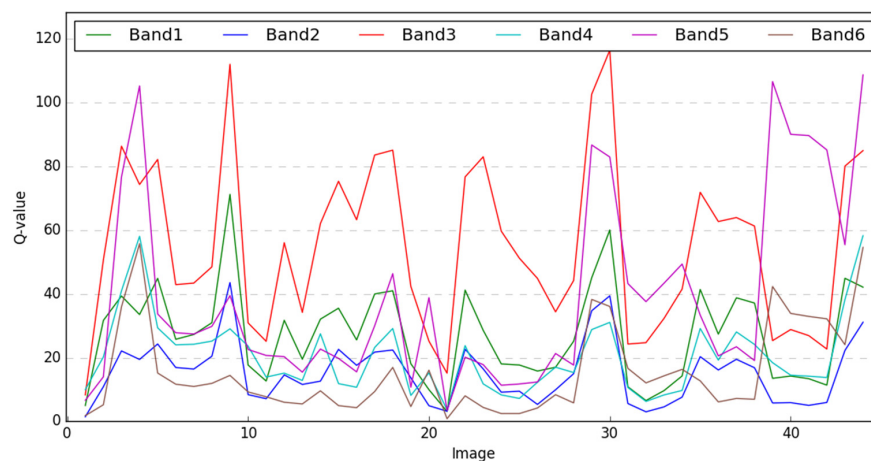


Figure 8. Q-value of dataset-popcorn. Band3 has the best performance but is not stable. In some images, the NIR band has better performance.

Table 2. Comparison of average Q-Values for each band.

Dataset Name	Band1	Band2	Band3	Band4	Band5	Band6
Dataset-urban	88.931	57.232	141.191	31.346	13.180	6.365
Dataset-popcorn	27.063	15.183	54.798	20.561	38.970	15.472

The Q-Value and the number of matched point pairs in the dataset-popcorn are listed in Table 3. The correlation coefficient of these two sets of values is 0.7470, which means the Q-Value and the number of matched pairs have a significant correlation with the confidence > 0.9 . The correlation coefficient result proved that the Q-value was an efficient way to select the optimal band for feature matching.

Table 3. Q-Value and number of matched point pairs, dataset-popcorn.

	Band1	Band2	Band3	Band4	Band5	Band6
Q-Value	27.063	15.183	54.798	20.561	38.970	15.472
Matched PointPairs	29506	26703	31751	20130	29288	23003

5.2. Mosaicking Results

The mosaicking results in the RGB mode of the two datasets, dataset-urban and dataset-popcorn, using our proposed method, are displayed in Figures 9 and 10 respectively. Figures 11 and 12 are mosaicking DOM results generated by the Pix4Dmapper [38]. The experiment was performed on a computer that has Core i7-2600 CPU, 16GB RAM.

As we can see in Figures 9 and 10, the results generated by our proposed method have no obvious distortion and malposition. This suggests that the proposed method can generate visually satisfactory results for both datasets. The flexible matching strategy and the simplified model enable the proposed method to get enough correct matches for the image alignment. Because the transformation of each image only depends on the image elements of the exterior orientation, the side and corner area can be utilized efficiently in mosaicking. On the contrary, Pix4Dmapper needs the 3D point cloud to generate the DOM, which means only the area inside the outermost 3D points can be used. As shown in Figure 13, in the mosaicking result of the dataset-urban generated by the Pix4Dmapper, the bottom left area has a slight malposition. For the dataset-popcorn, the Pix4Dmapper can produce a visually satisfactory result. Besides the significant distortion, the result of the Pix4Dmapper also has significant local breakages. Some examples are shown in Figure 14. These breakages are also caused in the sensitive 3D-reconstruction process. According to the official document of the Pix4Dmapper [39], the Pix4Dmapper is not quite in favor of MiniMCA6 because of the slow rolling shutter of MiniMCA6.



Figure 9. Image mosaic of the dataset-urban using the proposed method. The side and corner areas are efficiently used.

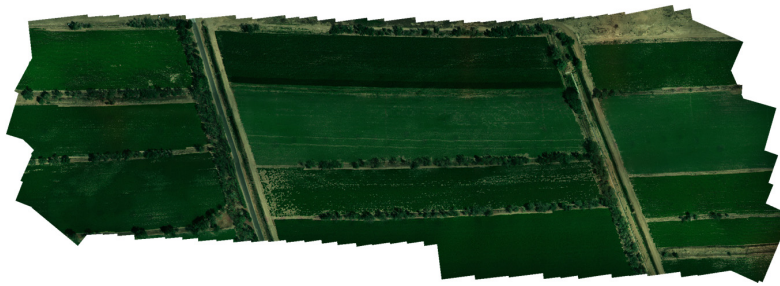


Figure 10. Image mosaic of the dataset-popcorn using the proposed method. The land cover type is mainly croplands and has limited feature level.



Figure 11. Image mosaic of the dataset-urban using Pix4Dmapper. The side and corner areas have considerable losses.

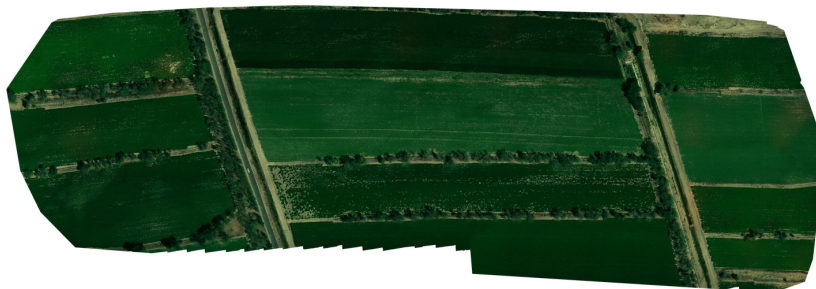


Figure 12. Image mosaic of the dataset-popcorn using Pix4Dmapper.



Figure 13. Close-up views of the malposition in the result by Pix4Dmapper. (a) The result generated by the proposed method; (b) the result generated by Pix4Dmapper. Areas in the white ellipses are the malposition parts.



Figure 14. Comparison of local mosaicking results of the two methods. The upper two regions are from the dataset-urban, the lower two regions are from the dataset-popcorn. (a) by the proposed method; (b) by Pix4Dmapper.

As shown in Figure 14, obvious geometric and spectral distortions in the transitional areas between the adjacent images can be observed in the mosaicking result by the Pix4Dmapper, indicating that the mosaicking result by our proposed method can produce the result that has a much better geometrical quality of the UAV MiniMCA6 image stitching.

6. Assessments and Discussion

6.1. Assessment of Processing Efficiency

To compare processing speeds, we divided the dataset-urban into several subsets that have different sizes. The size of each subset and the computing time by our method and the Pix4Dmapper are listed in Table 4. The time consumption of the Pix4Dmapper increased dramatically with the scale of the dataset. Meanwhile, the time consumption of our proposed method has a nearly linear growth. This is because the Pix4Dmapper tried to match all images exhaustively and recover the 3D structure of the ground surface. These two processes are time-expensive. In the comparison of the complicatedness of the models, the sparse 3D reconstruction process will take more time than the alignment process in the proposed method and, furthermore, the dense 3D reconstruction will take much longer than the sparse 3D reconstruction. On the contrary, the proposed method spent much less time in the reconstruction process by incorporating the simplified projection model. The implementation of the proposed method has not used GPU acceleration so far in the experiment, and the deeply optimized math library Numpy [40] has been used to perform the matrix operations. In the future, to implement the proposed method with GPU acceleration is indeed attractive.

Table 4. Comparison of time consumption in the mosaicking of the dataset-urban.

Dataset Size	Proposed Method (Time in S)	Pix4Dmapper (Time in S)
147	449	960
254	1047	3720
362	2117	11,400 s

6.2. Accuracy Assessment

Assessment of the mosaicking results can be done in two aspects: spatial and spectral accuracy. Higher spatial accuracy means the mosaicking leads to less geometric distortion while higher spectral accuracy means the mosaicking operation can keep better the spectral information from original images at each pixel location.

A SPOT 5 satellite image of Beitun city was used as a reference for assessing spatial accuracy. The satellite image is the 2.5 m resolution SPOT full-color image, captured on 21 November 2014. The matching points are selected manually and used to calculate the spatial root mean square error (RMSE). The matching points are illustrated in Figure 15.



Figure 15. Selection of matching points in spatial accuracy assessment. (a) By the proposed method; (b) by Pix4Dmapper.

The comparison in Table 5 shows that the proposed method has a smaller RMS error than the Pix4Dmapper, even if the latter should theoretically have higher spatial accuracy. This is because the 3D reconstruction model has more DOFs. Thus, it is more sensitive to matching noise. In local areas, the reconstruction process may be unstable due to the reasons like wrong matches or the rolling shutter effect. Since it has experienced simplification, our proposed model is more stable with the quality variation of

the data. For the dataset-popcorn, despite the fact that the Pix4Dmapper generated a visually satisfactory result, the point cloud of the result has considerable distortion (Figure 16) and the calibrated camera positions also have major distortion because the 3D reconstruction algorithm is more sensitive to image data quality. Without additional information such as the Ground Control Points (GCPs), the error will accumulate and finally form the arc of the point cloud shown in Figure 17, especially with datasets that have low texture such as the dataset-popcorn. The lack of matched point pairs will reduce the chance to correct such errors in the 3D-reconstruction process. On the contrary, the proposed method is more robust in such circumstances because of the hypothesis of planar ground and flying trajectory.

Table 5. Comparison of root mean square (RMS) errors with the dataset-urban.

RMS	Proposed Method (pixel)	Pix4Dmapper (pixel)
AVERAGE	12.84	38.89
MIN	4.34	14.58
MAX	23.36	80.51

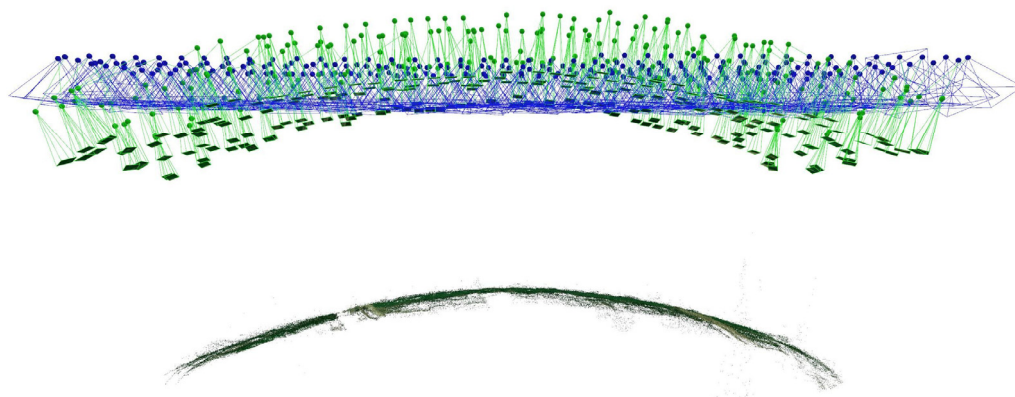


Figure 16. The point cloud for the dataset-popcorn generated by Pix4Dmapper. The point cloud and flight trajectory have significant distortion.

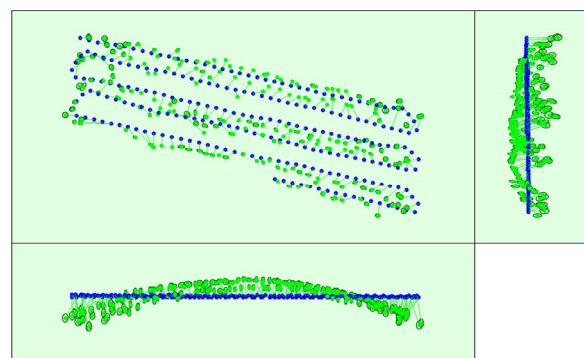


Figure 17. Calibrated camera positions by Pix4Dmapper. The blue dots indicate GPS positions from the unmanned aerial vehicle (UAV); the green ones are calibrated positions by Pix4Dmapper.

To evaluate the spectral accuracies of the image mosaic results, a comparative method based on feature matching is adopted. Because it is difficult to pick up sufficient points manually to perform the spectral comparison. Thus, we just chose several original images from the datasets and matched them with the two mosaicking results. The average RMS error of all the bands for one pixel is calculated based on the following equation:

$$e_{RMS} = \sqrt{\frac{\sum_{j=1}^{n_{bands}} (g_j^{original} - g_j^{result})^2}{n_{bands}}} \quad (9)$$

where n is the total number of the image bands, $g_j^{original}$ and g_j^{result} is the gray value of the j -th band at a matching point, and the n_{bands} are the total number of the image bands. In this case, the n_{bands} is equal to six.

To keep the same point set used to make the comparison, the point pairs that correspond to the same points on the original images were used to perform the comparison, i.e., the point pairs linked by solid lines in Figure 18. The positions of the original images in the mosaicking results are shown in Figure 19:

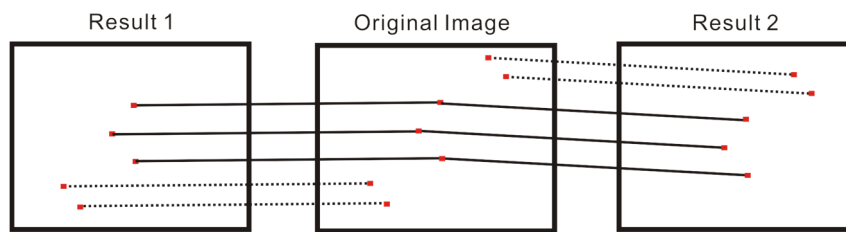


Figure 18. Matching original image with various results may generate different matched point pair sets. Only the common point pairs, as solid lines marked, are used for the comparison.

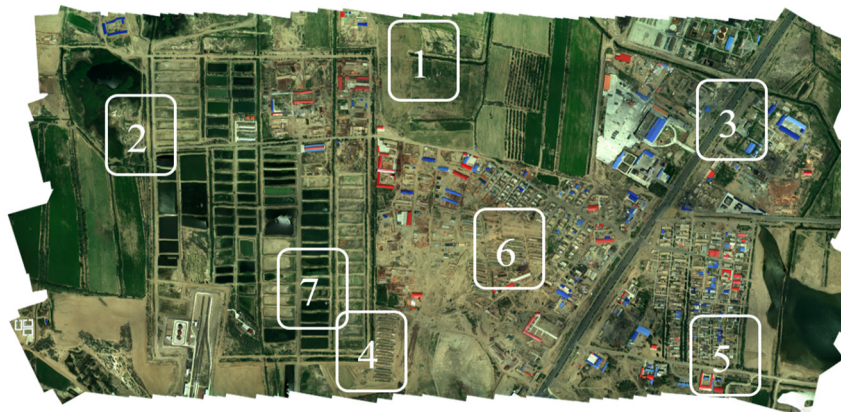


Figure 19. Positions of original images that were selected for the spectral accuracy assessment.

Table 6 shows the RMS errors of spectra at the seven areas on the mosaic images generated by the proposed method and the Pix4Dmapper. Z-Tests are performed to estimate the significance of the diversity of the two sets of the RMS errors. The result indicates that the proposed method has a smaller RMS error than the Pix4Dmapper in vegetated areas such as farmlands. In the bare soil area (site No.6) the two results have similar performances ($|Z| < 1.96$). Moreover, in the urban area (site No.3), the Pix4Dmapper performed better than the proposed method, but the diversity is also not significant. In sites No.1, 2, 4, and 7, the proposed method performs better than the Pix4Dmapper, and the diversity has major significance ($|Z| > 2.58$). In site No.5, the performance of the Pix4Dmapper is significantly better than the proposed method. Considering the proposed method mainly aims to be used in the farming areas, the result is acceptable. The spectral distortion of the mosaicking results is mainly caused by the resample process and the blending process. For the proposed method, an image will be resampled using the same translation and rotation parameters. On the contrary, the Pix4Dmapper will resample the pixels according to the 3D reconstruction results, indicating the resampling parameters will vary in different parts of the image, and this process will

bring more spectral distortion. Considering the agricultural areas always have a smaller elevation variation, the simplified projection model of the proposed method may have great potential in agriculture applications.

Table 6. Comparison of spectral distortion.

Sample Zone No.	Matched Point Number	RMS Error of Proposed Method	Standard Deviation of RMS Errors (Proposed Method)	RMS Error of Pix4Dmapper	Standard Deviation of RMS Errors (Pix4Dmapper)	Z-Test Value
1	321	7.9892	4.6523	8.8905	5.0781	−2.4201
2	775	6.3240	3.2269	7.8622	4.2996	−7.9659
3	233	9.4951	5.0966	8.8159	5.0751	1.4414
4	493	5.9284	3.5761	6.8598	4.0188	−3.8441
5	452	8.7438	5.0430	7.8676	4.9279	2.6420
6	634	7.3783	3.7951	7.7523	4.4444	−1.6114
7	336	6.6388	3.3675	7.6904	3.7820	−3.8064

For most 3D reconstruction methods, the feature points are used as reference points, which will have a lower precision loss in the resampling process. Meanwhile, the proposed method applies the same rotation parameter to one single image, which means such a comparison is not in favor of the proposed method. Even in such a comparison, the proposed method performs better than the state-of-the-art commercial software in agricultural lands. We can conclude that the proposed method can realize a lower precision loss in the spectral properties of the target.

The spectral distortion reduced by the proposed method can be discussed from two aspects. From one aspect, the proposed method completed the mosaicking of the two datasets in a rather shorter time than the Pix4Dmapper. This indicates that our proposed method may be a more efficient solution to conduct UAV image mosaicking in some emergency response applications such as disaster monitoring. From the other aspect, the spectral distortion brought by the blending process is not adequately studied in this research. Thus, a blending algorithm with less spectral distortion needs to be studied in the future. The lower spectral distortion with a conventional blending algorithm also indicates that the modified projection model is necessary to reduce the spectral distortion during the mosaicking process; thus, with a better blending algorithm, the spectral distortion may be further reduced.

6.3. The Impact of Altitude Accuracy

The simplification of the projection model depends on the steadiness of the working altitude of the UAV. Because the simplified projection model treats both the ground and the working trajectory of the UAV as planar, the unstable altitude of the UAV platform will result in a considerable spatial error or an intolerable image misplacement. Unfortunately, the accuracy of the GPS altitude is not good enough on most occasions, especially for onboard GPS receivers which suffer from severe limits by the weight and cost. So it is necessary to discuss whether the low positioning accuracy of the onboard GPS receivers has a strong impact on the result.

The altitude of a UAV depends on the GPS altitude. But since the GPS altitude is not quite reliable and may have a sudden change during the flight, the UAV also uses other sensors such as the accelerometer and the gyroscope to keep the pose and the altitude steady. Under this circumstance, the unstable GPS altitude can be supposed to vary around the true altitude. Therefore, an average of the GPS altitudes of the images is good enough to be an estimation of the H in Section 3.2.

6.4. The Impact of the Alignment of the Bands

In this research, it acquiesces that the bands of an image are already aligned because the alignment between the bands is beyond the topic of this study. However, during the experiment it is not possible to ignore this issue. The MiniMCA6 has an exclusive software package to correct the offsets between bands according to the factory calibration file. According to the FOV and the pixel size of the sensor, it can be calculated that when the UAV works 200 m high above the ground, the ground resolution

is 108.33 mm. This length is greater than the distance between the two farthest lenses. So the offset between the lenses can be ignored.

6.5. The Impact of the Other Reflection Characteristics of the Ground Targets

As we mentioned above, the environmental condition factors such as the atmosphere and the relative positions of the sun, the target, and the sensor may also affect the reflective properties. These factors indeed have a strong impact on the reflectivity of the ground targets, especially in agricultural areas because the plant canopy has anisotropic reflective properties. Although the environmental conditions may have an influence on the UAV images, the band selection method still can work very well, because the proposed method only evaluates the gray scale images of each band and uses the selected optimal band that has the highest average Q-Value to perform the feature matching of the image mosaicking. On the other hand, the environmental conditions mainly affect the remote sensing applications such as crop growth monitoring by changing the reflectance. A Teflon board can be used to calibrate the MiniMCA6 sensor right before every flight for the further applications using the mosaicking results.

7. Conclusions

This paper presents a simplified but efficient method for UAV multispectral image mosaicking in agricultural applications. In the proposed method, an algorithm for automatic image band selection was proposed and used in selecting an optimal band for feature matching, and a simplified projection model was also adopted to avoid the computationally expensive 3D reconstruction. The simplified projection is able to promote the efficiency of the mosaicking process and can reduce the spectral distortion by avoiding the irregular resampling in the 3D reconstruction process. The performance of the proposed method was demonstrated through a case study in Beitun City, Xinjiang Province, China. Results showed that the proposed method performed well on the two datasets and achieved better mosaicking results than the state-of-the-art commercial software Pix4Dmapper. The resultant mosaicked images by our proposed method have less spectral distortion and higher spatial accuracy, and the mosaicking process consumed much less time than the Pix4Dmapper package.

The limitation of this study is that the spectral distortion of the image blending process is not well considered. An image blending algorithm producing less spectral distortion has to be studied in future work. The evaluation process for feature matching will be very time consuming if the onboard sensor has a large number of bands (e.g., the hyperspectral sensor); thus, in future work, we should consider how to reduce the band number by merging bands or band selection based on the spectral features of the ground targets. Using GPU to accelerate the mosaicking process is also an attractive direction to study in the future.

Acknowledgments: This work has been funded by Key Technologies Research and Development Program of China (Grant Number 2016YFD0300601), the Department of Sciences and Technology, the Inner Mongolia Autonomous Region, China (Tectonic evolution, minerogenetic information and minerogenesis in Xing'an-Mongolia orogenic belt, 2015–2018), and the XPCC Geospatial Information Innovation Team (No. 2016AB001). We would like to thank anonymous reviewers and the editor in chief for their constructive comments, which greatly improved the quality of our manuscript.

Author Contributions: Xiang Ren and Min Sun conceived and designed the study; Xiang Ren performed the experiments; Lei Liu and Xiang Ren collected data; Xiang Ren and Xianfeng Zhang wrote the paper in English.

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

References

1. Feng, Q.; Liu, J.; Gong, J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sens.* **2015**, *7*, 1074–1094. [[CrossRef](#)]
2. Pajares, G. Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogramm. Eng. Remote Sens.* **2015**, *81*, 281–329. [[CrossRef](#)]

3. Colomina, I.; Molina, P. Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2014**, *92*, 79–97. [[CrossRef](#)]
4. Honkavaara, E.; Saari, H.; Kaivosoja, J.; Polonen, I.; Hakala, T.; Litkey, P.; Makynen, J.; Pesonen, L. Processing and assessment of spectrometric, stereoscopic imagery collected using a lightweight UAV spectral camera for precision agriculture. *Remote Sens.* **2013**, *5*, 5006–5039. [[CrossRef](#)]
5. Gomez-Candon, D.; De Castro, A.I.; Lopez-Granados, F. Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precis. Agric.* **2014**, *15*, 44–56. [[CrossRef](#)]
6. Sona, G.; Passoni, D.; Pinto, L.; Pagliari, D.; Masseroni, D.; Ortuani, B.; Facchi, A. UAV multispectral survey to map soil and crop for precision farming applications. *XXIII ISPRS Congr. Comm. I* **2016**, *41*, 1023–1029.
7. Nebiker, S.; Annen, A.; Scherrer, M.; Oesch, D. A light-weight multispectral sensor for micro UAV—Opportunities for very high resolution airborne remote sensing. *Int. Arch. Photogram. Remote Sens. Spat. Inf. Sci.* **2008**, *37*, 1193–1199.
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**, *4*, 1392–1410. [[CrossRef](#)]
9. Clapuyt, F.; Vanacker, V.; Van Oost, K. Reproducibility of UAV-based earth topography reconstructions based on structure-from-motion algorithms. *Geomorphology* **2016**, *260*, 4–15. [[CrossRef](#)]
10. Rau, J.-Y.; Jhan, J.-P.; Li, Y.-T. Development of a large-format uas imaging system with the construction of a one sensor geometry from a multicamera array. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5925–5934. [[CrossRef](#)]
11. Ramirez-Paredes, J.-P.; Lary, D.J.; Gans, N.R. Low-altitude terrestrial spectroscopy from a pushbroom sensor. *J. Field Robot.* **2016**, *33*, 837–852. [[CrossRef](#)]
12. Tang, X.; Zhang, G.; Zhu, X.; Pan, H.; Jiang, Y.; Zhou, P.; Wang, X. Triple linear-array image geometry model of ziyuan-3 surveying satellite and its validation. *Int. J. Image Data Fusion* **2013**, *4*, 33–51. [[CrossRef](#)]
13. Aguilar, M.A.; Saldana, M.d.M.; Aguilar, F.J. Assessing geometric accuracy of the orthorectification process from geoeye-1 and worldview-2 panchromatic images. *Int. J. Appl. Earth Obs. Geoinf.* **2013**, *21*, 427–435. [[CrossRef](#)]
14. Brown, M.; Lowe, D.G. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vis.* **2007**, *74*, 59–73. [[CrossRef](#)]
15. Xu, Y.H.; Ou, J.L.; He, H.; Zhang, X.H.; Mills, J. Mosaicking of unmanned aerial vehicle imagery in the absence of camera poses. *Remote Sens.* **2016**, *8*, 16. [[CrossRef](#)]
16. Kim, J.-I.; Kim, T.; Shin, D.; Kim, S. Fast and robust geometric correction for mosaicking UAV images with narrow overlaps. *Int. J. Remote Sens.* **2017**, *38*, 2557–2576. [[CrossRef](#)]
17. Gini, R.; Pagliari, D.; Passoni, D.; Pinto, L.; Sona, G.; Dosso, P. UAV photogrammetry: Block triangulation comparisons. *UAV-G2013* **2013**, *XL-1/W2*, 157–162.
18. Turner, D.; Lucieer, A.; Malenovsky, Z.; King, D.H.; Robinson, S.A. Spatial co-registration of ultra-high resolution visible, multispectral and thermal images acquired with a micro-UAV over antarctic moss beds. *Remote Sens.* **2014**, *6*, 4003–4024. [[CrossRef](#)]
19. Zhu, C.; Leow, W.K. Textured mesh surface reconstruction of large buildings with multi-view stereo. *Vis. Comput.* **2013**, *29*, 609–615. [[CrossRef](#)]
20. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
21. Ke, Y.; Sukthankar, R. *Pca-Sift: A More Distinctive Representation for Local Image Descriptors*; IEEE: Piscataway, NJ, USA, 2004.
22. Bay, H.; Tuytelaars, T.; Van Gool, L. Surf: Speeded up robust features. *Comput. Vis. ECCV 2006* **2006**, 3951, 404–417.
23. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G.; Ieee. Orb: An efficient alternative to sift or surf. In Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, 6–13 November 2011; pp. 2564–2571.
24. Mikolajczyk, K.; Schmid, C. A performance evaluation of local descriptors. *Pattern Anal. Mach. Intell. IEEE Trans.* **2005**, *27*, 1615–1630. [[CrossRef](#)] [[PubMed](#)]

25. Teke, M.; Temizel, A. Multi-spectral satellite image registration using scale-restricted surf. In Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR 2010), Istanbul, Turkey, 23–26 August 2010; pp. 2310–2313.
26. Teke, M.; Vural, M.F.; Temizel, A.; Yardimci, Y. High-resolution multispectral satellite image matching using scale invariant feature transform and speeded up robust features. *J. Appl. Remote Sens.* **2011**, *5*. [[CrossRef](#)]
27. Navia, J.; Mondragon, I.; Patino, D.; Colorado, J. Multispectral mapping in agriculture: Terrain mosaic using an autonomous quadcopter UAV. In Proceedings of the International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA, 7–10 June 2016; pp. 1351–1358.
28. Lucas, B.D.; Kanade, T. An iterative image registration technique with an application to stereo vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981.
29. Du, Q.; Raksuntorn, N.; Orduyilmaz, A.; Bruce, L.M. Automatic registration and mosaicking for airborne multispectral image sequences. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 169–181. [[CrossRef](#)]
30. Il, K.Y.; Yu-kyung, H.; Dong, H.; Wan, C.J. Mosaic image generation of aisa eagle hyperspectral sensor using sift method. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2013**, *30*, 165–172.
31. Zhengqi, L.; Isler, V. Large scale image mosaic construction for agricultural applications. *IEEE Robot. Autom. Lett.* **2016**, *1*, 295–302.
32. Li, X.; Yeh, A.G.-O. Analyzing spatial restructuring of land use patterns in a fast growing region using remote sensing and gis. *Landsc. Urban Plan.* **2004**, *69*, 335–354. [[CrossRef](#)]
33. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2003; p. 159.
34. Agarwal, S.; Snavely, N.; Seitz, S.M.; Szeliski, R. Bundle adjustment in the large. In Proceedings of the 11th European Conference on Computer Vision (ECCV 2010), Crete, Greece, 5–11 September 2010; Volume 6312, pp. 29–42.
35. Ren, X.; Sun, M.; Jiang, C.; Liu, L.; Zheng, H.; Li, X.D. A method of fast mosaic for massive UAV images. In Proceedings of the Conference on Land Surface Remote Sensing II, Beijing, China, 13–16 October 2014; Spie-Int Soc Optical Engineering: Beijing, China, 2014.
36. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment—a modern synthesis. *Vis. Algorithms Theory Pract. Int. Workshop Vis. Algorithms* **2000**, *1883*, 298–372.
37. Xiong, Y.L.; Turkowski, K.; Soc, I.C.; Ieee Comp, S.O.C. Registration, calibration and blending in creating high quality panoramas. In Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision—Wacv’98, Princeton, NJ, USA, 19–21 October 1998; pp. 69–74.
38. Pix4dmapper pro. Available online: <https://pix4d.com/product/pix4dmapper/> (accessed on 13 Spetember 2017).
39. Pixi4D. How to create and process a tetracam dataset. Available online: <https://support.pix4d.com/hc/en-us/articles/202560219-How-to-create-and-process-a-Tetracam-dataset> (accessed on 13 Spetember 2017).
40. Developers, N. Numpy. Available online: <http://www.numpy.org/> (accessed on 13 Spetember 2017).

