A Method for Building a Mosaic with UAV Images

Cheng Xing
School of Geodesy and Geomatics, Wuhan University, Wuhan, China
Key Laboratory of Precise Engineering and Industry Surveying, State Bureau of Surveying and Mapping, Wuhan, China
Email: c.xing.whu@gmail.com

Jinling Wang
School of Surveying and Spatial Information Systems, University of New South Wales, Sydney, Australia
Email: j.wang@unsw.edu.com

Yaming Xu
School of Geodesy and Geomatics, Wuhan University, Wuhan, China
Key Laboratory of Precise Engineering and Industry Surveying, State Bureau of Surveying and Mapping, Wuhan, China
Email: yamingxu@whu.edu.cn

Abstract—**At present, satellite and aerial remote sensing are common ways to collect data for territorial resources monitoring in most countries, but they are not effective or rapid enough. Compared with traditional ways of obtaining images, the UAV based platform for photogrammetry and remote sensing is a more flexible and easy way to provide high-resolution images with lower cost. So building UAV based platforms is becoming a hot field throughout the whole world. However, there are also some problems with UAV images, e.g. the views of UAV images from UAV are smaller than those of traditional aerial images, so these images with small views should be pasted together in order to increase the visual field. Therefore, mosaicking UAV images is a critical task. The homographies between sequence images will be affected by the accumulated errors, which will lead to drifts of the position of each image in the mosaic. In this paper, we introduce a two-step optimization method for mosaicking UAV sequence images which can correct the homographies and improve the position of each image in the mosaic. Experimental results will also be presented.**

Index Terms—*UAV, sequence images, mosaicking, optimization*

I. INTRODUCTION

Compared with manned aircraft, using Unmanned Aerial Vehicles (UAVs) has many advantages [1]: 1) UAVs do not need a qualified pilot on board; 2) UAVs can enter environments that are dangerous to human life; 3) It is easy to implement high-risk and high-tech missions; 4) No need of permission for airspace control for low-altitude flights in most countries (such as P. R. China); 5) Lower costs of the platforms than that of traditional airplanes; 6) High resolution images and precise positioning data. As a result of these advantages, building Earth observation systems based on UAV platforms has become a hot topic throughout the world. Therefore UAV-based photogrammetry is becoming a popular field of research.

Although the UAVs can provide images with high resolution in a portable and easy way, the images only cover small parts of the interesting areas. For some surveying missions, we need a method to put all the sequence images together to make a larger image to check the quality of the raw data.

The distance from the camera to the target objects is very much greater than the motion of the camera between views, so the relationship between neighboring sequence images can be described by a homographic model [2], [3], [4]. However, when we use all the sequence images to create a mosaic which can provide us a larger view, the accuracy of the homographies may be affected by the accumulated errors (Figure 1).
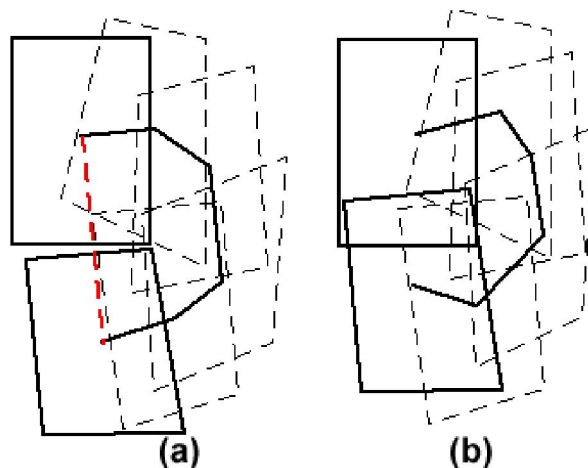

Figure 1. Accumulated errors cause poor alignment near the end of a loop. (a) the final image is not coincided with the first image, (b) a perfect mosaic with corrections.

In this paper, we introduce a two-step optimization method for mosaicking UAV sequence images. Firstly, overlap analysis of sequence images and some work of preprocessing should be done. Then stable features are extracted by using the SIFT (Scale Invariant Feature Transforms) algorithm, and these feature points should be matched. Thirdly, a local optimization method based on Extended Kalman Filter is applied to improve the homographies within local areas. Finally, a global optimization based on Levenberg-Marquardt algorithm is used to correct the position of each resampled image in the mosaic. The final results are shown in the section of tests.

II. HOMOGRAPHIC TRANSFORMATION COMPUTATION

A. Features extraction

A wide variety of scale and rotation invariant feature extraction algorithms have been proposed for detecting correspondences among images. Some experiments have been made to test the performance of existing feature extraction algorithms [5], and the results shows that the SIFT algorithm [6, 7] is better. So, in this paper, we use the SIFT algorithm to extract features from the sequence images (Figure. 2).
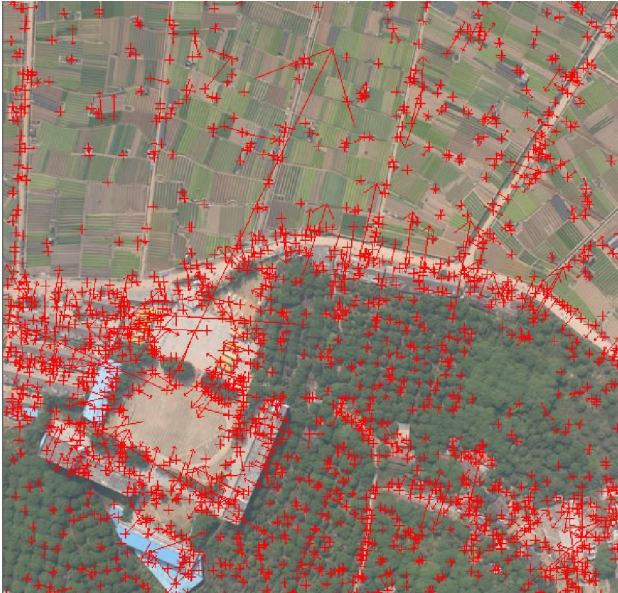

Figure 2. Feature points extracted by the SIFT

B. Perspective projection

A set of feature points can be extracted by the SIFT algorithm, and then we can match them. The next step is to compute the relationship between neighboring images.

The relationship between the two neighboring images can be considered as perspective projection, so the homographic transformation can be calculated by:

$$\begin{cases} x^{'} = \dfrac{m_1 x + m_2 y + m_3}{m_7 x + m_8 y + 1} \\ \\ y^{'} = \dfrac{m_4 x + m_5 y + m_6}{m_7 x + m_8 y + 1} \end{cases} \quad (1)$$

where (x, y) are the coordinates on image 1 and (x', y') are the coordinates on image 2, and M=[m_1, m_2, …, m_8, 1]^T are the 8 parameters of the homographic transformation which can also be expressed as a 3*3 matrix denoted by:

$$\overline{\mathbf{M}} = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{bmatrix} \quad (2)$$

So we can use at least 4 matched points to compute the transformation matrix.

C. Homography covariance estimation

Once the transformation matrix is computed, we require a measure of the estimation accuracy, and the here we can use the covariance matrix of the homography to measure the accuracy.

After getting a set of matched feature points, we can compute the covariance matrix by the way proposed by Hartley [8]. The matched feature points can be expressed as follows:

$$F = \left\{ \{F_1, F_1^{'}\}, \{F_2, F_2^{'}\}, …, \{F_n, F_n^{'}\} \right\} \quad (3)$$

Where $F_i = [x_i, y_i]^T$, $F_i^{'} = [x_i^{'}, y_i^{'}]^T$. Firstly, we should calculate the Jacobian of the transformation from F to F' with respect to the 8 parameters of $\overline{\mathbf{M}}$, and then we can get a $2n*9$ matrix $\mathbf{J}$. Secondly, the covariance of the residual ($\mathbf{C}_{F_i}$) between each matched points should be computed, and because the residuals are uncorrelated, we can get the following matrix $\mathbf{C}_F$:

$$\mathbf{C}_F = diag\left\{ C_{F_1}, C_{F_2}, …, C_{F_n} \right\} \quad (4)$$

Finally, we can use matrix $\mathbf{J}$ and matrix $\mathbf{C}_F$ to compute the covariance matrix of the homographic transformation by the following equation:

$$\mathbf{C}_{\overline{\mathbf{M}}} = (\mathbf{J}^T \mathbf{C}_F^{-1} \mathbf{J})^{-1} \quad (5)$$

D. Attitude correction

The attitude of the UAV cannot be controlled as easily as that of the traditional large airplane, so the same objects on the neighboring images may not be displayed with the same shape, and the mosaicking result will be affected by this kind of distortion. Therefore, some correction should be applied before calculating the relationship between each pair of sequence images, and the attitude information can be provided by the Position and Orientation System (POS) on the UAV.

Figure 3 shows the relationship between a horizontal image and a tilted image [9], where S is the camera center, p is a horizontal image and $p^\mathbf{0}$ is a tilted image, c is the isogonic point, $h_c h_c$ is the isometric parallel, f is the focal length, point a is on the image p, the point $a^\mathbf{0}$ is on the

image p^0, r_c and r_c^0 are the radius vectors, α is the angle of inclination, φ is a included angle between ca and h_c, and φ^0 is an another included angle between ca^0 and h_c. If the image p^0 is rotated around $h_c h_c$ to the image p, point a and a^0 must be on the same line, and the correction can be described as:
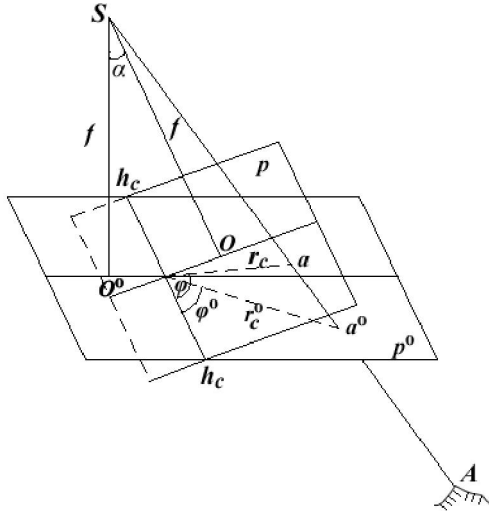
$$\delta_a = aa^0 = r_c - r_c^0 \qquad (6)$$

Figure 3. The relationship between a horizontal image and a tilted image

The relationship between r_c and r_c^0 is:

$$r_c^0 = \frac{f \bullet r_c}{f - r_c \sin\varphi \sin\alpha} \qquad (7)$$

Therefore,

$$\delta_a = r_c - r_c^0 = -\frac{r_c^2 \sin\varphi \sin\alpha}{f - r_c \sin\varphi \sin\alpha} = -\frac{r_c^{02} \sin\varphi \sin\alpha}{f + r_c^0 \sin\varphi \sin\alpha} \qquad (8)$$

The tilted image can be corrected by (8).

E. The reference plane

The first image of the first flight strip can be selected as the reference plane during the process of mosaicking. However, the distortion may be accumulated larger and larger with the increasing number of sequence images. Therefore, a proper reference plane should be selected to decrease the accumulated distortion, and the distortion can be distributed equally to each sequence images.

The final result is a just mosaicking image, so the reference plane can be selected randomly. Here we can choose a dynamic plane as a reference. For example, when we are processing the first and the second image, an average plane can be chosen as the reference, and the this plane can be calculated by using the average values of each pair of corresponding points. When we put the third image into the mosaic, we can calculate an another average plane with the result made by the previous two images and the third image, and this is a new reference plane, and the rest can be done in the same manner [10].

III. OPTIMIZATION METHODS FOR MOSAICKING

Suppose we have got the sequence images from two adjacent flight strips, and then we can compute the position of each image in the final mosaic through multiplying the current homography by all the previous homographies until the reference frame is reached. If the homographies are accurate, the mosaic will be perfect. However, there will be a drift between the last image and the reference image because of the accumulated errors. Therefore, we need to use some optimization method to improve the position of each image in the mosaic.

A. Local optimization

We can divide the whole sequence images into several groups (Figure 4) by using overlapping degree analysis method [11], and each image in a group should have an overlapping area with the reference image. In this paper, the overlapping degree is about 60%~70%, we set six images in a local group (shown as group A and B in Figure 4). Then a local optimization based on the Extended Kalman Filter [12] is used to improve the homographies within a local area. The homographies corrected in this step can be used as initial values for the global optimization step.
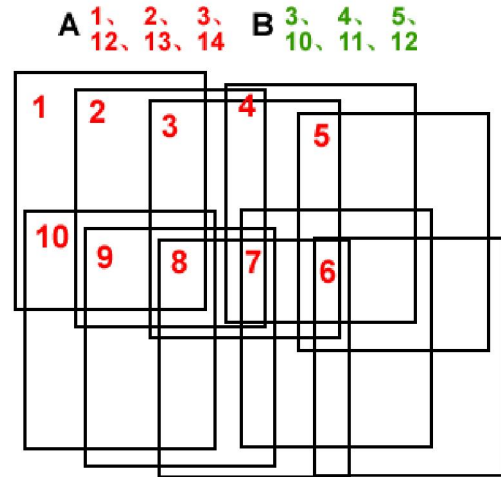

Figure 4. Groups of secquence images. With the analysis of overlapping degree, here we devide these two flight-strip sequence images into 3 groups denoted by A, B and C. The overlapping degree in our tests is about 60%~70%.

We suppose that the first image, denoted by $\mathbf{I}_1$, in a group is the reference image, and the last image in a group is $\mathbf{I}_n$. The relationship between an image and the reference image can be expressed as:

$$\bar{\mathbf{M}}_{k1} = \prod_{i=2}^{k} \bar{\mathbf{M}}_{(i-1)i} \qquad (6)$$

where $\bar{\mathbf{M}}_{k1}$ means the relationship between image $\mathbf{I}_k$ and the reference image $\mathbf{I}_1$. The a priori state vector will be composed by the n-1 transformations which align the n-1 images with $\mathbf{I}_1$ and can be defined by:

$$\mathbf{x}^- = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n-1}]^T = [\bar{\mathbf{M}}_{12}, \mathbf{x}_1 \bullet \bar{\mathbf{M}}_{23}, ..., \mathbf{x}_{n-2} \bullet \bar{\mathbf{M}}_{(n-1)n}]^T \quad (7)$$

where the operator " $\mathbf{g}$ " is the product of two 3*3 homographic matrices, and the result is expressed as a vector.

The prediction equation and the covariance matrix can be expressed as:

$$\mathbf{x}_i = \mathbf{x}_{i-1} \mathbf{g} \overline{\mathbf{M}}_{i(i+1)} \qquad (8)$$

$$\mathbf{C}_{\mathbf{x}_i} = \mathbf{A}\mathbf{C}_{\mathbf{x}_{i-1}}\mathbf{A}^{\mathrm{T}} + \mathbf{W}\mathbf{C}_{\overline{\mathbf{M}}_{i(i+1)}}\mathbf{W}^{\mathrm{T}} \qquad (9)$$

where matrix $\mathbf{A}$ can be obtained by computing the Jacobian of (8) with respect to the state vector and matrix $\mathbf{W}$ can be got by computing the Jacobian with respect to the variables of $\overline{\mathbf{M}}_{i(i+1)}$.

The measurement equation is defined by:

$$\mathbf{M}_{1i} = \mathbf{H}\mathbf{x} = \underset{9\times9}{\mathbf{I}}\ \underset{9\times1}{\mathbf{x}_{i-1}} \qquad (9)$$

To update the state vector we can use the following expression:

$$\mathbf{x} = \mathbf{x}^- + \mathbf{K}(\mathbf{M}_{1(i+1)} - \mathbf{H}\mathbf{x}^-) \qquad (10)$$

where $\mathbf{K}$ is the gain matrix. Finally, the relations among the images involved in the group can be updated with the measurement.

B. Global optimization

After optimizing the parameters within local areas, we need to use bundle adjustment to improve all the homographies globally. This step can solve the problems caused by the accumulated errors. All the sequence images are added to the bundle adjuster one by one, and the parameters can be updated by using the L-M (Levenberg-Marquardt) algorithm [13].

The objective function we use is a sum squared projection error. Each feature point is projected into every image in which it matches, and the sum of squared distances is minimised with respect to the parameters of the homographies. The squared residual can be computed by:

$$d_{ij}^k = d^2(F_i^k, F_{pij}^k) + d^2(F_j^k, F_{pji}^k) \qquad (11)$$

where F_{pji}^k is the projection from image j to image i of the feature point corresponding to F_i^k , and we can use the following equation to compute F_{pji}^k :

$$F_{pij}^k = \mathbf{M}_{ij}F_i^k \qquad (12)$$

The error function is the sum over all the images of the squared residuals:

$$D(\mathbf{M}) = \sum_{i=1}^{n}\sum_{j\in I(i)}\sum_{k\in S(i,j)} d_{ij}^k \qquad (13)$$

where n is the number of images, $\mathbf{I}(i)$ is the set of images matching to image $\mathbf{I}$, $S(i, j)$ is the set of matched features between images i and j.

This is a non-linear least squares problem which can be solved by using L-M algorithm. The iteration expression can be defined by:

$$= (\mathbf{J}^{\mathrm{T}}\mathbf{J} + \lambda\mathbf{I})^{-1}\mathbf{J}^{\mathrm{T}}d \qquad (14)$$

where are the corrections of all the parameters, d the squared residuals, $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ and $\mathbf{J}^{\mathrm{T}}d$ can be computed by:

$$\begin{cases} (\mathbf{J}^{\mathrm{T}}\mathbf{J})_{i,j} = \sum_{k\in S(i,j)} \dfrac{\partial d_{ij}^{k\mathrm{T}}}{\partial_i}\dfrac{\partial d_{ij}^k}{\partial_j} \\[2ex] (\mathbf{J}^{\mathrm{T}}d)_i = \sum_{i=1}^{n}\sum_{j\in I(i)}\sum_{k\in S(i,j)} \dfrac{\partial d_{ij}^{k\mathrm{T}}}{\partial_i} d_{ij}^k \end{cases} \qquad (15)$$

This iterative process needs initial values which can be provided by the local optimization step. The detailed step of the L-M algorithm can be described as:

(1) set k=0, and compute the initial transformation matrix $\mathbf{M}_i^k$;

(2) calculate the error function $\mathrm{D}(\mathbf{M}_i^k)$, and the get $\mathrm{D}(\mathbf{M})$;

(3) use formula (15) to compute $\mathbf{J}^{\mathrm{T}}\mathbf{J}$ and $\mathbf{J}^{\mathrm{T}}d$, and then compute using formula (14);

(4) if $\mathrm{D}(\mathbf{M}^k) < T$ (T is a threshold), go to step (6), otherwise compute $\mathbf{M}_i^{k+1}$ and the error function $\mathrm{D}(\mathbf{M}^{k+1})$;

(5) if $\mathrm{D}(\mathbf{M}^{k+1}) < \mathrm{D}(\mathbf{M}^k)$, we set $k= k+1$ and decrease the value of λ , and then turn to step (2), otherwise the transformation matrices should not be updated, and set $\mathbf{M}_i^{k+1} = \mathbf{M}_i^k$, and increase the value of λ , then go to step (4);

(6) over.

IV. TEST RESULTS

We use 10 sequence images from two fight strips in the test, and each fight strip includes 5 images. The resolution of each original image is 2592*3888. We take images No. 4 and No. 5 for example to show the processing results. Figure 5 shows the raw image No. 2, and Figure 6 shows the result of attitude correction with the POS information collected by the UAV.

The first step is the attitude correction for all the sequence images with the information provided by the POS, and then all the images can be similarly considered as horizontal images. The second step is the features extraction, we use the SIFT algorithm to extract features from each image in the sequence set, and then the feature points will be matched. Perhaps there are some wrong matches, so the wrong matches should be removed before calculating the homographies. Then, the local and global optimization methods described in this paper should be done to optimize the homographies between neighboring images, and finally we can put all the resampled images computed with the optimized homographies together to build a large mosaic.
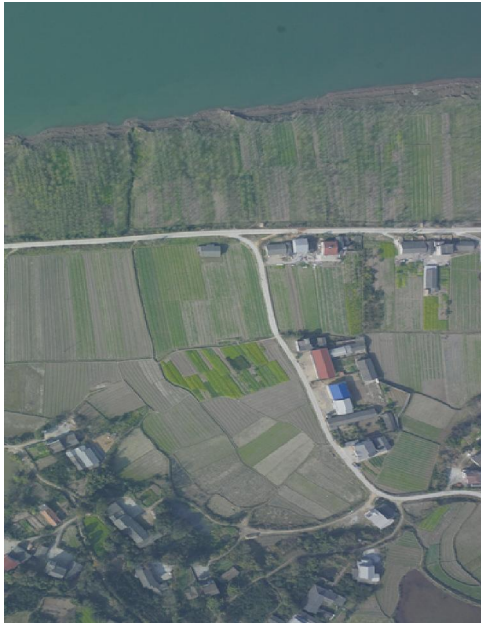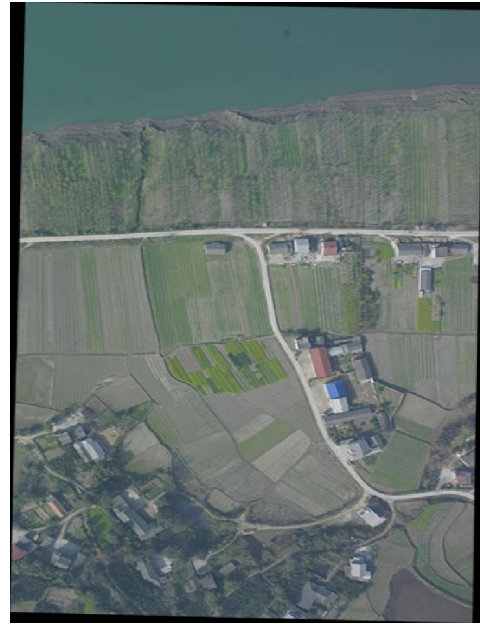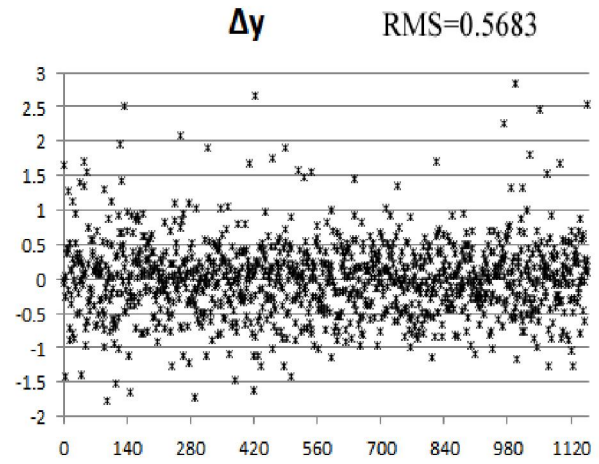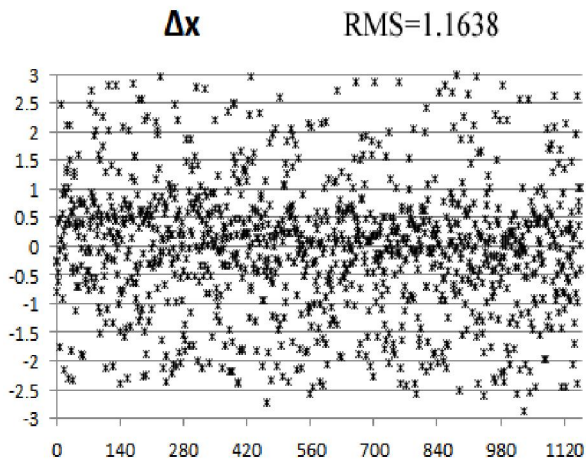
Figure 5. Image No. 2

Figure 6. Attitude correction of Image No. 2

Figure 7 shows the residuals of x and y of each corresponding points on the two images before optimization process. Figure 8 shows the residuals of x and y of each corresponding points on the two images after local optimization, and Figure 9 shows the final residuals. Figure 10 shows the final mosaic by using the steps described in this paper.

Figure 7. The residuals of x and y of each pair of corresponding points on two neighbouring images before optimization process.
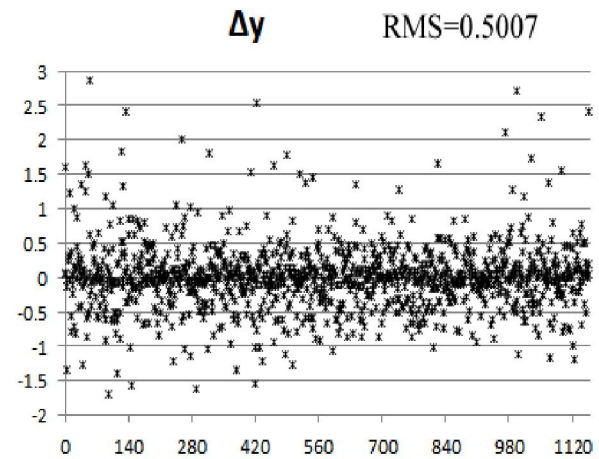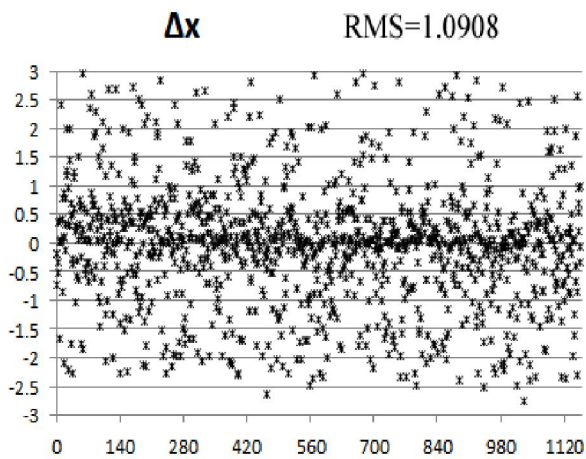
Figure 8. The residuals of x and y of each pair of corresponding points on two neighbouring images after local optimization process.
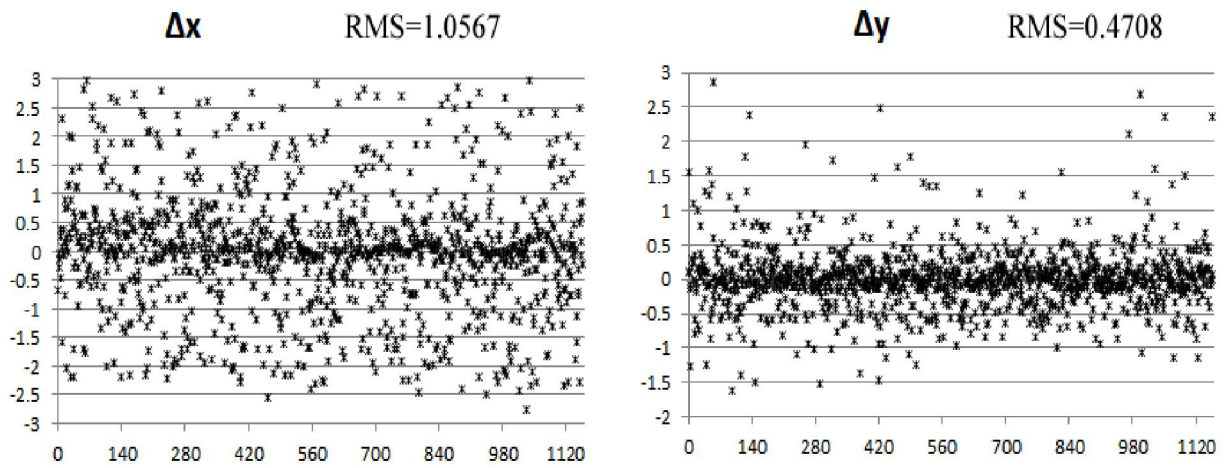
Figure 9. Final optimization results.

Figure 10. The final mosaic

V.   Conclusions

Figure 6 shows the result of attitude correction, and the distortion caused by the attitude of camera is corrected after this step. Compared with Figure 7, we can see that in Figure 8 the residuals of x and y between corresponding feature points are improved by the local optimization process, and the RMS of x and y are reduced from 1.1638 and 0.5683 to 1.0908 and 0.5007. In Figure 8, we can see that the residuals are further improved by the global optimization step, the RMS of x and y are reduced to 1.0567 and 0.4708. The final mosaic shown in Figure 10 denotes that this two-step optimization method is suitable for sequence images mosaicking.

References

[1]   J. Sun. "Principle and Application of Remote Sensing," Wuhan University Press, Wuhan, 2003.

[2]   J. Semple and G. Kneebone. "Algebraic projective geometry," Oxford University Press, 1952.

[3]   S. Ma and Z. Zhang. "Computer Vision," Sicience Press, Beijing, 2004.

[4]   O. Faugeras. "Three-Dimensional Computer Vision: A Geometric Viewpoint," MIT Press, Cambride, MA. 1993.

[5]   L. Van Gool, T. Moons, and D. Ungureanu. "Affine / photometric invariants for planar intensity patterns," Proceedings of the 4th European Conference on Computer Vision, Cambridge, UK, 1996, pp. 642-651.

[6]   D. Lowe. "Object recognition from local scale-invariant features," Proceedings of International Conference on Computer Vision, Corfu, Greece, 1999, pp. 1150-1157.

[7]   D. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer vision*, vol 60, pp. 91-110, Feb., 2004.

[8]   R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, 2nd ed, Cambridge University Press, 2004, pp. 258.

[9]   P. Wang and Y. Xu. Photogrammetry. Wuhan University Press, 2005, pp. 35-36.

[10]   Y. Wang. "Research on Key Technologies of Image Automatic Mosaic on Image Space," PH.D Dissertation, PLA Information University, 2008, pp. 57-60.

[11]   C. Xing, J. Wang and Y. Xu. "Overlap Analysis of the Images from Unmanned Aerial Vehicles," Proceedings of the International Conference on Electrical and Control Engineering, Wuhan, China, 2010, pp. 1459-1462.

[12]   F. Caballero, L. Merino, J. Ferruz, and A. Ollero. "Homography Based Kalman Filter for Mosaic Building. Applications to UAV position estimation," Proceedings of IEEE International Conference on Robotics and Automation Roma, Italy, 2007, pp. 2004-2009.

[13]   W. Press, S. Teukolsky, W. Vetterling and B. Flannery. "Numerical Recipes in C: the Art of Scientific Computing," 2nd ed., Cambridge University Press, Cambridge, 1992.

Cheng Xing, born in China, July, 1982. A Ph. D student in the School of Geodesy and Geomatics, Wuhan University. The major is digital close range photogrammetry.

Jinling Wang, born in China, 1961. He is a A/Professor in the School of Surveying and Spatial Information Systems, University of New South Wales. The major research fields are quality control and image processing.

Yaming Xu, born in China, July, 1964. He is a professor in the School of Geodesy and Geomatics, Wuhan University. The major research fields are digital photogrammetry and surveying engineering.