AUTOMATIC MATCHING AND GEO-REFERENCING OF HISTORICAL AERIAL IMAGES

I-Wen Chen[1], Hou-Ren Chen[2], Yi-Hsing Tseng[b]

Department of Geomatics, National Cheng Kung University, No.1, University Rd., Tainan City 70101, Taiwan

[1] sjasper1323@gmail.com, [2] P66024146@mail.ncku.edu.tw, [3]tseng@mail.ncku.edu.tw

Abstract

Nowadays, aerial images present a “bird’s-eye” view of geographical environment, and historical one provides the spatial information in the past. Through multi-temporal aerial images, we can analyze dynamic environmental changes. In Taiwan, Research Center for Humanities and Social Sciences (RCHSS) of Academia Sinica, has collected and scanned abundant historical maps and aerial images. By being processed through methods of computer vision, those materials can achieve greater value. Most of the historical aerial images haven’t been registered since there were no precise POS system for orientation assisting in the past. To handle the great quantity of images, we develop an automatic process to match historical aerial images by Scale Invariant Feature Transform (Lowe, 2004). This matching algorithm extracts extreme values in scale space, and becomes invariant image features, which are robust in rotation, scale, noise, and illumination. If two images have the same image feature point, we can use these points to do affine transformation or projective transformation for image alignment. Research that using feature points of SIFT for automatic registration of historical aerial images has proven feasible (Rau, 2014).

After image matching and alignment automatically, we only have the relative orientation of images. We still have to add control points manually for registration through least square adjustment based on collinear equation. Finally, we can use those feature points extracted by SIFT to build control image database in future work. Every new image will be query image and be extracted. If features of new points match with the point data in database, it means that the query image probably is overlapped with control images and then become new control data. After feature extracting, all computation is based on point data instead of image data, so the requirement of computation is low. With the growth of the database, more and more query image can be matched and aligned automatically. Also, further study such as multi-temporal environmental changes can be investigated by using this temporal spatial data system.

Key words：Historical aerial images; Automatic image matching; Image registration

1. INTRODUCTION

1.1. Multitemporal Aerial Photographs

Aerial photograph directly show the reality of geographical environment. With development of camera and aerial technology, not only aerial photographs but also remote sensing image data increase continuously. And multi-period aerial photographs can provide the evidence of environmental change. The Research Center for Humanities and Social Sciences (RCHSS), Taiwan Academia Sinica, conserves numerous historical maps and films. Some of them have georeferencing data, and have been georeferenced manually. But most portion of historical material doesn’t have georeferencing data, and they are too numerous to be processed manually for geo-referencing. But they could be processed by methods of computer vision after digital scanning. For example, image feature extraction of aerial photographs, image matching, image mosaicking, or using dense image matching for automatic reconstruction of buildings. If those historical aerial photographs have been georeferenced, they can display with other aerial photographs or satellite images. It's useful for some research, such as environmental changes, urban planning, and coastline monitoring.

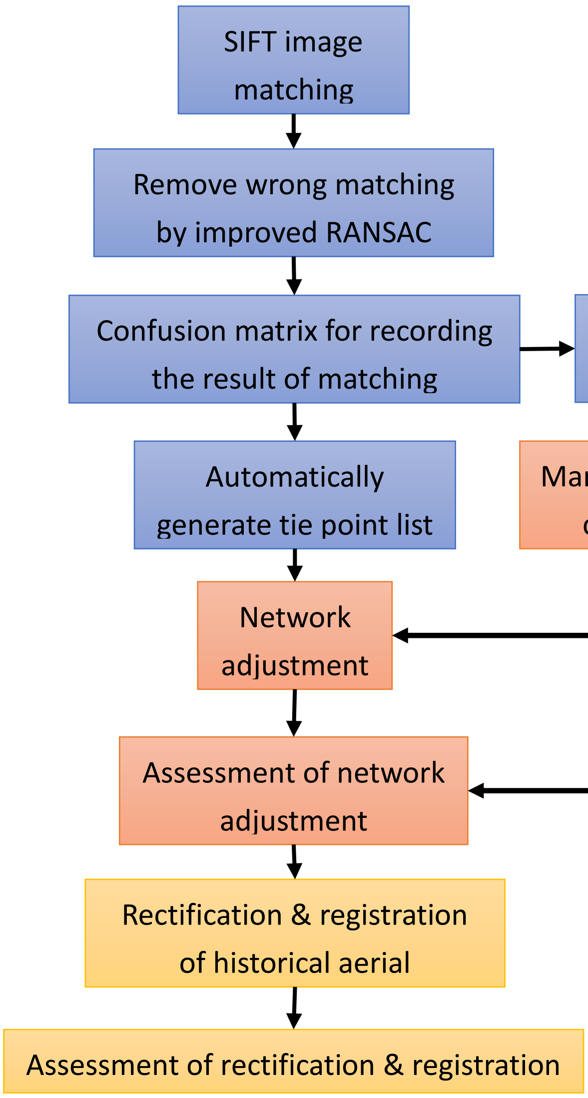
1.2. Image Feature

Image features are the specific structures in the image, such as point, edge or object. Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is one of the most popular method for extracting image feature. And these features are robust to rotation, scale, noise, illumination, and affine transformation. By comparing the descriptions of each feature, two or more images might have similar features. These matching features mean that images are over lapped. But there might be some wrong matching, then RANdom SAmple Consensus (RANSAC) is needed (Fischler and Bolles, 1981) to remove outlier. This algorithm estimates parameters for fitting mathematical model from a set of observed data by iterative method. After RNASAC, the matching points of historical aerial photographs can be used as tie points for network adjustment of photogrammetry. We also manually add some control points for coordinate transformation. Finally, we finished the geo referencing of historical aerial photograph by two kinds of coordinate transformation, 2D affine transformation and 2D projective transformation.

1.3. Research Framework

In this research, using image processing of computer vision shows the efficiency and feasibility of dealing with the large quantity of historical aerial photographs. Figure 1 shows the flowchart of this research. First, we use image feature extraction and matching for understanding the overlapped relationship of historical aerial photographs. Then we select some control points for coordinate transformation by calculation of network adjustment. We also consider two kinds of 2D coordinate transformation. Finally, we finished the georeferencing of historical aerial photographs. In this study, we developed an efficient way for matching historical aerial photographs and generating tie points automatically. We only need to manually add some control points for georeferencing of historical aerial photographs. Those georeferenced photographs became spatiotemporal data, and it's more useful for analysis of temporal environmental changes and some research about coastal, land use, and urban planning.

Figure 1. The flowchart of this study



2. DATA

2.1 Historical Aerial Photographs

Historical aerial photographs provide the truest evidences and plentiful information about land cover in the past. We can understand the environmental change and natural landscape evolution from multitemporal aerial photographs. The Research Center for Humanities and Social Sciences (RCHSS), Taiwan Academia Sinica, conserves numerous historical maps and aerial photography films which cover Taiwan and mainland China. They successively scanned these valuable historical photographs into digital images. Although the information of interior and exterior parameters had disappeared after digital scanning, those materials became more convenient via image processing. Figure 2 shows one example of scanned historical aerial photograph. Every image was scanned with a Kodak color control patches for representing the image quality. The same pattern of each image might influence the result of image matching. For reducing consumption of calculation and diminishing the errors of wrong matching owing to the color control patches. Those photographs had been cropped, as Figure 3 shows. We used this kind of edited photographs as test image in this study.

Figure 2. Example of scanned image

Figure 3. Example of cropped image

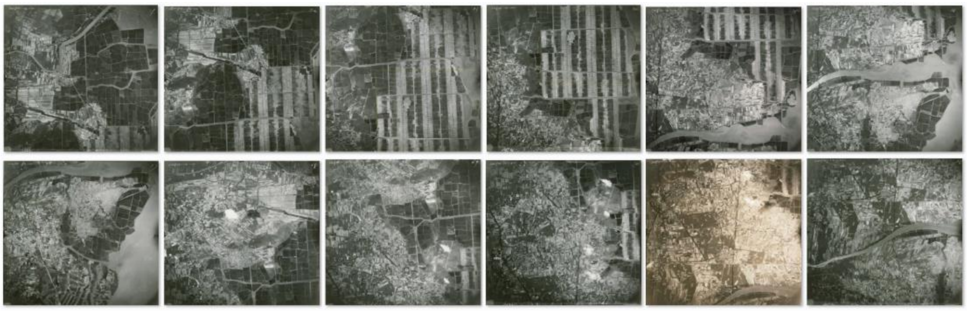
2.2 Recent Satellite Images

With the assistant of GPS (Global Positioning System) and IMU (Inertial Measurement Unit), recent aerial and satellite images are easily to be geo-referenced. In this study, we manually selected control points from FORMOSAT-2 satellite images aquired in 2015 for coordinate transformation. And we choose the Projected Coordinate Systems, TWD97, which was suitable for use in Taiwan. FORMOSAT-2 satellite is a remote sensing satellite which can capture images over the entire earth with 2 meters ground sampling distance in panchromatic band. Because natural landscape and human environment have changed significantly during decades, it’s not easy to find the corresponding control points between historical aerial photographs and satellite images. After laborious process for seeking corresponding control points, we added them to the network adjustment with tie points which derived from image matching of historical aerial photographs by SIFT. Finally, we can get the parameters for the georeferencing of historical aerial photographs.

3. METHOD

The test image datasets contain 12 photographs (2 flight lines) which were taken by US military in June 1956 and has 50% side overlap and 60% forward overlap. Figure 4 shows the test historical aerial photographs in this study. First, we used SIFT algorithm image feature extraction and matching. Second, we use the concept of confusion matrix for recording the result of image matching. And the matching points could be organized for tie points of network adjustment. We consider 2D affine transformation and 2D projective transformation as the mathematical model of network adjustment. Third, we manually selected control points from historical aerial photographs and modern satellite images for coordinate transformation. There are more detailed descriptions in the following paragraph.

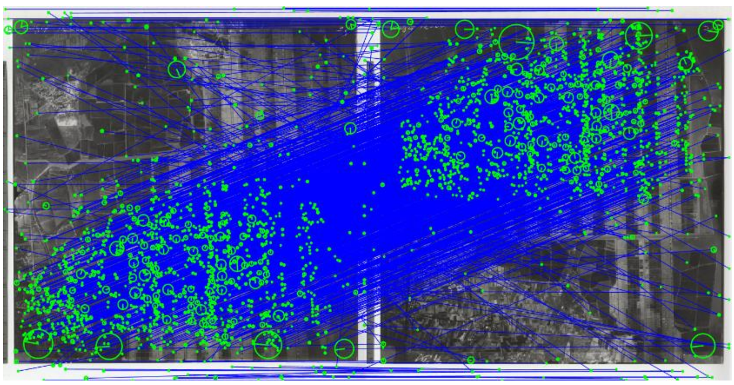
Figure 4. Test historical aerial photographs in this study



3.1. Image Feature Extraction and Matching

In order to automatically deal with the great amount of historical aerial photograph provided from RCHSS, we used SIFT for image feature extraction in this study. This algorithm is widely used in computer vision, and it has high probability against a large database of features and great recognition for object and scene. It uses Gaussian filter to build scale space for locating extreme values which are possible image features. After filtering by two kinds of threshold, every feature point is described by three dimensional histogram. If the descriptions of two features are similar, they can be matched as figure 5 shows. But there might be some similar image features that actually represent different points. So we use RANSAC algorithm for removing wrong matching. There is more detail introduction in the following section.

Figure 5. Image matching by SIFT



3.2. Improved RANSAC

RANSAC is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. Because most of the test historical aerial photographs are the view of the flat area, so we supposed 2D affine transformation as the mathematical model for removing wrong matching. Figure 7 shows the flowchart of RANSAC. First, randomly select a subset S(i) for calculating the mathematical model M(i). Then test all data and remove outliers with a given threshold to get the consensus set S*(i). Repeat the former step for given times, and we can obtain different quantity of consensus set S*. Finally, choose the largest consensus set to calculate the parameters of mathematical model. Then these parameters are most fitted for most of the origin data. It's is very useful for removing outliers of matching as figure 6 shows. But it's not efficient. And the final parameters are calculated by the consensus set which remove outliers only once. So we improved the original RANSAC for higher efficiency and precision. Figure 8 shows the flowchart of improved RANSAC. We not only test the randomly sampling matching points with a given threshold, but also consider the distribution of points in every trail of RANSAC. We did more randomly sample in in every trail in order to achieve the goal. After given times of trails, we also choose the largest consensus set for calculating parameters of mathematical model, and test the consensus set again. If it contains outliers, remove it and recalculate the parameters until no outlier in the consensus set. Then the final parameters is the most fitted and the consensus set has no outlier.

Figure 6. Image matching with RANSAC

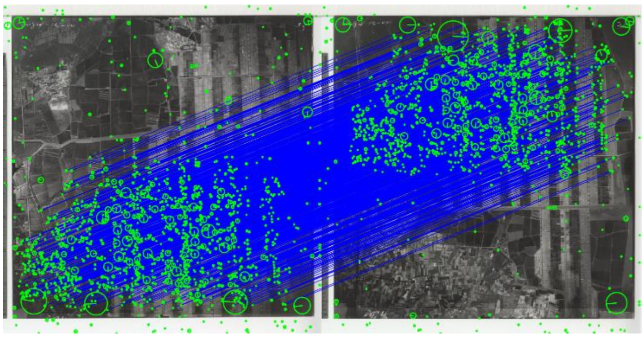


Figure 7. The flowchart of original RANSAC

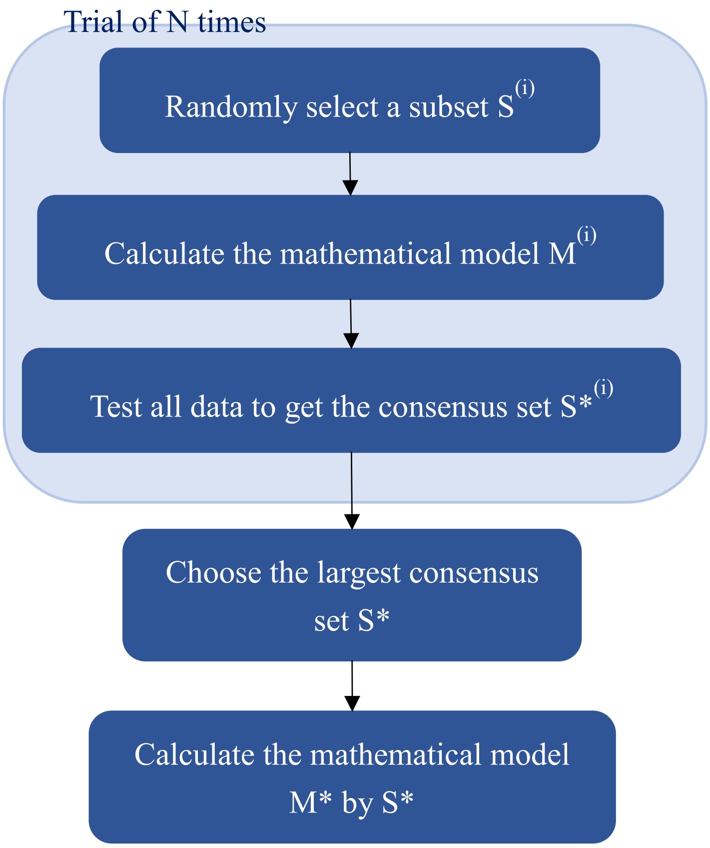
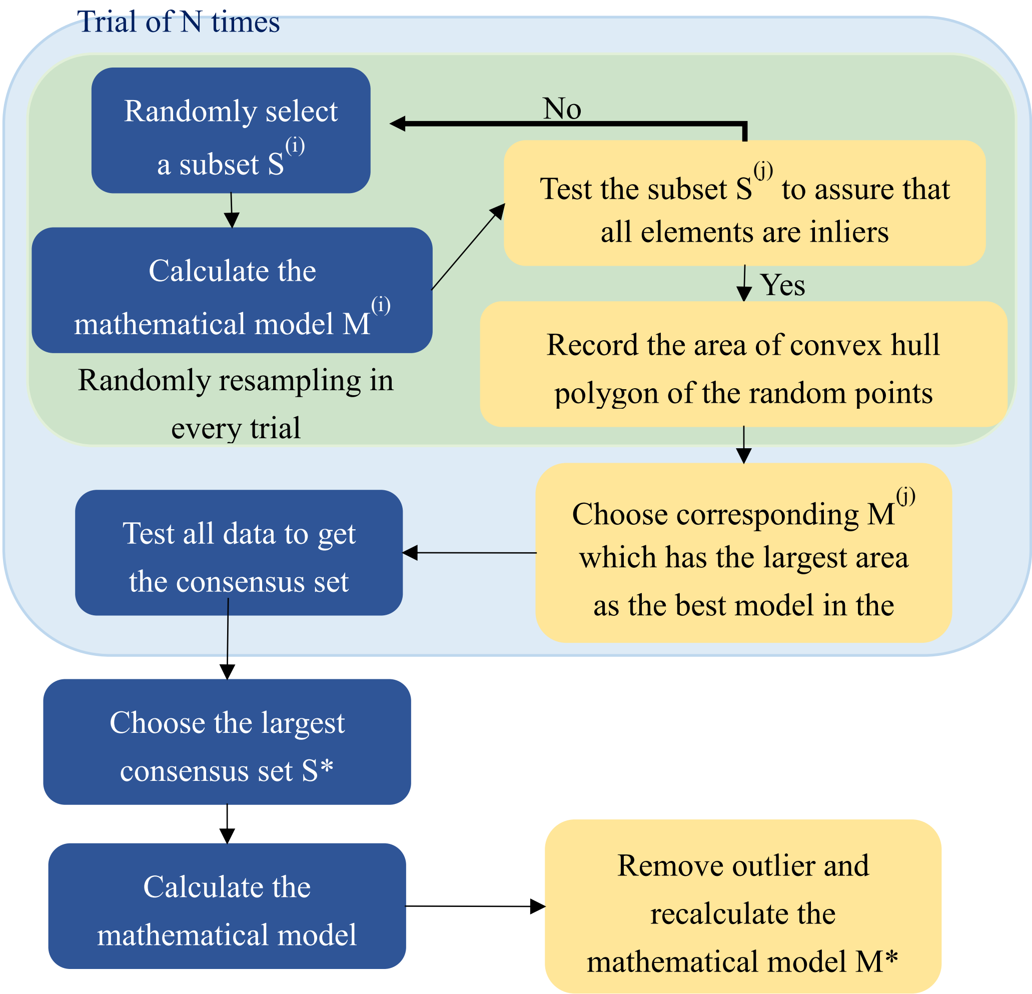


Figure 8. The flowchart of improved RANSAC



3.3. Confusion Matrix

After image matching for every image pairs, we use the concept of confusion matrix to record the result of image matching. Table 1 shows an example of confusion matrix. It's an upper triangular matrix, and every element means the quantity of matching points for corresponding image pair. We can understand the relationship of overlapped images by this kind of matrix.

Image number	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1		64	4	3	68	3	1
2			639	80	161	53	0
3				1564	181	175	174
4					43	443	525
5						1150	170
6							183
7							

Table 1. Example of confusion matrix

3.4. Automatically Generalize Tie Points

We obtained many image features from every image after image matching. And some of them might exist in more than two images. If one feature point exists in more than two images, the redundant observation could make the accuracy of network adjustment better. So we developed a program for automatically generalize these tie points and find how many images will a feature point appears. And the list of tie points could be used for network adjustment.

3.5. Network Adjustment

Due to the largest quantity of historical aerial photographs, it’s not efficiency if we calculate the coordinate transformation parameters of image one by one. After image matching and automatically generating tie points, we can use network adjustment to calculate the exterior parameters with adding control points.

3.5.1. *Manually Select Control Points and Check points*

For georeferencing of historical aerial photographs, we manually selected control points from both historical aerial photographs (1956) and FORMOSAT-2 satellite images (2015). So we can calculate the parameters of coordinate transformation. And the historical aerial photograph can be georeferenced to the same coordinate with other images. In this study, we also select some check points for the assessment of network adjustment. This step is the most laborious, because the land cover had changed through decades and that makes us not easy to select corresponding points for control points.

3.5.2. *Coordinate Transformation*

Because most of the historical aerial photograph are the flat area in Taiwan, we use two kinds transformation for coordinate transformation. First transformation, 2D affine transformation, contains six parameters which derived from scale, rotation, and offset of two directions. Equation (1) shows the expressed affine equation.

$$\begin{aligned} \mathrm{c} &= \mathrm{L}_1 E + L_2 N + L_3 \\ \mathrm{r} &= \mathrm{L}_4 E + L_5 E + L_6 \end{aligned} \qquad (1)$$

where (c, r) are image coordinates of control points of historical aerial photographs, (E, N) are the corresponding TWD97 coordinates of FORMASAT-2 satellite image, and L1 - L6 are parameters of 2D affine transformation.

This second transformation, 2D projective transformation, projects coordinate to another one which’s axes are not parallel to the former ones, so it has 2 more parameters than affine transformation. Equation (2) shows the expressed transformation equation. Because aerial photographs were taken with airplane, and the orientation is always changing during flight. So this transformation should be more accurate than 2D affine transformation theoretically.

$$c = \frac{L_1 E + L_2 E + L_3}{L_7 E + L_8 E + 1}$$

$$r = \frac{L_4 E + L_5 E + L_6}{L_7 E + L_8 E + 1}$$

where (c, r) and (E, N) are the same as mentioned before, and L1 - L8 are parameters of 2D projective transformation.

3.6. Georeferencing

After the calculation of network adjustment, we obtained two kinds of exterior parameters of all historical aerial photographs. And these photographs could be georeferenced to ground coordinate, TWD 97, by the process of image resampling.

4. RESULT

4.1. Image Matching with Improved RANSAC

Table 2 shows the comparison of original and improved RANSAC. We test three cases of different image resolution and five kinds of trail times for each case. The execution time of improved RANSAC is less than the original one, because the improved RANSAC didn't test all data for every trail. And the precision of improved RANSAC is also better than the other, because the mathematical model had been updated with removing outlier again and again.

Table 2. Comparison of original and improved RANSAC

Image resolution	Quantity of matching (SIFT)	Threshold	Original RANSAC				Improved RANSAC			
			Trail times	Quantity of matching	Execution time (s)	RMSE (pixel)	Trail times	Quantity of matching	Execution time (s)	RMSE (pixel)
1/10 Image resolution GSD: 6.9 m	471	1 pixel	100	168	0.6	0.686	10*10	139	0.1	0.523
			400	178	2.1	0.634	20*20	176	0.2	0.555
			900	194	4.6	0.629	30*30	180	0.3	0.538
			1600	185	7.8	0.586	40*40	182	0.5	0.553
			2500	185	12.3	0.669	50*50	188	0.8	0.611
1/3 image resolution GSD: 2.1 m	5220	3 pixel	100	2160	6.4	1.685	10*10	2017	1.9	1.568
			400	2218	23.7	1.787	20*20	2055	1.1	1.614
			900	2206	53.7	2.063	30*30	2107	2.3	1.639
			1600	2204	91.0	2.099	40*40	2065	4.0	1.580
			2500	2232	146.8	1.886	50*50	1990	4.9	1.651
Original image resolution GSD: 0.69 m	17949	10 pixel	100	6479	28.2	6.222	10*10	6658	27.3	5.327
			400	6357	94.6	6.603	20*20	5676	18.4	5.415
			900	6877	156.8	7.028	30*30	6216	36.7	5.459
			1600	6680	303.5	6.307	40*40	6334	15.3	4.889
			2500	7024	532.2	6.234	50*50	6686	16.0	5.162

4.2. Accuracy Assessment of Network Adjustment

We can expect that the accuracy of 2D projective transformation will better than the accuracy of 2D affine transformation. Figure 9 shows the residual vectors of check points in objective coordinate, and Figure 10 shows the residual vector of tie points in image coordinate. Both support the assumption. The length of residual vectors of 2D projective transformation are all less than 2D affine transformation, it means that the projective transformation is more suitable for the georeferencing of historical aerial photographs.

Figure 9. The residual vector of check points (objective coordinate)

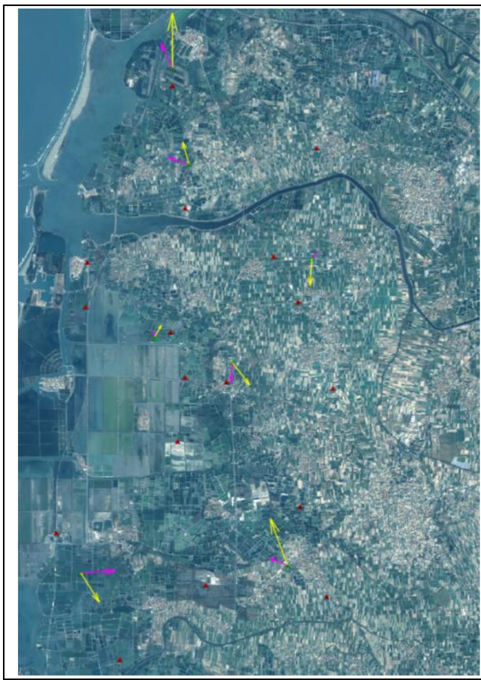
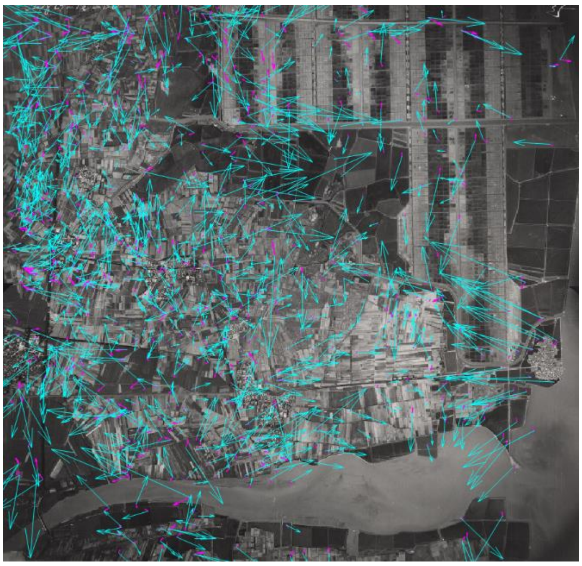
	2D affine transformation 2D projective transformation Magnification of residual vector: 20 times

Figure 10. The residual vector of tie points (image coordinate)

	2D affine transformation 2D projective transformation Magnification of residual vector: 10 times

4.3. Georeferencing of Historical Aerial Photographs

Both kinds of transformation can provide quite good georeferencing results after network adjustment as Figure 11 and 12 show. We zoom in to get a close-up view of georeferenced historical. Table 3 and table 4 show the comparisons of two kinds of transformation between historical aerial photographs. The continuity of line feature (like road or river) of 2D projective transformation are better than 2D affine transformation for the direction of latitude and longitude. Table 5 shows the comparisons of two kinds of transformation between historical aerial photograph and FORMOSAT-2 satellite image. The continuity of line feature of 2D projective transformation is also better than 2D affine transformation. It represents that the 2D projective transformation is more suitable for coordinate transformation than the other one.

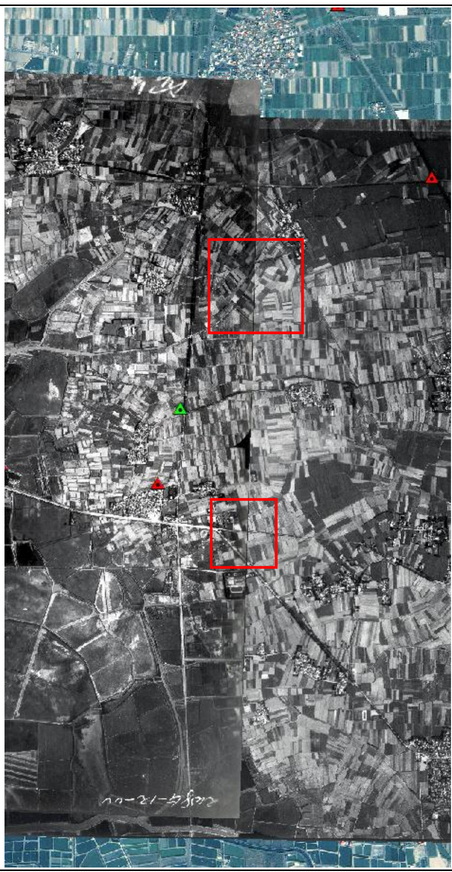
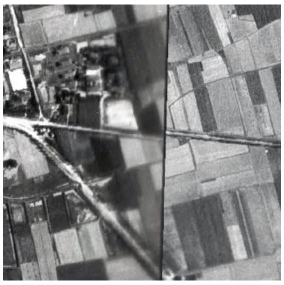
Figure 11. Result of 2D affine transformation

Figure 12. Result of 2D projective transformation

Table 3. Comparison between historical aerial photographs (longitude direction)

2D affine transformation	
2D projective transformation	

Table 4. Comparison between historical aerial photographs (latitude direction)

2D affine transformation		
		

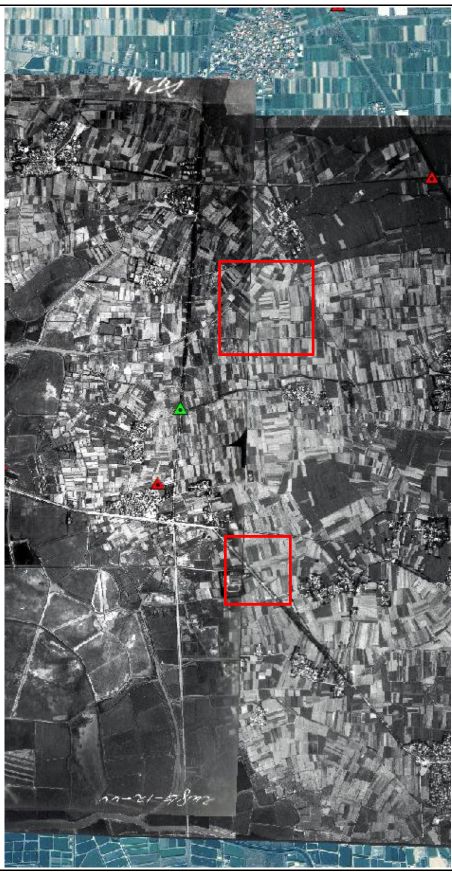
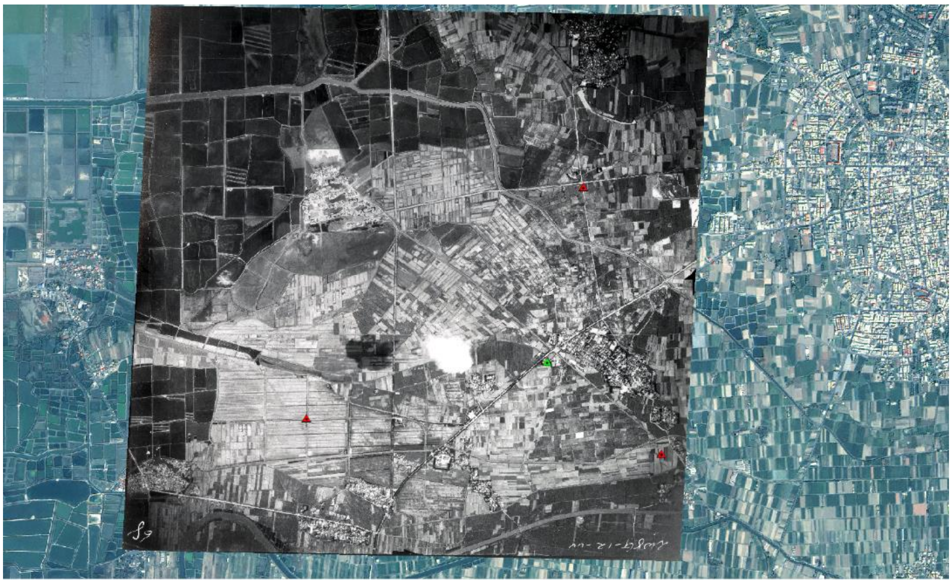
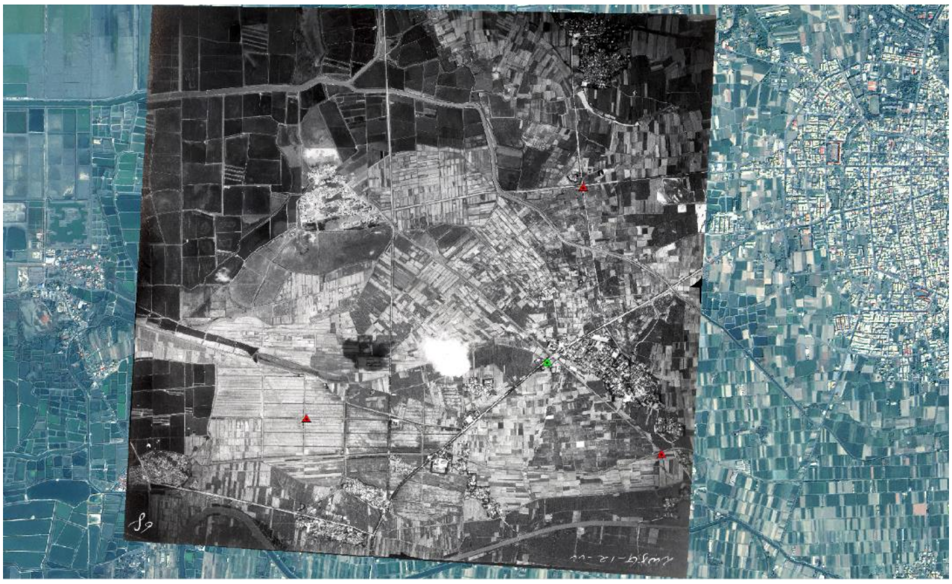
<table>
<tr><td rowspan="2">2D projective transformation</td><td rowspan="2"></td><td></td></tr>
<tr><td></td></tr>
</table>

Table 5. Georeferencing of historical aerial photographs

2D affine transformation	

2D projective transformation	

5. CONCLUSIONS

5.1. The Performance of Image Processing

Using SIFT and RANSAC is an efficiency way to automatically reuse those historical aerial photographs on extracting conjugate point, image matching, separating image groups, and generating the tie points table for network adjustment. Also, we make RANSAC more efficient and more precise. It's is useful if we need to do image matching for a lot of image. However, the only problem is that it's not easy to precisely select the same control point in both historical aerial photographs and modern satellite images due to the dramatic changes of the nature land cover and human residence from 1960's to nowadays.

5.2. The Accuracy of Coordinate Transformation

Using 2D projective transformation as the mathematical model for network adjustment is better than 2D affine transformation. For example, the residual vector of check points, and the residual vector of tie points. But the execution time of 2D projective transformation is much more than the other one.

We also finished the georeferencing of historical aerial photographs by the parameters derived from network adjustment. The parameters of 2D projective transformation can make the georeferencing of historical aerial photographs more accurate. The continuity of line feature is better than 2D affine transformation. Not only modern aerial photographs but also historical ones can be displayed in the same coordinate.

In this study, we build a highly automatic way for image matching, network adjustment, and georeferencing of historical aerial photographs. So it's possible that the numerous materials could be automatically geo-referencing via computer vision. In the future work, we want to use other transformation of more parameters or better accurate to register historical aerial photographs, like 3D projective transform or fundamental matrix.

6. ACKNOWLEDGEMENTS

Thanks to the financial aid (project number: MOST 104-2221-E-006-223-MY3) of Ministry of Science and Technology. Make us finish this study successfully. And thanks to RCHSS, Academia Sinica, who provided us great amount of historical aerial photographs which spent them a lot of time and manpower scanning and archiving. So we can try to make these materials more useful and valuable. Other researches which are relative to historical aerial photographs could process image data by our method.

7. REFERENCES

Adami, A (2015). 4D City transformations by time series of aerial images. The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 40(5): 339.

Baya, Herbert, T. Tuytelaars, and L. Van Gool (2008). Speeded-up robust features (SURF). Computer Vision and Image Understanding, 110(3), pp. 346-359.

Chiang, J. T (2014). *Land Use Changes of the Coastal Zone of Old Tainan City in the Past Hundred Years by Using Temporal Spatial Information.* National Cheng Kung University, Master thesis.

Fischler, Martin A., and Robert C. Bolles (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM 24(6): 381-395.

Harris, Chris, and Mike Stephens (1988). A combined corner and edge detector. Alvey vision conference. Vol. 15.

Jao, F. J. "Historical GIS Data Processing - Automatic Historical Aerial Image Registration using SIFT and Least-Squares." National Cheng Kung University, Master thesis, 2014.

Juan, Luo, and Oubong Gwun (2009). A comparison of sift, pca-sift and surf. International Journal of Image Processing (IJIP) 3(4), pp. 143-152.

Kadmon, R., & Harari-Kremer, R. (1999). Studying long-term vegetation dynamics using digital processing of historical aerial photographs. Remote Sensing of Environment, 68(2), pp. 164-176.

Kim, Jae Sung, Christopher C. Miller, and James Bethel (2010). "Automated Georeferencing of Historic Aerial Photography." Journal of Terrestrial Observation 2(1): 6.

Lowe, David G 1999. “Object recognition from local scale-invariant features.” Computer vision, 1999. The proceedings of the seventh IEEE international conference on. Vol. 2. Ieee.

Lowe, David G (2004). "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60(2), pp. 91-110.

Mikolajczyk, Krystian, and Cordelia Schmid (2004). "Scale & affine invariant interest point detectors." International journal of computer vision 60(1), pp. 63-86.

Nebiker, Stephan, Natalie Lack, and Marianne Deuber (2014). "Building change detection from historical aerial photographs using dense image matching and object-based image analysis." Remote Sensing 6(9), pp. 8310-8336.

Rosten, Edward, and Tom Drummond (2006). Machine learning for high-speed corner detection. European conference on computer vision. Springer Berlin Heidelberg.

Yu, Le, Dengrong Zhang, and Eun-Jung Holden (2008). "A fast and fully automatic registration approach based on point features for multi-source remote-sensing images." Computers & Geosciences 34(7), pp. 838-848.