AUTOMATIC IMAGE MATCHING AND GEOREFERENCING OF DIGITIZED HISTORICAL AERIAL PHOTOGRAPHS

I-Wen Chen[1], Hou-Ren Chen[2], Yi-Hsing Tseng[3]

[123] National Cheng Kung University, No.1, Daxue Rd., East Dist., Tainan City 701, Taiwan (R.O.C.)
[1]Email: sjasper1323@gmail.com, [2]Email: P66024146@mail.ncku.edu.tw, [3]Email: tseng@mail.ncku.edu.tw

KEY WORDS: Image registration; Alignment of images; Rectification

ABSTRACT: Historical aerial photographs can directly witness the historical landscape environment and link to the spatial information in the past. Which are important information sources to evaluate the temporal spatial evolution of zones of interest. In Taiwan, abundant historical aerial images have been acquired and archived by Research Center for Humanities and Social Sciences (RCHSS) of Academia Sinica. However, camera report of images was usually confidential. Most of the historical aerial images haven't been registered since there was no precise POS system for orientation assisting in the past.

To handle such great quantity of images, this study was processed through methods of computer vision. Image features were extracted and matched by Scale Invariant Feature Transform (SIFT). To improve matching accuracy and efficiency of conjugate points, an improved Random Sample Consensus (RANSAC) was promoted by considering the error and recording the largest areas covered by randomly sampling points. Then, according to adjacency matrix, the relationships of conjugate points were established and stored.

After image matching and alignment automatically, the relative orientation of images can be acquired. For registration, control points are needed to be added to do network adjustment based on 2D coordinate transformation (2D affine and 2D projective). Finally, those feature points matched by this procedure can be used to build control image database, and all computation is based on point data instead of image data. Also, further study such as multi-temporal environmental changes can be investigated by using this temporal spatial data system.

1. INTRODUCTION

1.1 Motivation and Objective

Historical aerial photographs provide the direct evidences and plentiful information about land cover in the past. The environmental change and natural landscape evolution could be analyzed from multi-temporal aerial photographs. During and after World War Ⅱ, abundant historical aerial photos have been acquired in Taiwan for a variety of missions which were taken by U.S military. Later acquired routinely by Taiwan Air Force and Aerial Survey Office. Due to the aerial images were considered as confidential information, there were neither calibration reports nor any camera parameters about the historical aerial images taken at that time.

To deal with these massive materials, an automatic georeferencing process needed to be conducted. Image processing of computer vision was used in this study. By SIFT and RANSAC, a higher precision and efficiency of image matching could be reached. And the concept of adjacency matrix for recording the result was used to realize the conjugate relationships among images. With manually adding control points for calculating parameters of coordinate transformation through least square adjustment, the georeferencing of hundreds of historical aerial photographs would be conducted and assessed.

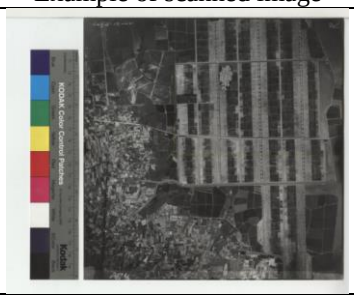
1.2 Reference to Related Work

Using multitemporal aerial images and going through the georeferencing progress, a vegitation classification was done and validated with the ground truth information to do statistical analysis. The result showed that historical aerial images can provide high resolution and large scale of landscape information which is proper to temporal spatial changes. (Kadmon and Harari-Kremer,1999) Scale-invariant feature transform (Lowe, 1999, 2004) approach equipped with a reliable outlier removal procedure provides an accurate, robust, and efficient process to deal with various remote sensing optical and synthetic aperture radar images taken at different situations (multispectral, multisensor, and multitemporal) with the affine transformation model. (Gong, et al., 2014) With these methodologies, we can run the image registration and georeferencing in an automatic and potent way.

1.3 Experimental Materials

1.3.1 Historical Aerial Photographs: The Research Center for Humanities and Social Sciences (RCHSS), Taiwan Academia Sinica, had conserved numerous historical maps and aerial photography films which covered Taiwan and mainland China. They successively scanned these valuable materials into digital images. Table 1 shows the 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. And then used this kind of edited images as test materials in this study.

Table 1. Example of scanned and edited image (Copyright by RCHSS, 1956)

Example of scanned image	Example of cropped image
	

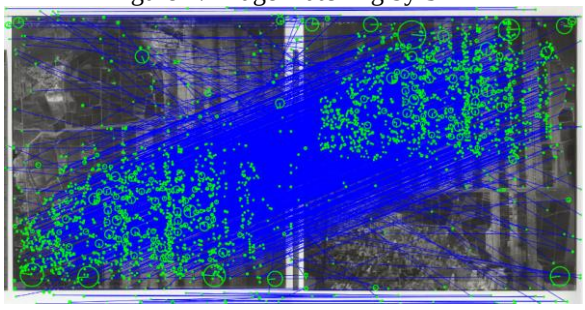
1.3.2 Recent Satellite Images: With the assistant of Global Positioning System and Inertial Measurement Unit, nowadays aerial and satellite images are easy to be georeferenced. In this study, projected coordinate system, TWD97 was chosen. We manually selected control points from FORMOSAT-2 satellite images acquired in 2015 for coordinate transformation. FORMOSAT-2 is a remote sensing satellite which can capture panchromatic band images with 2 meters ground sampling distance. 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, they were added to do network adjustment with tie points derived from image matching by SIFT. Finally, we got the parameters for historical aerial photographs georeferencing.

2. MATCH AND GROUP THE HISTORICAL AERIAL IMAGES

2.1 Feature Extraction

In order to automatically deal with the great amount of historical aerial photographs provided from RCHSS, SIFT for image feature extraction was used. This algorithm has been widely used in computer vision, which 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 1 shows. But there might be some similar image features that actually represent different points. So, RANSAC algorithm was used for removing wrong matching.

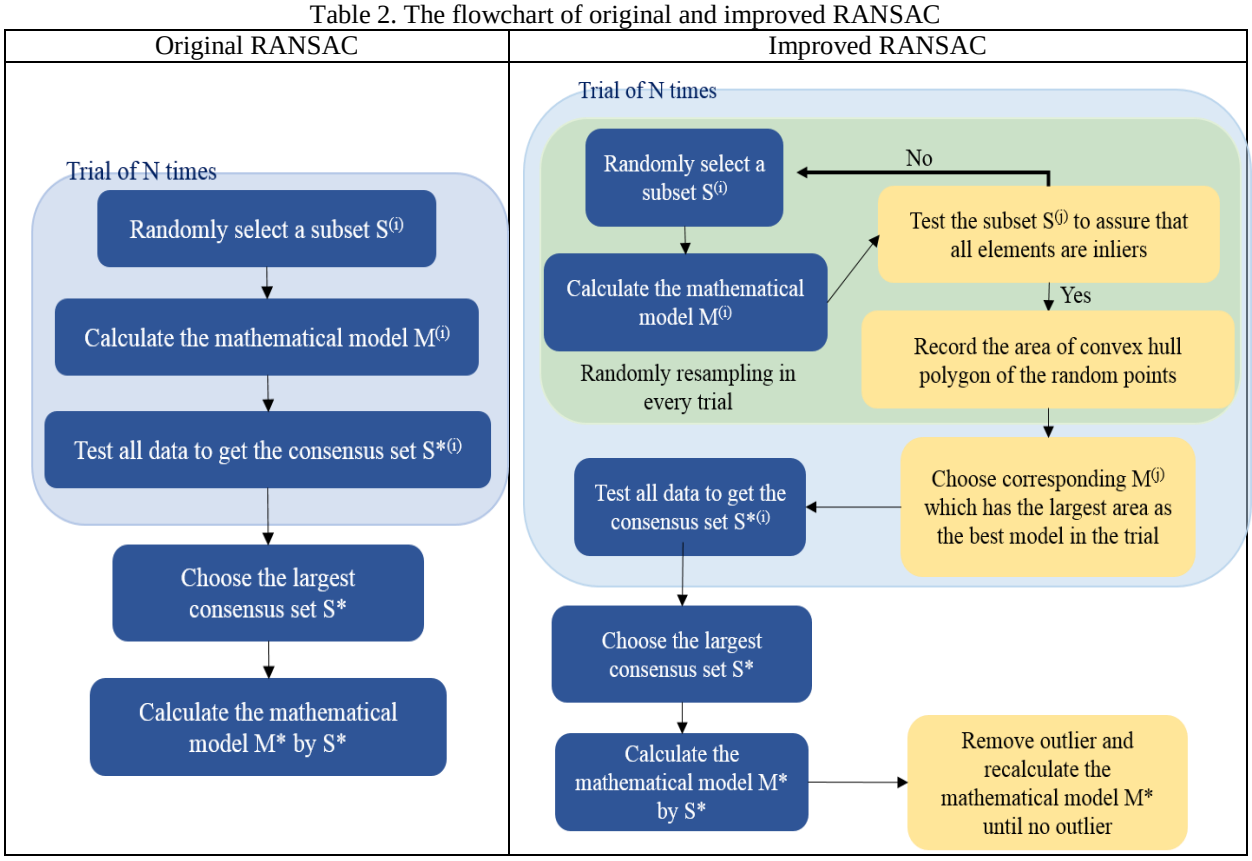
Figure 1. Image matching by SIFT



2.2 Improved RANSAC

RANSAC is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. Table 2 shows the flowchart of original and improved 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. Since the final parameters are calculated by the consensus set which removed outliers only once, the original RANSAC is not efficient. So we improved it for less computation time and higher precision. The random sampling matching points were tested with a given threshold, and also considered the distribution of points in every trail. After giving times of trails, the largest consensus set was selected for calculating parameters of mathematical model. If it contained outliers, removed it and recalculated the parameters until no outlier in the consensus set. With this iteration, the final parameters were the most fitted and eliminated the outlier.

Table 2. The flowchart of original and improved RANSAC



2.3 Adjacency matrix

After image matching for every image pairs, the concept of adjacency matrix was used to record the result of image matching. It was an upper triangular matrix, and every element meant the quantity of matching points for corresponding image pair. So the relationship of overlapped images could be indicated. For example, as table 3 shown, image number 3 and 4 had the most matching points, treated as the initial images group. And found the next maximum number in the third and fourth column and row, which happened in the image pair of image number 2 and 3. This meant that there was overlapped areas among image 2, 3 and 4 so they could merge into a bigger image. Repeating this procedure, all seven images could be merged into a whole image without manually selecting tie points. Even if there were different group of images or isolated images, the relationship sill could be discovered through this matrix.

Table 3. Example of adjacency 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							

3. GEOREFERENCING AND NETWORK ADJUSTMENT OF HISTORICAL AERIAL IMAGES

3.1 Manually Select Control Points

For georeferencing of historical aerial photographs, we manually selected control points from both historical aerial photographs and FORMOSAT-2 satellite images. 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.

3.2 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 added control points.

Because most of the historical aerial photograph are the flat area in Taiwan, two kinds transformation were used for coordinate transformation. First, 2D affine transformation, contains six parameters which derived from scale, rotation, and offset of two directions. Equation (3.1) shows the expressed affine equation. 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.

$$\begin{aligned} c &= L_1 E + L_2 N + L_3 \\ r &= L_4 E + L_5 E + L_6 \end{aligned} \tag{3.1}$$

Second, 2D projective transformation, projects coordinate to another one whose axes are not parallel like the former ones. So it has 2 more parameters than affine transformation. Equation (3.2) shows the expressed transformation equation. 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 - L8 are parameters of 2D projective transformation.

$$\begin{aligned} c &= \frac{L_1 E + L_2 E + L_3}{L_7 E + L_8 E + 1} \\ r &= \frac{L_4 E + L_5 E + L_6}{L_7 E + L_8 E + 1} \end{aligned} \tag{3.2}$$

Since the 2D projective transformation is a nonlinear equation, we needed to linearize the observation equations, and gave initial approximation for the unknowns. Then, solved the linearized equations by least squares, and update the initial approximations. Last, do the iteration until convergence condition is match. Because aerial photographs were taken with airplane, and the orientation was always changing during flight, this transformation should be more accurate than 2D affine transformation theoretically.

3.3 Resampling

After network adjustment, the transformation parameters could be got, and then the image resampling was needed to georeference historical aerial images into the same coordinate system. First, the corners of original images coordinate were transformed onto the object coordinate to defined the corresponding object boundary. Second, used the defined ground sampling distance to divide this region as the size of the resampling image. Then, used eight parameters transformed the center coordinates od the object grid into the historical aerial images. Last, the bilinear interpolation was used to interpolate corresponding color to each grid.

$$\begin{bmatrix} c \\ r \\ 1 \end{bmatrix} = \begin{bmatrix} L_1 & L_2 & L_3 \\ L_4 & L_5 & L_6 \\ L_7 & L_8 & 1 \end{bmatrix} \begin{bmatrix} E \\ N \\ 1 \end{bmatrix} = R_8 \begin{bmatrix} E \\ N \\ 1 \end{bmatrix} \quad (3.3)$$

$$\begin{bmatrix} c' \\ r' \\ 1 \end{bmatrix} = \begin{bmatrix} L_1 & L_2 & L_3 \\ L_4 & L_5 & L_6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} E \\ N \\ 1 \end{bmatrix} = R_6 \begin{bmatrix} E \\ N \\ 1 \end{bmatrix} \quad (3.4)$$

Equation (3.4) means the transformed relationship between the historical image coordinates and object coordinates. Which (c', r') means the historical aerial image coordinates, and (E,N) means the object coordinates under TWD97 coordinate system. And rotation matrix, R8, contains eight projective parameters, which is homogeneous matrix. While equation (3.3) means the transformed relationship between the resampled image coordinates and object coordinates. Which (c', r') means the historical aerial image coordinates after resampling. And rotation matrix, R6, means six affine parameters, which only contain scale and translation.

4. EXPERIMENT RESULT AND ANALYSIS

4.1 Automatic images matching result

In this study, two improvements were made to refine the efficiency and accuracy. First, tested the residual of check points, calculated by the mathematical model which acquired from the randomly selected subsets, passing the threshold. Also, considered the distribution of randomly selected points in every trail of RANSAC. Second, the final parameters were not calculated by the consensus set which remove outliers only once, but recalculated the parameters until no outlier in the consensus set. Table 4 shows the comparison of original and improved RANSAC. Three cases of different image resolution were tested, and five kinds of trail times for each case. The execution time of improved RANSAC was less than the original one, because it didn't test all data for every trail. And the precision of improved RANSAC was also better than the original one, because the mathematical model had been updated with removing outlier again and again.

Table 4. 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 Network Adjustment Test

We expected that the accuracy of 2D projective transformation will better than the accuracy of 2D affine transformation. Figure 2 shows the residual vectors of check points in objective coordinate, which supports the assumption. The length of residual vectors of 2D projective transformation are all less than 2D affine transformation, which means that the projective transformation is more suitable for the georeferencing of historical aerial photographs. In table 5, the nonlinear equation and iteration caused the execution time of 2D projective transformation was obviously longer than the 2D affine one. Because the aerial photographs were taken with changing orientation in the flight which had been mentioned in section 3.2, 2D projective transformation highly enhance the accuracy on the N axis.

Table 5. Comparison of the accuracy of check points with different transformation model and resolution

Transformation Method	Network Adj. Result	Original GSD: 0.7 m	1/3 resolution GSD: 2.1 m	1/10 resolution GSD: 6.9 m
2D Affine	Execution time	1 h 7 m	1 h 11 m	1h 10 m
	RMSE_E	±15.641 m	±17.185 m	±20.315 m
	RMSE_N	±45.503 m	±46.876 m	±45.326 m
2D Projective	Execution time	14 h 23 m	14 h 29 m	14 h 15 m
	RMSE_E	±22.250 m	±22.271 m	±22.947 m
	RMSE_N	±17.036 m	±15.260 m	±13.378 m

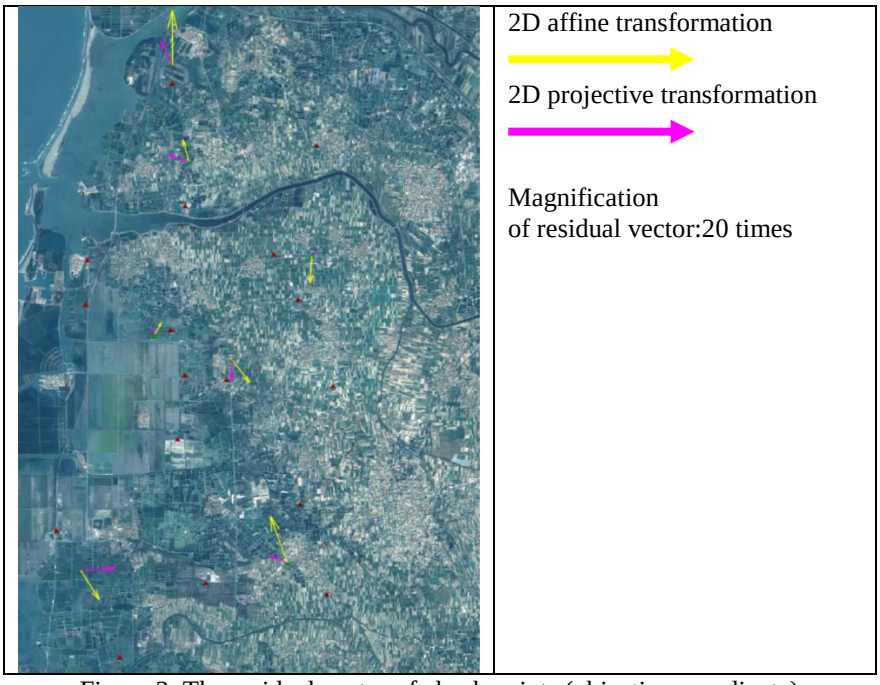


Figure 2. The residual vector of check points (objective coordinate)

4.3 Georeferencing of Historical Aerial Photographs

Both kinds of transformation can provide quite good georeferencing results after network adjustment as Figure 3 and 4 show. We zoom in to get a close-up view of georeferenced historical. Table 6 and 7 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. 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 3. Result of 2D affine transformation

Figure 4. Result of 2D projective transformation

Table 6. Comparison between historical aerial photographs (longitude direction)

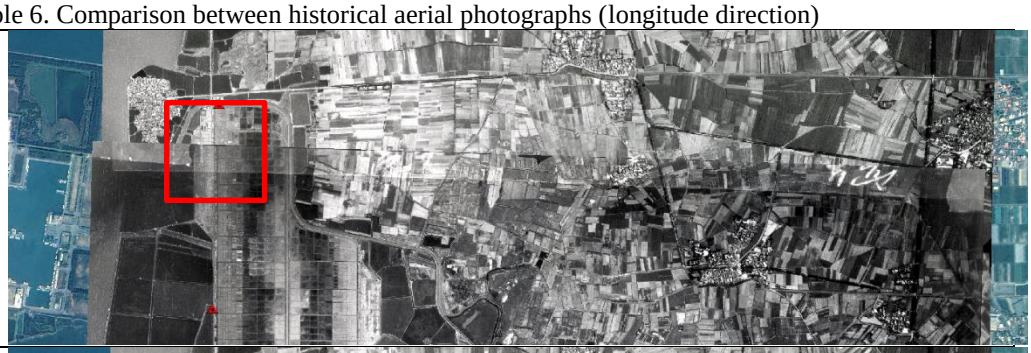
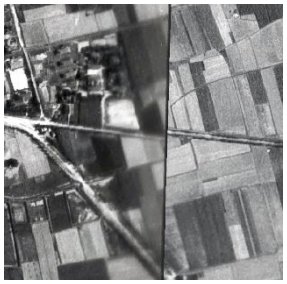
2D affine transformation	
2D projective transformation	

Table 7. Comparison between historical aerial photographs (latitude direction)

2D affine transformation		

2D projective transformation		

5. CONCLUSION

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 images. 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. REFERENCE

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

Gong, M., Zhao, S., Jiao, L., Tian, D., & Wang, S. (2014). A novel coarse-to-fine scheme for automatic image registration based on SIFT and mutual information. IEEE Transactions on Geoscience and Remote Sensing, 52(7), pp.4328~4338.

Juan, L., & Gwun, O. (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, J. S., Miller, C. C., & Bethel, J. (2010). Automated Georeferencing of Historic Aerial Photography. Journal of Terrestrial Observation, 2(1), pp.6.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. Proceednig of International Conference on Computer Vision, pp.1150~1157.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), pp.91~110.

Mikolajczyk, K., & Schmid, C. (2004). Scale & affine invariant interest point detectors. International Journal of Computer Vision, 60(1), pp.63~86.

Nebiker, S., Lack, N., & Deuber, M. (2014). Building change detection from historical aerial photographs using dense image matching and object-based image analysis. Remote Sensing, 6(9), pp.8310~8336.

Ping L., Yung-Chung C., Chih-Da W., Jihn-Fa J., Hsiung-Ming L. (2011). Application of Multi-source Remote Sensing Images in Detecting Shoreline Change along Yilan Coast. Journal of Geographical Research, 55, pp.47-68.