A MULTI-VIEW IMAGE MATCHING METHOD FOR FEATURE POINTS BASED ON THE MOVING Z-PLANE CONSTRAINT

Jingxue Wang *, Weidong Song, Fanqiu Bu

School of Geomatics, Liaoning Technical University, 123000, Fuxin, Liaoning, China
xiaoxue1861@163.com; song_wd@163.net

Commission Ⅲ/4

KEY WORDS: Multi-View Image Matching, Feature Points Matching, Moving Z-Plane Constraint, Grid Cell, Occlusion

ABSTRACT:

Focusing on the serious occlusion problem in city images, this paper makes full use of the advantage of multi-view image matching, and proposes a reliable multi-view image matching method based on the moving Z-Plane constraint. It supposes a fictitious plane in the object space, and the plane is divided to regular grid cell (small plane element) by a certain interval ($\geq$ image resolution). By moving the plane to different elevation positions, this algorithm makes feature point projection ray in overall images intersect with the plane, and constrains the candidate points by grid cells in the plane. Feature points which come from different images projection ray in the same grid cell on the plane may be regarded as the matching candidates. It selects the images which matching candidate points by gray similarity constraint to avoid the effect from occlusion image. According to the number of projection ray in the grid cell, this algorithm adopts hierarchy matching strategy of "the best candidate will be matched in the first instant", and uses initial matching results as constraint condition in the latter matching process. Finally, the validity of the algorithm proposed in this paper is verified by the experiments using four UltraCamX (UCX) digital aerial images and the algorithm is shown to have reliable matching results.

1. INTRODUCTION

Image matching is one of the key technologies to obtain 3D space information from 2D airborne/spaceborne image by digital photogrammetry. Along with the increasing use of new digital sensors, it becomes more and more easier to acquire large overlap digital images covering the same area, the multi-view image matching approach has attracted wide interests in both photogrammetry and computer vision. The traditional image matching is restricted to the imaging abilities of stereo sensors, which based on the matching of "single stereo-pair", and therefore is a challenging and "ill-posed" problem. Multi-image matching translates the single-stereo mode matching in image space to the multi-image matching by combining the image space and object space. Comparing with the single stereo-pair matching, multi-view image matching has advantages in below: Firstly, it improves the reliability of matching utilizing the redundant multi-view image information to effectively solve mistake matching with repetition texture, broken feature etc; Secondly, it maximally reduces the information blind area of image, and solves the occlusion problem in matching.

The algorithm of multi-view image matching can be divided into two categories by matching modes: multi-image can be matched in pairwise mode or in simultaneously multiple-matching mode. The former is based on the traditionally pairwise stereo matching. It separately matches the stereo pairs one by one, integrates all of the matching results in object space, and then obtains the correct matching results. For example, Pateraki (2005) proposes the algorithm of adaptive multi-image

matching (AIM), which is used for ADS40 image. It separately establishes stereo pairs with reference image and searches image at first. Then the algorithm makes quality check to matching result in single stereo pair, and makes the least square matching with multi-image according to corrected matching result by the former matching. Yuan and Ming (2009) introduce a multi-image matching algorithm by integrating image and space information. This algorithm realizes the integration of the matching results in each pairwise in the object space through the multi-ray intersection with the function of gross error detection by iterative predictive approach. Matching result in pairwise mode is uncertain because it does not combine multi-image redundancy information in the process of matching. So it needs to be consistently constraint in object space or filter method to be integrated into the matching results of multiple stereo pairs in the object space, which simultaneously increases the complexity of algorithm. Matching to all images simultaneously by adopting object geometry constraint mode, which obtains the corresponding point and the space coordinate of the corresponding at the same time. Zhang (2005) and Zhang, et al. (2008) propose the geometrically constrained cross-correlation (GC^3) algorithm, which chooses the nadir-viewing image of linear array image as the reference image, and extracts features from the reference image, and then searches the corresponding feature in the search image. It comes through the course of the from image space (reference image) to object space and to image space (search image). Because it is restraint to features extracted in the reference image, it does not adapt to the image obtained by the center projective in the areas with large gurgitation of the earth surface. If a certain space area is

---

* Corresponding author: Jingxue Wang, Ph.D, E-mail: xiaoxue1861@163.com

not imaged in the reference image because of occlusion but imaged in the other images, this area will not participate in matching. This problem may be solved to take turns to select each image as the reference image, but it increases the computation. Vertical line locus (VLL) is used by image correlator in DSR-11 mix digital photogrammetry workstation (Zhang & Zhang, 2002). VLL utilizes "ground element" in the object space as matching primitives, which have been used multi-image matching (Zhang, et al., 2007; Ji, 2008). It is uncertain because it integrates the similarity measure function in all stereo pairs in the course of matching, which reduces the effect from incorrect matching caused by occlusion and repetitive texture by right matching.

Focusing on the serious occlusion problem in city images, this paper makes full use of the advantage of multi-view image matching, and proposes a reliable multi-view image matching algorithm supported by the moving Z-Plane constraint (MZPC). It introduces geometric constraint used in the Space-Sweep method (Collins, 1995), and makes multi-image matching simultaneously for feature points. Based on Space-Sweep method, the MZPC proposes the multi-image selective matching strategy under the grey similarity constraint. This algorithm can simultaneously carry on the matching of multi-image feature points under "the best candidate will be matched in the first instant" matching strategy and plane grid height constraint.

2. MULTI-VIEW IMAGE MATCHING FOR FEATURE POINTS UNDER THE MOVING Z-PLANE CONSTRAINT

The multi-view image matching algorithm under the moving Z-Plane constraint is based on the basic photogrammetry principle of forward intersection that the corresponding points in the different images will always intersect to the same object point in the object space. In this paper, supposing that the position where intersected by different image projection rays of feature points may be the space position of the imaged feature, one new object constraint mode to multi-view image simultaneously matching based on feature points can be established according to this hypothesis.

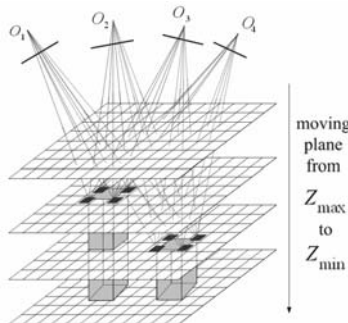### 2.1 Constraint by the Moving Z-Plane

Figure 1. Sketch map of moving Z-Plane constraint

Suppose a fictitious plane in the object space, the direction of this plane is vertical to the direction of vertical line in the object space, and the size of the plane contains the areas in the all images in the object space with $XY \in (X_{min} \sim X_{max}, Y_{min} \sim Y_{max})$. In Fig. 1, the plane is divided into regular grid cells (small plane elements) by a certain interval ($\geq$ image resolution), which can be seen as a

grid of ground element or "groundels" in the scene. The largest height value Z_{max} and the smallest height value Z_{min} of ground surface are determined, and moves the plane along the object vertical line with a certain step in the range of $Z_{min} \sim Z_{max}$. The size of step is directly in ratio to the size of grid cell.

This paper extracts the feature points in all images. First, it moves the plane to the largest height value position, and the plane equation is $Z = Z_{max}$. Using the inverse solution of collinearity equation, it makes the point projection rays in all image intersect with the plane, and obtains the object point $P_i(X_i, Y_i, Z)$ in the plane, i=1,2,...,N, N is the total number of projection feature points. Then it separately statistics the number of projection rays in the each grid cell, and if $number > T$ in a certain cell (T is the threshold), the position of this grid cell will be regarded as the feature point position in the object space. Finally, it records all the grid cells which with $number > T$ in the plane of this height, and considers them as the grid cells to be matched.

In the height plane, for each grid cell to be matched, its corresponding feature points in different images will be simultaneously recorded, which means that the projection rays pass through the same grid cell. This algorithm selects images having feature points, and performs the grey correlation matching. If the computed correlation coefficient is beyond a certain threshold, it regards the position of grid cell as the object feature point position, and the corresponding feature points in different image will be the corresponding points. Then the grid cell is evaluated with height value of the plane position, which will not participate in the latter matching. This process is called as grid cell in the plane matching.
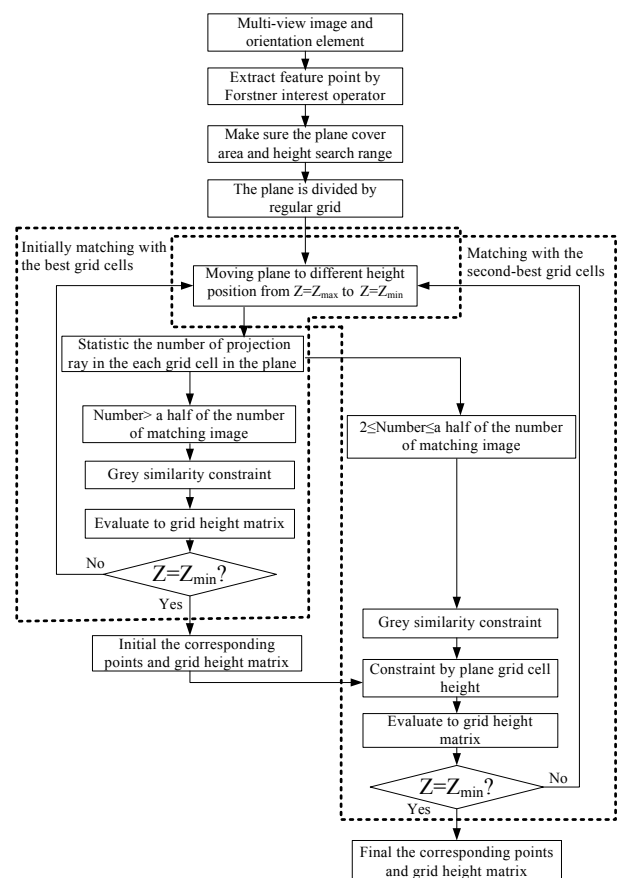

Figure 2. Flow chart of the overall algorithm

After that, this algorithm deletes the feature points successfully matched, and moves the plane to the next height position. Then it initializes the number of projection rays pass through all grid cells in the plane $number = 0$, and repeats the process above until reaches the lowest height value position.

By moving the plane to different height positions, this algorithm utilizes the positions of grid cells in the plane to constraint the range of projection rays for feature points in different images, and determines the matching candidates. This algorithm is called as the moving Z-Plane constraint. The flow chart of overall algorithm is shown in Fig. 2, involving the key technology as follows:

2.2 Hierarchy Matching Strategy

Considering the occlusion problem in multi-view image matching, and some feature points in some images are not extracted in the process of feature extraction, our algorithm matches all the grid cells whose number of projection rays are above 1. The more the projection rays passing through a grid, the more reliable matching result will be obtained. According to the matching principle of "the best candidate will be matched in the first instant", this paper adopts a hierarchy matching strategy. First, it matches the best grid cells in the plane at each height position, namely matching the grid cells with $number > T_q$, and T_q is half of the number of matching image.

Second, it matches the second-best grid cells in the plane at each height position, namely matching the grid cells with $2 \leq number \leq T_q$. In order to enhance the reliability of matching results, this process utilizes the matching results of the best grid cells to constraint the matching results of the second-best grid cells.

2.3 Grey similarity constraint

For the feature points in different images whose projection rays pass through the same grid cell in the plane, this paper selects images that have the same feature points, and adopts the grey similarity constraint to match these feature points. This process involves three key problems as follows:

⑴ Correlation window transform strategy of from object space to image space. First, the object window with the center of grid cell to be matched in the plane is determined, and is marked by W. In this process, it assumes that the ground is flat, and all points in this window are having the same height values, the same as the height value of plane location. Then, this paper projects the four corners of the object window onto all images to be matched according to the mathematical model of the sensor imaging, and obtains four corner points of corresponding region correlation windows in different images. Finally, according to the sizes of correlation windows, it calculates the number of pixels within the correlation windows, and inserts the pixel gray values in the corresponding positions.

⑵ Reference image selection. For each grid cell to be matched, the MZPC algorithm selects the image having a nearest distance from projection center in the plane to the grid cell to be matched as reference image.

⑶ Similarity measurement calculation. The MZPC algorithm takes correlation window in the reference image as a standard. Then it separately calculates the normalized cross-correlation (NCC) based on the grey similarity measure in correlation windows between the reference image and other images, and finally obtains average of the *NCC* (*ANCC*) according to calculate the average of all the *NCC*.

$$NCC_i(Z, Grid_{row,col}) = \frac{\sum_{s \in W}(I_0(S_0(s)) - \overline{I_0})(I_i(S_i(s)) - \overline{I_i})}{\sqrt{\sum_{s \in W}(I_0(S_0(s)) - \overline{I_0})^2}\sqrt{\sum_{s \in W}(I_i(S_i(s)) - \overline{I_i})^2}}$$

(1)

$$\overline{I_0} = \frac{1}{m \times n}\sum_{s \in W}I_0(S_0(s)), \quad \overline{I_i} = \frac{1}{m \times n}\sum_{s \in W}I_i(S_i(s))$$ (2)

$$ANCC(Z, Grid_{row,col}) = \frac{1}{k}\sum_{i=1}^{k}NCC_i(Z, Grid_{row,col})$$ (3)

where Z = plane height position

$Grid_{row,col}$ = grid cell with *row* and *col*

W = the object correlation window

$S_0(s)$, $S_i(s)$ = separately denote to transform from the object correlation window to image, which are obtained in corresponding pixel coordinates in the reference image and other image.

$I_0(\bullet)$, $I_i(\bullet)$ = separately mean the pixel grey in the reference image and other images

k = the number of stereopairs

2.4 Constraint by plane grid cell height

The MZPC algorithm initializes the height of all grid cells in the plane as zero, and records the grid cell height using a matrix having the same dimension with the grid cells in the plane, marked by grid height matrix $R_{m \times n}$, where m and n separately mean the number of column and the number of rows of grid cell in the plane. Each value in the matrix means the height of corresponding position grid cell in the plane.

After matching the best grid cells, the MZPC algorithm assigns the height values to every successfully matched grid cells to constrain the latter matching. In the process of matching with the second-best grid cell, for the grid cells meeting the grey similarity constraint, it assigns the plane height to this grid cell, and compares with the heights of other grid cells in a certain neighborhood range. According to the surface smooth principle in the local range area, if there are some grid cells having the similar height value within this neighborhood, it will be considered as the correct matching result, otherwise will be abandoned.

3. EXPERIMENTS ANALYSIS

In the experiments, this paper selectes three UCX digital aerial images in the same strip, whose pixel size is $7.2 \mu m$, the corresponding ground resolution is 0.049m, and the along-track overlapping is more than 80%. The precise orientation elements of each image are obtained by the triangulation using the VirtuoZo. In these images, there are some high buildings, which produce different form occlusions to those surrounding surfaces.

3.1 Determine the grid plane

The coverage area of the image is determined as the range of moving Z-Plane. The height range of the area in image as show is 3 m ~ 93 m, and the plane moving step is 1m.

3.2 Feature point matching under the moving Z-Plane constraint

Focusing on experimental images, this paper adopts the Forstner operator to extract feature points in all images. The number of feature points extracted in image L1, image L2, and image L3 are 34343, 32901, and 28053, respectively.

⑴ Initially matching of the best grid cells. It moves the plane from high to low, and matches the grid cells with *number* = 3 at each height position, the matching results of grid cells on Z=31 position is shown as Fig. 3. The threshold of *ANCC* is 0.85 in the grey similarity measurement calculation. The result of initially matching of the best grid cells is shown as Fig. 4, and the number of the homologous points is 3009, which is denoted by red cross, and the corresponding positions in the grid cell is shown as Fig. 5.
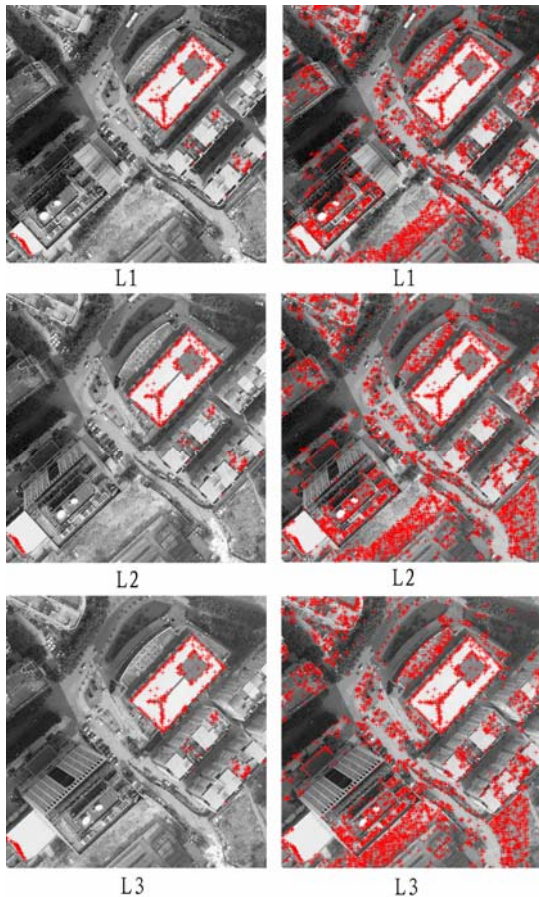
⑵ Matching of the second-best grid cells. After the matching of the best grid cells, this algorithm matches the grid cells with *number* = 2 , and the number of successfully matched grid cell is 5853, which corresponds to the homologous points in different images (Fig. 6). This paper adopts different colors to express the matching results of different images combination, for example, the yellow dots express the matching result of image L1~L2, the blue dots express the matching result of image L2~L3, and the red dots express the matching result of image L1~L3. The number of the homologous points obtained by matching with different image combination separately are 2871, 2566, and 416. Obviously, the longer baseline among multi-view images, and the bigger intersection angle, the more difficult of the matching. The ellipse area in image L1 and image L2 in Fig. 6 shows that this area does not match the homologous points because of the occlusion in the image L3 in the initial matching process. Through the matching of the second-best grid cells, this algorithm automatically selects the image L1 and image L2 to match according to the projection rays, and obtains the correct matching results, as shown in the homologous points by yellow color in the ellipse area, which effectively avoids the influence of the occlusion area in image L3.
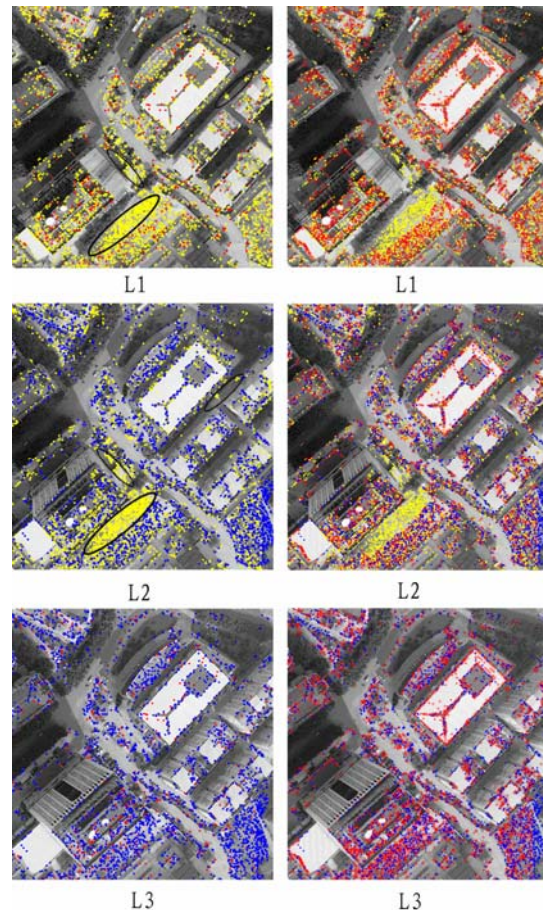
Figure 3. The matching results of grid cells as number=3 on Z=31 position

Figure 4. The matching results of the best grid cells as Number=3

Figure 5. Grid cells of initial matching results

Figure 6. (left) The matching results of the second-best grid cells as number=2

Figure 7. (right)The final results

After the matching of the best grid cells and the second-best grid cells, the total number of the successfully matched grid cells is 8862 (Fig. 8), which corresponds to the homologous points in different images (Fig. 7). The meaning of different colors and different signs is the same as the front. It can be see

that the second-best grid cells matching enhances the dense of initial matching results, more abundantly expresses the object details, and gets the homologous points in two overlapping area and occlusion area in the multi-view image.

Figure 8. Grid cells of final matching results

3.3 Result evaluation

This paper makes forward intersection with the homologous points, and obtains the discrete 3D points (Fig. 9). From this figure it can be seen that there are a few mistakes. As shown in Fig. 9, the ellipse regions includes the three error matching points obtained in initially matching of the best grid cells, and the seven error matching points obtained in matching with the second-best grid cells. So it needs to adopt a reliable filter to reject the mismatching results.
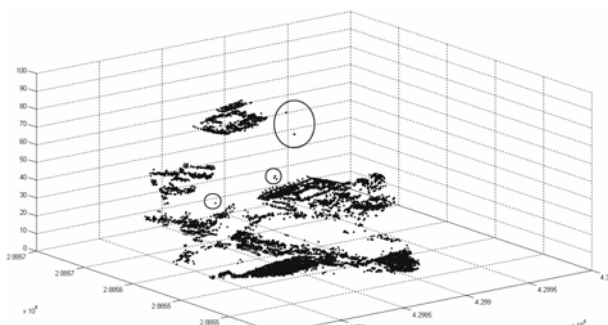

Figure 9. Discrete 3D points

3.4 Contrast analysis with different constraint conditions

The existing multi-view image matching object constraint models, such as GC³ algorithm (Zhang, 2005), modified vertical line locus algorithm (MVLL) (Ji, 2008), constrain the image space search range by points to be matched moving in the range of an approximate height value along a certain linear direction (projection ray locus, vertical line locus) in the object space, and reduces the search range from two-dimensional to one-dimensional. The traditional standard normalized cross-correlation (*NCC*) is expressed as a function of height value Z, in the approximate height value ranges $Z \in [Z_0 - \Delta Z, Z_0 + \Delta Z]$. The traditional algorithm calculates the sum of *NCC* (*SNCC*) values iteratively, and selects the height value Z corresponding to the maximum *SNCC* value as the object correct height value. The formulas are,

$$NCC_i(p_0, Z) = \frac{\sum_{s \in W}(I_0(s) - \overline{I_0})(I_i(S_i(s)) - \overline{I_i})}{\sqrt{\sum_{s \in W}(I_0(s) - \overline{I_0})^2}\sqrt{\sum_{s \in W}(I_i(S_i(s)) - \overline{I_i})^2}} \quad (4)$$

$$SNCC(p_0, Z) = \frac{1}{n}\sum_{i=1}^{n} NCC_i(p_0, Z) \quad (5)$$

Taking the point P in the image L1 as an example, and according to the geometric constraint principle in GC³ algorithm, this paper restricts the height search range to 3~10m, and the search step is 0.1m. The calculated cross-correlation curve is shown as Fig. 10. The final calculated result shows that the $SNCC < 0.6$ and obtains the maximum at the point p' which should be the point p. The reasons include (1) points to be matched in the image L1 are occluded in image L3, and the $NCC_{L1-L3} < 0.4$, which is too low and (2) the NCC_{L1-L2} at point p and point p' appear peak values because of the repetitive texture in image L2, which means the local grey distribution at point p and point p' is similar.

Fig. 11 is the matching result using the MZPC algorithm. In the matching of the second-best grid cells, it matches the grid cells with *number* = 2 on the plane $Z = 7$, and selects image L1 and image L2 according to the projection rays passing through the grid cells, and avoids the occlusion in image L3. The *ANCC* value is 0.94825602, and the correct matching result is obtained. Comparing with the existing multi-view image matching algorithm based on the object constraints, the MZPC algorithm limits search distance in the image space under traditional epipolar constraints or the search distance in object space under geometric constraints to a grid cell position in the plane, which avoids the appearance of multiple peaks in the cross-correlation curve caused by similar texture, and reduces the probability of mismatches. Simultaneously, the MZPC algorithm carries on the selective matching for multiple images according to projection rays in the grid cells to avoid the effects of occlusions, and thereby improves the reliability of the matching.
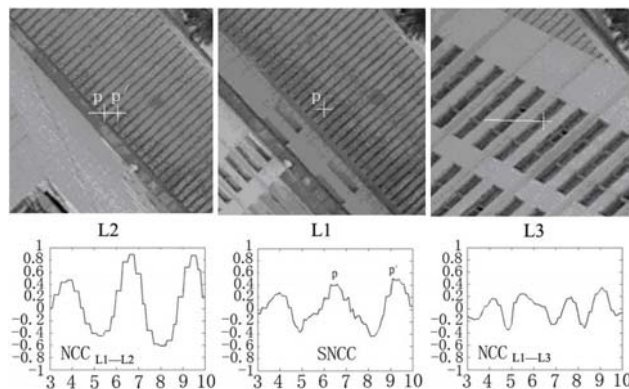

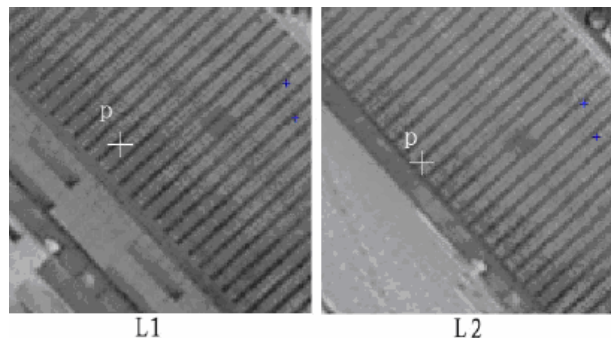Figure 10. Experiment results by projection rays constraint matching

Figure 11. Experiment results under moving Z-Plane constraint

4. CONCLUSION

Focusing on the buildings occlusion problem in matching with city image, as well as the multiple solutions in matching due to repetitive texture, this paper summarizes the existing multi-view image matching algorithms, and proposes a multi-view image matching algorithm for feature point supported by the moving Z-Plane constraint. This paper selects three UCX digital aerial images of a typical building area in the same strip for matching experiments, and verifies the validity of the proposed algorithm. The conclusions are as follows.

⑴ The matching algorithm proposed in this paper can simultaneously match with any number of multi-view image, and obtains the matching results in any overlapping areas in multi-view image ;

⑵ Based on the selective matching, the MZPC algorithm effectively avoids the effects by occlusion image, and improves the reliability of the matched results;

⑶ The MZPC algorithm does not need the iterative calculation of height, and avoids the appearance of multiple peaks in the cross-correlation curve caused by similar texture, and reduces the probability of mismatches.

⑷ The MZPC algorithm is entirely based on the matching with feature points. The uniformity and the density of feature points distribution directly determine the density of matching results. It needs to further research the dense matching with other matching primitives.

⑸ The experimental results also have a few mismatched homologous points. It needs to research a high reliable method to reject the mistakes.

REFERENCES

Collins, R.T. , 1995. A space-sweep approach to true multi-image matching. *Conference on Computer Vision and Pattern Recognition*, San Francisco, pp. 358-363.

Ji S., 2008. Multi-view matching theories and methods for automatic DSM extraction from linear array images. PLA Information Engineering University for the Degree of Master of Engineering, Zhengzhou, pp. 24-25.

Pateraki M., 2005. Adaptive Multi-image matching for DSM generation from airborne linear array CCD Data. Institute of Technology Zurich, ETH Zurich, Zurich, pp. 58-61.

Yuan X. X., Ming Y., 2009. A novel method of multi-image matching using image and space synthesis information. *Acta Geodaetica et Cartographica Sinica*, 38(3), pp. 216-222.

Zhang L., 2005. Automatic digital surface model (DSM) generation from linear array images. Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich, pp. 81-99.

Zhang L, Zhang J. X., Wang S. C., 2008. Multi-image matching for DTM generation from SPOT-5 HRS/HRG and IRS-P5 imagery for the project of West China topographic mapping at 1:50,000 scale. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, chian, XXXVII, Part B1, pp. 1109-1116.

Zhang Y. S., Fan D. Z., Ji S., 2007. Multi-view matching algorithm model for ADS40 sensor. *Journal of Zhengzhou Institute of Surveying and Mapping*, 24(2), pp. 83-86.

Zhang Z. X., Zhang J. Q., 2002. *Principles of Digital Photogrammetry*. Wuhan University Press, Wuhan, pp. 111-114.