

A HIERARCHICAL IMAGE MATCHING METHOD FOR STEREO SATELLITE IMAGERY

Fei Yan ^{a,*}, Weian Wang ^{a,b}, Shijie Liu ^{a,b}, Wen Chen ^{a,b}

^a College of Surveying and Geo-Informatics, Tongji University, Siping Road, Shanghai, China –
yanfei19891023@126.com

^b Center for Spatial Information Science and Sustainable Development Applications, Tongji University, Siping Road,
Shanghai, China – weian@tongji.edu.cn, liusjtj@gmail.com, chenwen_ecnu@126.com

Commission VII WGVI/6

KEY WORDS: Image Matching, Hierarchical Strategy, Scale-invariant Feature Transform, Stereo Satellite Imagery

ABSTRACT:

Image matching is an essential and difficult task in digital photogrammetry and computer vision. This paper presents a triangulation-based hierarchical image matching algorithm for stereo satellite imagery. It uses a coarse-to-fine hierarchical strategy and combines feature points and grid points to provide a dense, precise and reliable matching result. First, some seed points are extracted at the top level of image pyramid using the SIFT algorithm with RANSAC approach to remove mismatches and enhance robustness. These points are used to construct an initial triangulation. Then, feature point and grid point matching are conducted based on the triangle constraint. In the process of hierarchical image matching, the parallaxes from upper levels are transferred to levels beneath with triangle constraint and epipolar geometrical constraint. At last, outliers are detected and removed based on local smooth constraint of parallax. Also, bidirectional image matching method is adopted to verify the matching results and increase the number of matched points. Experiments with ALOS images show that the proposed method has the capacity to generate reliable and dense matching results for surface reconstruction from stereo satellite imagery.

1. INTRODUCTION

Image matching is an essential and difficult task in digital photogrammetry and computer vision. It is the foundation of computer vision applications, such as camera calibration, three-dimensional reconstruction, intelligent monitoring and motion analysis. Image matching is used for finding corresponding pixels in a pair of images, which allows 3D reconstruction by triangulation. It is to study how to choose some features and similar standards based on the reference image and searching image to search for strategies for correlation computing and to determine the best space responding point for matching. Its main issues focus on the feature space, similarity measurement and searching strategies. The fatal step is to ascertain effective matching methods; a good matching method requires high reliability, small error, fast speed and good real-time.

Image matching is relatively easy when encountered with good image texture conditions. However, on relatively poor textural images, image matching is a difficult and challenging problem. Most of the traditional digital photogrammetry systems require lots of human interactions to remove the errors in the matching results when dealing with poor textural images.

Despite the algorithms and the matching strategies used may be different from each other, the accuracy performance and the problems encountered are very similar in the major systems and the performance of commercial image matchers does by far not live up to the standards set by manual measurements (Gruen et al., 2000). The main problems in image matching are encountered with (Zhang et al., 2004.)

- (1) Little or no texture
- (2) Repetitive texture
- (3) Local object patch is non planar face

- (4) Distinct object discontinuities
- (5) Occlusions
- (6) Moving objects, incl. shadows
- (7) Multi-layered and transparent objects
- (8) Radiometric artefacts like specular reflections and others

In this paper, a triangulation-based hierarchical image matching algorithm for stereo satellite imagery is described. It uses a coarse-to-fine hierarchical strategy and combines feature points and grid points to provide a dense, precise and reliable matching result. First, some seed points are extracted at the top level of image pyramid using the SIFT algorithm with RANSAC approach to remove mismatches and enhance robustness. These points are used to construct an initial triangulation. Then, feature point and grid point matching are conducted based on the triangle constraint. In the process of hierarchical image matching, the parallaxes from upper levels are transferred to levels beneath with triangle constraint and epipolar geometrical constraint. At last, outliers are detected and removed based on local smooth constraint of parallax. Also, bidirectional image matching method is adopted to verify the matching results and increase the number of matched points. Experiments with ALOS images show that the proposed method has the capacity to generate reliable and dense matching results for surface reconstruction from stereo satellite imagery.

In the process of projection from three-dimensional world to two-dimensional images, a lot of information is lost and it is an ill-posed problem. So we must make the best of constraints included in the problem to be solved to limit the size of solution space. After the computation of matching, in reality, there are complex situations such as occlusion, shadows, poor texture, atmospheric dust and steep terrain. So there may be some wrong

* Corresponding author.

correspondences and constraints are necessary to eliminate blunders. For now, the common constraints for image matching in digital photogrammetry are mainly as follows:

1) Epipolar geometrical constraint. For any point in the left image, its matching point in the right image must lie on the corresponding epipolar line. In this paper, we use a projection track method to calculate the approximation of epipolar line and the problem of 2D conjugate search can be reduced into 1D one. As a result, the processing time is minimized drastically and the accuracy increased.

2) Similarity constraint. Corresponding points are assumed to have similar intensity or colour. So intensity is the main information used in stereo matching.

3) Continuity constraint. Surfaces of objects are assumed to be smooth, which means its parallax varies continuously.

4) Uniqueness constraint. Uniqueness constraint is that searching for matching point on the right image, taking left image as reference, matching point with right image and reference point with left image should be consistent.

2. REVIEW ABOUT MATCHING METHODS

In the last few decades, a lot of efforts have been devoted in the field of photogrammetry and computer vision to improve the reliability, automation, and efficiency of image matching which can be generally divided into two classes based on the matching primitives. One is area-based matching and the other is feature-based matching.

Area-based Matching: Area-based matching usually works directly on local image windows, and it can acquire dense correspondences. It uses the grey value of the whole image to measure the similarity of two images directly. And a certain method is employed to search the point where the similarity measurement is the biggest. There are many area-based matching methods such as Maximization of Mutual Information, correlation method, conditional entropy method, joint entropy method and so on. Although area-based matching is the most widely used, there exists some shortcomings such as huge computation, long time of matching and sensitivity to rotating, scaling and distort.

Feature-based matching: The common image feature includes point feature, straight line, edge, shape, closed area, statistical moment, etc. By far, feature extraction algorithm can be divided into three main classes: one is point feature extraction operator such as F \ddot{u} rstner operator, Harris operator and SUSAN operator, the second is linear feature extraction operator (such as Canny operator, LoG operator), and the third is surface extraction operator mainly through region segmentation. Generally speaking, feature-based matching has the advantage of being simple to operate, rapid matching speed and high precise matching rate, but it also requires human intervention and the obtaining of feature points is a bit difficult. Besides, it is only suitable for simple images with significant geometric features.

According to the above analyses, feature-based image matching obtains robust but sparse matching results, while area-based matching can obtain dense matching results but the matching reliability may depend on the texture conditions of the images. Therefore, this paper presents a hierarchical image matching method that combines the advantage of both the feature-based matching and the area-based matching methods and produces reliable and dense matching results with high efficiency and automation.

After the matching primitives are selected, the next task is to measure the similarity of the corresponding points in the image pairs, simulating human eyes by means of similarity

measurement. The similarity measurement will be the matching score to judge the corresponding points or will be used for a global strategy to judge the corresponding points. There have been many similarity measurement for image matching presented in the past decades, such as the normalized cross-correlation (NCC) (Helava, 1978), sum of squares difference, normalised mutual information (Knops et al., 2006). These are simple computationally, but not distinct enough, and are sensitive to geometrical distortion and discontinuity problems. When using normalized cross-correlation to measure the similarity between the features on stereo images, an important problem is the selection of an appropriate window size to calculate the correlation values. So in this paper, NCC is selected as the matching score to measure the similarity between corresponding points of a stereo pair and the selection of window size will be discussed in experimental section.

3. MATCHING APPROACH

3.1 Overview of the approach

The inputs for this approach are the images and rational polynomial coefficient (RPC) parameters. The workflow as shown in Fig. 1 includes the following steps:

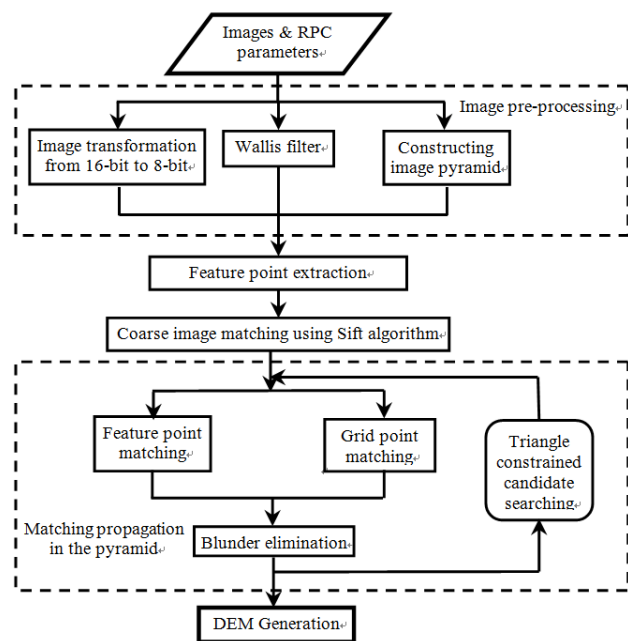


Figure 1. Workflow of our image matching procedure

(1) Image pre-processing including image transformation from 16-bit to 8-bit, the Wallis filter and constructing image pyramid.

(2) Feature point extraction using the F \ddot{u} rstner. This step is performed to provide the interest points and edges for later image matching.

(3) Coarse image matching using Sift algorithm. This step is employed to generate the initial triangulation for image matching.

(4) Matching propagation in the image pyramid. This step includes feature point and grid point matching.

(5) Blunders elimination in each level including local smooth constraint of parallax and bidirectional image matching.

(6) Least squares image matching in original image level. By means of correcting radiometric and geometric distortion for

images, not only is the matching precision improved but also enhance robustness by the rejection of blunder.

This paper will carry on the analysis from the following several important aspects about the approach.

3.2 Image pre-processing

The quality of matching results relies heavily on the quality of the images. Poor contrast would reduce the number of interest points extracted. The Wallis filter is employed to strongly enhance and sharpen the already existing texture patterns and increase the signal-to-noise ratio, which is beneficial for feature extraction.

At present, the grey values of panchromatic remote sensing images are stored by 16-bit. However, in our research, 8-bit stored images are necessary for the computation of image matching and the display of images in order to decrease calculation amount. So the 16-bit images are mapped to 8-bit images using a principle of auto levels.

After that, we must construct the image pyramid before matching. There are two kinds of methods for generating image pyramid: the direct method and the filtering method. The direct method is to transform 2 pixels by 2 pixels to 1 pixel with an average value. Filtering method, which is adopted in this paper, varies from the first one in that low-pass filtering is used to replace the average. Specifically, starting with the original image, each subsequent level of the image pyramid is created by sub-sampling the previous level image and smoothed by a Gaussian filter. Lower level images have coarser resolution and details are lost due to smoothing. The pyramid level number is a pre-defined value which could be either a user-input or can be determined according to the height range of the imaging area.

3.3 Matching through the image pyramids

The hierarchical searching algorithm is by narrowing the search scope to achieve the purpose of reducing the computational complexity to improve matching speed. It is proposed just as people find things with the coarse-to-fine strategy.

The hierarchical image matching method firstly uses a SIFT algorithm and RANSAC approach to obtain a few reliable correspondences, and then construct an initial triangulation. The SIFT algorithm (Lowe, 1999) is proved to be able to produce robust but relative sparse corresponding points invariant to moderate scale changes or distortions, which is ideal for the purpose of generating a certain number of well distributed matching points for the initial triangulation. In the SIFT descriptor, each interest point is characterized by a vector with 128 unsigned eight-bit numbers generated from a local region, which defines the multi-scale gradient orientation histogram. The matching is performed by measuring the similarity between the two vectors associated with the two matching points.

The RANSAC approach is used to detect and eliminate possible blunders from the previous SIFT matching results. It starts by randomly selecting a subset of the matched corresponding points. From the chosen matched points, a fundamental matrix can be calculated based on which a model is then built. This model is evaluated by determining whether each pair of corresponding points fit reasonably well to it. This is used as a criterion to determine the best model which has the largest number of correct corresponding points. This process is repeated to find the best model. Those matched points which do not fit for the final best model are considered as blunders and eliminated from the initial matching point set (Wu, et al., 2012).

After the seed points are extracted at the top level, an initial triangulation can be constructed and an area-based image matching with feature points and grid points is conducted at the top level again.

In the process of hierarchical strategy, image matching is first conducted on the lowest resolution. The matched points are then transferred to the next level (of higher resolution) where additional feature points could be matched. This process repeats until it reaches up to the original image level. At a subsequent level, points from upper level are matched again to achieve higher precision. A TIN (Triangulated Irregular Network) surface of parallaxes is generated from these matched points using the Delaunay triangulation. This TIN is used to estimate the correspondence of additional feature points (Hwangbo J. W., 2010).

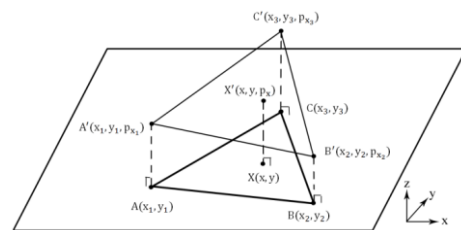


Figure 2. Interpolation of x parallax using TIN surface

As shown in Fig. 4, a point $X(x, y)$ is inside of a Delaunay triangle formed by points $A(x_1, y_1)$, $B(x_2, y_2)$, $C(x_3, y_3)$, the x parallax is interpolated based on the TIN surface formed by $A'(x_1, y_1, p_{x1})$, $B'(x_2, y_2, p_{x2})$, $C'(x_3, y_3, p_{x3})$. The x parallax of point $X(p_x)$ is interpolated as the z coordinate of point X' , which is the intersection between the 3-D plane $A'B'C'$ and the line XX' parallel to z axis. Parallax in y direction (p_y) can be calculated using the same strategy, only using y parallax as z coordinates. Finally, the estimated corresponding point coordinate (x', y') is defined as:

$$\begin{aligned} x' &= x + p_x \\ y' &= y + p_y \end{aligned} \quad (1)$$

3.4 Searching along epipolar line

For linear push-broom imagery, every scan line has its own projection center and orientation elements, so there is no strict definition of epipolar line. The main methods for generating approximate epipolar line include polynomial fitting (Zhang et al., 1989) and projection track method (Jiang et al., 2008). The former method needs a great number of corresponding image points which is not available before image matching, while the latter only needs the exterior orientation elements. Therefore, under the condition of having known RPCs which are in high accuracy, this paper uses projection track based approximate epipolar line as a constraint for image matching.

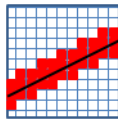


Figure 3. Epipolar line in the search window

After obtaining an approximate corresponding image point in the searching image with triangle constraint, we create a search window centered on it. Then an epipolar line is generated in this window and image searching is conducted along the line widened by 2 pixels considering the calculation error of epipolar line.

3.5 Feature point and grid point matching

Förstner operator is one of the most famous operators to extract feature point and it has high speed and positioning accuracy. Förstner operator generates a series of feature points by identifying distinctive variations of pixel brightness values. The detected feature points including corners and round dots can reflect the shape of the terrain in the images. Feature points are typically found around shadow, where significant change of brightness values occurs. These points are related to the shape of terrain or objects such as ridges and rocks.

Feature point matching is very efficient and suitable in texture-rich regions with grey value variation. On the other hand, in image regions with poor texture or no texture information, few or even no feature points can be extracted. Thus, feature point matching will lead to holes on the DEM in these areas. To solve this problem, grid points can be used and grid point matching has been introduced.

Grid points are points determined at given positions that are generally uniformly distributed over the whole image. Compared to the feature points, the choice of grid points is blind and thus many grid points may lie in poorly textured regions or occluded areas. The search for the match of a given grid point has a higher possibility of yielding an ambiguous match and even no matching candidates. The grid points are also matched using cross-correlation following the coarse-to-fine strategy.

3.6 Blunder elimination

Even though the constraint of epipolar line and triangle is adopted in the process of image matching, still it cannot guarantee all the matching results are right. So it is important to employ an algorithm for detecting and removing outliers.

At first, bidirectional image matching algorithm is imbedded in our program. The basic idea of this strategy is simple as follows: First, a point in left image is taken as target point and find out the point of maximum similarity in right image. Then take the matched point (the point in right image) as target point and search for the best point in left image in the same way. Finally, we accept it if the two results overlap, or give it up as mismatch.

Although bidirectional image matching method eliminates some wrong matches, there still exist some mismatched points. Therefore, local smooth constraint of parallax is adopted in this research. The logic behind the strategy is that the terrain is continuous and spatially correlated in a small area. Speaking specifically, for every pair of corresponding image point, the adjacent corresponding points within a certain distance are used to fit an optimal plane of parallax. Outliers are the values that are statistically far from most others in a set of data. Since the

spatial distribution pattern of the parallax of corresponding points could vary depending on the type of local terrain, the parallax offset from the fitted plane is compared with the standard deviation of the neighboring points. If the offset is beyond 3σ , the matched point is considered as a blunder and discarded from the result.

4. EXPERIMENT RESULT

We have applied the proposed matching approach to ALOS stereo images for DEM extraction in a test site around Hongkou, Sichuan, China. The resolution of the images is 2.5 m. The scene covers a total area of 36km by 40 km and consists of a variety of land cover types, including plain, urban housing and mountains.

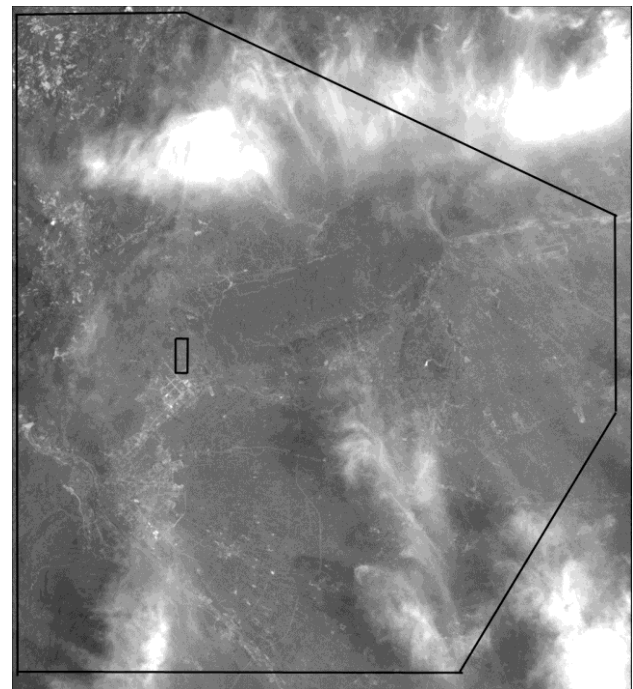


Figure 4. ALOS image of Hongkou, Sichuan, China

Controlling the matching parameter setting is important for high quality DEM generation. Since the stereo matching result is highly dependent on the properties of input data, there is no single set of parameters that is perfect for every image (Hwangbo J. W., 2010). So Table 1 shows some comparatively optimal parameters we selected through many experiments.

Level	1	2	3	4
Image scale	1/8	1/4	1/2	1
Matching window size	11×11	13×13	15×15	15×15
Search window size	11×11	9×9	7×7	7×7
Correlation coefficient threshold	0.8	0.8	0.8	0.7

Table 1. Selected parameters for hierarchical image matching

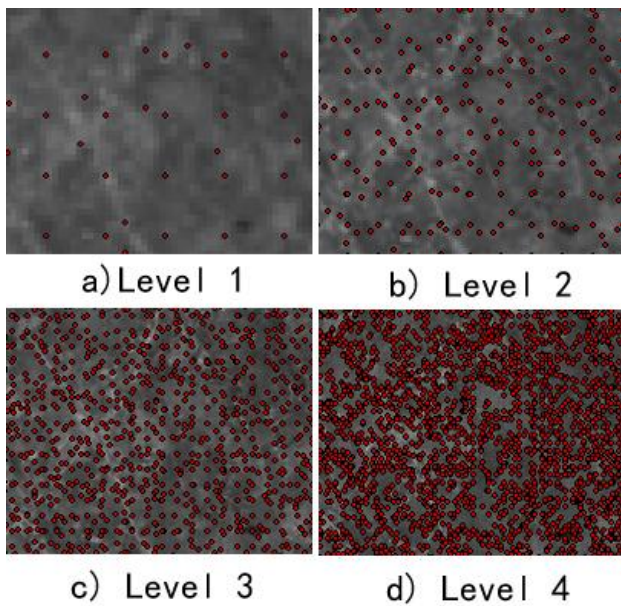


Figure 5. Matched points in hierarchical image pyramid

Fig. 5 shows the matched points including feature points and grid points at each hierarchical matching level. They provide robust structure of terrain from coarse level with a few points to finer level with an increased amount of detail. Fig. 6 illustrates the generated DEM with a ground sampling distance of 5 meters using the proposed image matching method and it is located in the big polygon indicated in Fig.4. Because many areas have cloud cover or significant changes by landslide between the stereo images, there are many blunders in these areas. This DEM is generated from the matched points using triangular linear interpolation approach. Visual inspection reveals that good results have been achieved.

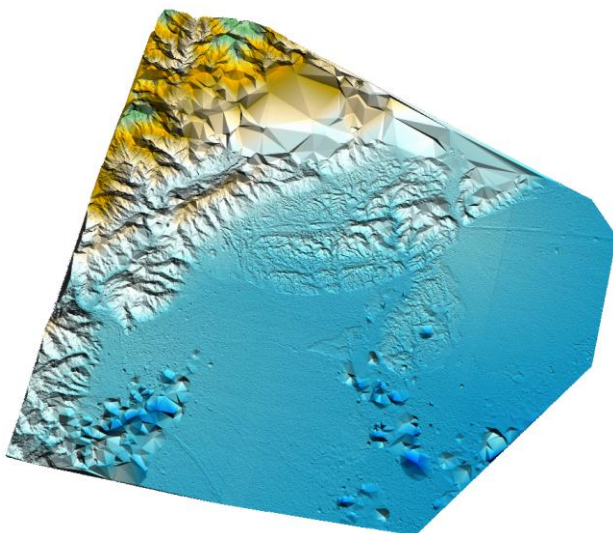


Figure 6. A view of the generated DEM from ALOS stereo imagery over Hongkou, Sichuan, China

Finally, an area of about 600 m by 2400 m (rectangle in Fig.4) was selected and checked to ensure that these automatically matched points describe the topographic details in a DEM. In this area, points were manually matched in the stereo images for generating a manual DEM. Then manual DEM was

compared with the DEM generated from automatically matched points in the form of profile. Three profiles in this area were inspected (Figure 7). Overall, the elevation profiles from the two methods showed high consistence. The differences in elevation between manually and automatically generated DEM were in the range of $\pm 2.5\text{m}$ maximum with standard deviation less than 0.6m. The random elevation errors of such magnitude are reasonable, considering that the pixel resolution of the raw images is about 2.5m.

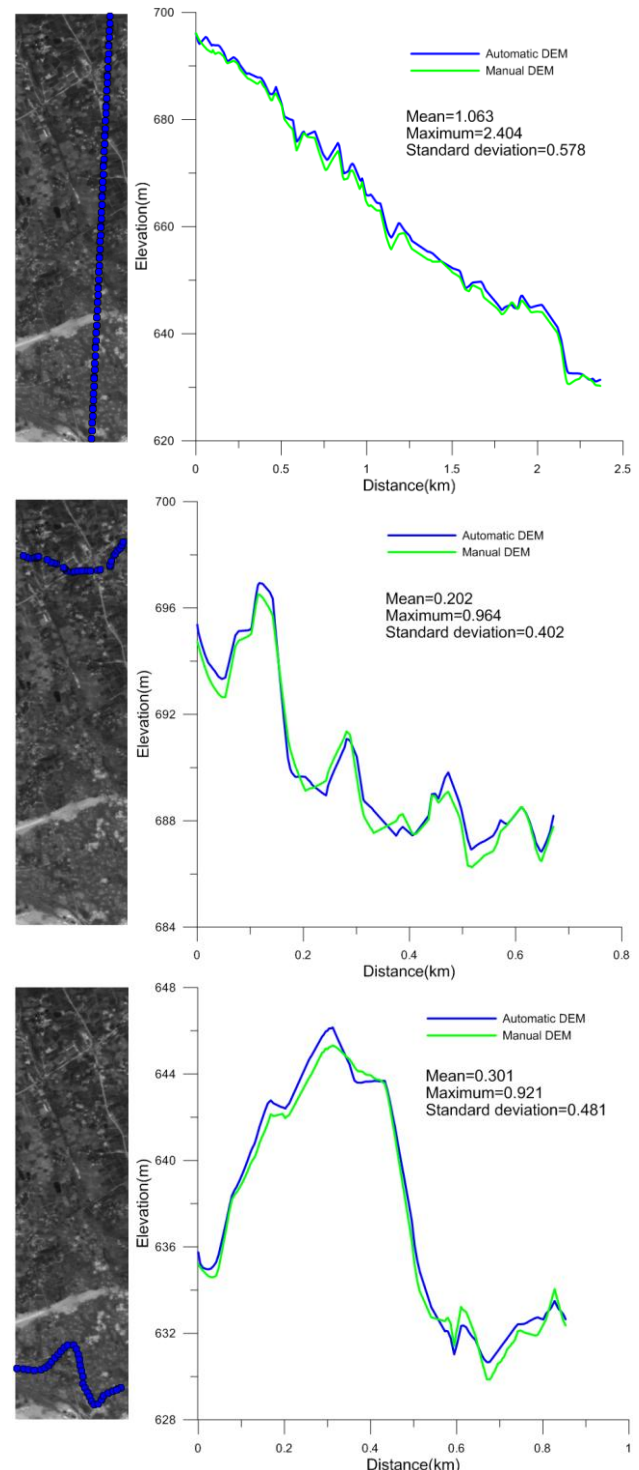


Figure 7. Elevation profile of automatically and manually generated DEM of three districts

5. DISCUSSION AND CONCLUSIONS

This paper presented a triangulation based hierarchical image matching method for stereo satellite imagery. Table 2 gives the DEM accuracy evaluation results in the whole rectangular region mentioned above. About 94% of the region is within three standard deviations though the maximum absolute difference is 4.841m.

Statistics	Value
RMSE(m)	1.354
Mean(m)	1.140
Abs max(m)	4.841
Standard deviation(m)	0.730
Percentage over three standard deviations	6%

Table 2. Statistics of the difference between automatically and manually generated DEM from the region in Figure 7

Though the experiment using ALOS images, we draw the following conclusions:

(1) The SIFT algorithm with RANSAC approach can produce reliable but sparse corresponding points which is ideal for obtaining the initial values for hierarchical image matching.

(2) The hierarchical matching strategy based on triangulation constraint proved to be effective and reliable especially in some area with poor and repetitive texture. More importantly, for stereo satellite imagery, the hierarchical search strategy can improve the speed of image matching; otherwise it will be very time consuming to search corresponding points in the original image.

(3) The epipolar geometrical constraint can help improve the search efficiency and alleviate the matching ambiguity, particularly in the region with repetitive features.

(4) The experiment results showed that local smooth constraint of parallax can eliminate most of outliers effectively, although there still exist some mismatches. Future work will be focused on this problem to better refine the matching results.

ACKNOWLEDGEMENTS

This work has been supported by National Key Basic Research Program of China (No. 2012CB957704) and Shanghai Municipal Natural Science Foundation (No.11ZR1439000). Comments to the manuscript from the referees are greatly acknowledged.

REFERENCES

Hwangbo J. W., 2010. Integration of Orbital and Ground Imagery for Automation of Rover Localization. *PhD Dissertation*, The Ohio State University, pp. 115-152.

Jiang, W., Zhang, J., Zhang, Z., 2008. Simulation of three-line CCD satellite images from given orthoimage and DEM. *Geomatics and Information Science of Wuhan University*, 33(9), pp. 943-946.

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

Poon, J., Fraser, C., Zhang, C., Zhang, L., Gruen, A., 2005. Quality Assessment of a Digital Surface Model Over Urban

Areas Generated from IKONOS Stereo Pair. *The Photogrammetric Record*, 20(110): 162-171.

Tang, L., Wu, C., Chen, Z., 2002. Image dense matching based on region growth with adaptive window. *Pattern Recognition Letters*, 23, pp. 1169-1178.

Wu, B., Zhang, Y., Zhu, Q., 2012. Integrated point and edge matching on poor textural images constrained by self-adaptive triangulations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 68, pp. 40-55.

Wu, B., Zhang, Y., Zhu, Q., 2011. A Triangulation-based Hierarchical Image Matching Method for Wide-Baseline Images. *Photogrammetric Engineering & Remote Sensing*, 77(7), pp. 695-708.

Zhang, L., Armin G., 2006. Multi-image matching for DSM generation from IKONOS imagery. *ISPRS Journal of Photogrammetry & Remote Sensing*, 60, pp. 195 - 211.

Zhang, L., Gruen, A., 2004. Automatic DSM generation from linear array imagery data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXV. Part B3, pp. 128-133.

Zhang, C., Fraser, C., 2008. Generation of digital surface model from high resolution satellite imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Beijing, China, Vol. XXXVII. Part B1, pp. 785-790.

Zhang, C., Baltsavias, E., 2000. Knowledge-based image analysis for 3D edge extraction and road reconstruction. *International Archives of Photogrammetry and Remote Sensing*, Amsterdam, Netherlands, Vol. XXXIII, Part B3/2, pp. 008-101.

Zhang, C., Fraser, C., 2009. An improved approach for DSM generation from high resolution satellite imagery. *Journal of Spatial Science*, 54(2), pp. 1-13.

Zhang, Z., Zhou, Y., 1989. Epipolar line of SPOT images using fitting method. *Journal of Wuhan University of Surveying and Mapping Technology*, 14(2), pp. 20-24.