Building Height Extraction
Based on Satellite GF-7 High-Resolution Stereo Image

Zijin Tian [1], Yan Gong [2]

[1] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China – zijintian@whu.edu.cn
[2] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China – gongyan@whu.edu.cn

Keywords: GaoFen-7 Satellite (GF-7), Roof Matching, Rational Function Model, Building Height Estimation.

Abstract

High-resolution remote sensing images can distinguish objects of smaller size, so as to more clearly express the texture features and structural information of objects, and provide a data source for the development of large-scale mapping, high-precision stereo measurement and other fields. The purpose of this paper is to complete the height estimation of the buildings by analyzing the stereoscopic observation formed by the front and rear view images of the Gaofen-7 line array CCD. After determining the roof profile of the building on the rear-view image, assuming a series of object elevations of the building, that is, searching for elevations with a certain step distance within a certain elevation search range, adopt the object-based image matching VLL algorithm, Through the RFM imaging model of the Gaofen-7 sensor, the rear-view contour is projected to the front-view image, and then the PSNR is selected as the similarity measure of the window, and the similarity of the image block area corresponding to the front- and rear-view contour is calculated. Corresponds to the hypothetical object elevation as the estimated height of the building. Under the technical route of this paper, the height of buildings on high-resolution images can be estimated to a level within 3 meters of accuracy.

1. Introduction

Urban areas serve as primary sites for high-density population aggregation and various production activities. Urban buildings constitute a major component of the urban substrate, and the accurate and rapid extraction of their height information is a key research focus (An et al., 2010; Song et al., 2014). As the urban construction areas' rapid expansion and the replacement of buildings, among the various elements of geographic databases, the expansion, demolition, and other changes of buildings occur the most, and are also the major areas that require the fastest updates. With the rapid development of sensor technology and continuous progress in computer information technology, the application scope of high-resolution satellite imagery is expanding. Compared to on-site measurements, leveraging remote sensing images for three-dimensional reconstruction of buildings offers advantages in terms of wide coverage, rapid efficiency, and high precision.

The technology for extracting building heights using remote sensing imagery has evolved into two main categories. One involves extracting from stereo image pairs, and the other utilizes shadow information in a single remote sensing image to infer building heights. Extracting building height information from a single remote sensing image is approached partly by restoring geometric relationships in the imaging process (You et al., 2001; Zhao et al., 2010) and partly by calculating building height based on the shadow length formation process (HAN and CHEN, 2005; Xie and Li, 2004). The latter requires providing accurate building shadow length, obtaining various constraint information such as the orientation of the building as prior conditions, which greatly complicates the utilization and calculation of the model and hinders the rapid acquisition of building height (He et al., 2001). Extracting building heights through stereo image pairs often requires the use of image matching techniques (Cao and Chen, 2006; Hong and Chen, 2004; Xiao et al., 2010). In feature-based stereo matching techniques, Lowe (Lowe, 2004) introduced the widely applied Scale-Invariant Feature Transform (SIFT) operator, while in region-based stereo matching techniques, Pratt (Pratt, 1990) proposed the influential cross-correlation method.

As a crucial component capable of achieving sub-meter spatial stereo mapping in high-resolution Earth observation systems, the High-Resolution Imaging Satellite-7 (GF-7) stereo camera exhibits a large imaging convergence angle, leading to significant projection differences, building occlusions, and adhesion phenomena in the forward-looking images. These factors pose challenges to the extraction of building heights.

To leverage the capabilities of the GF-7 high-resolution Earth observation system, this study employs stereo image matching techniques to estimate building heights. Images from GF-7 satellite for Xi'an area, including forward and backward-looking images, are selected. An area with buildings of varying heights, shapes, and projection differences is chosen for study. The contour information of building roofs is obtained from the backward-looking image, and using a rational function model, the corresponding contour positions of the buildings on the forward-looking image are projected to obtain their heights. Image matching of corresponding building contour parts in forward and backward-looking images yields similarity degree calculation results, enabling the determination of the most likely object height. The objective is to achieve three-dimensional information extraction from two-dimensional high-resolution remote sensing images for target buildings.

The structure of this paper is organized as follows: Section 1 introduces the implementation of rational function model projection and the building matching methods selected. Section 2 provides an overview of the study area and data sources. Section 3 introduces the experimental process and results. Section 4 concludes the findings, offering prospects for future research.

2. Methodology

As illustrated above, the overall technical process can be decomposed into two key modules: 1) projection of the roof contour using the rational function model, and 2) calculate the matching degree to determine the optimal height.

2.1 Satellite Sensor Model

2.1.1 Imaging Model for GF-7

Remote sensing satellites employ various sensor models, typically categorized into strict sensor models and general sensor models. The latter avoids the physical geometric processes of sensor imaging and directly establishes mathematical relationships between image and ground coordinates. These models do not require consideration of the physical imaging process and do not demand comprehensive sensor information, making them particularly advantageous for processing line-array images. Although the theoretical expression of such methods may lack strict precision, their mathematical simplicity and computational efficiency make them suitable for applications in areas with relatively small terrain variations.

In the case of the selected High-Resolution Imaging Satellite-7 (GF-7) line-array CCD push-broom remote sensing imagery, achieving high spatial resolution requires designing sensor parameters with longer focal lengths and narrower field of view angles. If a strict sensor model is used to express the coordinate relationship between image and ground coordinates at this point, the directional parameters exhibit strong correlations due to their physical significance. This significantly compromises the accuracy and robustness of the orientation results, diminishing the advantages of high-resolution imagery in providing precise results. Choosing a general sensor model over a strict one under such circumstances allows for temporarily concealing certain sensor information from users, thus ensuring the confidentiality of some technical parameters. Moreover, it effectively accommodates the diverse development of sensor imaging methods and addresses emerging requirements. The GeoTIFF imagery selected for this study comes with Rational Polynomial Coefficients (RPC), a set of parameters defined under the Rational Function Model (RFM) within the general sensor model.

2.1.2 The Forward and Inverse Solution Formulas of RFM

The forward solution formula for the Rational Function Model (RFM) is as shown in Equation (1):

$$\begin{cases} r_n = \frac{p_1(X_n,Y_n,Z_n)}{p_2(X_n,Y_n,Z_n)} \\ c_n = \frac{p_3(X_n,Y_n,Z_n)}{p_4(X_n,Y_n,Z_n)} \end{cases} \tag{1}$$

If we denote the image pixel coordinates as (r,c) and the corresponding ground point coordinates as (X,Y,Z), the forward solution of the Rational Function Model (RFM) can be expressed as the ratio of polynomials in image coordinates to ground coordinates. The above Equation (1) represents the forward solution formula of the RFM.

In Equation (1), (r_n, c_n) and (X_n, Y_n, Z_n) represent the normalized coordinates after translation and scaling according to Equation (2). The purpose of normalization is to minimize computational errors and enhance numerical stability. The function $p(X, Y, Z)$ is a cubic polynomial in terms of coordinates, as shown in Equation (3), where p_2 and p_4 can be either equal or unequal and $(i + j + k) \leq 3$, a_{ijk} represents the Rational Function Coefficients (RFCs) with a total of 20 coefficients. This implies that Equation (1) has 80 coefficients. If the constant term a_0 is canceled out from both the numerator and denominator, reducing one coefficient, we can consider that Equation (1) contains 78 coefficients. In Equation (3), the linear terms are used to express pixel distortions caused by optical projection. The quadratic terms approximate corrections for Earth curvature, atmospheric refraction, lens distortion, etc. The cubic terms express unknown errors with high-order components, such as camera vibrations.

$$\begin{cases} r_n = \frac{r-r_0}{r_s} \\ c_n = \frac{c-c_0}{c_s} \\ X_n = \frac{X-X_0}{X_s} \\ Y_n = \frac{Y-Y_0}{Y_s} \\ Z_n = \frac{Z-Z_0}{Z_s} \end{cases} \tag{2}$$

$$p = \sum_{i=0}^{m_1}\sum_{j=0}^{m_2}\sum_{k=0}^{m_3}\sum_{i=0}^{m_1} a_{ijk}X^iY^jZ^k = a_0 + a_1Z + a_2Y + a_3X + a_4ZY + a_5ZX + a_6YX + a_7Z^2 + a_8Y^2 + a_9X^2 + a_{10}ZYX +$$

$$a_{11}Z^2Y + a_{12}Z^2X + a_{13}Y^2Z + a_{14}Y^2X + a_{15}X^2Z + a_{16}X^2Y + a_{17}Z^3 + a_{18}Y^3 + a_{19}X^3 \tag{3}$$

In Equation (2), $(r_0, c_0, X_0, Y_0, Z_0)$ represents the normalized translation parameters, calculated according to Equation (4). $(r_s, c_s, X_s, Y_s, Z_s)$ represents the normalized scale parameters, calculated according to Equation (5).

$$\begin{cases} r_0 = \frac{\sum r}{m} \\ c_0 = \frac{\sum c}{m} \\ X_0 = \frac{\sum X}{m} \\ Y_0 = \frac{\sum Y}{m} \\ Z_0 = \frac{\sum Z}{m} \end{cases} \tag{4}$$

$$\begin{cases} r_s = max(|r_{max} - r_0|, |r_{min} - r_0|) \\ c_s = max(|c_{max} - c_0|, |c_{min} - c_0|) \\ X_s = max(|X_{max} - X_0|, |X_{min} - X_0|) \\ Y_s = max(|Y_{max} - Y_0|, |Y_{min} - Y_0|) \\ Z_s = max(|Z_{max} - Z_0|, |Z_{min} - Z_0|) \end{cases} \tag{5}$$

The inverse form of the Rational Function Model (RFM) is to determine the ground plane coordinates (X,Y) from image plane coordinates (r,c) and the object space elevation Z. The inverse solution formula for the RFM is given by Equation (6).

$$\begin{cases} X_n = \frac{p_5(r_n,c_n,Z_n)}{p_6(r_n,c_n,Z_n)} \\ Y_n = \frac{p_7(r_n,c_n,Z_n)}{p_8(r_n,c_n,Z_n)} \end{cases} \tag{6}$$

2.2 Roof Matching Based on PSNR

Peak Signal-to-Noise Ratio (PSNR) is an engineering term defined by a ratio, describing the relationship between the maximum possible power of a signal and the power of destructive noise affecting the accuracy of that signal. Given that the dynamic range of most signals can be very wide, it is common to use logarithmic decibel units to express PSNR, as shown in Equation (7).

$$PSNR = 10log_{10}(\frac{MAX_I^2}{MSE}) \qquad (7)$$

where MSE (Mean Square Error) is the numerical value of the mean square error between two images. In other words, for a given digital image of size M×N and a reference image, the formula to calculate the PSNR for the given image is as shown in Equation (8), where MSE (Mean Square Error) represents the numerical value of the mean square error between the two images. For a given digital image f(x，y) of size M×N and a reference image f_0(x，y), the formula to calculate the PSNR for the image f is shown in Equation (15):

$$PSNR = 10log_{10}(\frac{f_{max}^2}{\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}[f(x,y)-f_0(x,y)]}) \qquad (8)$$

The traditional evaluation methods for image quality can often be divided into two categories: subjective evaluation and objective evaluation. Among them, when using objective evaluation methods to measure the quality of the measured image, the evaluation standard used is the error of the measured image deviating from the original reference image. The full reference image quality evaluation method is one of the methods belonging to the objective evaluation model of image quality. This method requires the original undistorted image as the reference benchmark when evaluating image quality, so that the quality difference of the tested image can be determined based on comparative analysis with the original data, and the final evaluation result can be obtained. The evaluation methods based on error sensitivity, pixel error, and structural similarity are the three most typical full reference image quality evaluation methods. The evaluation method for the quality of the tested image is based on pixel error by comparing the differences between each pixel of the evaluated image and the benchmark image, and comprehensively analyzing these differences. Mean squared error (MSE) and peak signal-to-noise ratio (PSNR) are the most commonly used objective evaluation methods for image quality based on pixel errors. The difference between the original image and the test image can be represented by mean square error, and the fidelity of the test image can be reflected by peak signal-to-noise ratio. The traditional objective quality evaluation methods mainly have the following characteristics: firstly, they are generally simple in algorithm, easy to implement, and fast in operation; Secondly, they all use the difference in grayscale between the test image and the original image as a measurement method; Again, the main approach is to use airspace based methods; Finally, the quality indicator is represented by the numerical value of the distance between the standard image and the quality space.

The calculation basis of peak signal-to-noise ratio (PSNR) is based on the statistical and average grayscale values of image pixels. It is one of the commonly used indicators to measure whether the signal is distorted and the degree of distortion. The larger the PSNR, the less distortion the signal has, and the better the quality of the generated image. However, in research on image quality evaluation based on the subjective perception

scores (MOS) of humans, PSNR often fails to take into account the characteristics of the Human Visual System (HVS) and the content factors of the video sequence that affect distorted visibility. The evaluation method of peak signal-to-noise ratio often does not match the subjective experience results of humans, resulting in lower performance (Sheikh et al., 2006).

The Peak Signal-to-Noise Ratio (PSNR) is widely used in the field of image quality assessment. It is an important parameter reflecting the performance of images and serves as one of the key indicators for full-reference image quality evaluation. PSNR holds significant guiding significance in image applications, particularly in the evaluation of image quality. The computation of Peak Signal-to-Noise Ratio (PSNR) is based on the statistical analysis and averaging of the grayscale values of image pixels. It is a commonly used metric for assessing whether a signal is distorted and to what extent. A higher PSNR indicates less distortion in the signal and better image quality.

Applying PSNR to compare the similarity of images involves calculating the Peak Signal-to-Noise Ratio for two matched images. A smaller value of PSNR indicates greater similarity between the two images, and when PSNR = 0, it signifies that the two images are identical. Generally, when the PSNR of two images is less than 30, it is considered that these two images are relatively similar.

3. Materials

3.1 Overview of the Study Area

The research utilizes high-resolution imagery from the High-Resolution Imaging Satellite-7 (GF-7), focusing on the region of Xi'an, China. The selected imagery corresponds to a specific area located at approximately 109.0 degrees east longitude and 34.4 degrees north latitude and its location in Xi'an city is shown in Figure 1. The chosen region within Xi'an encompasses a diverse range of building features, including variations in height, architectural shapes, and degrees of spatial aggregation. The inclusion of these diverse elements in the study area adds practical significance to the research on estimating the heights of urban buildings.
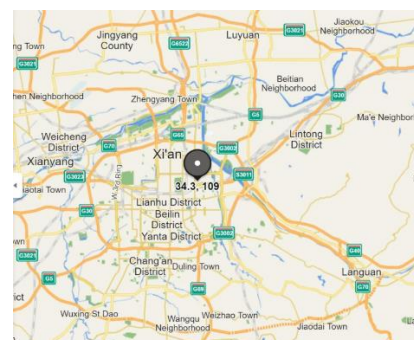

Figure 1. The corresponding ground position of the image area

In the chosen GF-7 imagery, a specific region featuring buildings of varying heights and shapes was selected for experimentation. The backward-looking image and forward-looking image of the selected area are separately presented in Figure 2 and 3. As observed from the images, the region encompasses a total of 34 buildings, showcasing a diverse range of architectural characteristics. These include regular buildings with rectangular roof shapes, as well as irregular buildings with roof shapes ranging from arcs to polygons. The buildings exhibit variations in height, with some having taller structures and larger shadow

areas, while others have fewer floors. Additionally, there are differences in roof brightness, ranging from brighter roofs to darker ones or roofs situated in shaded areas. The location of buildings varies from open and independent structures to denser clusters. This experimental area is chosen to effectively cover various real-world building scenarios.

Figure 2. Backward-looking Image of the Selected Area

Figure 3. Backward-looking Image of the Selected Area

3.2 Data sources

The choice of different data sources and image resolutions significantly determines the divergence in the adopted solutions. The selected data source is highly correlated with the methods chosen for building stereo matching and height estimation.

The single remote sensing image chosen for this study is from the High-Resolution Imaging Satellite-7 (GF-7), captured on February 17, 2020. The imaging location corresponds to approximately 109.0 degrees east longitude and 34.4 degrees north latitude. The imagery used is a panchromatic image (PAN) with a spatial resolution better than 0.8 meters. The ground elevation in the Xi'an area, corresponding to this imagery, ranges from approximately 300 to 400 meters. The dataset includes both forward-looking and backward-looking images captured over the same ground area.

The GF-7 satellite's stereo camera has a backward-looking panchromatic image with a tilt of 5 degrees and a forward-looking panchromatic image with a tilt of 26 degrees. The dataset encompasses buildings with a comprehensive range of heights, featuring various shapes such as regular rectangles, polygons with multiple vertices, and arc shapes. This diversity in building characteristics provides a robust experimental and validation environment for testing the effectiveness of building height estimation algorithms.

4. Experiment

4.1. Experimental Process

Firstly, the experiment requires obtaining contour data of buildings in the rear-view image. Determining the contour of buildings is crucial preparation for estimating their heights. In this paper, the determination of the contour of building roofs involves two main methods. One method is to obtain open-source data of building contour vectors on OpenStreetMap as shown in Figure 4, thereby acquiring the latitude and longitude coordinates of building corners. The other method involves manually drawing and marking the contours of building roofs using ENVI software as shown in Figure 5, exporting the vector data of building contours, and subsequently obtaining the row and column coordinates of corners on the image. Through these methods, especially the latter, the precise location of the research object on the image is determined, and its pixel coordinates are obtained.

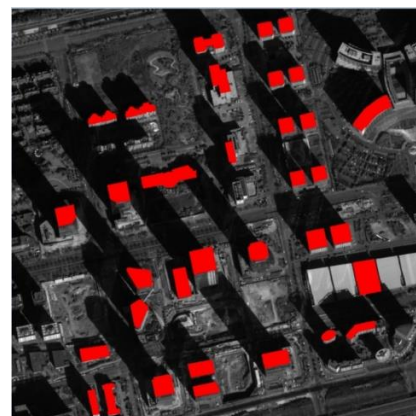Figure 4. OpenStreetMap open-source contour vector data

Figure 5. ENVI self-marked contour vector

Secondly, in the technical workflow of this paper, the RFM process needs to be completed to project the known building contours from the rear-view image to the front-view image. This provides the corresponding image information for the subsequent matching process. To achieve this step, the paper utilizes the relevant functions of the open-source raster spatial data conversion library GDAL.

GDAL 3.0 is compiled with x64 static library debug version in the WIN10 environment using VS2019. The program is written in C++. Firstly, the program reads the vector data file of building contours marked by ENVI on the rear-view image, obtaining the row and column coordinates (pixel coordinates - r, c) of the roof corners of the buildings on the rear-view image. Assuming a certain object space elevation, the RFM is used to transform pixel coordinates into corresponding ground point coordinates.

As analyzed earlier, the Rational Function Model (RFM) is universally adaptable to various coordinate systems. By performing RPC transformations under this function, the results obtained are the latitude, longitude, and geodetic height of ground points. Further modification of function parameters allows for RPC inverse transformations, obtaining the row and column coordinates of the corresponding points on the front-view image based on ground coordinates. This completes the localization of contours from the rear-view image to the front-view image, providing conditions for the next step of roof matching.

The assumption of the object space elevation for buildings is based on the ground elevation in the Xi'an area and the manually measured elevation values of some buildings in ENVI. The ground elevation in the Xi'an area in the selected image is roughly in the range of 300 meters to 400 meters. Combining this with the manually measured values, for all buildings in the selected research area, it can be considered that their elevation ranges from 300 meters to 750 meters. The paper conducts experiments with a 1-meter search step within this elevation range, obtaining contours of buildings on the image corresponding to these elevations. Figure 6 shows the contours of a specific building at representative elevations in the front-view image.
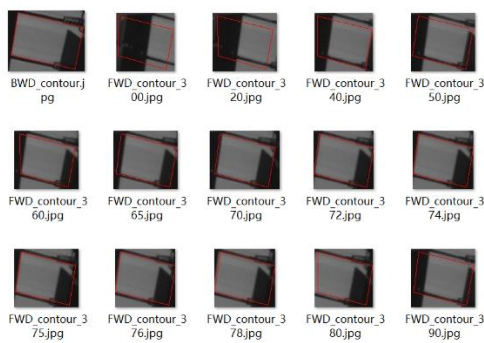

Figure 6. Front-view building contours corresponding to a certain building at various elevations

Finally, based on the previous extraction of building roof contours in the rear-view image, manual stereoscopic elevation measurement, and the implementation of the Rational Function Model (RFM) for forward and inverse projection, the positions of building roofs in the front and rear-view images under different assumed heights were obtained. The next step involves conducting roof matching experiments.

Due to the diverse shapes and heights of buildings, their contours in the front and rear-view images are not always perfectly regular. Matching roofs under such varied contour features is not universally applicable and is challenging to achieve through grayscale correlation in the surrounding areas of the extracted contours. Therefore, this paper chooses to use the minimum bounding rectangle of the roof contour (with edges parallel to the coordinate axes of the image's image plane coordinate system) as the matching window. By calculating the correlation coefficients NCC, grayscale histogram, perceptual hash algorithm PHA, and Peak Signal-to-Noise Ratio (PSNR) for the two windows corresponding to different heights, the paper selects the building height corresponding to the maximum value of the similarity measure as the elevation estimate for that building. This process is repeated for all buildings in the region, searching for the highest similarity value to determine the elevation for each building.

4.2. Results and Analysis

When using a search step of 1 meter and selecting PSNR as the similarity measure, the elevation estimation experimental results for all selected buildings in the region are shown in Table 1.

As shown in the table, there are a total of 34 buildings in the study area, and statistical analysis of their elevation estimation results can be obtained as shown in Figure 7.

Building ID	Elevation by manual measurement	Elevation estimated	Δ
1	564.565	566	-1.435
2	703.384	702	1.384
3	592.809	592	0.809
4	618.217	616	2.217
5	614.165	614	0.165
6	454.479	454	0.479
7	567.774	568	-0.226
8	515.138	748	-232.862
9	456.724	454	2.724
10	456.403	454	2.403
11	453.515	454	-0.485
12	454.789	456	-1.211
13	455.120	458	-2.88
14	455.42	456	-0.58
15	456.724	456	0.724
16	456.724	454	2.724
17	435.542	434	1.542
18	435.542	436	-0.458
19	442.282	442	0.282
20	441.961	442	-0.039
21	374.564	376	-1.436
22	434.420	434	0.42
23	380.341	378	2.341
24	565.208	564	1.208
25	461.218	460	1.218
26	458.651	676	-217.349
27	460.897	460	0.897
28	461.539	426	35.539
29	455.441	454	1.441
30	375.206	368	7.206
31	377.132	374	3.132
32	374.564	374	0.564
33	432.333	432	0.333
34	432.975	534	-101.025

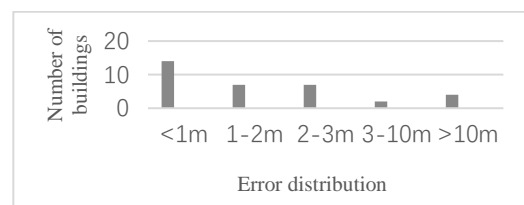Table 1. Estimated elevation values for each building

Figure 7. Error distribution of building elevation estimation

Under the technical route of this paper, there are a total of 14 buildings whose elevation estimates can be approximated as true values with a difference of less than 1m compared to manually observed elevations. There are 22 buildings whose elevation

estimation errors are less than 2m, 29 buildings whose elevation estimation errors are less than 3m, and only 4 buildings whose estimation results show significant deviations. The accuracy and reliability meet the requirements, and the probability of obvious errors only accounts for 10%. It can be considered to discover through automatic gross error detection.

5. Conclusion

This paper's discussion is based on high-resolution satellite remote sensing images, specifically the high-resolution panchromatic stereo images from the High-Resolution Satellite 7's linear CCD push-broom imaging system. Addressing the significant grayscale differences and noticeable geometric deformations between stereo image pairs caused by different imaging conditions in the linear CCD push-broom remote sensing stereo images, the study utilizes a Vertical Line Locus (VLL) approach based on assumed object heights for image matching. This is complemented by a matching scheme based on PSNR similarity measure, identifying the maximum value to estimate the three-dimensional height information of buildings on the images. This method, built upon the correct projection of building contours from front to rear-view images, considers the utilization of pixel grayscale information from building roof contour image blocks. The results from this method meet the requirements for building height estimation in terms of accuracy, speed, and reliability. Further research can be conducted to automate the method for determining building roof contours. Additionally, to address the potential issue of building rotation in front and rear-view images caused by stereo imaging, solutions include generating kernel line images to transform the height search into a disparity search or applying an affine transformation to correct the study area for rotation effects.

References:

An, J.Y., Cheng, P.G. and Ding, B.F., 2010. Method for acquiring buildings height from 2D images in Google earth. *Geography and Geo-Information Science*, 26(6): 31-37.

Cao, Z. and Chen, H., 2006. Quick 3D reconstruction of the IKONOS-satellite stereo images. *Computer and Digital Engineering*, 34(11): 9-11.

HAN, Y. and CHEN, Y., 2005. An exploration of 3D reconstruction of house with single image. *Journal of Shandong Institute of Architecture and Engineering*, 20(3): 17-22.

He, G., Chen, G., He, X.Y., Wang, W. and Liu, D.S., 2001. Extracting buildings distribution information of different heights in a city from the shadows in a panchromatic SPOT image. *Journal of image and graphics*, 6(5): 425-428.

Hong, L. and Chen, G., 2004. Segment-based stereo matching using graph cuts, *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE*, pp. I-I.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60: 91-110.

Pratt, W.K., 1990. *Digital Image Processing*, A Wiley-Interscience Publication, 1991. P. Amorese and J. Bloomfield,"A slew of standard camera systems", Datacube World Review: 40-49.

Sheikh, H.R., Sabir, M.F. and Bovik, A.C., 2006. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, 15(11): 3440-3451.

Song, R.B., Li, W.H. and Wang, X.Y., 2014. A method of extracting urban building heights from double images. *Engineering of Surveying and Mapping*, 23(9): 66-69.

Xiao, J., Xia, L. and Lin, L., 2010. Segment-based stereo matching using edge dynamic programming, *2010 3rd International Congress on Image and Signal Processing. IEEE*, pp. 1676-1679.

Xie, J.F. and Li, Y.M., 2004. The extraction of building distribution information of different heights in a city from the shadow in a IKONOS image. *Remote Sens. Land Resour*, 4: 4-6.

You, H. et al., 2001. Fast acquiring urban DSM image and displaying 3D image. *JOURNAL OF REMOTE SENSING-BEIJING-*, 5(1): 12-15.

Zhao, M., Li, X. and Lang, R., 2010. Improved adaptive stereo matching using double dynamic programming with correlation of row and column, *2010 2nd International Conference on Signal Processing Systems. IEEE*, pp. V3-95-V3-100.