Research on Building Height Extraction Method from High-resolution Image

Tang Hao1*, Du Linlin1, Wang Chun1, Li Pengfei1, Ge Shuangquan1, Meng Dezhuang1

1Computer Application Institute, China Academy of Engineering Physics, Mianyang, 621900, China
*Corresponding author’s e-mail: tanghao747@163.com

Abstract: In recent years, there are many researches on calculating the height of buildings by extracting the shadow of buildings, but this method is difficult to be realized in the case of dense buildings or complex ground. To solve this problem, we can use the side information of the building to retrieve the height of the building. In this paper, Otsu Algorithm and LVQ neural network are used to extract the side information of buildings from high-resolution remote sensing images, calculate its length, and then calculate the height of the building according to the physical model of imaging. Compared with the measured building height, the calculation results of this method can meet the accuracy requirements of building height calculation, and this method can complement the results of shadow extraction method. Finally, the height information of most buildings in the area can be obtained.

1. Introduction

The key of this work is to obtain the building height information when building three-dimensional modeling of large area buildings in areas where basic surveying and mapping data are scarce, such as a certain area abroad and poor and backward areas. The current method of obtaining building height information can be divided into the following categories: (1) using stereo image pairs of images to calculate building height according to the principle of aerial triangulation. This method has high accuracy in theory, but requires two or more images with a certain overlap; (2) by obtain building shadow information and calculate shadow length, simulate the spatial geometric relationship between shadow and building according to Sun's altitude and azimuth, this method only needs one image, which is relatively low cost, but cannot obtain complete shadow information when the building is dense or the ground objects are complex; (3) Using Lidar (Light Detection And Ranging) data to obtain height information of buildings, which is less applicable[1-2].

The research group has done many years of shadow inversion of building height [3-6] and found that this method has some limitations. For example, when the building is dense, the shadow of the front row building may be projected to the back row building, resulting in incomplete shadow of the extracted building.

Therefore, the paper attempts to use depth learning and other methods to obtain the side information of buildings, calculate their length, and then estimate the height of buildings according to the spatial geometric relationship between satellites and buildings. In order to make up for the defect of using shadow information to calculate building height information. At home and abroad, some literatures have been studying this problem[17], but the research content mainly involves a relatively simple scene, and has not been compared with the shadow extraction method.
2. Research methods and principle

The following prerequisites are required for research of this methods:

1. Remote sensing images with high spatial resolution. Spatial remote sensing images with high resolution can obtain more pixels, such as sub-meter Quick Bird images or higher resolution aerial images.

2. Non-positive projection imaging. When using the side information of the building to estimate the height of the building, it’s necessary to see the side information of the building in the remote sensing image, but when the satellite height angle is 90 degrees or close to 90 degrees, the side image is not obvious. Therefore, the length of the visible side should not be too short, otherwise the side information of the building cannot be identified.

3. When the sun and satellite are on the same side information of the building, the spectral features of the building side are obvious, the shape and texture features is beneficial to the extraction of the building side.

4. The structure outside the building is regular and tidy. If it’s irregular, it is necessary to use more complex physical models to calculate the height of the building. The irregular external structure of the building will affect the extraction accuracy of the side information of the building, and prevent inversion of building height.

5. The side information of the building is less sheltered. When the buildings are dense, the adjacent buildings will block each other, or the shadow of the adjacent buildings will be projected on the side of another building; if there is more vegetation on the ground, it will also block the building side.

Subject to the above five premises, the spatial location of satellites and buildings is shown in figure 1:

![Figure 1. Schematic diagram of space position relationship between satellite and buildings](image)

It can be seen from figure 1 that the height information of the building is closely related to the azimuth of the satellite and the height angle of the satellite. Because the remote sensing image is a two-dimensional image, it is necessary to restore the image to a three-dimensional graph to calculate. The relationship between the two-dimensional graph and the three-dimensional graph is shown in figure 2. We can see that the side information of the building on the image is B, while the side of the real building is A. The relationship between the actual height of the building and the length of the side information of the building is shown in (1).

\[
H = \frac{L}{\cos \beta}
\]  

(1)

The H represents the actual height of the building, the L represents the length of the side information of the building on the image, the \( \beta \) represents the height angle of the satellite.
Figure 2. 3-D and 2-D diagram of the building side

In fact, in most cases, the parallel line of sight taken by the satellite is not perpendicular to the orientation of the building, and the actual shape is shown in plane C. In order to simplify the algorithm, only need to consider the edge direction of the building, the edge line is orientation of the building side. As shown in figure 3, the angle between the edge line and Y direction of the screen obtained, and calculate the actual length of the side according to the triangular relationship.

Figure 3. Building edges

3. Side information extraction based on deep learning method

Experimental data is a Quick Bird image after fusion. There are two methods for building side extraction: (1) threshold segmentation (Ostu algorithm) \(^8\) and mathematical morphology; (2) using deep learning which is LVQ neural network \(^9\) to extract building side information. The following is a brief introduction to the processing process of the second method.

LVQ neural network used in this paper, that is, Learning Vector Quantization, a kind of forward neural network, which is mainly used to train competitive neural network, and include supervised learning methods. Therefore, LVQ neural networks can be classified like neurons, which can overcome the classification errors by unsupervised learning algorithms. For example, input a picture, set its discriminant condition as whether this picture contains an aircraft, if there is an aircraft, you can define this picture as 1 flag, if there is no aircraft in this picture, the picture is defined as 0 flag. By using this feature, the LVQ neural network is regarded as a classifier, to divide the picture into several blocks. The image block which feature satisfies the side of the building is defined as 1, and the image block which does not satisfied is defined as 0. Then the image block defined as 1 is set as foreground and 0 is set as background. finally, side information of the building can be achieved.

(1) Select training samples

A large number of images with a pixel size of 256×256 can be cut from the remote sensing image, which has clearly sight of building side and mark it. Turn this picture into a gray image, and set the side
information of the building as the foreground and the other places as the background. Rotate the marked pictures in different angles to double the number of training samples. One of the label diagrams is shown in figure 4.

![Figure 4. Label diagram](image)

(2) Set sampling parameters
Determine the pixel intervals and window size for block sampling, loop into the original picture and label picture, calculate the number of blocks according to the pixel interval and window size of block sampling, and calculate the positive and negative sample number in each block. The positive and negative samples are evenly balanced.

(3) Set input vectors
For most buildings, the building material of side is the same as top, which may cause the spectral features of the building side and top to be similar, so we can not only rely on the spectral features as feature vectors. The geometric and texture features of the building side must also be added to form the input vector. Geometric features contain three indexes: The ratio of square root to circumference in extracted area, the ratio of circumference to the number of vertex, and the area ratio of characteristic primitive and external rectangle, the above forms the eigenvectors of 3D shapes. The formulas are shown in formulas (2), (3) and (4). Texture features can reflect the arrangement rules in a local area in image, but texture features can only reflect few features of the object’s surface. Texture description can use four characteristic descriptors of the second-order gray level co-occurrence matrix of histogram and pixel relative position as input vectors. the four descriptors are: consistency, entropy, contrast, and self-correlation. The formulas are shown in (5), (6), (7) and (8).

\[
S = \frac{\sqrt{A}}{P} \quad (2)
\]

\[
ML = \frac{P}{Pnts} \quad (3)
\]

\[
Cmp = \frac{A}{EA} \quad (4)
\]

\[
\sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij}^2 \quad (5)
\]

\[
-\sum_{i=1}^{K} \sum_{j=1}^{K} p_{ij} \log_2 p_{ij} \quad (6)
\]

\[
\sum_{i=1}^{K} \sum_{j=1}^{K} (i - j)^2 p_{ij} \quad (7)
\]
\[
\sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(i - m_r)(j - m_c)p_{ij}}{\sigma_r \sigma_c}
\]  \tag{8}

The \( A \) is the side area of the building, the \( P \) is the side circumference of the building, the \( Pnts \) is the number of vertex of the extracted building side, \( p_{ij} \) is the total elements of \( ij \) term of the symbiotic matrix divided by the symbiotic matrix. \( m_r \) and \( m_c \) are the mean values of the normalized symbiotic matrix along the row and column directions, \( \sigma_r \) and \( \sigma_c \) are the standard deviations of the normalized symbiotic matrix along the row and column directions.

(4) Normalized inputs

In training a model, in order to speed up the learning process and reduce the experimental time, the input feature vector is normalized. The normalized input is mainly divided into two steps, the first step is zero value equalization, that is, the variance of the difference between the training value and the average value, and the second step is to normalize the variance data according to the previous step.

(5) Training and testing

The training data are obtained according to the learning efficiency settings, the number of iterations, the number of neurons and the connection weights, and then extraction results of experimental area A and B can be obtained by testing. Which shown in Figures 5 and 6.

![Figure 5. Experimental area A](image1)

(a) Results extracted  (b) Actual images

![Figure 6. Experimental area B](image2)

(a) Results extracted  (b) Actual images
4. Precision analysis of experimental results

Three parameters are used to evaluate the extraction accuracy: side missed detection rate, side false detection rate and side total error rate. The calculated expressions are shown in formulas (9), (10) and (11).

\[ F_{SOC} = \frac{F_N}{F_Y} \]  
\[ F_{SCE} = \frac{F_E}{F_Y} \]  
\[ F_{STE} = \frac{F_N}{F_Y} + \frac{F_E}{F_Y} \]

The \( F_{SOC} \) represents the side missed detection rate, the \( F_{SCE} \) represents side false detection rate, the \( F_{STE} \) represents side total error rate, \( F_N \) represents the total number of pixels undetected on the side information of the building. \( F_E \) represents the total number of non-building side pixels mistakenly detected as building side pixels, \( F_Y \) represents the real total number of building side pixels. The above formula is used to calculate the three evaluation parameters, and the results are shown in the following table.

| Total number of side pixels extracted | Manual plotting of total side pixels | Side missed detection rate (%) | False detection rate on shade side (%) | Total Side Error (%) |
|--------------------------------------|--------------------------------------|--------------------------------|----------------------------------------|-------------------|
| A Area 54309                         | 55203                                | 1.619477                       | 0.724598                               | 2.344             |
| B Area 25354                         | 25623                                | 1.049838                       | 0.507357                               | 1.557             |

| Total number of side pixels extracted | Manual plotting of total side pixels | Side missed detection rate (%) | False detection rate on shade side (%) | Total Side Error (%) |
|--------------------------------------|--------------------------------------|--------------------------------|----------------------------------------|-------------------|
| A Area 54869                         | 55203                                | 0.605040                       | 0.362300                               | 0.967             |
| B Area 25508                         | 25623                                | 0.448816                       | 0.390274                               | 0.839             |

From the above two tables, it can be seen that the results extracted by the depth learning method are convenient and accurate than traditional method (Ostu algorithm). It shows that this extraction method is feasible.

5. Estimate for building’s actual height

The estimation of the actual building’s height is divided into two steps: (1) calculating the length of the extracted side information corresponding to building’s height;(2) using the method shown in formula (1) to calculate the actual height of building.

In this paper, the side of the building in the experimental area is mostly parallelogram, and the length of the building side can be obtained by direct measurement. A fishing net method[3] is used calculate the side length of the building. According to fishing net method, the grid line is generated on the binary image, and the cut line is obtained by logic negation operating. The cut line is counted in an array in the form of connected components. Filter with pauta criterion, the number of pixels of the remaining line is calculated, the average value is obtained, and the final result of the side length is obtained.

After calculating the side length, the height of the building is estimated by formula (1) method. The calculated height and actual height of 10 buildings are compared. The results are shown in Table 3 and Table 4.
Table 3. Accuracy evaluation of inversion results extracted from building sides based on traditional methods

| Building ID | Side length(m) | Calculated Building height(m) | Measured Building Height(m) | Absolute error(m) | Relative error(%) |
|-------------|----------------|-------------------------------|----------------------------|------------------|-----------------|
| 1           | 8.1            | 22.29163                      | 23.4                       | -1.10837         | 4.7366          |
| 2           | 7.8            | 21.46601                      | 23.4                       | -1.93399         | 8.2649          |
| 3           | 8.1            | 22.29163                      | 23.4                       | -1.10837         | 4.7366          |
| 4           | 8.0            | 22.01643                      | 23.4                       | -1.38357         | 5.9127          |
| 5           | 7.9            | 21.74122                      | 23.4                       | -1.65878         | 7.0888          |
| 6           | 7.9            | 21.74122                      | 23.4                       | -1.65878         | 7.0888          |
| 7           | 8.0            | 22.01643                      | 23.4                       | -1.38357         | 5.9127          |
| 8           | 7.9            | 21.74122                      | 23.4                       | -1.65878         | 7.0888          |
| 9           | 7.7            | 21.46601                      | 23.4                       | -1.93399         | 8.2649          |
| 10          | 8.1            | 22.29163                      | 23.4                       | -1.10837         | 4.7366          |

Table 4. Accuracy evaluation of inversion results extracted from building sides based on deep learning method

| Building ID | Side length(m) | Calculated Building height(m) | Measured Building Height(m) | Absolute error(m) | Relative error(%) |
|-------------|----------------|-------------------------------|----------------------------|------------------|-----------------|
| 1           | 8.2            | 22.29163                      | 23.4                       | -0.83316         | 3.5605          |
| 2           | 7.9            | 21.46601                      | 23.4                       | -1.65878         | 7.0888          |
| 3           | 8.1            | 22.29163                      | 23.4                       | -1.10837         | 4.7366          |
| 4           | 8.0            | 22.01643                      | 23.4                       | -1.38357         | 5.9127          |
| 5           | 8.2            | 22.29163                      | 23.4                       | -0.83316         | 3.5605          |
| 6           | 8.1            | 21.46601                      | 23.4                       | -1.10837         | 4.7366          |
| 7           | 7.9            | 21.46601                      | 23.4                       | -1.65878         | 7.0888          |
| 8           | 8.1            | 21.74122                      | 23.4                       | -1.10837         | 4.7366          |
| 9           | 8.1            | 22.29163                      | 23.4                       | -1.10837         | 4.7366          |
| 10          | 8.0            | 21.46601                      | 23.4                       | -1.38357         | 5.9127          |

In order to compare the error comparison of the two methods intuitively, the line chart comparison of absolute error with two methods is constructed, and the results are shown in figure 7.

Figure 7. Comparison of absolute error
It can be seen from the diagram that there are some errors in both methods and need to be improved continuously. In general, the absolute accuracy of the depth learning method is slightly higher than the threshold segmentation method.

6. Conclusion
It is found that the following conditions will affect the effect of this method: (1) the vegetation at the bottom of the building is too lush, which will lead to the boundary line of the building contact with the ground cannot be directly selected; (2) the structure material of roof and side are consistent, which will interfere with side extraction; (3) the top structure of the building is complex and uneven. (4) Over-dense buildings with fewer exposed sides.

It is found that it is feasible to estimate building height by using building side information in high resolution remote sensing images, and can complement each other with shadow calculation method, which can make up for some defects of shadow calculation, such as dense building group area. At the same time, some of the difficult situations in this method can be realized in the shadow method, such as when the material is consistent with side and the top of building. The degree of complementation between this method and the shadow extraction method needs further verification in the subsequent research.

References
[1] Chen hui, Zhang Ka, Su Dong. Method of building height estimation under the constraint of profile contour lines from high resolution remote sensing images[J]. Bulletin of Surveying and Mapping, 2019, 9:34-37.
[2] Xue Wang, Peijun Li. Extraction of urban building damage using spectral, height and corner information from VHR satellite images and airborne LiDAR data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 159:322-336.
[3] Xie Yakun. Research on Building Height Information Extraction Method Based on Shadow of High-resolution Image in Multiple Scenes[D], Southwest Jiaotong University, 2018.
[4] WU Yiyao, FENG Dejun, HU Minjun. Extraction of Building Area Shadow Based on Aerial Images[J]. Geomatics World, 2019, 26(6): 44-48.
[5] Li Qiang;Feng Dejun;Hu Minjun. Shadow Detection of Integrated Characteristic Components for GF-2 Image[J]. Remote Sensing Technology and Application, 2019, 34(6): 1252-1260.
[6] Zhang Heng, Research on Buildings Shadow Detection Method and Height Inversion with High Resolution Sensed Image[D], Southwest Jiaotong University, 2017.
[7] Song Xudong, Cao Wenfeng, Feng Dejun. Building Side Information Extraction and Height Calculation Based on High-Resolution Remote Sensing Imagery[J]. Beijing Surveying and Mapping, 2020, 34(3): 296-300.
[8] Hu Weijian. Research of 3D Visilization of Medical Data Based on Volume Rendering[D], Beijing Institute of Technology, 2016.
[9] Parini, Mawengkang H, Efendi S. Hybrid learning vector quantization (LVQ) algorithm on face recognition using webcam[J]. IOP Conference Series: Materials Science and Engineering, 2018, 420(1).