MULTI-VIEW 3D CITY MODEL GENERATION WITH IMAGE SEQUENCES

Zuxun Zhang, Jun Wu, Yong Zhang, Yongjun Zhang, Jianqing Zhang

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430079, P.R. China
(zxzhang, wujun, yongzhang, yjzhang, jqzhang)@Supresoft.com.cn

Commission V/WG 4 and IntCommission V/III

KEY WORDS: 3D city model, Image sequence, Automatic aerial triangulation, Coarse-to-fine building modelling, Texture mapping

ABSTRACT:

Rapid texture mapping of buildings and other man-made objects is a key aspect for reconstruction of 3D city landscapes. An effective approach by the way of coarse-to-fine 3D city model generation based on digital photogrammetric theory is proposed. Three image sequences, two oblique photography to buildings' walls and one vertical photography to building's roof, acquired by digital video camera on a helicopter, coarse 2D vector data of buildings and LIDAR data are used as sources of information. Automatic aerial triangulation technique for high overlapping image sequences is used to give initial values of camera parameters of each image. The correspondence between the space edge of building and its line feature in image sequences is determined with a coarse-to-fine strategy. Hybrid point-line photogrammetric technique is used for accurate reconstruction of buildings. Reconstructed buildings with fine textures superimposed on DSM and orthoimage are visualized realistically.

1. Introduction

Three-dimensional (3D) graphics has many advantages over video or 2D graphics. 3D interactive environment is an intuitive and user-friendly way to view location-based information. For many applications, 3D representation of city model improves the usability of data. People are becoming increasingly conscious about their technical, environmental and commercial possibilities (Forstner 1999, Brenner 1999), such as car navigation and service browsing, tourism and marketing, architecture and town planning, city climate, noise propagation and environmental research, and new landmark-based navigation systems. Interest in 3D city model generation and quickly update has increased over the past years (Masafumi 2001, Brenner 2001).

3D city model is usually comprised of a description of terrain, streets, buildings, vegetations, and other man-made objects. The main workload is for the building modelling. Rapid texture mapping of buildings and other man-made objects is a key aspect for reconstruction of 3D city landscapes. It is clear that photogrammetry is able to provide a means to collect the required three-dimensional information.

Research on building reconstruction can be subdivided into two large classes, namely automatic systems (Henricsson and Baltsavias 1997, Baillard et al. 1999) and semiautomatic systems (Grun 1998, Brenner 1999).

Fully automatic systems working with aerial images are not considered for practical use (Brenner 2001), although tremendous progress has been made. Companies actually collect 3D models relying almost exclusively on manual data acquisition, even though, measured models are not very detailed. One of the great disadvantages of systems working on the basis of DSM is the usually lower point density than LIght Detection And Ranging (LIDAR) data, that is the laser scanning data. This leads to a poor lateral accuracy. Combining DSM with existing vector data for automatic reconstruction has shown good results (Brenner 1999).

It can be doubted that automatic systems can achieve success rates comparable to human operators within the next few decades (Forstner 1999), while semiautomatic systems can be used in large practical projects. They can also be taken as kernel of systems where fully automatic algorithms can be incrementally integrated. Researchers have put forward approaches to model buildings using a small number of CSG models (Heuvel 2000). But a lot of interactions make the modelling process quite lower efficient. It is desirable to have semiautomatic systems trying to reduce interactions per building. Moreover, one would like to have a system reconstructing a considerable number of buildings without any operator interference at all.

Therefore, current photogrammetric system is not the most efficient solution to create city model, since it is made to measure points rather than structured objects mainly composed of line segments, and until now, even semiautomatic approach for building geometry modeling from aerial images is still far away from real practice. On the other hand, the vertical walls of buildings are usually invisible in normal aerial image.

Several important techniques by the way of coarse-to-fine 3D city model generation, based on digital photogrammetric theory, are introduced in this paper. In a project for searching a method on quickly producing 3D data for car navigation, the techniques are used. Three image sequences, one vertical photograph to building's roof and, two oblique photographs to buildings' walls, acquired by digital video camera on a helicopter, coarse 2D vector data of buildings and LIDAR data are used as sources of information in the project.

Automatic aerial triangulation technique for high overlapping image sequences is firstly discussed to give initial values of camera parameters of each image. As to building modelling, a multi-view interactive coarse-to-fine reconstruction algorithm with low-accuracy 2D vector data is proposed, for determining the correspondence between the space edge of building and its line feature in image sequences then. Hybrid point-line photogrammetric technique is used for accurate reconstruction of buildings. Rectification and automatic noise removing

methods are adopted to generate realistic textures of buildings and roads. Reconstructed buildings with fine textures superimposed on DSM and orthoimage are visualized realistically. The experiment results of 3D city model generation show that the proposed approach has a promising future in both efficiency and effect. The conclusion with outlook of future developments is also given.

2. AUTOMATIC AERIAL TRIANGULATION

Traditional aerial photographs overlapped about 60 percent in a strip. While image sequences taken with digital video camera has the advantages of high overlapping and redundancy of corresponding features, which has a well potential for automatical 3D reconstruction. This will surely improve the speed of 3D reconstruction and also reduce the manpower significantly.

In the project, coarse 2D vector data of buildings, which are for 2D car navigation and with un-negligible errors, LIDAR data and image sequences of city are used as sources of information. Image sequences are taken by a digital video camera on the helicopter. The camera is fixed in a stabilizer, which can keep the camera pose stable while the helicopter flying. There are totally three image sequences, one with optical axes looking downward to get the images for roofs and roads (level image), other two sequences with oblique view-angles of about 45 degree to acquire the texture of walls of buildings. The image in the first sequence is nearly level (called level image in the later part of this paper) and the others in the sequences with oblique photography are oblique (called oblique image in the later part of this paper). Three flying routes are shown in Figure 1. The practical image sequences are extracted from original image sequences and each one is taken from every ten images. The practical image sequences are with about 95% overlapping and 0.08 m ground resolution.
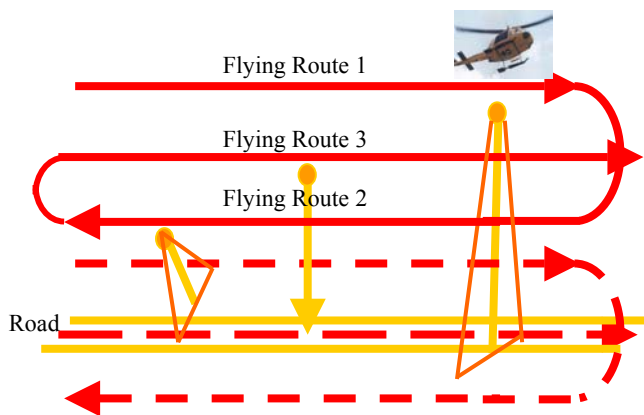

Figure 1. Three Flying Routes

The most important thing of 3D-model generation is the corresponding problem among images and back-projected spatial data. It is well known that, the vector data for car navigation are often not accurate, so both camera parameters and model of buildings must be determined. Automatic Aerial Triangulation (AAT) technique can be adopted to acquire initial values of camera parameters. The three image sequences are processed with VirtuoZo AAT software as single strip separately. The output of AAT is a free network taking the first image as reference coordinates system, which is different from

the world coordinates system. So the camera parameters of the free network need to be transformed into world coordinates system in order to give reliable initial values of the image poses in each image sequences. Two images near the beginning and end part of the image sequences are selected. Then some control points are selected from the vector data and LIDAR data, and their corresponding image points are given interactively in the selected images. Thus camera parameters of the two selected images can be obtained with space resection. Then all parameters of the strip can be transformed into world coordinates system. The pass points are also calculated, which will be used in hybrid point-line photogrammetry to ensure the stability of bundle adjustment.

3. COARSE-TO-FINE BUILDING MODELLING

Nowadays, the 2D vector data of buildings and other man-made objects are normally available in most cities. Although the 2D vector data are often not as so precise as image resolution, they can be used as initial values for building reconstruction. An iterative coarse-to-fine process, composed of projecting, drawing, matching and bundle adjustment, is adopted in the procedure of the building reconstruction.

3.1 Initial Geometry Model of Building

The initial geometry model of the building is determined by 2D vector and LIDAR data, which provides the approximate height value (The detail is not introduced in this paper). The initial model with low accuracy will be back-projected onto the level images according to AAT results to acquire initial values of edges of buildings on image.

3.2 Initial Wire-frame of Building on Image

For each building, all relative images, on which building should be fully included, are automatically retrieved from the whole image sequence, by comparing the building location with every image range on ground, calculated through image pose parameters. The best image, on which centre the building should be about, is auto-selected too.

After the initial geometry model of the building is projected backward to the relative images, the initial wire-frame of building is displayed on image. If the error is too large to process later, the corners of building should be drawn to its actual position on the image. The initial frame-wire on image is also maybe rotated and scaled. Building height can also be corrected by drawing the frame-wire on the image. These operations can be easily performed on the best image through user interface, and corresponding operation will automatically be implemented on all relative images.

As shown in Figure 2, there are not only offsets but also scale factors between the projected wire-frame of buildings and image edges. Further more, the geometric shape of the left building in Figure 2 is different from the image significantly because of the data errors, although offset and scale factor are removed by the manual way, as shown in Figure 3.

3.3 Line Extraction

The extraction of line feature on images is only performed in the neighbourhood of projected initial frame-wire of the roof. At the beginning, one rectangle image patch, centering at the projected position, is used as target region for searching possible image edges. The main purpose is to gain the long line feature.

Common steps of line extraction are edge detection, edge tracing and edge fitting (Forstner, 1982; HU,2000; Zhang, 2001) based on image pyramid. Although the image line features, extracted in this stage, are not so good in geometry precision, it's still good enough for line intersection and space forward intersection to improve the roof geometry model. After the roof geometry model is improved with the adjustment of space forward intersection later, the line extraction algorithm based on Least Squares Template Matching-LSTM (Ackermann, 1983; Forstner., 1982; Gruen, 1985) to a narrow rectangle image can be applied to gain the final line feature..
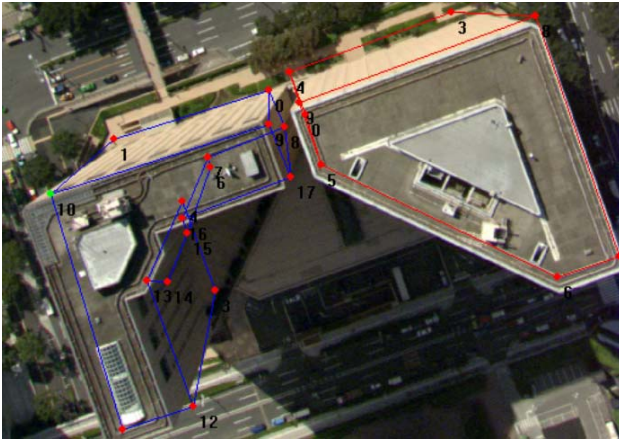

Figure 2. Initial projections of buildings

Figure 3. Imprecise 2D ground map of Building

3.4 Matching Space Edge to Corresponding Image Edge

After extracting the line features on the best image, system can automatically match all visible lines of the corresponding building on each image.

3.4.1 Matching of Space Edge for Roof: Since there are several homologue points in different images for each roof

vertex by intersecting the two adjacent candidate line, an indirect least squares adjustment with the flat roof constraint is used in space forward intersection to restore the roof's vertex. The collinear equation

$$x_i - x_0 = -f \frac{a_1(X_i - X_s) + b_1(Y_i - Y_s) + c_1(Z_i - Z_S)}{a_3(X_i - X_s) + b_3(Y_i - Y_s) + c_3(Z_i - Z_S)}$$
$$y_i - y_0 = -f \frac{a_2(X_i - X_s) + b_2(Y_i - Y_s) + c_2(Z_i - Z_S)}{a_3(X_i - X_s) + b_3(Y_i - Y_s) + c_3(Z_i - Z_S)} \quad (1)$$

(Where (X_S, Y_S, Z_S) is the camera center; f, x_0, y_0 are the focus and principal point; a_i, b_i, c_i, $i=1,2,3$ are the elements of the rotation matrix; (X_i, Y_i, Z_i) is the vertex of the roof in the space; (x_i, y_i) is the vertex of roof on the image.), and conditional equation

$$Z_i - Z_j = \Delta Z_{i,j} = 0 \quad (2)$$

(Z_i and Z_j are the heights of two corners) are used.

However, for some difficult situation such as dense buildings and building with double edges, it is possible to produce some misidentification. A homologue line matching, based on three-view geometry constraint, is applied for automatically getting the corresponding line features in all target images.

Let P', P and P'' be the projection matrix of left, middle and right images respectively. There is a perspective transformation to P', P and P'' resulting in;

$$P' = [A \,|\, a] \qquad P = [I \,|\, 0] \qquad P'' = [B \,|\, b] \quad (3)$$

where A and B are 3 x 3 matrix, a and b are 3 x 1 vector. Let a_i and b_i be the ith ($i=1,2,3,4$) column for projection matrix P' and P''; $l'(l_1',l_2',l_3')$, $l(l_1,l_2,l_3)$ and $l''(l_1'',l_2'',l_3'')$ be corresponding image line feature for space line L in left, middle and right image; S', S and S'' be the interpretation plane for l', l and l'' (Figure 4). Then, equation (4) can be derivate.

$$l_i = l'^T(a_i b^T)l'' - l'^T(ab_i^T)l'' \quad (4)$$

Let matrix

$$T = [\, T_1 \ T_2 \ T_3 \,], \quad T_i = a_i b^T - ab_i^T \quad (5)$$

Then, the 3D tensor T should fit (Mendon, 1998):

$$l = l'^T \, T \, l'' \quad (6)$$

Based on three-view geometry constraint, all corresponding line feature on images for any space edge can be automatically determined.
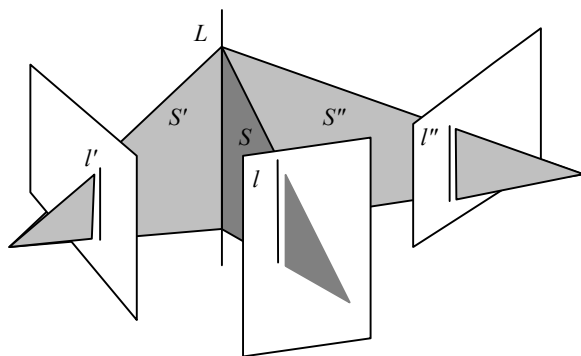
Figure 4. Three interpretation planes

3.4.2 Matching of Space Edge for Wall: After the correspondence between space edge of roof and its line feature on image is determined, building geometry model is also improved. When the building geometry model is projected onto oblique images, the correspondence between space edge of wall and its line feature on image can easily set up simply by adjusting the height of building's bottom or top.
Figure 5 shows the matched line segments with initial values shown in Figure 3. It can be seen that the matched image segments are well fitted to the real image edges.
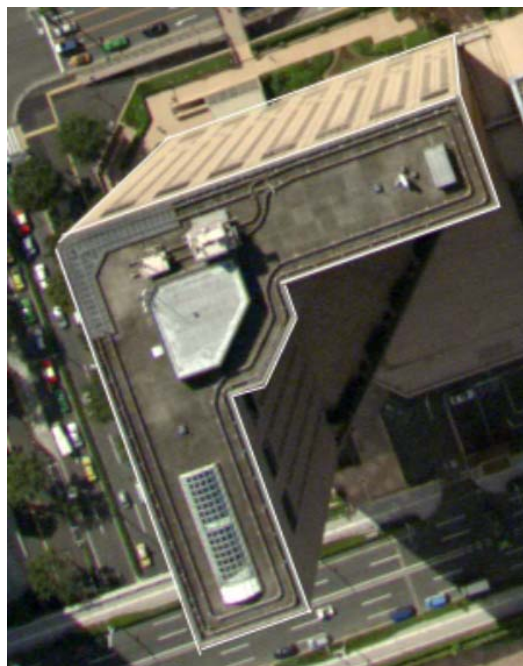

Figure 5. Matched image lines of building

3.4.3 Iterating from Coarse to Fine: After one building is processed by interaction and automatic matching technology mentioned above, the bundle adjustment will be repeated, and camera parameters of the image sequences are also updated according to bundle adjustment results. The projections of other buildings will also more closer to its actual positions in images. One example is shown in the right part of Figure.6. If there is a building, which projection on the image is with un- negligible error, the interaction and automatic matching will be performed for them. The bundle adjustment will be repeated until there is no such building, which projection on the image is with un-negligible error. The user intervention will decrease very

quickly. Finally a large part of buildings can be reconstructed automatically together with precise camera parameters.
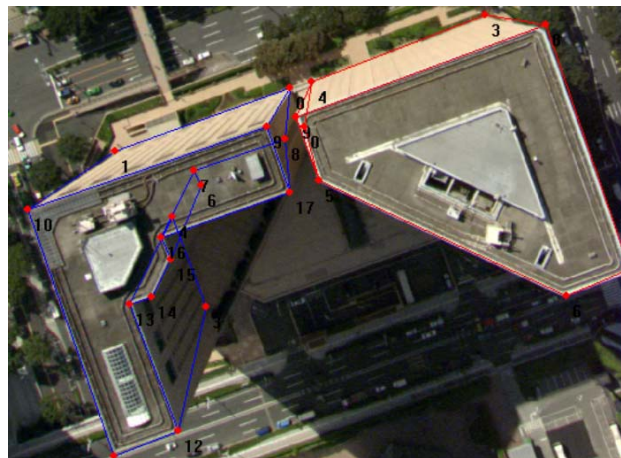

Figure 6. Result projections with coarse-to-fine strategies

Usually, the model of interested building will significantly improve after line matching and bundle adjustment. Therefore the errors in vector data can be corrected considerably, as shown in the left part of Figure 6, compared with Figure 2 and Figure 3.

4. TEXTURE AND 3D VISUALIZATION

Since one of the most important applications of virtual city model is the generation of realistic visualizations, a proper representation of geometry and texture has to be provided for each building and ground.

4.1 Texture of Building

When all buildings are successfully reconstructed, the precise camera parameters of each image can also be acquired. This information can provide a simple back-projection of the reconstructed 3D building to the aerial image. The corresponding image patch for each visible plane of a building is available and can be used for texture mapping. The geometry resolution of image patches of walls from level images is often lower than oblique one, so the best texture of visible plane is auto-selected and mapped. Texture for the roof is mapped from the level images. Usually, it is taken from the middle image among all the target images. Texture for the wall is mapped from the oblique images. When a wall appears in multiple images the image with maxmium projected area will provide the corresponding texture. Image textures are rectified with image processing techniques, as shown in Figure 7.

4.2 Texture of Ground

For visualization of the ground, the orthoimage of the relative region is needed. Firstly, digital surface model (DSM) of the whole region is generated automatically by filtering the LIDAR data. Then, DSM, precise camera parameters and level image sequence are used to generate orthoimage by digital ortho-rectification. Figure 8 is a part of the orthoimage.
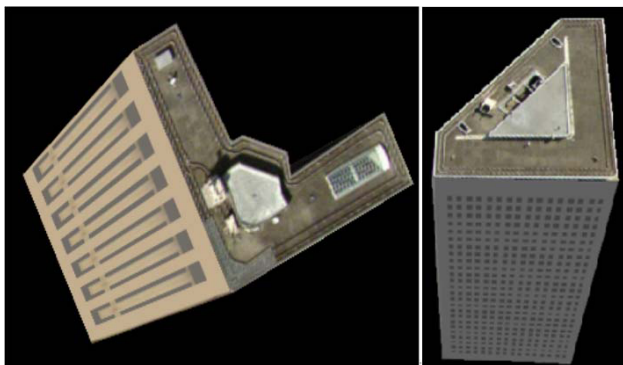
Figure 7. Texture and 3D visual model of two buildings

Figure 8 A part of the orthoimage

Reconstructed buildings are merged into textured ground generated from DSM and orthoimages. These 3D models are visualized realistically as shown in Figure 9. Figure 10 is the partly flying view of one crossway 3D scene along the road.

Figure 9. 3D city model visualization

5. CONCLUSION

An effective approach by the way of coarse-to-fine 3D city model generation is proposed. The approach is fully based on the digital photogrammetric theories. Three image sequences acquired from a digital video camera on the helicopter, coarse 2D vector data of buildings and LIDAR data are used as source of information. Automatic aerial triangulation technique of high-overlapping image sequences is used to generate initial values of camera parameters. The correspondence between space edge of buildings and its line segment in image sequences is determined with a coarse-to-fine strategy. Hybrid point-line photogrammetric technique is used for accurate reconstruction of buildings. Reconstructed buildings with fine textures superimposed on DSM with orthoimage are visualized realistically. The experiments of 3D city model generation shows that the proposed approach has a promising future in both efficiency and effect.

Figure 10. A partly flying view of one crossway 3D scene

There are often trees on both sides of roads. Their 3D models should be reconstructed also. The trees occlude road on level images and walls on oblique images. For realistic visualization, tree occlusions and cars on the roads must be removed automatically. But transportation marks on roads should be remained. 2D vector of roads can be projected onto images and corresponding image patches should be extracted and processed automatically. All of the items mentioned above will be the focus of the future work.

References

Ackermann F. 1983, High precision digital image correlation. *39th Photogrammetry Week, University of Stuttgart*.

Baillard, C., Zisserman, A., 1999. Automatic Reconstruction of Piecewise Planar Models from Multiple Views. *IEEE Conference on Computer Vision and Pattern Recognition, pp. 559-565*.

Brenner, C., 1999. Interactive Modeling Tools for 3D Building Reconstruction. *In: D. Fritsch & R. Spiller (Eds), Photogrammetric Week'99, pp. 23-34, Wichmann Verlag, Heidelberg.*

Brenner, C., Haala, N. and Fritsch, D, 2001. Towards fully automated 3D city model generation. *In Automatic Extraction of Man-Made Objects from Aerial and Space Images III. 2001.*

Forstner W., 1982. On the Geometric precision of digital correlation, *IAPRS CommIII Vol 24, 176-189.*

Forstner, W., 1999. 3D-City Models: Automatic and Semiautomatic Acquisition Methods. *In: D. Fritsch & R. Spiller (Eds), Photogrammetric Week'99, pp. 291-303, Wichmann Verlag, Heidelberg.*

Gruen A., 1985. Adaptive least square coorelation: a powerful image matching technique, *South African Journal of Photogrammtry and Remote Sensing.,(14)3:175-187*

Grun, A., Wang, X., 1998. CC Modeler: A Topology Generator for 3-D City Models. *ISPRS Commission IV Symposium on "GIS - Between vision and application", Stuttgart/Germany, IAPRS. Vol. 32, Part 4, pp. 188-196.*

Henricsson, O., Baltsavias, E., 1997. 3-D Building Reconstruction with ARUBA: A Qualitative and Quatitative Evaluation. *In: Grun, A., Baltsavias, E., Henricsson, O. (Eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), pp. 65-76.*

Heuvel, F.A., 2000. Trends in CAD-based Photogrammetric Measurement. *International Archives of Photogrammetry and Remote Sensing, vol XXXIII, part B 5/2, commission V, ISSN 0256-1840, pp. 852-863.*

Masafumi, N., Ryosuke S., 2001. Study on Making City Model with Image and Laser Range Data. *22nd Asian Conference on Remote Sensing, 5-9, November 2001, Singapore, Vol. 2, pp. 1108-1112.*

Mendon, P.R.S. and R. Cipolla.,1998. Analysis and computation of an affine trifocal tensor. *Proc. British Machine Vision Conference, Southampton, pp.125-133.*

Hu, X., Z. Zhang, J. zhang, 2000. An approach of semiautomatic road extraction from aerial image based on template matching and neural network. *International Archives of Photogrammetry and Remote Sensing Vol XXXIII Part B3: 2000,Amsterdam: 994-999.*

Zhang, Z; X. Hu, 2001. Object-Space Based Interactive Extraction of Manmade Object from Aerial Images. *Proceedings of the Society of SPIE, Vol. 45520-8194-4280-1, p.274-279.*