

KNOWLEDGE-BASED TOPOLOGICAL RECONSTRUCTION FOR BUILDING FAÇADE SURFACE PATCHES

Yixiang Tian^{a,b,*}, Qing Zhu^b, Markus Gerke^a, George Vosselman^a

^a International Institute for Geo-Information Science and Earth Observation (ITC), Hengelosestraat 99, P.O.Box 6, 7500AA, Enschede, the Netherlands - (ytian, gerke, vosselman)@itc.nl

^b State Key Lab of Information Engineering in Surveying Mapping and Remote Sensing, Wuhan University, 129 Luo Yu Road, Wuhan, Hubei, 430079, P.R. China – zhuq66@263.net

KEY WORDS: topological, façade, reconstruction, video image sequence

ABSTRACT:

3D city models constructed from ground based data are becoming an interesting and challenging problem as they present the realistic facades, which contain more details than the models constructed from aerial data. Such kind of information is interesting for quite a lot of applications. This paper presents a new method for connecting building façade surface patches that generated from video image sequence, which integrates building structure knowledge into reconstruction. Therefore reasonable and correct topological relationship can be built up between them even when some surface patches are not observed or wrongly detected. The results show our method correctly set up topological relationship between generated surface patches and getting reasonable structure model about area with occlusions.

1. INTRODUCTION

Nowadays, reconstructing 3D city models has become an interesting and challenging problem in computer vision and photogrammetry. This is because of the increasing demand for 3D models from several society fields such as urban planning, architectural design, emergency response and virtual tourism. Commonly, ground based object extraction has mainly relied on manual operations with the support of some commercial 3D modeling packages such as 3D Studio Max, and finally the 3D model is textured by manually selecting certain parts from images. Due to the huge number of urban objects in a city and variety of shapes, manual reconstruction of a city is rather time-consuming and expensive procedure (Brenner, 2005). Automated methods for reliable and accurate 3D reconstruction of manufactured objects are essential to many users and providers of 3D city data. Many processes have been reported in the areas of semi-automated and fully automated reconstruction (Werner and Zisserman, 2002; Dick et al., 2004; Mayer, 2008; Remondino et al., 2008). However, they are not always success in practical applications.

In general, there are two remote sensing techniques for 3D reconstruction of real objects and scenes that are commonly used. One is based on active range data (e.g., laser scanning), and the other is based on camera or video images. Active range based modeling methods directly capture the 3D geometric information of an object. They provide a highly detailed and accurate representation of shapes. On the other hand, they can be cumbersome and expensive. Image-based 3D modeling generally requires some user's interaction in the different steps of the modeling pipeline. Because recovering reliable three dimensional information from two dimensional images is still a problem (Remondino et al., 2008). However, economic and flexible data acquisition procedures together with the automatic

structure from motion approach are attractive advantages of using video image sequence as the source data for object reconstruction. In recent years, there has been intensive research activity on the reconstruction of 3D objects and scenes from image sequences or image sets, especially in the field of computer vision (Mayer and Reznik, 2007; Cornelis et al., 2008; Pollefeys et al., 2008). (Cornelis et al., 2008) presented a system for real-time city modeling that employs a simple model for geometry. More common way is using textured polygonal meshes to present the objects. As a majority of buildings in existence nowadays satisfy the assumption that they can geometrically be modelled as an ensemble of planar polygonal surface patches, using polyhedral models seem to be a relatively simple and efficient way to present building structures (Werner and Zisserman, 2002). Such representations that building models with detailed roof structures and planar facades are for example sufficient for simulations or visualizations at small or medium scale.

However, from video image sequence that captured from monocular camera even other ground based images, there are always some edges or surface patches cannot be observed. And occlusions of buildings or buildings parts by themselves or other objects in front of them cause failure in complete surface patch generation. Therefore, only the actually observed/detected features are connected if using a data-driven method only. However this may not fit the actual situation. At the other hand, the enormous variations in the structure and shape of the building facades prevent to use too tight constraints to recover the structure.

In this paper the knowledge of building structure is integrated into reconstruction step. The integration of knowledge does not complicate the processing rather it simplifies the reconstruction process. The term knowledge is widely used in many image

* Corresponding author

analysis methods and it may describe any kind of information (Baltsavias, 2004). In this paper, knowledge is rules or constraints that are retrieved from a general building façade structure. Topological properties are not metrical, but concern such things as connectivity and dimensional continuity. Such character makes them useful to be considered in building modeling (Ameri and Fritsch, 2000; Heuel and Förstner, 2001). So, our building models contain both geometric and topological information and our emphasis is on quickly recovering the shape of building.

The paper is organized as follows: Section two describes the previous work of this research. In section three, the method to set up topological relationship between building façade surface patches are introduced. Section four presents and discusses some building reconstruction results. Conclusions are presented in section five.

2. PREVIOUS WORK

The method for extracting point and edge features from video image sequence is presented in (Tian et al., 2008). A brief introduction is given here for it's the basic of our work and the feature points and edges are used in the following steps. Starting from 3D points tracked from video image sequence, point accuracy is analyzed first to obtain reliable matched points. Only edges near the reliable matched points are considered as edge candidates and all the edge candidates are used to estimate a 3D edge. Based on the estimated variance factor, only good 3D edge estimation are accepted, which ensure the accurate position of matched 3D edges. In order to introduce more constraints for reconstruction and fill the gaps in 3D point clouds, 3D edges are also used as primitives for reconstruction.

An automatic approach to generate surface patch is presented in (Tian et al., 2009). This step has strong relevance with building topological reconstruction. Our method to get the surface patches can be divided into four steps. First, plane hypotheses are formulated from cues based on point-cloud segmentation and on some of the 3D edges derived beforehand. The hypotheses then are verified by incorporating unused 3D edges. Afterwards, plane parameters are obtained by all the edges and points in the plane. The one with the least residual RMS is chosen as the best fit. The final step is to define the surface patches' outline. In our approach, walls and roofs are main structures of buildings. Extrusions, like windows and doors, which are attached to the façade planes, are not considered in the surface patch generation step. Figure 1 shows the surface patch generation result.

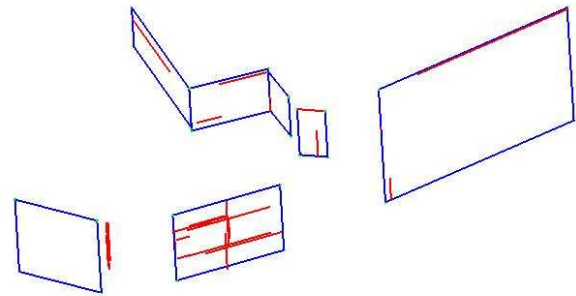


Figure 1. Surface generation result, 2D view (above), 3D side view (down), reliable points (dot), extracted edges (finite line), surface patches (polygon)

The above rules and processing steps can reasonably group extracted sparse 3D points and edges. Although the reconstructed surface patches present the building façade geometry, three aspects need to be mentioned. First, if there is no point or edge extracted from a surface patch, the pure geometric reconstruction fails to determine it. Secondly, small surface patches appear between surface patches which should be adjacent. Third, the outlines still need to be modified. Figure 1 also illustrates these aspects.

3. TOPOLOGICAL RECONSTRUCTION

This section presents the method for connecting building façade surface patches being generated from a video image sequence, integrating building structure knowledge into reconstruction. From the structural information of building we specify our assumption about building models and divide knowledge into two different kinds, the obligatory one and the preferential one. Based on the obligatory knowledge, building types are restricting to some extent as this method only aims at ordinary buildings not specific architectures. The knowledge applied for this purpose is:

- All the building facades are planner.
- Walls are vertical.
- Roof intersect with walls

As we known, there are many different kinds of buildings in the world. Some of them will be neglected in this approach if they are not satisfying above rules.

During finding the adjacent surface patch and further making model hypothesis based on two adjacent surface patches, the preferential knowledge provides essential guidance when there is no other evidence:

- Building ground plane mostly has rectangular angles.
- Structural regularity, often as result of economical, manufacturing, functional, or aesthetic considerations, results in simple models can describe basic building models.
- Repeating structures are also an essential component in architecture design, such as windows of one building are similar.

Our method for topological reconstruction can be divided into three steps. First, surface patch neighbourhood relationship are set up. The local model hypotheses then are made based on adjacent surface patches. Afterwards, all local models are

connected to form a whole building model. The details are presented in the following.

3.1 Searching the neighbourhood surface patches

There are two kinds of neighbourhood relation between surface patches. One is the two surface patches can form a volume, such as surface 1 and 2 in figure 2. The other one is one surface patch is attached to another, e.g., surface 2 is attached to surface 3 in figure 2. These two types can be separated by the location of intersecting line and surface patches.

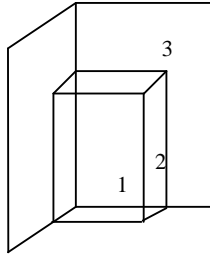


Figure 2. Surface patches' neighbourhood relation

Obviously, if two surface patches contain the same edge, they must have some relation. Due to the imperfect surface patch generation result, the boundaries of extracted surface patches are not always consistent with the actual case. Some surface patches need to be extended or subscribed into small parts. Most of time, the adjacent surface patches are searched along boundary edges that formed by two boundary points.

If two surface patches are adjacent, as shown in figure 2, the projections of surface patches to their intersecting line must overlap and the intersecting line is perpendicular to the normal vectors of surface patches. So for two intersecting surface patches, the overlap for their projections on intersecting line is compute first. The angle between object boundary edge and other surface patches is one parameter used to evaluate the found surface patch. As there are usually some parallel surface patches in building façade, the distance between object boundary edge and intersecting line is another evaluator that needs to be considered. So the one that with smallest distance and is almost perpendicular to the object boundary edge is considered as the adjacent surface from that boundary edge.

3.2 Constructing local model

This step only considers surface patches with the first kind of neighbourhood relation.

3.2.1 Surface coherence

If two surface patches form a volume, as we defined as the local model, one important thing need to be considered is surface coherence that means the normal vectors of them should be consistent to point interior or exterior to the volume. As shown in figure 3, two adjacent surface patches are coherent along their common edge if:

$$[s, v_1, n_1] \cdot [s, v_2, n_2] < 0 \quad (1)$$

Where $[a, b, c]$ stand for the triple vector product.

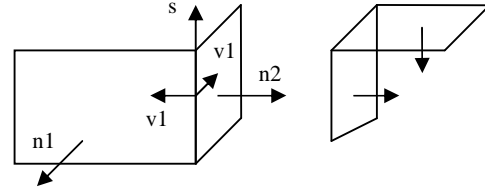


Figure 3. Coherent adjacent surfaces

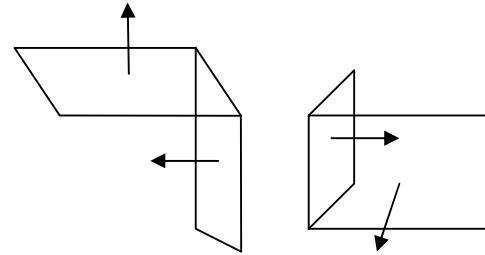


Figure 4. Non coherent adjacent surfaces

Figure 4 shows some counter examples. If two nearby surface patches are not coherent, one of them needs to be modified and a new surface patch hypothesis can be made based on them to form a suitable local model.

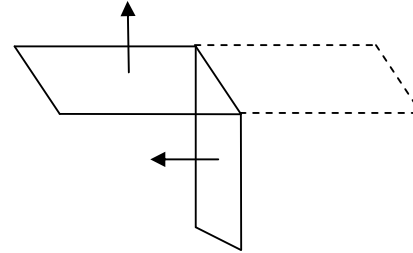


Figure 5. New surface patch hypothesis

Figure 5 shows one example how to make a new surface patch hypothesis. We extend one surface patch versa the common edge as the preferential knowledge shows nearby things are similar compare to far away ones. The new surface patch and the unmodified surface patch must coherent and their normal vectors should point to exterior of the volume.

3.2.2 Surface patch verification

These new surface patch hypotheses need to be verified by intensity similarity over the image sequence as we didn't find evidence from extracted points and edges in those areas. If a surface patch is visible in the images, we can get a high cross-correlation result for the points on it between any two images. The hypothesis is accepted or rejected according to the average cross-correlation value for points in valid images.

In more detail, given the plane \mathcal{P} , there is a homographic represented by 3×3 matrix H_i between the first and i th frame, so that corresponding points are mapped as

$$x_i = H_i x_0 \quad (2)$$

Where x_0 and x_i are image points represented by homogeneous 3-vectors.

The homography matrix is obtained from 3×4 camera projection matrices for each frame. For example, if the projection matrices for the first and i th frame are $P_0 = [I | 0]$ and $P_i = [A | a]$, and a plane defined by $\pi^T X = 0$ with $\pi = (v^T, 1)^T$, then the homography induced by the plane is:

$$H_i = A - av^T \quad (3)$$

In our process, we consider the first frame that the surface patch is visible as the reference frame and compute the cross-correlation between the reference frame and frames within visibility frame range. The homography for $x_i = H_{ri}x_r$ is:

$$H_{ri} = H_i H_r^T \quad (4)$$

Where i presents i th frame,
 r presents reference frame

The points of interest are endpoints of 2d edge extraction result within the projected image area. If the number of points is too few, some points are chosen regularly in that image area. The similarity score for the average cross-correlation value for points in the valid images is:

$$sim = \sum_{frame \in valid} \int_{POI} Cor^2(x, H_{ri}x) \quad (5)$$

If the similarity score is higher than the threshold (0.8 for experience from images), the hypothesis is accepted. Otherwise, the hypothesis is rejected and plane sweeping method is used to find a more reliable surface patch. Based on the common edge, the optimal angle is computed by searching for the maximum of function sim over a range $[-\frac{\pi}{6}, \frac{\pi}{6}]$ with 1° each time.

3.2.3 Basic local model

As the video is captured from ground, only part of roof structure can be observed. Compare with building façade, there may be less features extracted from roof, due to the roof part is simpler than walls it connects with. Therefore, in our approach roof surface patches are considered before wall surface patches. Following the sequence of topological connection, if a roof contains the same edge with a wall is dealt first. Then we will search other walls connecting with it if there are un-dealt boundary edges after local model generation. The steps for two adjacent walls are more or less similar. For those roofs that can not be observed from ground, we assume they are horizontal.

After two coherent adjacent surface patches are found, not only their topological relation can be built, but also a local model is constructed based on them. During this step, surface patches are searched to support the hypothesis. The searching step is similar to the above step that searching the neighbourhood surface patches. But here are two boundary edges should be considered. If three surface patches that form a local model are all found, the local model is defined by them. Otherwise, we chose simple block types, based on building structure defined before, as shown in figure 6 to fit them. During this fitting step, if the new

surface patch can not be observed from image sequence, we also accept the hypothesis.

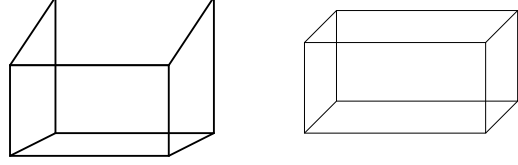


Figure 6. Example for basic local models

3.3 Connecting different local models

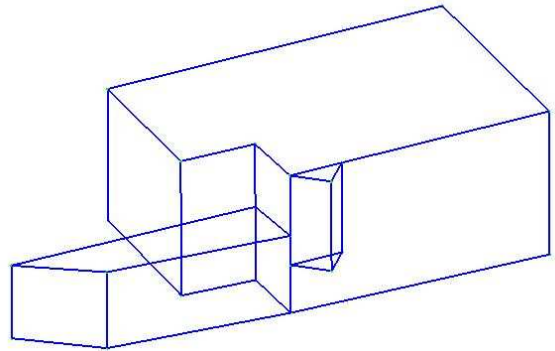
For each surface patch, a collection of point numbers following the boundary sequence is recorded. The topological relationship with other surface patches can be shown by edges from other surface patches that connected with each boundary edge.

For surface patches that are attached to other surface patches, they also have two different types. If the surface patch is located inside the local model, it is considered as the intrusion that is attached to one surface. If the surface patch is located outside the local model, it must connect two different local models.

So the whole building model is reconstructed during connecting local models one by one and presented as a collection of points, point numbers in sequence and topological connections.

4. EXPERIMENT

Figure 7 shows the reconstruction result for the building façade in figure 1. The video was captured by a hand-held SONY camera. The images have a resolution of 640×480 pixels and a frame rate of 30 frames per second. There are 189 frames in total in this case.



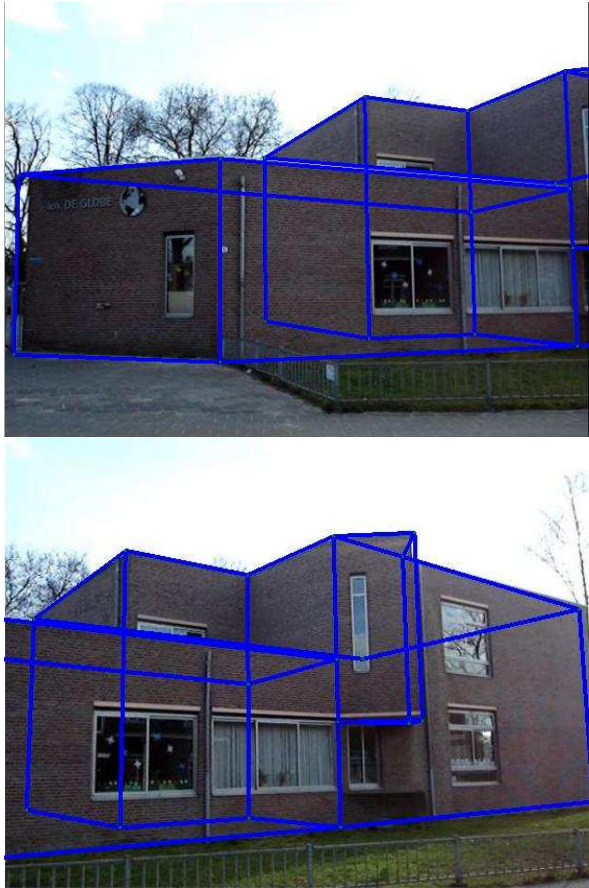


Figure 7. Reconstruction result of building façade in figure 1, 3D view (top), projection on first frame (middle), projection on last frame (down)

The building has three main parts that can be observed from images. One part only has one floor. Another part has two floors and one big extrusion is attached to it. The result shows the structure of this building is successfully recovered as we seen from the images. Some unobserved parts are generated according to assumptions, which may be different from reality. However, if there are images viable for these parts, the models can easily update to the real case.

As in this approach, the main attention is to set up topological relationship between generated surface patches and quickly recover the building shape. The location of each surface patch has small change when intersecting it with other surface patches. So the accuracy for surface patches need to be improved in the further work.

5. CONCLUSIONS

The major advantage of the strategy is the easy and natural way in which various types of knowledge are integrated into the data-driven processing. The integration of knowledge does not complicate the processing rather it simplifies the reconstruction process. And such model restrictions make the reconstruction much more robust. Meanwhile, our method does not restrict ourselves to basic structures. We can also identify complex

objects when there are observations contradict the preferential knowledge.

In fact, the method described should be more or less independent from the original image source. However, this paper presents part of our work named building structure recovering from video image sequence. So our tests are based on video image sequence. The results show our method correctly set up topological relationship between generated surface patches and getting reasonable structure model about area with occlusions.

ACKNOWLEDGEMENTS

The work described in this paper is supported by National Natural Science Foundation of China (40871212 and 40671158), and National High Technology Research and Development Program of China (2006AA12Z224).

REFERENCE

- Ameri, B. and Fritsch, D., 2000. Automatic 3D building reconstruction using plane roof structures, ASPRS Congress, Washington, DC.
- Baltsavias, E.P., 2004. Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(3-4): 129-151.
- Brenner, C., 2005. Building reconstruction from images and laser scanning. *International Journal of Applied Earth Observation and Geoinformation*, 6(3-4): 187-198.
- Cornelis, N., Leibe, B., Cornelis, K. and Van Gool, L., 2008. 3D Urban Scene Modeling Integrating Recognition and Reconstruction. *International Journal of Computer Vision*, 78(2): 121-141.
- Dick, A., Torr, P. and Cipolla, R., 2004. Modelling and Interpretation of architecture from several images *International Journal of computer vision*, 60(2): 111-134.
- Heuel, S. and Förstner, W., 2001. Topological and geometrical models for building extraction from multiple images. In: E.P. Baltsavias, A. Grün and L. VanGool (Editors), *Automatic Extraction of Man-Made Objects from Aerial and Satellite Images III*. A.A.Balkema Publishers, Ascona, Switzerland, pp. 13-24.
- Mayer, H., 2008. Object extraction in photogrammetric computer vision. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(2): 213-222.
- Mayer, H. and Reznik, S., 2007. Building facade interpretation from uncalibrated wide-baseline image sequences. *ISPRS Journal of Photogrammetry and Remote Sensing*, 61(6): 371-380.

- Pollefeys, M. et al., 2008. Detailed Real-Time Urban 3D Reconstruction from Video. *International Journal of Computer Vision*, 78(2): 143-167.
- Remondino, F., El-Hakim, S.F., Gruen, A. and Li, Z., 2008. Turning images into 3-D models. *Signal Processing Magazine, IEEE*, 25(4): 55-65.
- Tian, Y., Gerke, M., Vosselman, G. and Zhu, Q., 2008. Automatic Edge Matching Across an Image Sequence Based on Reliable Points. In: J. Chen, J. Jiang and W. Förstner (Editors), *the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science Beijing*, pp. 657-662.
- Tian, Y., Gerke, M., Vosselman, G. and Zhu, Q., 2009. Automatic Surface Patch Generation from a Video Image Sequence, *3D Geo-Information Sciences: requirements, acquisition, modeling, analysis, visualization*. Springer, Berlin, pp. 235-246.
- Werner, T. and Zisserman, A., 2002. New techniques for automated architectural reconstruction from photographs, *Seventh European Conference on Computer Vision*, pp. 541-555.