BUILDING POINT GROUPING USING VIEW-GEOMETRY RELATIONS

I-Chieh Lee[1], Shaojun He[1], Po-Lun Lai[2], Alper Yilmaz[2]
[1]Mapping and GIS Laboratory
[2]Photogrammetric Computer Vision Laboratory
Dept. of Civil & Environmental Engineering & Geodetic Science, The Ohio State University
470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210-1275
lee.3007@osu.edu
he.119@osu.edu
lai.138@osu.edu
yilmaz.15@osu.edu

ABSTRACT

Interest points on the building façade are the basic element for 3-D building modeling and texturing. Grouping these points to the same or separate buildings is a fundamental process for establishing building models and detecting building boundaries. The grouping process is generally achieved by analyzing the geometric relation and the distances between the points in the object space, which requires precise interior and exterior orientation camera parameters. In this paper, we propose a method for grouping points on buildings in the image space and focus on the close-range stereo image problems. The highlight of this method is the capability to work with un-calibrated cameras and even with images gathered from the internet, such as Google Street View^{TM}.

Every plane in an image pair, such as building façade, has a unique homography matrix, which we exploit as the basis of grouping points lying on this plane. Three major steps in this method: Fundamental matrix estimation, grouping points to planes, and assigning planes to building façades. Projective geometry between the two images can be represented by the fundamental matrix, which can be solved using point correspondences. We apply 2D Delaunay triangulation to define a set of patches in image space. Each patch that reside on the building façade has a corresponding homography matrix can be estimated from the fundamental matrix. We group the patches based on the similarity of the estimated homography matrices for the patches and forming points into planes. Buildings are usually connected and obscured in the photograph taken in a building rich circumstance. Grouping planes into building façade require two additional steps: 1) Separating buildings that are connected in the image but not in the real world, and 2) connecting two or more planes that belong to the same building façade. These resulting building façade can then be transform to 3-D using the geometric constraint or GCPs.

INTRODUCTION

Different methods have been employed to reconstruct the 3-D building model, and finding union of building façades is a common approach used by researchers. The data used in this method to unite the facades usually includes close-range images (terrestrial images) (Dick et al., 2001, Schindler and Bauer, 2003) / Video (Pollefeys et al., 2007), terrestrial LiDAR point clouds (Frueh et al., 2005, Becker and Haala, 2007, Pu and Vosselman, 2009a) or combination of these methods (Pu and Vosselman, 2009b). General procedures in these methods involve projection of the image points into 3-D, fitting models from the projected points in 3-D, and constructing surfaces in 3-D. Although working in 3-D is a straight forward approach, sophisticated and calibrated equipments are needed. Recently, the points from LiDAR scanner are getting denser, but they are usually accompanied with imaging sensors to increase accuracy of 3D point cloud. In the Mobile Mapping System (MMS), the high point spacing is too luxury to reach. We are proposing a method for grouping building points into building façade in 2-D using multiple images, which can be used to generate 3-D building models.

In this research, a new technique using multiple view geometry to group the objects is introduced. In the 3D or world coordinate frame, surfaces that belong to same building have certain characteristics that connect them together. For example, the patches of the same plane on stereo images have the same homographic transformation. However, for irregular surfaces, the complexity of the problem is too high to be solved with one single algorithm. Therefore, to simplify the problem, in this study we focus mainly on the objects which contain flat surfaces, i.e., building planes.

The first step in the proposed method is to group the points belonging to the same plane. One way to describe the relationship between two corresponding planes on two images is through planar homography in the projective space. By definition, any pixel on a plane of one image can be transformed onto the corresponding pixel in the other image using homography of the plane. This characteristic gives the basic constraint for connecting and separating planar and non-planar image segments. Moreover, two different plane in two images have different homography. However, under certain circumstance two individual planes have the same homography, for instance, two facets of two unconnected building facing the street with the same distance from the street. This can be detected by the in-continuity of homography at some location between the two facets. The next step is to group those planes which belong to the same object entity. To accomplish the task, we add one more constraint by introducing the line or edge information into the homography transform. Finally the image pixels are properly grouped into planar and non-planar segments and those planes of the same object are combined together as sub-groups.

By extending the usage of 2 or 3 images to a series of images, the resulting grouped building points and façade can be applied to varieties of applications. For example, building façade generation from MMS, near real-time change detection of buildings from MMS, mobile tunnel inspection system, mobile device localization form GIS map or satellite image, etc.

BUILDING POINT GROUPING PROCEDURE

The flow diagram of the proposed method is depicted in Figure 1. The data acquisition and fundamental matrix estimation is standard procedures in the field in photogrammetry. Our new findings rest on the last four steps. We will discuss every step in detail in the following text.

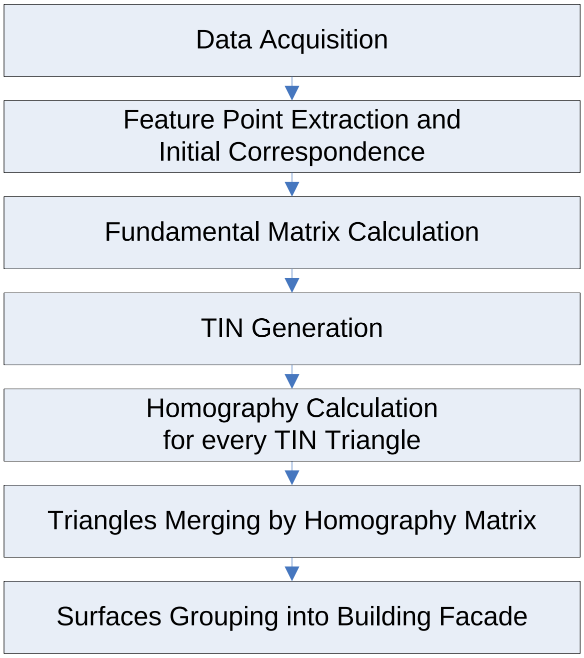


Figure 1. Flow chart of the building point grouping procedure

Data Description

A series of images of building facades can be captured by a regular digital camera. The overlap between any two consecutive images of a stereo pair is over 60 percent due to the short baseline of cameras to consecutive time instants, and the large overlap gives advantage on reliably obtaining the point correspondences. The scenes of the images mainly contain two or more separated buildings. Usually, the spatial relationship between two facades that belong to two separated buildings always falls into two categories in the image space: either being partial occlusion or being disconnected. As shown in Figure 2, 2.a is facades of two buildings, but plane A is partially occluded by plane B. Figure 2.b is an example of two separated planes of two buildings. In the final grouping result, the algorithm should have the ability to separate those facades that are adjacent on image space (occlusion) but disconnected in object space. On the other hand, the algorithm also divide those planes of the same homography but actually disconnected, e.g., the two planes A and B in figure 2.b.

(a) (b)

Figure 2. Two common cases of relationship between two facets belonging to two buildings.

Feature Points and Initial Correspondence

In this step, two methods are applied to test the effect of extraction of feature points and initial matching. First method is based on the Harris corner points. In order to cover the whole image, a sufficient number of interest points with reasonable distribution are extracted on both stereo images by the Harris detector (this requirement is critical for estimating the fundamental matrix). Then, putative correlation matching is made to establish the initial correspondences. Although the positions and the attitudes of stereo images are close, it is not guaranteed that all the interest points are correctly matched by the automatic correlation. The percentage of incorrectly matched pairs of all interest points may as large as 20%. This is reasonable for the images used in this research because the image of wall has a huge number of similar textures, such as the corners of the bricks or the frame of windows, which brings difficulty to the matching based on gray level. Meanwhile, another method, SIFT (Scale-invariant feature transform) gives another alternative to combine feature extraction and finding point correspondence, although some incorrect match still exist and need to be removed in the next step. By comparing the two sets of result, we found that there is no significant difference for the imageries used in this research.

Estimating the Fundamental Matrix

For interest point matching, the epipolar constraint is applicable in reducing the search range and eliminating the incorrect matches. Since the epipolar geometry can be decomposed from the fundamental matrix, the other reason for using fundamental matrix is to eliminate outliers as well as reducing the computation of homography matrices. Because outliers in point matching are remained from the correlation, RANSAC is employed in this algorithm to compute the correct fundamental matrix, from which the epipolar constraints can be decomposed. However, as mentioned in the previous step, there are many similar features on the image, which would also regularly distribute along the epipolar line. Consequently, a small number of incorrect matches remained in the final matching result. Fortunately, the residual error caused by the incorrect matching makes a small part in the final matching pairs and does not cause large error in calculating the fundamental matrix. Further, these incorrect matches can be eliminated in the final grouping step. Because the incorrectly matched pair has the inconsistent homographic transformation to other correctly matched pairs in the neighborhood, the inconsistency will create a hole in the grouping area. In addition, the fundamental matrix can be also used as an input parameter to calculate the homographic transformation of a plane, which is utilized for expanding the triangles in next step.

After the correspondence with most correct matches, a group of any four pairs of interest points result a homographic transformation. However, if the four pairs are selected randomly, it has two problems for the following grouping work. First, the possibility that four points resting on different planes is much higher than the points on the same plane, thus, the homographic matrix calculated from four randomly selected points is more possible unsuitable for any plane. The second problem is that the randomly selected points are more likely to form planes not existed in the object space. In order to overcome these two problems, TIN (Triangular Irregular Network) is established in the next step.

Triangulation

Triangulated Irregular Network (TIN) has been used in geometric representation in the field for a very long time, for each triangle as a basic surface element. Since one of the assumptions of a TIN is that the points used to form the TIN rest on a continuous surface, two adjacent triangles have high possibility that they are on the same plane. If the fundamental matrix of two images are known, homography can be calculated from any three-point group. For two triangles in a TIN, if they are of the same facets, same homography result from the vertices of these two triangles respectively. Therefore, Instead of computing the homography by points selected randomly, the homography can be obtained by points which have strong connections in TIN. The problem is traversed from the domain of 2-D point cloud to 2D planes. By the constraint of topography in TIN, the blindness for expanding the grouping area is reduced. In addition, the continuous expanding reassure that two separate plane of same homography would not be merged as one plane.

Merging the Triangles and Computing Homography Matrices

When the TIN is established, following steps merge the triangles to group the triangles of the same homography belonging to the same plane.

1) The triangles are ordered according to area size, the larger triangles are selected as seed triangles. For each seed triangle, the homography matrix *H* can be computed by equation (1) (Hartley and Zisserman, 2004) using the fundamental matrix and three corresponding points (x_1, x_2, x_3) of the triangle.

$$H = A - e'(M^{-1}b)^T \tag{1}$$

Where

e' is the epipole of stereo image, F is the fundamental matrix,

$A = [e']_\times F$,

$b_i = (x'_i \times (Ax_i))^T (x'_i \times e') / \|x'_i \times e'\|^2 \ (i = 1,2,3)$, and

$$M = \begin{bmatrix} x_1^T \\ x_2^T \\ x_3^T \end{bmatrix}.$$

2) Merge the adjacent triangles, namely expand the area of triangles of the same homography. The criterion for merging two triangles is: translating the vertex of the seed's neighbor triangle by the homographic matrix *H* computed in step 1 from seed points. For example, point x and point x' are conjugate points on image *I* and image *I+1* respectively. If the difference between the translated point Hx and x' is less than the threshold, say, 2 pixels, the point should be merged into the seed points, and the triangles formed by the seed points are of the same facade. Once a new point is added, update the homographic matrix *H* as well as the epipolar geometry by the points in the current seed.

3) Continuously merge triangles until the expansion ends, namely the residual of all candidate points is larger than threshold.

4) Select a new triangle as seed from the ungrouped triangles and repeat step 2 to 3. The final merging result is the merged polygons composed by the triangles having the same homography.

Grouping

Two grouping procedure have been used in this research dealing with two categories of façade relationship described previously.

The first process is to distinguish two facets that are separated. The concept we use is projective depth ρ referenced to a facade. The reason is: if a facet of a building connects to another facet of the same building, the facet should have all its points resting on one side of the connected facet. That is if facet A divides the points of another facade B into three regions with $\rho>0$, $\rho=0$, and $\rho<0$ according to equation (2) (Hartley and Zisserman, 2004), these two planes are not connected to each other. Figure 3 shows the concept of this approach. As can been seen, the facet on the background building will be divided into three regions (Plane A, B and intersecting red line). It represents that these two planes are not connected in object space. On the contrary, if they really connected to each other in image and in real world (if the left plane extends to red region) the background building facet should only be divided into two regions, ρ>0 and ρ=0 or, ρ<0 and ρ=0.

$$x' = Hx + \rho e' \tag{2}$$

The second is to determine two facets belong to one building. This step is done by finding an edge that connects the two facets and satisfies the homography of both planes. If these two planes are connected by the intersecting line within a threshold, this line would be the boundary line and these two planes are belonging to one object. Figure 3 shows an example of planes intersecting in a line. As can be seen in this figure, although line 1 satisfies homographies of facets A, B, and C, only A and B are connected and belong to the same object. That is, an additional distant constraint should be considered.

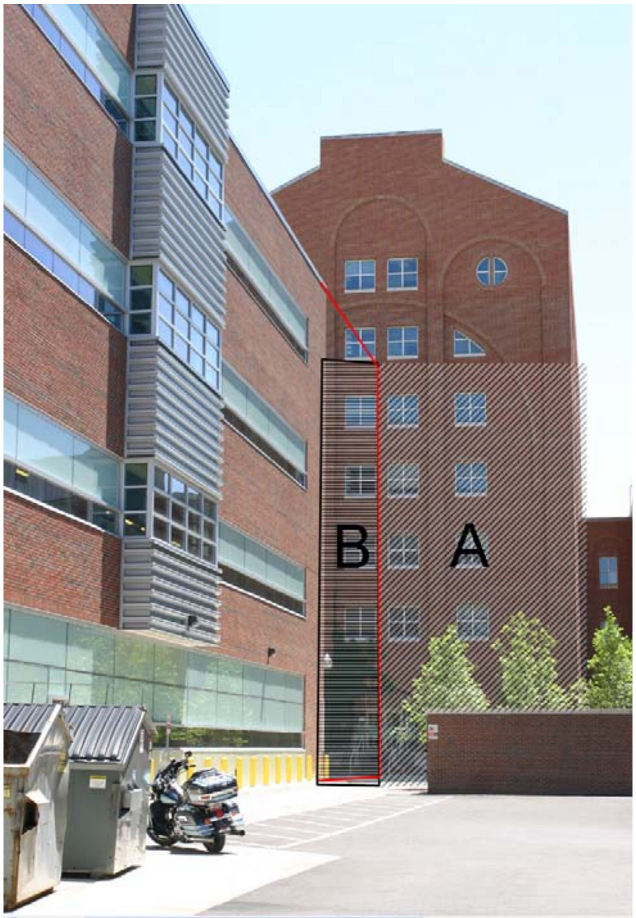


Figure 3. Separating two planes which are apart in object space but adjacent in image space

Figure 4. An example used to determine two planes are connected or not. If Line 1 fits both the homography of plane A and plane B, plane A and B are adjacent in object space.

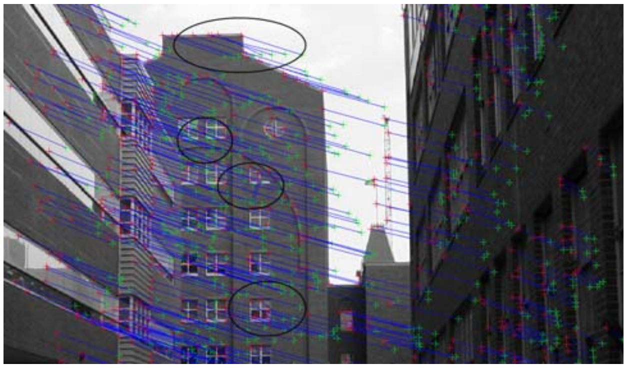


Figure 5. Image matching result

EXPERIMENTAL ANALYSIS

Matching Result with RANSAC and Fundamental Constraint

Figure 5 shows the matching result of one pair of stereo images. The red and green crosses represent the position of extracted interest points on two stereo images respectively. It is clear that most of the matched points have the consistent parallax, however some points, as marked on image, are not correctly matched. These false matches would make them having the different homography from the adjacent points, and consequently, make them of different homography from the plane they rest during the procedure of merging the triangles.

Merging the Triangles

Figure 5 shows the TIN with the matched points. Figure 7 displays a triangle merging result. The extracted triangles are marked in green, and the threshold used to group the points falling into the seed triangles is 2 pixels. It is clear that the merged area crosses the boundary between two planes, and has incorrectly merged some points on another plane. Furthermore, for those points that are incorrectly merged, the position difference between the transformed point and corresponding point is less than 2 pixels. The range of the incorrectly merged points is large, and consequently the parallax of these points is relatively small, and further makes the points insensitive to the homography. To solve this problem, the most fitted homography for the triangle should be determined. For those triangles that would be merged by more than one seed triangle, only one homographic matrix best fits the triangle. The criterion for choosing the best homographic matrix is the residual of least position.

Figure 8 is the final merging result. By employing the best homographic matrix, the points that are incorrectly merged to the plane in green in Figure 7 are correctly merged to the plane in yellow in Figure 8.

However, in Figure 8, it is obvious that holes exist in the façade formed from yellow triangles. By comparing the image matching in Figure 5, it is easy to conclude that those holes are exactly the mismatched points.

Figure 6. Generated TIN

Figure 7. An example of merging result from one seed triangle.

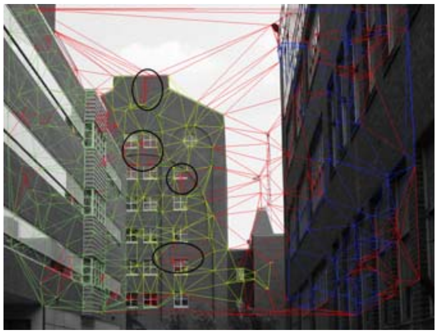


Figure 8. The merging result of the image.

Determining the Connectivity Among the Planes

Two datasets have been used in this research to test the method for determining the connectivity among facades. The first dataset is a pair of photos that contain 2 buildings, one building in the foreground, the other in the background. Figure 8 shows the images of the first dataset. The two buildings in the left part of the images are separated from each other in real world; however, they are adjacent in image. The goal of this dataset is to determine these two building facet belong to different buildings.

By taking the facet with triangles in green in Figure 8 as the reference facet to check the projective depth of each point in TIN to this facet with green triangles, the two facets with green and yellow triangles respectively connected on image can be separated. As shown in Figure 9, all the interest points are represented by three colors according to their projective depths. Points in green are in front of reference facet; blue points are on the reference façade within the threshold; the red points are at the backside of the reference facade. In another word, the points on the plane formed with yellow triangles in Figure 8 can be divided into three groups, which have different topographic relations with the facet of green triangles in Figure 8 if these facets are connected. Therefore, plane A and B should be separated.

The second dataset is used to check whether two planes belong to one building. If two facets are connected in object space, commonly there should be a line connecting these two planes and holds the homography of each plane. If such a line can be found near the adjacent boundary of these two planes, then the two planes are connected. Figure 10.a is the merging result of an image, and Figure 10.b shows the edge that fit both homography of the adjacent plane.

DISCUSSION AND CONCLUSIONS

According to the experiment result, the algorithm presented in this paper is applicable for grouping points and establishing building facades by analyzing the homography of triangle patches without knowing the image orientation. But there are still issues need to be improved.

The parallax plays a very important role in the triangle merging and determination of the boundary between two planes. The less the parallax is, the more unstable the calculated homographic matrix. If the area of boundary in image has small parallax, the merging criteria would make the area crossing the true boundary. In the future, a method to evaluate the robustness of the calculated homography is needed.

As discussed in the experiment result, the incorrect matching brings homographic inconsistency to the points that are really on the plane. Solving this problem has two directions. One is to improve the feasibility of the image matching; the other is to eliminate those false matched points according to their topographic to the merged plane. How to improve the matching quality is another key issue for a robust result.

Currently, only the points are used in the algorithm. In fact, lines, which also have implicit information as well as the points. Consequently, exploration of using lines in matching should be one the future task of this research.

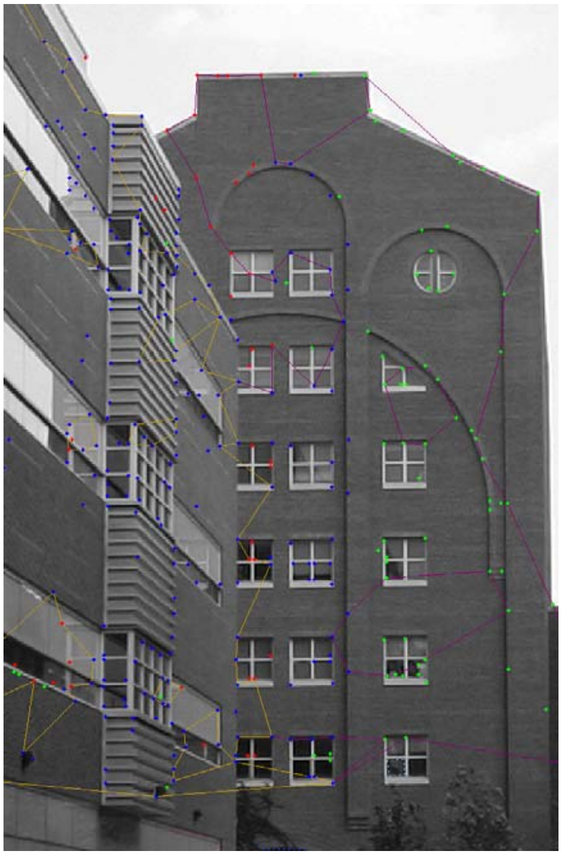
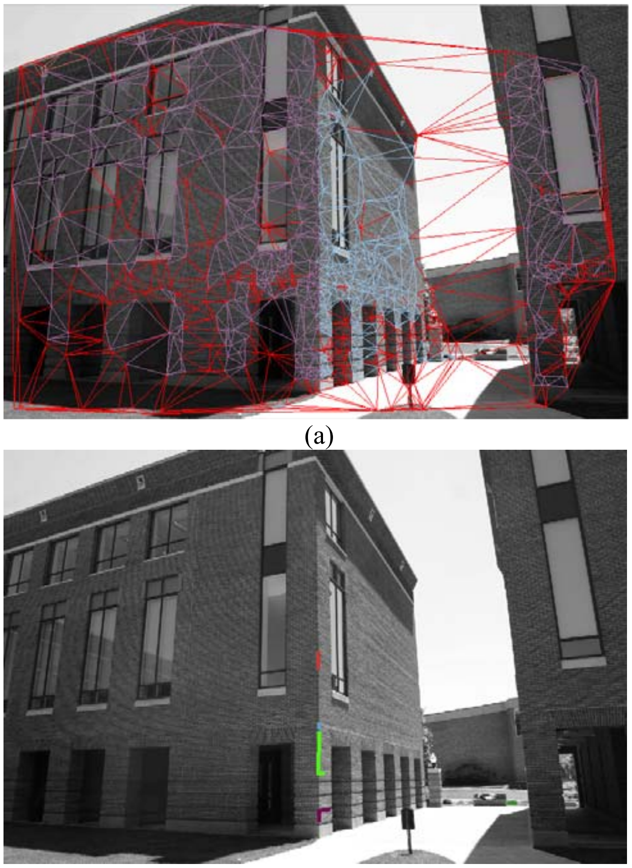


Figure 9. The grouping result (colored polygon) based on this method.



(a)

(b)

Figure 10. The connecting line located between two planes.

REFERENCES

Becker, S. and N. Haala, 2007. Combined Feature Extraction for Facade Reconstruction. In *Proceedings of the ISPRS Workshop Laser Scanning 2007 and SilviLaser 2007*, Espoo, Finland, pp. 241–247.

Dick, A., P.Torr, S. Ruffle and R. Cipolla, 2001. Combining Single View Recognition and Multiple View Stereo for Architectural Scenes, In *Proc. 8th IEEE International Conference on Computer Vision (ICCV'01)* Vancouver, Vol I, pp 268-274.

Frueh, C., S. Jain, and A. Zakhor, 2005. Data processing algorithms for generating textured 3D building facade meshes from laser scans and camera images. *International Journal of Computer Vision, 61*:159–184.

Hartley, R., and A. Zisserman, 2004. *Multiple View Geometry in Computer Vision*, Cambridge University Press.

Pollefeys, M., D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, H. Towles, 2007. Detailed real-time urban 3D reconstruction from video. *International Journal of Computer Vision, 78*(2-3):143-167.

Pu, S. and G. Vosselman, 2009a. Knowledge based reconstruction of building models from terrestrial laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing 64*(6):575-584.

Pu, S. and G. Vosselman, 2009b. Building facade reconstruction by fusing terrestrial laser points and images. *Sensors 9*(6):4525-4542.

Schindler, K. and J. Bauer, 2003. A model-based method for building reconstruction. In *First IEEE International Workshop on Higher-Level Knowledge in 3D Modeling and Motion Analysis, HLK 2003*, 2003, pp. 74–82.