3D Building Reconstruction by Map Based Generation and Evaluation of Hypotheses

Ildiko Suveg, George Vosselman
Photogrammetry and Remote Sensing
Delft University of Technology
Delft, Thijsseweg 11, 2629 JA, The Netherlands
{I.Suveg,G.Vosselman}@geo.tudelft.nl

Abstract

This paper presents a knowledge-based approach for automatic 3D reconstruction of buildings from aerial images. By combining the image analysis with information from GIS maps and specific knowledge of the buildings the complexity of the building reconstruction task can be greatly reduced. The building reconstruction process is described as a tree search in the space of possible building hypotheses. Hypotheses derived from outlines of building footprints from the map are fit against image pixel gradients. To guide the search of the tree an evaluation function based on information theory principles is defined. The proposed evaluation function defines the score of matching between a hypothesised building model and the image pixel gradients. It uses a mutual information measure and MDL criterion to select the best fit to image data in the tree search.

1 Introduction

3D building reconstruction has been an active research topic in computer vision in recent years. Most approaches have focused on the reconstruction of specific building models: rectilinear shapes [12, 14], flat roofs [7, 8] or parametric models [4]. But buildings show a much wider variety in their shape. Other approaches employ a generic roof model that assumes planar roof surfaces [3, 11, 15]. These 3D roof planes are generated by grouping the coplanar 3D lines or corners computed from the images. However, the feature extractors can fragment or miss boundary lines, due to low contrast, occlusions, and bad perspective. To overcome these problems, the image data has to be combined with other data sources, for example fusing images with scanned [10] or digital maps [6]. These approaches represent the newest trend in 3D building reconstruction.

Our strategy for 3D reconstruction of buildings combines pairs of stereo images with large-scale Geographic Information System (GIS) maps and domain knowledge as additional information sources. The 2D GIS map contains the outline of footprints of the buildings. The knowledge about the problem domain is represented by a building library containing primitive building models. Although, buildings reveal a high variability in

BMVC 2001 doi:10.5244/C.15.66

shape, even complex buildings can be generated by combining simple building models with flat, gable or hip roof.

This paper is organized as follows: Section 2 presents a brief overview of the steps involved in our method. The next section describes the generation of building hypotheses. Section 4 describes an evaluation function based on mutual information for determining the best building hypothesis. Section 5 presents some results. The conclusions and future work are discussed in the final section.

2 Method Overview

One of the important issues in the proposed method is separation of the building detection process from that of building reconstruction. The fact that the reconstruction process is focused on one building reduces the complexity of the reconstruction by a large amount. The localization of buildings in the images can be performed based on the ground plan of the buildings contained in the map. This method was described in detail in [16]. The actual building reconstruction process is formulated as a multi-level hypothesis generation and verification scheme and it is implemented as a search tree.

Because a complex building can be described as an aggregation of simple building models, the first step is the partitioning of the buildings into simple building-parts. These building parts might correspond to one of the building models defined in the building library. In the first attempt the partitioning is done using only the ground plan of a building defined in the GIS map. Each of the partitioning schemes will start up a branch in the search tree.

As basic building models we can consider a flat roof, a gable roof and a hip roof building. The approach of modelling buildings using a set of basic building models (primitives) suggests the usage of Constructive Solid Geometry (CSG) representation for building description.

Next, the tree will be expanded with a level corresponding to the different building hypotheses generated for each building-part obtained after partitioning of the ground plan. The building hypotheses can be verified by back projecting them into the images and then matching with the information extracted from the image. The matching has to define a score function that will be used to guide the search in the tree. Finally the CSG tree representing a building will be given by the best fit of the building models corresponding to the building partitions.

3 Generation of Building Hypotheses

3.1 Partitioning into Building-parts

The first step of the actual reconstruction process is the partitioning of the building in simple building parts, which might correspond to a building model defined in the building library. First, the partitioning is done using only the ground plans of a building defined in the GIS map. If the ground plan of the building is not a rectangle, then it can be divided in rectangles, called partitions. Then, a partitioning scheme can be defined as a subdivision of a building into disjoint partitions. A building can have multiple partitioning schemes (figure 1). Each of these partitioning schemes will start up a branch in the search tree.

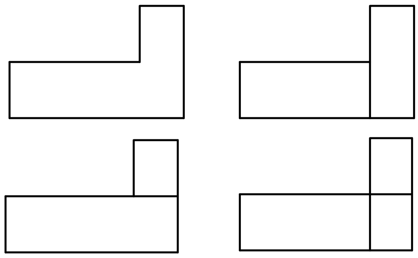


Figure 1. Ground plan of a building and possible partitioning schemes of it

To avoid a blind search method of the tree, the Minimum Description Length (MDL) principle can be used. This principle provides a means of giving higher priority to the partitioning schemes with a smaller number of rectangles.

If all the partitioning schemes are rejected by the tree search then the partitioning has to be refined using image information as well. This process will start up a new branch in the search tree and the whole process is repeated.

3.2 Generating Building-part Hypotheses

3.2.1 Library of Building Primitives

To cope with the complexity of aerial images we have to incorporate specific knowledge about buildings. Since most buildings can be describe as an aggregation of simple building types, the knowledge about the problem domain can be represented in a building library containing the simple building models. In this way a complex building can be seen as a CSG tree, where the leaf nodes contain primitive building models and the internal nodes contain boolean operations such as union, intersection, difference.

The basic building models in the building library are described by parametric models having pose and shape parameters. For instance to describe a flat roof building 6 parameters are necessary: width, length, height, x, y coordinates of the building reference point and the orientation in the xy-plane. For a gable roof an extra parameter, the height of the ridge has to be considered.

3.2.2 Parameter Estimation

The parameters of the model are estimated in a two-step method. First an approximate estimation is done based on the information from the map and 3D information extracted from images.

The x, y coordinates and the orientation of a building primitive are given by the ground plan of the building. The parameters width and length are the width and the length of the rectangle corresponding to the ground plan of the building part. The height of the building primitive is computed taking into account the heights of the reconstructed 3D corners of the building part. For a gable roof the height of the ridge is considered as the height of the reconstructed 3D top line if the top lines were detected in both images and the 3D line could be reconstructed. Otherwise, the approximate position of the projected ridge in the images can be deducted taken into account the symmetry of a gable roof. Then the 3D ridge can be reconstructed by matching these two approximate line segments. At this stage the estimation is influenced by uncertainties of the knowledge sources. The uncertainties are due to the accuracy of the GIS map, the roof extensions, and estimated height [16].

In order to handle these uncertainties, a more precise estimation of the parameters is obtained using a fitting algorithm. This algorithm fits the edges of the projected wire frame of the model to gradients of the pixels from both images simultaneously [19]. This algorithm is similar to the one described by Lowe [9].

4 Evaluation of Building Models

The 3D reconstruction of a building can be seen as a tree search. The search space for the best fit building model can be represented as a tree with the nodes of the tree representing the different building primitive hypotheses. The tree is generated incrementally by the search method.

The root node of the tree represents the initial state, where only the ground plan of the building is known. The first level of the tree contains all the possible partitioning schemes of a building. The second level contains the partitions corresponding to each partitioning scheme. Next, the search tree is expanded with a level corresponding to different building hypotheses generated for each building partition. The a priori knowledge about the building types from the data set, if it is available, can guide the process of building hypotheses generation. Consequently, the more frequently occurring building models are treated first.

There are two problems that have to be considered at the search of the tree:

- Definition of an evaluation function to guide the search to the best solution
- Definition of a stop criterion that speeds up the search by reducing the search space.

In this paper we describe the definition of an evaluation function in detail.

From the possible hypotheses of a matching between an object model and an image, one wants to select the hypothesis that maximizes some appropriate evaluation function. Therefore, an evaluation function is defined to measure for the quality of the match. Usually, an evaluation function is based on error models that describe how an image feature may differ from what the object model has predicted. Two main categories of approaches for defining evaluation functions can be distinguished. Ad hoc evaluation functions were used by Ayache [1], Beveridge [2], Grimson [5]. With this approach, components of the evaluation function are combined using trade-off parameters that are determined empirically. Other class of evaluation functions is based on statistical theory. Match quality measures are often defined using Bayesian probability theory ([13], [18]). Our evaluation function belongs to this latest category, using a mutual information based measure.

4.1 Mutual Information

To guide the search of the tree we have developed an evaluation function based on information theory. This evaluation function can be used to compare different building hypotheses in order to choose the best one from a set of building hypotheses.

The evaluation function defines the score of matching between the hypothesized building model and the images. Matching can be seen as a communication problem, where the model description $M = \{m_1, m_2, \dots\}$ is transmitted through a communication channel into the image $D = \{d_1, d_2, \dots\}$. The image data will be similar to the model data but sometimes it is corrupted due to occlusions, noise, etc. The similarity between the two descriptions can be measured by the mutual information $I(M; D)$.

The mutual information is defined as the difference between the self-information and the conditional information [20]:

$$I(m_i; d_j) = I(d_i) - I(d_j \mid m_i) \tag{1}$$

where $I(d_j) = -\log P(d_j)$ and $I(d_j \mid m_i) = -\log P(d_j \mid m_i)$

Thus, the mutual information can be written as:

$$I(m_i; d_j) = \log \frac{P(d_j \mid m_i)}{P(d_j)} \tag{2}$$

The descriptions of the model and image data depend on the level of abstraction chosen. A lot of work has been done on computing matching scores. Generally, one can distinguish between feature based and intensity based approaches. The feature-based methods require segmentation of the images before the matching process. But, usually the segmentation needs selection of a threshold. In addition, the extracted features are influenced by noise, bad contrast and occlusions in the image. To overcome these problems we do the matching between the model and the images and the evaluation of the matching at the lowest level of abstraction, namely at pixel level. The attributes dealt with at this level are gradients.

Thus, the mutual information between an image pixel and the corresponding model point is given by:

$$I(point_m; point_i) = \log \frac{P(grad_i \mid point_m)}{P(grad_i)} \tag{3}$$

Another advantage of our evaluation function is the simplicity. The distribution of the gradients at random image points and also the conditional distribution of the gradients along the projected roof edges can be determined by training.

Our evaluation function gives a positive response where points match with high confidence, a negative response where there is a clear mismatch, and zero response in the points where there is neither evidence for match nor evidence against a match.

The mutual information for a model line is found by taking the sum of the points of the line:

$$I(line_m; line_i) = \sum_{point_m \in line_m} I(point_m; point_i) \tag{4}$$

The total information for a building model in both images is given by the sum over all points on all projected model lines in all images:

$$I(M; D) = \sum_{k=1}^{2} \sum_{line_m} \sum_{point_m} \log \frac{P(grad_i \mid point_m)}{P(grad_i)} \tag{5}$$

4.2 Minimum Description Length Principle

The goal is to select the model M_i from a list of models $M = \{ M_1, M_2, \dots \}$, which best fits the image data, knowing the transformations.

If all the models have the same complexity, this goal could be achieved by choosing the model with the highest mutual information. But the mutual information between a model and the image data increases with the complexity of a model. Therefore, the mutual information between different building models and image data are not directly comparable and cannot be used as an evaluation function.

The problem can be solved using the MDL principle. This principle selects the model M_i with the shortest complete description of the data, thus the model that minimises $L(D \mid M_i) + L(M_i)$

If the code used for the description is optimal, the length of the description is equivalent to its information content.

$$L(M_i) = I(M_i) \text{ and } L(D \mid M_i) = I(D \mid M_i) \tag{6}$$

Thus, the minimum description length principle minimizes:

$$I(D \mid M_i) + I(M_i) \tag{7}$$

By using the definition of the mutual information:

$$I(D;M_i) = I(D) - I(D \mid M_i) \tag{8}$$

the formula (7) can be expressed as:

$$I(M_i) + I(D) - I(D \mid M_i) \tag{9}$$

Since $I(D)$ is constant, the expression $I(D;M_i) - I(M_i)$ has to be maximized. Therefore, it follows that the best model is given by:

$$M_{opt} : \max_i (I(M_i;D) - L(M_i)) \tag{10}$$

and the expression $Score(M_i) = I(D;M_i) - I(M_i)$ can be used as an evaluation function for the matching between building model and image data.

4.2 Relation MAP and MDL

It can be shown that the MAP (maximum a posteriori) and the MDL principle lead to the same solution. The maximum a posterior strategy selects the model M_i that maximizes the conditional probability of the model given the data D, $P(M_i \mid D)$.

By using Bayes' formula:

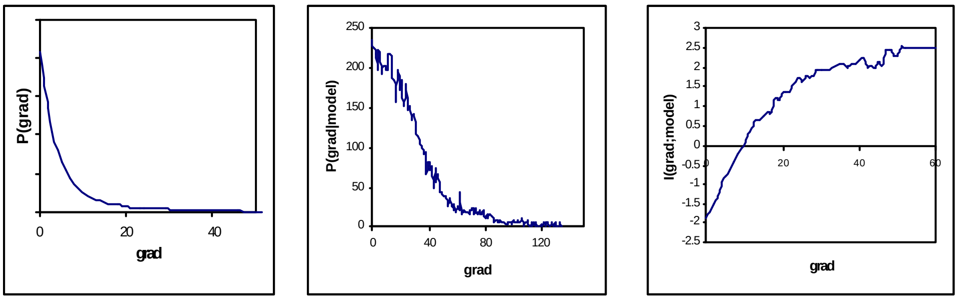
$$P(M_i \mid D) = \frac{P(D \mid M_i)P(M_i)}{P(D)} \tag{11}$$

Since $P(D)$ is constant, MAP states that $P(D \mid M_i)P(M_i)$ has to be maximised. The minimum description length principle minimizes: $I(D \mid M_i) + I(M_i)$

This can be written as:

$$\begin{aligned} I(D \mid M_i) + I(M_i) &= -\log P(D \mid M_i) - \log P(M_i) \\ &= -\log P(D \mid M_i)P(M_i) \Rightarrow \min \end{aligned} \tag{12}$$

Therefore, MDL maximises $P(D \mid M_i)P(M_i)$ like MAP.



a) b) c)

Figure 2. a) Gradient distribution $P(grad_i)$ b) Conditional probability density of the gradient $P(grad_i|point_m)$ c) Mutual information $I(point_m; point_i)$

4.3 Computation of the Gradient Distribution

In order to calculate the score function for matching between a building model and the image data as defined in (10), we need to know the a priori probabilities $P(grad_i)$ and the conditional probabilities $P(grad_i \mid point_m)$.

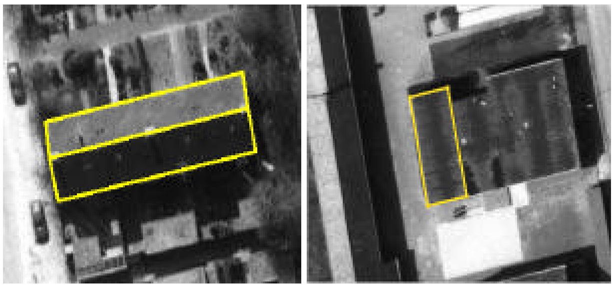
The probabilities of the gradient at random image points can be obtained directly from the images. The gradient distribution is determined as the histogram of the gradient values in the regions of the images where there is a building. The delineation of these regions in the images was described in [16]. The obtained a priori probability $P(grad_i)$ is shown in figure 2a. In [15] Sullivan also estimated the gradient distribution from histogram for identification of cars in traffic scenes.

The conditional probability density function of the gradient along the projected roof edges can be determined from training matches by analysing the probabilities of gradients in these training matches. Some image lines corresponding to model lines are selected manually. Next, the histogram of the gradient values along these lines is computed. The obtained conditional probability density function $P(grad_i \mid point_m)$ is shown in figure 2b.

Knowing these two distributions, the mutual information can be computed using (3) and this is shown in figure 2c.

5 Results

The test data consists of high-resolution aerial images. The scale of the images is 1:3000 and they are scanned at 600 dpi. Therefore, one pixel in the image corresponds to about 12.7 cm in object space. Two images with 60% overlap are used. The interior orientation parameters of the camera and also the exterior orientation parameters of the images are known. A 2D GIS map containing the ground planes of the buildings is given. In our current implementation, three hypotheses are generated corresponding to a flat roof building primitive and two gable roof primitives with different orientations. Therefore we can reconstruct only flat roof buildings, gable roof buildings or buildings formed by combining these two building types. However, the building library can be easily extended with other primitive building models. Also, we assume that the buildings have only 90^o corners. This is actually a limitation of the models described in the building library, since both flat roof building and gable roof building model require rectangular base.



Initial scores (before fitting):

$Score_{flat} = -274.5$	$Score_{flat} = -157.2$
$Score_{gable} = -325.3$	$Score_{gable} = -467.7$

Final scores (after fitting):

$Score_{flat} = -185.3$	$Score_{flat} = 251.6$
$Score_{gable} = 618.7$	$Score_{gable} = -365.2$

Figure 3. Reconstructed roof models projected back into the image

The first experiment was to generate and evaluate building hypotheses for simple buildings composed by only one building primitive. First building hypotheses derived from outlines of building footprints from the map are generated corresponding to the building models from the building library. Next, the building hypotheses are fit to the image data. The scores computed for matching the hypotheses against the images are used to choose the best model. The resultant building models projected back into one of the images are presented in figure 3.

Next, we tested our approach on complex buildings (figure 4). First, the partitioning of the building into building primitives based on the ground plan was performed. Then, for each resultant building primitive, hypotheses are generated. Evaluating the partition schemes we found that the partitions presented in figure 4 are the best ones.

The results from the proposed approach are encouraging. The method worked well even in difficult conditions (figure 4a, one of the sides of the building is shadowed), where feature based approaches would have failed. Problems can appear in case of building parts with very small size. To overcome this problem, constraints, which describe geometric relationships between building primitives, were incorporated in the fitting algorithm (figure 4c).

6 Conclusions

A knowledge-based approach for automatic 3D reconstruction of buildings from aerial images was presented. The 3D reconstruction of buildings was described as a search tree. The generation of the search tree containing the multiple consistent building primitive hypotheses was described. To guide the search of the tree an evaluation function using mutual information was defined. The mutual information is computed directly from the image gradients. This evaluation function allows comparison of different building hypotheses. The robustness of this evaluation function is assured by

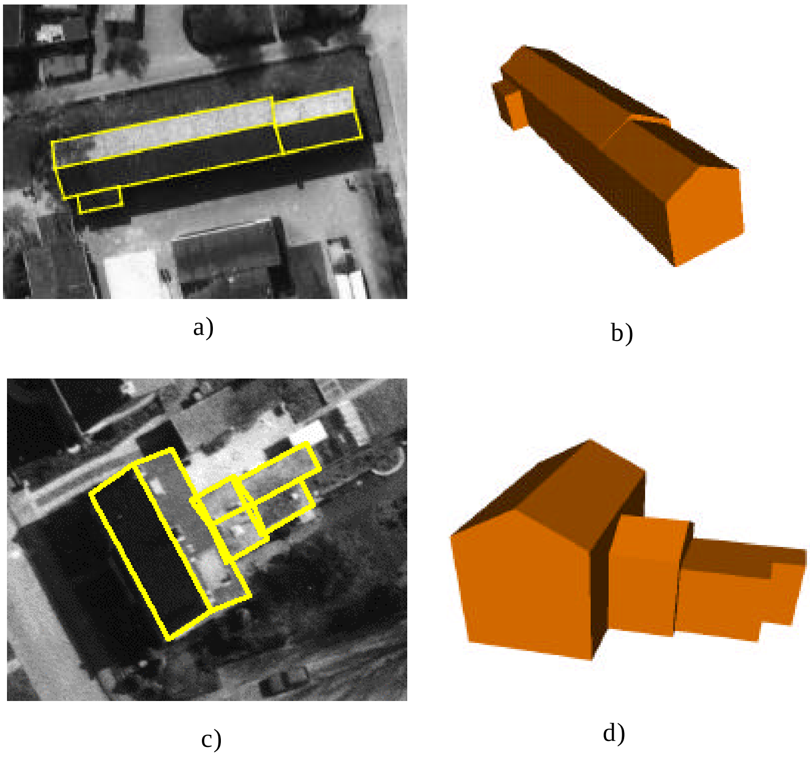


Figure 4. Reconstruction of complex buildings. a, c) Partitioning schemes with the highest scores projected back into images. b, d) Vrml models corresponding to these partitioning schemes

working directly with the image gradients. In this way the problems encountered in feature based measures are avoided.

Future work will be directed towards the definition of the expected amount of mutual information needed for a reliable matching. This will to be used as stop criterion for the tree search in order to speed up the search. The search can be further speeded up by sorting the building hypotheses based on a priori knowledge about the building types from the data set.

Next, the resultant building hypothesis described as a CSG tree has to be verified by back projecting it into the images and fitting it to the image data. The fitting algorithm will consider only the visible edges of the wire-frame model, which can be determined by a hidden line analysis algorithm.

References

[1] N.Ayache, O.D.Faugeras, *HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects,* IEEE PAMI, vol.8, no.1, pp. 44-54, Jan 1986
[2] J. Beveridge, R.Weiss, E.Riseman, *Optimisation of 2D Model Matching,* Proc. Image Understanding Workshop, pp. 815-830, 1989

[3] F. Bignone, O. Henricsson, P. Fua, M. Stricker, *Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery*, Computer Vision - ECCV'96, Springer Verlag, vol.1, pp. 85-96, 1996

[4] A. Fischer, T.H. Kolbe, F. Lang, A.B. Cremers, W. Förstner, L. Plümer, V. Steinhage, *Extracting Buildings from Aerial Images Using Hierarchical Aggregation in 2D and 3D*, Computer Vision and Image Understanding, vol. 72, no. 2, Nov. 1998

[5] W.E.L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints*, MIT Press, 1990

[6] N. Haala, M. Hahn, *Data fusion for the detection and reconstruction of buildings*, Automatic Extraction of Man-Made Objects from Aerial Images, pp.211-220, Birkhauser Verlag, Basel, 1995

[7] C. Jaynes, M. Marengoni, A. Hanson, E.Riseman, H.Schultz, *Knowledge-Directed Reconstruction from Multiple Aerial Images*, Proc. ARPA Image Understanding Workshop, New Orleans, LA, Vol. II, pp. 971-976, 1997

[8] C. Lin, A. Huertas, R. Nevatia, *Detection of Buildings Using Perceptual Grouping and Shadows*, Proc. CVPR, IEEE Computer Society Press, pp. 62-69, 1994

[9] D.G. Lowe, *Fitting Parameterized Three-dimensional Models to Images*, IEEE PAMI, vol.13, no.5, pp. 441-450, May 1991

[10] H.Maitre, I. Bloch, H. Moissinac, C. Gouinaud, *Cooperative Use of Aerial Images and Maps for the Interpretation of Urban Scenes*, Automatic Extraction of Man-Made Objects from Aerial Images, pp.297-306, Birkhauser Verlag, Basel, 1995

[11] T.Moons, D.Frere, J.Vandekerckhove, L.Gool, *Automatic Modeling and 3D Reconstruction of Urban House Roofs from High Resolution Aerial Imagery*, ECCV98, pp. 410-425, 1998

[12] S. Noronha, R. Nevatia, *Detection and Description of Buildings from Multiple Images*, Proc. CVPR, pp.588-594, 1997

[13] A.R. Pope, D.G. Lowe, *Modeling Positional Uncertainty in Object Recognition*, University of British Columbia, Computer Science, TR-94-32, Nov. 1994

[14] M. Roux, D.M. McKeown, *Feature Matching for Building Extraction from Multiple Views*, Proc. ARPA IUW, Menterey, CA, pp.331-339, 1994

[15] C. Schmid, A. Zisserman, *Automatic Line Matching Across Views*, Proc. CVPR, pp. 666-671, 1997

[16] I. Suveg, G. Vosselman, *3D reconstruction of Building Models*, Int. Archives of Photogrammetry and Remote Sensing, vol XXXIII, part B2, pp 538-545, 2000

[17] G. D. Sullivan, *Visual Interpretation of Known Objects in Constrained Scenes*, Phil. Trans. Royal Society London, B, 337, pp 361-370, 1992

[18] I. Weiss, *Statistical Object Recognition*, Ph.D. dissertation, Cambridge MIT, 1993

[19] G. Vosselman, H. Veldhuis, *Mapping by Dragging and Fitting of Wire-Frame Models*, Photogrammetric Engineering & Remote Sensing, vol. 65, no.7, July, pp. 769-776, 1999

[20] G. Vosselman, *Relational Matching*, Lecture Notes in Computer Science 628, Springer-Verlag, 1992