Automatic 3D Building Reconstruction

Ildiko Suveg, George Vosselman

Delft University of Technology, The Netherlands
Section of Photogrammetry and Remote Sensing
{I.Suveg, G.Vosselman}@geo.tudelft.nl

ABSTRACT

We present a knowledge-based system for automatic 3D building reconstruction from aerial images. Our approach relies on combining pairs of stereo images with 2D GIS map and domain knowledge. Since most buildings can be described as aggregation of simple building types, the domain knowledge is represented in a building library containing building primitives (flat, gable, and hip roof building). The approach of modeling buildings using a set of basic building models suggests the usage of Constructive Solid Geometry representation for building description.
The building reconstruction process is formulated as a hypothesis generation and verification scheme. It starts with the partitioning of a building in simple building parts based on the ground plan defined in the map. For each building partition different building hypotheses are generated corresponding to the building primitives defined in the building library. The evaluation of the generated building models is based on the formulation of the mutual information between the model and the images. The CSG tree representing a building is given by the best fit of the building models corresponding to the building partitions.
We used this method to reconstruct buildings in suburban and urban scenes. The method worked well even in difficult conditions (noise, shadow).

Keywords: 3D reconstruction, minimum description length, model fitting, evaluation, mutual information

1. INTRODUCTION

3D reconstruction of buildings has been an active research topic in computer vision in recent years. 3D reconstruction of buildings from aerial images is becoming of increasing practical importance. Traditional applications are those of cartography and photo interpretation. Newer applications include urban planning, construction, environment, communication, transportation, energy and property management, tourism, and virtual tours of cities.
Photogrammetric methods are well established but show inefficiencies due to the extensive amount of data. Manual 3D processing of aerial images is very time consuming and requires highly qualified personnel and expensive instruments. Therefore, speeding up this process by automatic or semiautomatic procedures has become a necessity. The current state of automation in the reconstruction of buildings from aerial images is still surprisingly low. A lot of algorithms and systems have been proposed towards this problem. However, a versatile solution to the automatic reconstruction has not been found yet, with only partial solutions and limited success in constrained environments being the state of art.
The difficulty in obtaining a general solution to this problem can be attributed to the complexity of the reconstruction itself, as it involves processing at different levels: low level processing (feature extraction), middle level processing (representation and description of building models) and high level processing (matching and reasoning). The success of a reconstruction system depends upon the succeeding at all these levels and in combining these levels of processing.
Most approaches have focused on the reconstruction of specific building models: rectilinear shapes [13, 15], flat roofs [8, 9] 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, 12, 16]. These 3D roof planes are generated by grouping the coplanar 3D lines or corners computed from the images. Nevertheless, 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 [11] or digital map [7]. These approaches represent the newest trend in 3D building reconstruction.

Three-Dimensional Image Capture and Applications V, Brian D. Corner, Roy P. Pargas,
Joseph H. Nurre, Editors, Proceedings of SPIE Vol. 4661 (2002) © 2002 SPIE · 0277-786X/02/$15.00

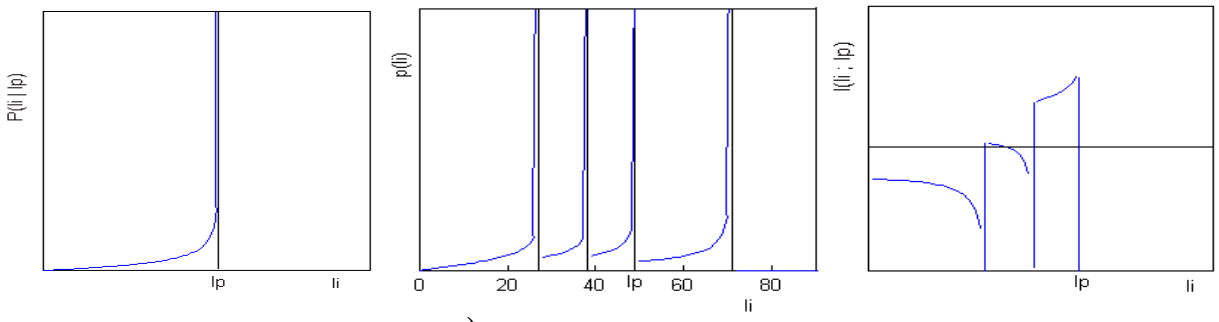
Our strategy for 3D reconstruction of buildings combines pairs of stereo images with large-scale Geographic Information System (GIS) maps and domain knowledge as additional information sources. The 2D GIS map contains the outline of footprints of the buildings. The knowledge about the problem domain is represented by a building library containing primitive building models. Although, buildings reveal a high variability in 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 an overview of the steps involved in the proposed method for 3D reconstruction of buildings. The partitioning of a building and ranking of the resultant partitioning schemes are discussed in section 3. The next section describes the generation of building hypotheses. Section 5 presents an evaluation function based on mutual information for determining the best building hypothesis. The refined reconstruction of a complete building is described in section 5. Section 6 shows some experimental results. The conclusions and future work are discussed in the final section.

2. METHOD OVERVIEW

The building reconstruction process is formulated as a multi-level hypothesis generation and verification scheme and it is implemented as a search tree. The tree is generated incrementally by the search method.
To cope with the complexity of aerial images specific knowledge about buildings have to be integrated in the reconstruction process. Since most buildings can be described 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 (flat roof, gable roof, and hip roof building). The approach of modeling buildings using a set of basic building models (primitives) suggests the usage of Constructive Solid Geometry (CSG) representation for building description. 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 reconstruction process starts with the partitioning of a building in simple building parts. All possible partitioning schemes of a building are represented on the first level of the search tree. The second level of the tree contains the partitions corresponding to each partitioning scheme. Next, the tree is expanded with a level corresponding to different building hypotheses generated for each building partition. Corresponding to each building primitive defined in the building library, building hypotheses are generated. The estimation of the parameters of the building model is performed using a fitting algorithm.
The building hypotheses are verified by back projecting them into the images and then matching with the information extracted from the images. The matching defines a score function that is used to guide the search in the tree. The CSG tree representing a building is given by the best fit of the building models corresponding to the building partitions. In the final verification step the complete CSG tree is fit to the image data. To improve the results, constraints, which describe geometric relationships between building primitives, are incorporated in the fitting algorithm.

3. PARTITIONING OF A BUILDING

The first step of the actual reconstruction process is the partitioning of a building in simple building parts, which might correspond to the building models defined in the building library. First, the partitioning is done using only the ground plan of the building defined in the GIS map. If the ground plan of the building is not a rectangle, then it is divided in rectangles, called partitions. The line segments created by partitioning are labeled as internal lines, while the original house segments are labeled as external lines. A partitioning scheme can be defined as a subdivision of a building into disjoint partitions. These partitioning schemes are generated by sequential merging of the partitions. Two partitions can be merged if they share a common edge. A building can have multiple partitioning schemes (fig. 2), more or less likely to occur in reality.
Each of these partitioning schemes will start up a branch in the search tree. The number of possible partitioning schemes increases with the complexity of a building. Due to the exponential complexity of the tree search, searching all the paths would be very time consuming. Therefore, we consider only the partitioning schemes that have some support in the images. The goal is to define a measure of this support and to rank the partitioning schemes based on this measure. Optimizing the order, in which the partitioning schemes are considered in the search tree, can reduce the number of partitioning schemes that have to be verified and, thus, the reconstruction time. In this way the most likely partitioning schemes are evaluated first and then the search is stopped when an optimum building model is found, thus minimizing the number of unnecessary verification of partitioning schemes.



a)

Figure 1. a) Conditional probability density $P(l_i \mid l_p)$ b) Probability density function of the image line length $P(l_i)$ c) Mutual information $I(l_i; l_p)$

For ranking partitioning schemes, a metric based on mutual information has been developed. This metric employs image information, namely image lines. The image lines are extracted using the Foerstner operator [5]. The metric measures the support that the image lines give to a partitioning scheme. This support should be high for a good partitioning scheme, and it should be negative if there is no image information to support a partitioning scheme. Therefore, the best partitioning scheme can be defined as the one with the highest mutual information.

The formulation of mutual information is largely credited to Shannon (1948). Since then, there have been many uses of mutual information especially in communication theory. In computer vision mutual information was used for relational matching [21] and for image registration [22].

The mutual information between two descriptions $M = \{m_1, m_2, \ldots\}$ and $D = \{d_1, d_2, \ldots\}$ is defined as the difference between the self-information and the conditional information [23]:

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

where

$$I(d_j) = -\log P(d_j) \quad \text{and} \quad I(m_i \mid d_j) = -\log P(m_i \mid d_j)$$

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 metric used for ranking of partitioning schemes is based on the formulation of mutual information between the length of the lines of partitioning schemes and the length of the lines extracted from images. To calculate the mutual information between an image line length l_i and a partition line length l_p, we have to know the conditional probability density of the image line length given the partition line length $P(l_i \mid l_p)$. This expression can be given by:

$$p(l_i \mid l_p) = \begin{cases} \dfrac{1}{l_p} \dfrac{\dfrac{l_i}{l_p}}{\sqrt{1-\left(\dfrac{l_i}{l_p}\right)^2}}, & \text{if } 0 \le l_i \le l_p \\ 0, & \text{else} \end{cases} \tag{3}$$

This function is shown in figure 1a for the building from figure 2. The length of the image line can be at most length of the projected partition line l_p.

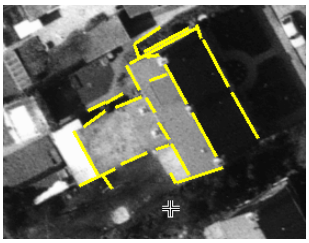
Since the probabilities of the lines length in the partitioning schemes are known, the probability density function of the image line length can be calculated by:

$$p(l_i) = \sum_{l_p} p(l_i \mid l_p) P(l_p) \tag{4}$$

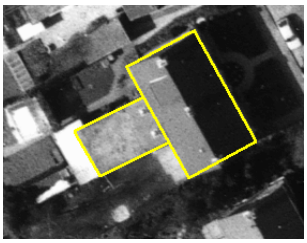
Figure 1b shows the probability density function of the image line length, without the knowledge of the projected partition line length.

Conform (1) the mutual information between an image line length and a partition line length can be calculated as:

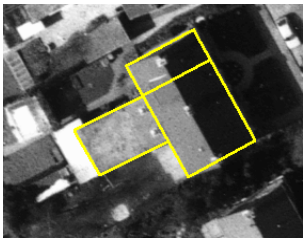
$$I(l_p; l_i) = \log \frac{p(l_i \mid l_p)}{p(l_i)} \tag{5}$$



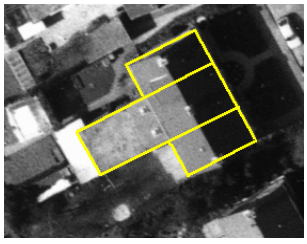
a)



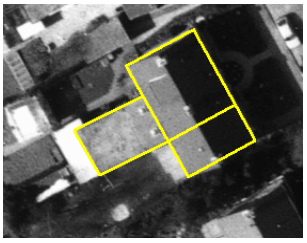
b) Rank = 4.21



c) Rank = 3.35



d) Rank = -1.06



e) Rank = -5.02

f) Rank = -8.07

Figure 2. a) Lines extracted from image
b) – f) Possible partitioning schemes

Figure 3. Best ranked partitioning schemes selected from the list of all possible partitioning schemes

and it is shown in figure 1c.

The rank of a partitioning scheme is given by the sum of the mutual information between its lines and image lines:

$$I(P;ID) = \sum_{l_p} I(l_p \mid l_i) \qquad (6)$$

Clearly, the external house lines do not provide any information about a partitioning scheme, therefore their support can be considered null. On the other hand, the internal house lines are more important than the external house lines and thus will have a large impact on the rank measure. Their support can be computed using (5).

Another criterion that has to be considered at the ranking of the partitioning schemes is the simplicity, expressed by Minimum Description Length (MDL) criterion. This criterion provides a means of giving higher priority to the partitioning schemes with a smaller number of partitions. The measure of the simplicity then becomes the number of bits necessary to describe the partitioning scheme. For an optimal description this is given by: $-\log_2 n$ where n is the number of partitions composing the partition scheme.

Hence, the rank of a partitioning scheme is given by:

$$Rank(PS) = I(P;ID) - \log n \qquad (7)$$

If the rank of a partition scheme is high enough then the partition can be classified as standalone partition and it is not used in further partition merging operations. Thus the number of partitioning schemes is reduced. Using this ranking method for ordering the possible partitioning schemes, we can avoid a blind search of the tree.

An example is given in figure 2. There are five possible partitioning schemes for that building. These partitioning schemes together with their ranks are presented in figure 2 b-f. The highest rank was assigned to the first partitioning scheme. Other examples are presented in figure 3. Here only the best ranked partitioning schemes are shown.

4. GENERATION OF BUILDING HYPOTHESES

After the partitioning process the next step is the generation of building hypotheses. For each partition of the partitioning schemes different building hypotheses are generated corresponding to the building models defined in the building library. To avoid multiple generation of building hypotheses for a partition included in more partitioning schemes a dynamic programming method is used. The generated building hypotheses are represented on the last level of the search tree.

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.

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. Afterwards the values of the parameters are improved by fitting the model to the images.

5.1 Reconstruction of 3D corners and 3D lines

For the approximate estimation of the parameters of a building model 3D information are extracted from images. This 3D information consists of 3D corners and 3D lines obtained by matching 2D features extracted from different images. To cope with the combinatorial complexity of the matching, many constraints have been incorporated in the matching process.

The constraints used for corner points matching are:

- Epipolar geometry: the epipolar constraint is applied to restrict the search for correspondences along one line, the epipolar line.
- Height: The 3D point obtained by triangulation from two 2D corner points must have a height between the minimum and maximum height of the building. This problem is identical to the determination of the disparity search range along the epipolar line.
- Ground plan of a building: The 3D point must lie inside or sufficiently close to the ground plan of the building defined in the map.

The problem of matching line segments is slightly more complex than that of matching points, because the line segment extraction algorithm often produces different results in the two images. Hence, two line segments generally do not correspond globally and only contain a subset of homologous points.

The constraints used for line segments matching are:

- Epipolar geometry: The line segment in one view must lie at least partially within a bean defined by the epipolar geometry and the building height constraints. Consider figure 4. By the epipolar constraints for points, the match for A must lie on the epipolar line ep1', defined by A. Similarly, a match for B must lie on ep2'. By knowing the height range of building, the search space can be reduced to the segments on the epipolar lines defined by extreme height values. Hence, the search space for matching line segments is limited to a four-corner polygon. Each line segment at least partially inside of this polygon can be matched with the initial line segment.

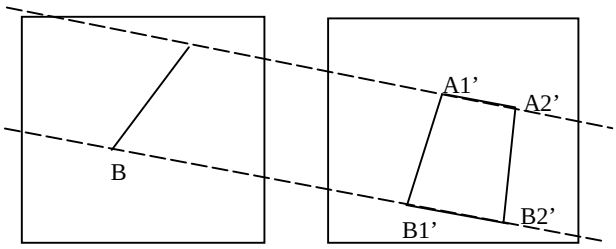


Figure 4. Epipolar constraint for matching line segments

- Uniqueness: This constraint imposes that one line segment from the first image has, at most, one corresponding line segment in the second image (a symmetric constraint applies to a line segment of the second image).
- Order: This constraint assumes the preservation of the order of corresponding lines from the two images.

5.2 Estimation of building model parameters

In the initial approximation the x, y coordinates and the orientation of a building primitive are given by the ground plan of the building. The width and length parameters 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 [17].

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 [20]. This algorithm is similar to the one described by Lowe [10].

The fitting method is described as an iterative least-square algorithm. It estimates the changes of the parameter values that have to be applied in order to minimize the square sum of the perpendicular distances of the image pixels to the nearest wire frame edge. An observation equation is set up for each image pixel within some range of a wire frame edge. The linearized observation equation for a pixel j can be written as:

$$E\{\Delta u_j\} = \sum_{i=1}^{K} \frac{\partial u_j}{\partial p_i} \Delta p_i \qquad (8)$$

where Δu_j is the observed perpendicular distance of the pixel to the nearest edge of the wire frame, p_i are the model parameters, K is the number of parameters, and Δp_i are the changes of the parameters that have to be estimated. In order to ensure that the pixels with higher gradients dominate the parameter estimation, the squared gradient of the pixel can be used as a weight to its observation equation.

5. 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 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 the building ordered by their ranks. The second level contains the partitions corresponding to each partitioning scheme. The last level of the search tree contains the different building hypotheses generated for each building partition.
Two problems have to be considered at the search of the tree:

- Definition of an evaluation function to guide the search for the best solution
- Definition of stops criteria, which speed up the search by reducing the search space.

In this section, we describe the definition of an evaluation function.
From the possible hypotheses of a matching between an object model and an image, one wants to select the hypothesis that maximizes some appropriate metric. 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 [6]. 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 ([14], [19]). Our evaluation function belongs to this latest category, using a mutual information based measure.
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, …} is transmitted through a communication channel into the image D = {d_1, d_2, … }. 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).
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. However, 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. Another advantage of our evaluation function is the simplicity. The distribution of the gradients at random image points $P(grad_i)$ and the conditional distribution of the gradients along the projected roof edges $P(grad_i \mid point_m)$ can be determined by training (see section 5.3). These distributions are shown in figure 5a, b. Thus, the mutual information between an image pixel and the corresponding model point is given by:

$$I(point_m; point_i) = \log \frac{P(grad_i \mid point_m)}{P(grad_i)} \tag{9}$$

and it is shown in figure 5c.
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:

$$I(line_m; line_i) = \sum_{point_m \in line_m} I(point_m; point_i) \tag{10}$$

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:

$$I(M;D) = \sum_{k=1}^{2} \sum_{line_m} \sum_{point_m} \log \frac{P(grad_i \mid point_m)}{P(grad_i)} \quad (11)$$

5.1 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 had the same complexity, this could have been 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 which minimizes $L(D|M_i)+L(M_i)$.

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

$$L(M_i) = I(M_i) \text{ and } L(D \mid M_i) = I(D \mid M_i) \quad (12)$$

Thus, the minimum description length principle minimizes:

$$I(D \mid M_i) + L(M_i) \quad (13)$$

By using the definition of the mutual information:

$$I(D; M_i) = I(D) - I(D \mid M_i) \quad (14)$$

the formula (13) can be expressed as:

$$L(M_i) + I(D) - I(D; M_i) \quad => \min \quad (15)$$

Since $I(D)$ is constant, the expression $I(D; M_i) - L(M_i)$ has to be maximized.

Therefore, it follows that the best model is given by:

$$M_{opt} : \max_i (I(M_i;D) - L(M_i)) \quad (16)$$

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

5.2 Relation MAP and MDL

It can be shown that the MAP (maximum a posteriori) and the MDL principle lead to the same solution [18].

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:

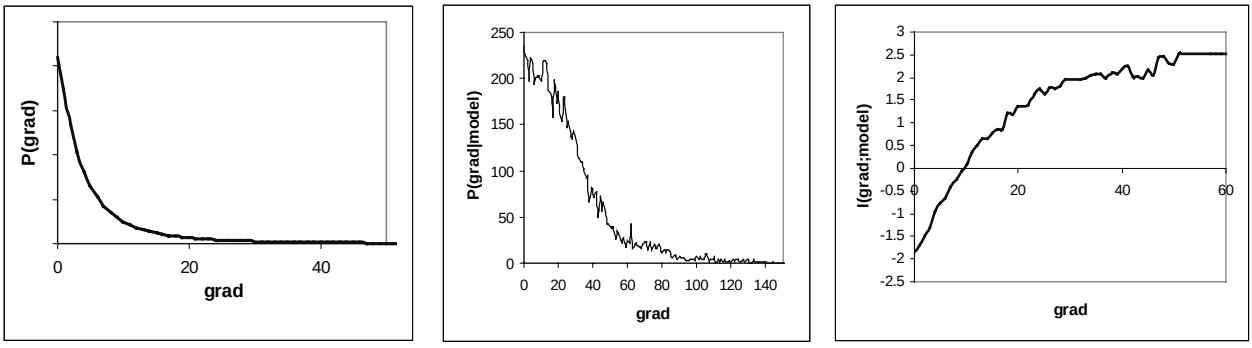
$$P(M_i \mid D) = \frac{P(D \mid M_i)P(M_i)}{P(D)} \quad (17)$$

since $P(D)$ is constant, MAP states that $P(D \mid M_i)P(M_i)$ has to be maximized.

The minimum description length principle minimizes (13)

$$I(D \mid M_i) + I(M_i)$$

This can be written as:



a) b) c)

Figure 5. a) Gradient distribution $P(grad_i)$ b) Conditional probability density of the gradient $P(grad_i|point_m)$ c) Mutual information $I(point_m; point_i)$

$$I(D \mid M_i) + I(M_i) = -\log P(D \mid M_i) - \log P(M_i)$$
$$= -\log P(D \mid M_i) P(M_i) \Rightarrow \min \quad (18)$$

Hence, MDL maximizes $P(D \mid M_i)P(M_i)$ like MAP.

The first term of (17), $P(M_i \mid D)$ can be compute, while the second term $P(M_i)$ cannot be specified easily. The simplest specification of the prior probabilities is that they are all the same, i.e. $P(M_i)$ is a constant. This leads to the maximum likelihood strategy of choosing M_i that maximizes $P(M_i \mid D)$. In case that no information about the a priori probabilities is available, instead of MAP criterion MDL criterion can be used, since MDL does not require the definition of a priori probabilities.

5.3 Density estimation

In order to calculate the score function for matching between a building model and the image data as defined in (16), we need to know the a priori probabilities $P(grad_i)$ and the conditional probabilities $P(grad_i \mid point_m)$.

In estimating the densities we don't know the shape of the density function which describes the data. Therefore we have to use non-parametric techniques. The easiest and the fastest method is histogramming. In this method, the frequency of sample points in each bin are counted and then normalized so that the histogram frequencies all add to one. One alternative to the histogramming could be Parzen window density estimation [22].

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 [17]. The obtained a priori probability $P(grad_i)$ is shown in figure 5a .

The conditional probability density function of the gradient along the projected roof edges can be determined from training matches by analyzing 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 5b.

Knowing these two distributions, the mutual information can be computed conform (9) and this is shown in figure 5c.

6. CSG TREE FITTING

In the final verification step the complete CSG tree will be fit to the image data. In the previous stages of our method, the building models corresponding to different building partitions were treated as isolated objects. The results can be further refined if contextual information is utilized. The fact that the building models contained in the CSG tree form a complex building can be seen as contextual information. Between the building models of a complex building many geometric relationships can be identified, which constitute very valuable information. Therefore, the global fitting algorithm performs a simultaneous adjustment of the building models contained in the CSG tree taking into account the geometric relationships between them.

The geometric relationships between building models can be represented by constraints. In the parameter estimation process, these constraints mean that the parameters of different building models are correlated with each other. The usage of the constraints reduces the degree of freedom of some parameters, therefore the precision of the parameter estimation is increased.

Consider the building from figure 9a, composed from a main part and a small attached shed. The estimation of the parameters of the small building extension can produce problems. By imposing the constraint that the shed is connected to the main building, its parameters can be estimated more precisely, since much information can be derived from the main building.

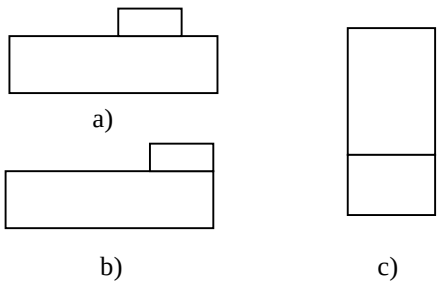


Figure 6. Constraints a) Connection constraint b) Corner constraint c) Extension constraint

In our building reconstruction system the following types of constraints are used:

- *Parameter constraint*: establishes a relation between two parameters of two building models. For example, two building models have the same orientation.
- *Connection constraint*. One edge of a building model lies on one of edges of the other building model (figure 6a).
- *Corner constraint*. Two building models share a common corner (figure 6b).

- *Extension constraint*. Two building models share a common edge (figure 6c).

Constraints can be implemented as either hard constraints or soft constraints. A hard constraint imposes a relation in any condition, while a soft constraint allows small relaxation in the specification of the constraint.

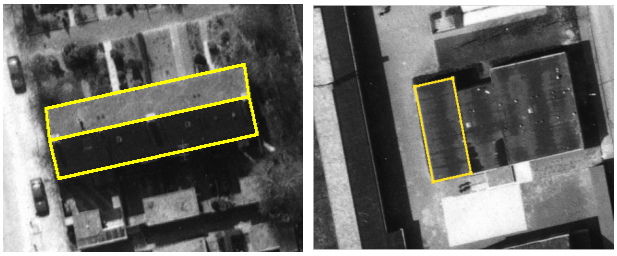
Our goal is to integrate the constraints in the fitting algorithm (8) used to estimate the parameters of the building models. We have chosen to implement the constraints as soft constraints. A soft constraint can be implemented in a least-square adjustment as weighted observation with a standard deviation. The weight specifies the strength of the constraint in the adjustment.

At each iteration of the parameter estimation, the constraints are linearized in the neighborhood of the current estimate and then included together with the observations equations corresponding to the image pixels in the estimation of the parameters of the building.

7. RESULTS

The test data consists of high resolution aerial images with the scale of 1:3000. The interior orientation parameters of the camera and 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 defined in the building library, since the current models require rectangular base.

The first experiment was to generate building hypotheses for simple buildings composed by only one building primitive and to select the right building 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.



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

Figure 7. Reconstruction of simple buildings

a) b)

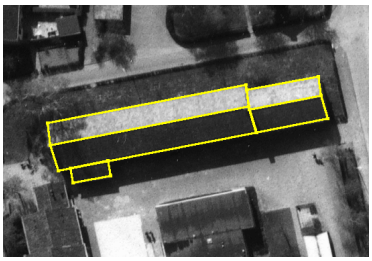
Figure 8. Refined reconstruction of a building.
a) Reconstruction without constraints Score = 794.1
b) Reconstruction with constraints Score = 851.1

The resultant building models projected back into one of the images are presented in figure 7. The evaluation function based on mutual information reliably selects the correct model. The building models are reconstructed correctly even though some of the roof edges have very bad contrast.

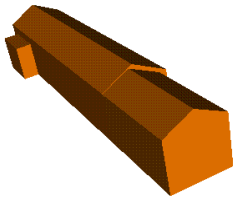
Next, we tested our approach on complex buildings. First, the partitioning of the building into building primitives based on the ground plan was performed. Then, for each resultant building primitive, hypotheses were generated. Evaluating the partition schemes we found that the partition schemes presented in figure 8, 9 are the best ones. Figure 8 shows the improvement gained by integrating geometric constraints between building models in the fitting. When fitting without constraints the two building models are not connected to each other, there is slight difference in their orientation (figure 8a). This problem is solved by introducing the connection constraint between the two building models. The evaluation score increased from 794.1 at the fitting without constraints to 851.1 at constrained fitting.

Figure 10 shows the results obtained by applying our approach on an entire scene. Artificial vertical walls were added to the automatically extracted 3D models of the roofs.

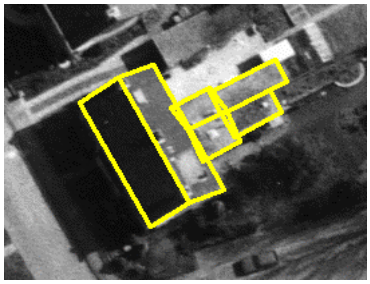
The results from the proposed approach are encouraging. The method worked well even in difficult conditions, where feature based approaches would have failed. The constraints incorporated in the primitive building models



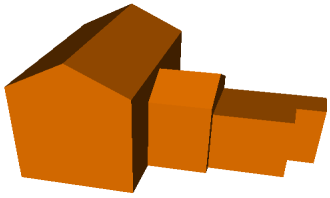
a)



b)



c)



d)

Figure 9. 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

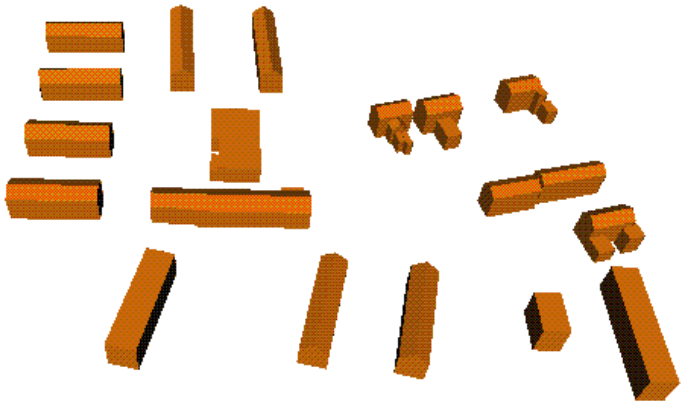


Figure 10. 3D model of a reconstructed scene

contribute to the robustness of the method. Hence the reconstruction has a high redundancy and can succeed even in difficult conditions.

8. 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 search the tree two metrics based on mutual information were defined. A metric orders the building partitioning schemes such that the most likely is presented first for verification to the search tree. Another metric allows comparison and evaluation of different building hypotheses. This metric is unique in that it compares 3D building models directly to raw images. No preprocessing or edge detection is required. The metric has been rigorously derived from information theory. The results showed the usefulness of the mutual information as evaluation metric. This technique works well in domains where edge based methods have difficulties.

Future work will be directed towards the definition of the expected amount on mutual information needed for a reliable matching. This is going to be used as stop criteria for the tree search in order to speed up the search. The reconstruction can be further speed up by incorporating more knowledge for partitioning the buildings.

References

1. N.Ayache, O.D.Faugeras, “HYPER: A New Approach for the Recognition and Positioning of Two-Dimensional Objects”, *IEEE PAMI*, 8, 1, pp. 44-54, Jan 1986.
2. J. Beveridge, R.Weiss, E.Riseman, “Optimisation of 2D Model Matching”, In 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. Foerstner, “A Framework for Low Level Feature Extraction”, *Computer Vision - ECCV94*, vol.2, Springer Verlag, Berlin, 1994, pp.283-394.
6. W.E.L. Grimson, “*Object Recognition by Computer: The Role of Geometric Constraints*”, MIT Press, 1990

7. 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.
8. 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.
9. C. Lin, A. Huertas, and R. Nevatia, “Detection of Buildings Using Perceptual Grouping and Shadows”, *Proc. CVPR*, IEEE Computer Society Press, Los Alamitos, CA, pp. 62-69, 1994.
10. D.G. Lowe, “Fitting Parameterized Three-dimensional Models to Images”, *IEEE PAMI*, 13, 5, pp. 441-450, May 1991.
11. 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
12. 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
13. S. Noronha, R. Nevatia, “Detection and Description of Buildings from Multiple Images”, *Proc. CVPR*, pp.588-594, 1997.
14. A.R. Pope, D.G. Lowe, “Modeling Positional Uncertainty in Object Recognition”, *University of British Columbia, Computer Science, TR-94-32*, November 1994.
15. M. Roux, D.M. McKeown, “Feature Matching for Building Extraction from Multiple Views”, *Proceedings of the ARPA IUW*, Menterey, CA, pp.331-339, 1994
16. C. Schmid, A. Zisserman, “Automatic Line Matching Across Views”, *Proc. CVPR*, pp. 666-671, 1997.
17. I. Suveg, G. Vosselman, “3D reconstruction of Building Models”, *Int. Archives of Photogrammetry and Remote Sensing*, vol XXXIII, part B2, pp 538-545, 2000.
18. I. Suveg, G. Vosselman, “3D Building Reconstruction by Map Based Generation and Evaluation of Hypotheses”, *BMVC2001*.
19. I. Weiss, “Statistical Object Recognition”, Ph.D. dissertation, *Cambridge MIT*, 1993
20. 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
21. G. Vosselman, “Relational Matching”, *Lecture Notes in Computer Science 628*, Springer-Verlag, 1992
22. P. Viola, W. Wells, “Alignment by Maximization of Mutual Information”, *Proceedings of 5th International Conference on Computer Vision*, pp. 16-23, 1995.
23. T. M. Cover, J. A. Thomas, “*Elements of Information Theory*”, John Wiley & Sons, 1991.