3D RECONSTRUCTION OF BUILDING MODELS

Ildiko SUVEG, George VOSSELMAN
Technical University of Delft, The Netherlands
Photogrammetry and Remote Sensing
{I.Suveg, G.Vosselman}@geo.tudelft.nl

Working Group IV/2

KEY WORDS: buildings, data fusion, GIS, image matching, reconstruction.

ABSTRACT

This paper describes a knowledge-based approach to automatic 3D reconstruction of buildings from aerial imagery. The approach relies upon the use of 2D GIS map and knowledge about the imaging geometry and acquisition parameters. We show that by integrating knowledge from the images and the GIS map, the complexity of the building reconstruction process can be greatly reduced. In the first stage the buildings are localized in the images based on information about the ground planes of the buildings contained in the GIS map. This restricts the processing at all following stages to one building structure. Further, a building can be partitioned into more simple building parts, corresponding to some basic building models. By merging the separately reconstructed building parts the initial building model is found. The building reconstruction process is described as a tree search in the space of possible building hypotheses.

1 INTRODUCTION

The 3D reconstruction of buildings has been an active research topic in Computer Vision as well as in Digital Photogrammetry in recent years. Three dimensional building models are increasingly necessary for urban planning, tourism, etc.
Manual 3D processing of aerial images is very time consuming. Therefore, speeding up this process by automatic or semiautomatic procedures has become a necessity. In principle the use of aerial images alone for automatic reconstruction of buildings should be sufficient. However the images contain a large amount of information and it is difficult to separate the useful information from irrelevant details. Also, useful information for the reconstruction task can be lost because of occlusions, low contrast, and bad perspective. To overcome these problems, the image data has to be combined with other data sources. This additional information source can be a scanned or a digital GIS map. GIS maps are widely available for most well developed countries. By combining the images with a GIS map, the specific strengths of both the images (high resolution, accuracy, large information content) and the map (relatively simple interpretation) can be exploited.

The goal of this project is to develop a method for 3D reconstruction of buildings with application to the data set from The Netherlands, consisting of pairs of stereo images and a large scale GIS map. The advantage of our method is its ability to robustly combine information from the images and a GIS map to generate 3D information.

This paper is organized as follows: Section 2 presents a brief review of different strategies which have been proposed to reconstruct buildings. The following section gives a brief overview of the steps involved in our method. Section 4 describes the localization of a building and localization of wall primitives in an image. Section 5 presents the 3D reconstruction of points and line segments, which are necessary in the next steps. The next section describes the generation of building part hypothesis. Each step is also illustrated using a real example. The conclusions and future work are discussed in the final section.

2 RELATED WORK

A variety of approaches have been suggested for the reconstruction of buildings from aerial images. Since manual 3D processing of aerial images is time-consuming, the development of automatic or, at least, semi-automatic techniques becomes a necessity.

In the semiautomatic methods the user needs to actively interfere to delineate and / or support the process of generating building hypothesis. In the image the user often places the given object models, which are then refined by the system. The system does the measurement task, whereas the user handles the interpretation and modeling tasks (Lang et al, 1995).
The automatic systems usually impose constraints upon the geometry of the buildings (rectangular shape, flat roof, etc). They use generic models, which require a well populated object library, but this is difficult to acquire. An alternative could be a model database containing building parts, since a variety of building models can be generated from a relatively small class of predefined building parts. The reconstruction is then performed using a two step process of hypothesis generation and verification. First, 3D feature points are grouped and related in building part primitives. In a second step, these parts are combined into a building model (Fischer et al, 1997).
In another approach, the geometry of a building roof is described as a collection of line segments which form planar structures. In this way a roof is modeled as a set of planar polygons, each of which encloses a compact area with consistent photometric and chromatic properties (Bignone et al, 1996).

Another way to classify the existing systems concerns the input data used.
There are a lot of systems that work solely with monocular images. These systems exploit shadows either to infer the third dimension or to verify the generated hypothesis. Often the buildings are assumed to be rectangular or rectilinear flat roofs (Huertas, Nevatia, 1988), (Irvin, McKeown, 1989), (Collins et a1, 1995).
Other systems use widely available stereo images. Stereo images allow the determination of the third dimension by epipolar matching of different features extracted from both images. Multi-view strategies are advantageous in providing redundant information and improving the accuracy of the reconstruction (Roux, McKeown, 1994).

The newest trend consists of exploiting information other than images to support the analysis; for example fusing images with scanned (Maitre et al, 1995), or digital maps (Pasko, Gruber, 1996), (Haala, Hahn, 1995). The information about the ground plans of the buildings available in the GIS map are used to define parametric building models. These models are verified and the unknown parameters are determined by matching the extracted lines from the images against the lines of the building models.

The approach presented in this paper belongs to this latest trend, integrating knowledge from images and a GIS map. The difference between our method and those previously described is the way in which building models are generated. While the previous approaches defined building models using only the information available in the map, our approach generates building models by deriving 3D information from the stereo images. The information contained in the map is used to focus the search and to constrain the possible building hypotheses.

3 METHOD OVERVIEW

One of the important issues in the proposed method is separation of the building detection from the building reconstruction process. 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.

Buildings reveal a high variability in shape, but even complex buildings can be generated by a combination of simple building models. Therefore, it is useful if a complex building is partitioned into simple building parts, each of them corresponding to a basic building model. This process can be done by using the ground plan information from the map. The ground plan of a building can be divided in rectangles - each of the rectangles representing the ground plane of a building part.
As basic building models we can consider a flat roof, a gable roof and a hip roof building (figure 1). This approach of modeling buildings using a set of basic

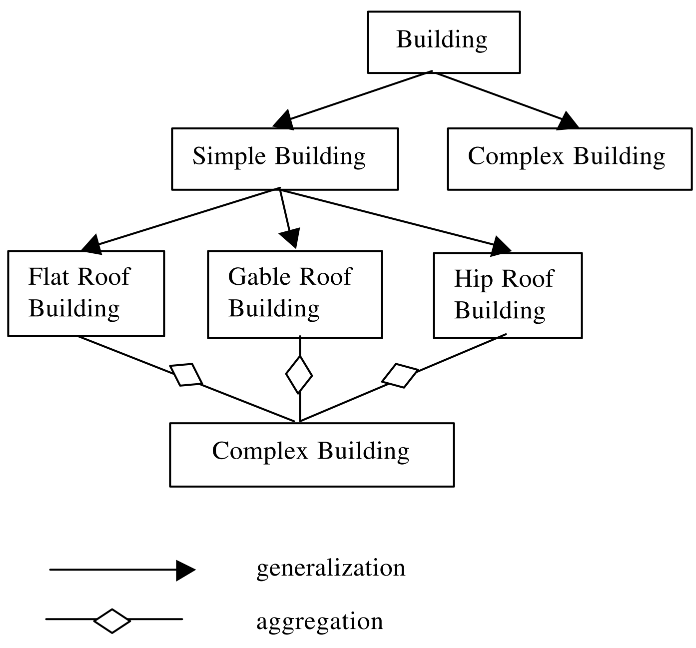

Figure 1. Building hierarchy

building models (primitives) suggests the usage of the Constructive Solid Geometry (CSG) representation for describing buildings.
Next, corresponding to each rectangle obtained after the partitioning process, several building model hypotheses are generated by combining the 2D information of a rectangle (shape and pose) with the 3D information extracted from the images, namely the height. These building primitive hypotheses are then verified by fitting them to the images.
The 3D reconstructed building is found by merging the building primitives with the best fitting results.

4 LOCALIZATION OF BUILDING PRIMITIVES IN AERIAL IMAGES

4. 1 Localization of a building in aerial images

Information about the ground plan of the building contained in the GIS map can be used to delineate a building in an image. The GIS map used is a two-dimensional digital map representing the ground plans of buildings. The buildings in the map are described by a list of the coordinates of their corner points.
The 3D reconstruction of the buildings can be improved by integrating the available knowledge sources. These knowledge sources should be as robust as possible but none of them can be expected to be error free. These errors / uncertainties can influence the reconstruction process.
In the building localization process the uncertainties are due to the unknown height of the buildings, the accuracy of the GIS map, the roof extensions, and the feature extraction.

In order to handle these uncertainties we designed a two-step method. In the first step the uncertainty due to the unknown height of the buildings is handled, by assuming the height of a building to be between two extreme values. By projecting the ground plan of the building into the image for each of these values two contours are obtained. These contours have to be concatenated in order to get the area where the building is located. In the next step the contour obtained after the concatenation process has to be dilated, taking into account the other uncertainties mentioned above.

4.1.1 Concatenation of the contours

If the GIS data contained 3D information, i.e. height information, then it would be easy to find the exact position of a building in an image. Using the known orientation parameters of the camera, the building could be projected into the image.
However, the GIS map contains only 2D information, therefore an assumption about the missing third dimension has to be made. In the area of our test data the height of the buildings could be assumed to be between 3m and 15m above ground level. For these extreme height values, the ground plan of the building contained in the GIS map can be projected into the image using the known orientation parameters obtaining two slightly shifted contours for a building (figure 2).

Supposing that the real height of a building is between the two extreme values chosen, the position of the building in the image will be in an intermediate position between these two contours. In the next step, these two contours have to be concatenated in order to get the maximal possible area in which the building could be found (figure 3).

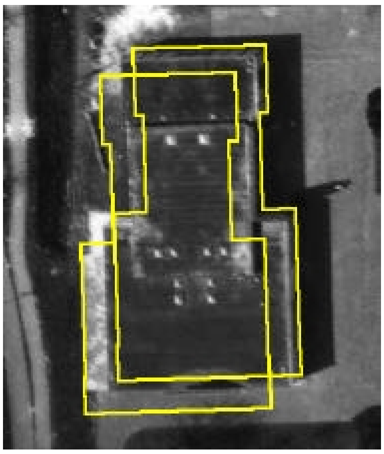

Figure 2. Two shifted contours

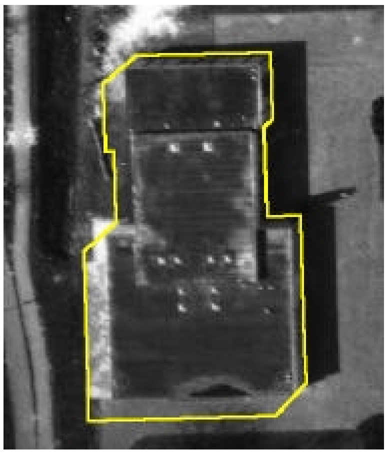

Figure 3. The concatenated contour

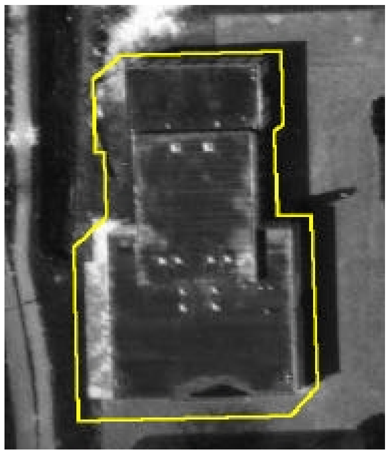

Figure 4. The dilated contour

4.1.2 Dilation of the Contour

The resulting contour would represent the location of the building if all knowledge sources were error free. However, we also have to handle the uncertainties due to:

1. Accuracy of map data - Among the uncertainties in the map data, the positional accuracy of the map is important. In fact, positional accuracy of GIS data is only a minor source of uncertainty. The uncertainty introduced by the map, described by the standard deviation, is 20 cm, which, in our case, corresponds to about 2 pixels in image space ($\sigma_1 = 2$).
2. Roof extension - Usually the map contains the ground plan of a building, which is smaller than the roof base of a building, because of the overhang. On the other hand, in the image the roof base can be seen. The value of this difference could be up to 50 cm, corresponding to about 5 pixels, which results in a mean of 2.5 pixels ($\mu_2 = 2.5$) and a standard deviation of 1.25 ($\sigma_2 = 1.25$) considering a 95% confidence interval.
3. Feature extraction - The features are extracted using the Foerstner operator (Foerstner, 1994), which can be used for interest point extraction as well as straight line extraction. This operator assures subpixel accuracy for the position of the extracted points and lines. Hence the standard deviation of the error from feature extraction is $\sigma_3 = 1$. However, the quality of the extracted features can be influenced by the acquisition of the image (high noise). This can affect the number, the shape, and the position of the extracted features.

To deal with these uncertainties the contour obtained after the concatenation has to be dilated. The value used for dilation can be computed by from the uncertainties introduced by the three error sources mentioned above. The total standard deviation can be obtained using the formula:

$$\sigma_t = \sqrt{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}$$

Considering a 95% confidence interval, the interval is [μ - 2σ, μ + 2σ]. In this way the dilation parameter will be 9 pixels. After the dilation a new contour is obtained inside of which lies the building of interest (figure 4).

4.2 Localization of wall primitives

Based on the information contained in the map, not only a building but also the wall primitives, namely corners and segments, can be localized in the images. Actually, the localization of wall primitives is a more constrained case of a building localization process. The same uncertainties have to be considered but dealing with them is much easier. For a wall corner point two points are obtained by projecting a corner from the map into the image for the considered extreme height values. The concatenation process is thus reduced to the connection of the two points by a segment Next, dilating a segment means building a rectangle around the segment (figure 5). Also, for a wall segment, instead of a polygonal contour, we have to work with a trapezoidal form (figure 6).

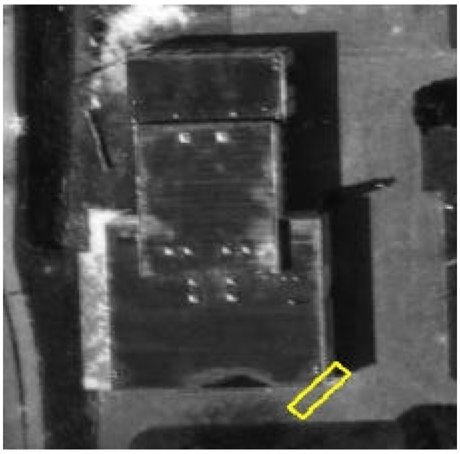

Figure 5. Localization of wall corners

After the locations of the wall primitives are found the image, features, which may correspond to these wall primitives, have to be determined.
The image features which will be used as primitives for the building reconstruction are corner points and lines. The image features, which are inside of a contour defining the location of a wall primitive, are determined and labeled as a wall primitive. In that way corner points are classified as wall corners having the label of the corresponding map corner and roof corners. Also, the line segments are labeled as wall segments or roof segments.

Figure 6. Localization of wall segments

5 3D RECONSTRUCTION OF CORNER POINTS AND LINE SEGMENTS

5.1 Matching corner points

The 2D corner points from different images have to be matched in order to get 3D corners. To cope with the combinatorial complexity of the matching, many constraints have been incorporated in the matching process. To assure independence from lightning conditions, the intensity of features are not considered in the matching. The constraints used for the corner point matching are:

- Corner point label: a corner point labeled as a wall corner in one of the images can only be matched with a corner point corresponding to the same wall corner, or with a roof corner of the other image. But roof corners can be matched with any kind of corner points.
- Epipolar geometry: the epipolar constraint is applied to restrict the search for correspondences along one line, the epipolar line.
- Height: the 3D points obtained by forward intersection of the two 2D corner points must have a height between the considered extreme values. This problem is identical to the determination of the disparity search range along the epipolar line.
- Ground plan of a building: the 3D points must lie inside or sufficiently close to the ground plan of the building defined in the map.

The order in which the constraints are applied was chosen in such a way as to reduce the total amount of computation required to match two corner points.

5.2 Matching line segments

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

The usage of different constraint can reduce the search space of the matching.

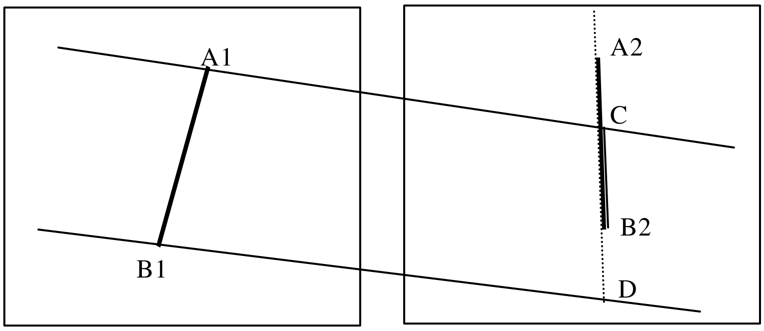

Figure 7. Modified epipolar constraint for line segments

- Line segment label: this is one of the most powerful constraint which is incorporated in the matching process. In the same way as for the corner points matching, a line segment labeled as a wall segment in one of the images can only be matched with a segment corresponding to the same wall segment, or with a roof segment of the other image. But roof segments can be matched with any kind of segments.
- Epipolar geometry: this is used to further reduce the possible correspondences between line segments. To check the correspondences between two lines, a modified epipolar constraint has to be applied for both end points of the line. The modified epipolar constraint (figure 7) imposes that the line segment A_2B_2 in the second image must have a common part with the line segment CD created by the intersection of the epipolar lines through the end points of the line segment A_1B_1 in the first image, with the line through the segment in the second image and vice versa.
- Order constraint. The order constraint assumes the preservation of the order of corresponding lines of the two corners (figure 9). If S_1 matches S'_2 then S_2 cannot match S'_1.

Figure 8. The order constraint

6 GENERATION OF BUILDING HYPOTHESES

6.1 Building Models

A building can be represented in various ways: volumetric, boundary or wire-frame depending on the processing context.

Volumetric representation: CSG is well suited to describing complex shapes, which can be composed from a set of primitives. Within this approach buildings are described by combining a set of basic primitives. As basic primitives we use box, wedge, and rectangular pyramid. The primitives are combined by means of the boolean operations union, intersection, and difference. Finally a building will be described as a CSG tree, where the leaves of the tree contain primitives and the internal nodes contain boolean operations.

Boundary representation describes the object in terms of its boundary elements. The representation is a directed graph containing face, edge and vertex nodes. Vertices constitute the basic geometric information, while edges and faces are defined through topological relations. This type of representation is frequently used for visualization purposes since the visualization algorithms and the representation are well suited.

Wire frame representation describes the object as vertices and edges, but not surfaces or faces. It is used to project the primitive into the image.

6.2 Search Tree

The 3D reconstruction of a building can be seen as a tree search algorithm, because 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 search tree is generated incrementally by the search method.

First, the partitioning of a building into building parts has to be performed. This can be done using the ground plan of the building provided by the 2D GIS map. If the ground plan of the building is not a rectangle, then it can be divided in rectangles. A ground plan can have multiple rectangle partitioning schemes (figure 9). Each of these partitioning schemes will start up a branch in the search tree. 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. Currently the partitioning of the building into building parts is done manually but in the future will be automated. If all building model hypotheses generated by partitioning based on the ground plans fails are rejected then the partitioning has to be refined using image features.

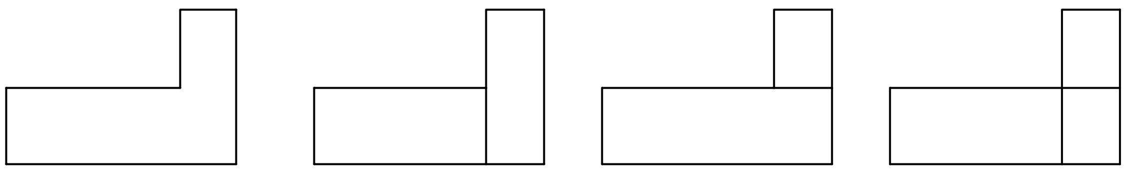

Figure 9. Building ground plan and possible partitioning schemes of it

Each of the rectangles from a partitioning scheme correspond to a building part. Next the search tree will be expanded with a level corresponding to the hypotheses of the building parts. The a priori knowledge about the building types from the data set, if it is available, can guide the process of building hypotheses generation. As a consequence, the more frequently occurring building models are treated first.

For generating building part hypotheses, the basic building primitives of the CSG representation are described by parametric models having pose and shape parameters. In our current approach we only deal with two parametric models: flat roof and gable roof building. To describe a flat roof 6 parameters are necessary: width, length, height, x, y coordinates of the buildings reference point and the orientation in the xy-plane. For a gable roof an extra parameter, the height of the ridge has to be added.

The building primitive hypotheses can be verified using a fitting algorithm. The application of the fitting method requires approximate values for the building primitive parameters. These approximate values can be obtained 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 this purpose a simplified ranking scheme was developed. The height is given by the median of the 10% minimal and maximal values as estimations of the minimum and maximum. For a gable roof the height of the ridge is considered the height of the reconstructed 3D top line; if the top lines were detected in both images and the 3D line can be reconstructed. Otherwise, it is considered to be higher, with a given value (3m), than the height of the building.

The verification of the building primitive hypothesis and also the more precise estimation of the pose and shape parameters is obtained by a fitting algorithm. We use the fitting algorithm described in (Vosselman, Veldhuis, 1999) which uses least square to fit the edges of the projected wire frame to gradients of the pixels from both images simultaneously.

Figure 10. Reconstructed gable roof buildings projected back into the image

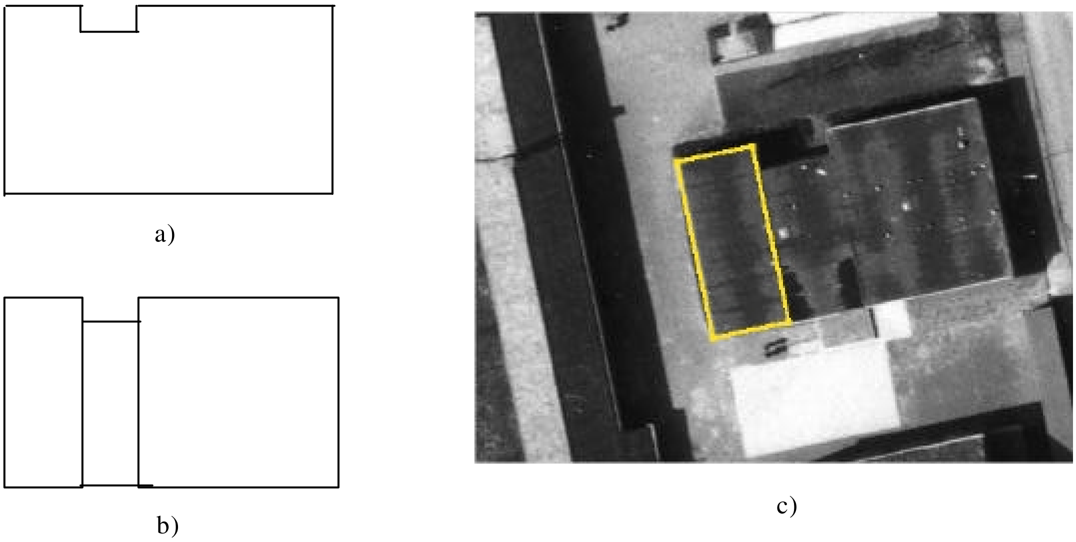

Figure 11. Reconstruction of a complex building. a) Ground plan of the building. b) Partitioning scheme of the ground plan. c) Reconstructed building primitive corresponding to the first rectangle of the partitioning scheme

7 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 considered as known. A 2D GIS map containing the ground planes of the buildings is given.

The first experiment was to generate building hypotheses for simple buildings composed by only one building primitive. In our current implementation two hypotheses are generated corresponding to a flat roof building primitive and a gable roof primitive. The resultant building models projected back into one of the images are presented in figure 10.

Next, we tested our approach on complex buildings (figure 11). The partitioning of the building in building primitives based on the ground plan is currently performed manually. Then, for each resultant building primitive, hypotheses are generated.

8 CONCLUSIONS

A method, which reconstructs 3D corners of buildings integrating information from aerial images and a 2D GIS map, was described in detail. We showed how the map information can be actively integrated in the reconstruction process. The map information was used to locate building primitives in images and to constrain the search for correspondences at the 2D corner points as well as line segments matching. Also, the ground plan of a building contained in the map can guide the its partitioning into building primitives.
A method for building a search tree containing the multiple consistent building primitive hypotheses was described.

Future work will be directed towards the development of methods for :
- automation of the partitioning process of a building
- definition of an evaluation function for measuring the quality of the reconstructed building primitive. This function will be used for searching in the tree.
- generation of building hypotheses by merging the reconstructed building primitives.

REFERENCES

1. Bignone, F., Henricsson, O., Fua, P., Stricker, M., 1996. Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery, Computer Vision -- ECCV'96, Springer Verlag, vol.1, pg.85-96.
2. Collins, R.T., Hanson, A.R., Riseman, M.R., Schultz, H., 1995. Automatic Extraction of Buildings and Terrain from Aerial Images, Automatic Extraction of Man-Made Objects from Aerial Images, pp.169-178, Birkhauser Verlag, Basel.
3. Fischer, A., Kolbe,T. H., Lang, F., 1997. Integration of 2D and 3D Reasoning for Building Reconstruction using a Generic Hierarchical Model, Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps SMATI '97, Bonn 5/97.
4. Foerstner, W., 1994. A Framework for Low Level Feature Extraction, Computer Vision - ECCV94, vol.2, pp.283-394, Springer Verlag, Berlin.
5. Irvine, R.B., McKeown, D.M., 1989. Methods for Exploiting the Relationship Between Buildings and Their Shadows in Aerial Imagery, IEEE, Transactions on Systems, Man, and Cybernetics, vol.19, no.6, pp.1564-1575.
6. Haala, N., Hahn M., 1995. Data fusion for the detection and reconstruction of buildings, Automatic Extraction of Man-Made Objects from Aerial Images, pp.211-220, Birkhauser Verlag, Basel.
7. Huertas, A., Nevatia, R., 1988. Detecting Buildings in Aerial Images, CVGIP, vol.41, pp.131-153.
8. F. Lang, T. Loecherbach, W. Schickler, 1995: A one-eye Stereo System for Semi-Automatic 3D-Building Extraction, Geomatics Info Magazine, June.
9. Maitre, H., Bloch, I., Moissinac, H., Gouinaud, C., 1995. 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.
10. Pasko, M., Gruber, M., 1996. Fusion of 2D GIS Data and Aerial Images for 3D Building Reconstruction, International Archives of Photogrammetry and Remote Sensing, vol.XXXI, part B3, pp.257-260.
11. Roux, M., McKeown, D.M., Feature Matching for Building Extraction from Multiple Views, Proceedings of the ARPA IUW, Menterey, CA, 13-16 Nov, pp.331-339.
12. Vosselman, G., Veldhuis, H., 1999. Mapping by Dragging and Fitting of Wire-Frame Models, Photogrammetric Engineering & Remote Sensing, vol. 65., no.7, July, pp. 769-776.