FINDING 3D-STRUCTURES IN MULTIPLE AERIAL IMAGES USING LINES AND REGIONS

Håkan Wiman, Peter Axelsson
Department of Geodesy and Photogrammetry, Royal Institute of Technology
S-100 44 Stockholm, Sweden
e-mail: hakanw@geomatics.kth.se, pax@geomatics.kth.se

ABSTRACT

A framework for autonomous generation of 3D structures using multiple aerial images is presented. Objects of interest are man-made structures consisting of planar surfaces delineated by straight lines, for example buildings. The task is subdivided into several stages, including increasing object knowledge. By moving from the image domain to object space when searching for correspondences between image features, the advantages of using multiple images are made possible. One of these advantages is the added redundancy and reliability to the result. Also, by avoiding image-to-image processing, the search space is increased only linearly with the number of images. The Minimum Description Length principle is used both for feature extraction and for clustering of features in object space. Currently, only buildings with rectangular roof wings can be described. An example is presented for a buildings covered by six images.

KEYWORDS image understanding, object reconstruction, object space modeling

1. INTRODUCTION

Automated photogrammetric map compilation has become one of the largest research topics in the photogrammetric community. Most efforts have been made to automatically localise and describe man made object, *s.a.* buildings and roads. One of the reasons for the desire to automate map compilation is that it is labour, and thus cost, intensive. The fundamental change of medium for geographic information, from paper sheets to computer data bases, and the interactive way maps thereby can be created, also make new demands on geographic data capture. Such demands are made for three dimensional city models, which are asked for by many users for *e.g.* city planning and tele communications. A 3D city model requires the entire building volumes to be mapped, which, without the aid of automatic mapping, is a most elaborate exercise.

2. OBJECTIVES OF THE STRATEGY

A strategy for autonomous generation of 3D structures was developed with the following basic objectives:

- to use multiple images
- to use parallel search strategies for evidence
- to use object knowledge to constrain the search domain
- to work in object space when possible

When using multiple images, more information becomes available for the evaluation task. If evidence is weak in one image it may be found in another. Also, the problem of occlusion may be reduced significantly. A drawback with multiple images is however the complexity of the correspondence problem. Moving to object space when analysing more than two images simultaneously has several advantages, *e.g.*, the strength of geometrical constraints, but the major one seems to be that the complexity of the correspondence search can be treated in an efficient and rational way. More images just adds more information while the search space remain the same. This is still only possible if the search procedures are designed as being independent of the order in which the evidence is collected, *i.e.*, parallel procedures. The advantages of parallel procedures are thus twofold: (*i*) it is easier to avoid local minima since all evidence are treated simultaneously and (*ii*) the search space is in principle independent of the number of evidence. For complex structures, s.a., buildings a number of generalisations and constraints must be imposed on data if such parallel search procedures are to be designed. These constraints come from the object knowledge and excludes objects outside a defined category.

Following this discussion, we believe that moving to object space when solving the corresponding problem task is a necessity if the goals of such a system are to be met.

3. FINDING 3D-STRUCTURES

3.1 System Outline

We will present a system for autonomous 3D description of buildings. There are no semantic rules, like e.g. windows are surrounded by walls, so in principle any imaged 3D structure fulfilling the postulated geometric and radiometric criteria could be described. We will however refer to buildings, since the main objective is to describe them in 3D. In order to reduce the complexity of the task, the following criteria have been formulated:

- One building at a time is analysed.

- Approximate localisation of the building is assumed known.
- The building must be oriented in three main directions; one vertical, and two horizontal and perpendicular.

The last requirement, does *not* imply that the building must consist *exclusively* of boundaries in the three main directions, only that such boundaries must exist. In the current state, parallel, horizontal lines may only be connected by lines perpendicular to them. In effect, this means that 3D rectangles (tilted as well as horizontal) can be found. General 3D parallelograms can not be found, since neither of two opposite lines is perpendicular to the other two.

The main parts of the system are shown in Fig. 1. The implementation of the described approach is not done as a streamlined production tool, but more as a loosely connected system of related programs. For this reason, no calculation times or efficiency numbers are presented.

The main image feature used for the correspondence task are straight lines. Straight lines with high precision can be found by standard methods. In general, these methods give the end points of an isolated line. If regional descriptors, like average grey level, or the topology of the lines are desired, a region segmentation must be performed. We believe that both straight lines and regional information are needed, and use a region segmentation, that uses straight lined boundaries of the regions (Wiman 1995).

The vast majority of all buildings fulfil the criteria that some lines are horizontal, and that these lines are oriented in either of two perpendicular angles when projected to the *XY* plane. Most additions to buildings, that may be added to a geographic data base in an automated map revision process, obey these rules as well. The lines extracted by the region segmentation are therefor first examined with respect to object space orientation. The two main, horizontal and perpendicular directions are determined by the examination.

The lines that have contributed to the definition of the main directions are then analysed in a clustering algorithm (Axelsson 1994), which accumulates intersections of these lines. Once again, evidence from each image is accumulated in object space and then analysed in object space. The major clusters, which have contributions from all images, form in principal endless horizontal lines in either of two perpendicular orientations. These endless lines are currently truncated based on expected size of the object.

Each pair of parallel lines forms a plane, which is hypothesised as an object plane. Each 2D line that is inside the projected window of a hypothesised plane is analysed whether it (*i*) fits to the plane and (*ii*) intersects the parallel 3D lines in right angles. If so, their intersection points on the 3D lines are computed. For true vertical structures, the intersection points will be approximately the same for a large number of 2D lines, which thus form a cluster. False planes do not have any pronounced clusters. The outlines of a true plane is determined by finding the strongest two clusters. This is the third time an image-to-object accumulation followed by an object space analysis is performed. In separating false plane from true and defining the outlines of the true planes, radiometric evidence, collected from the original segmentation, is used in combination with the geometric evidence.

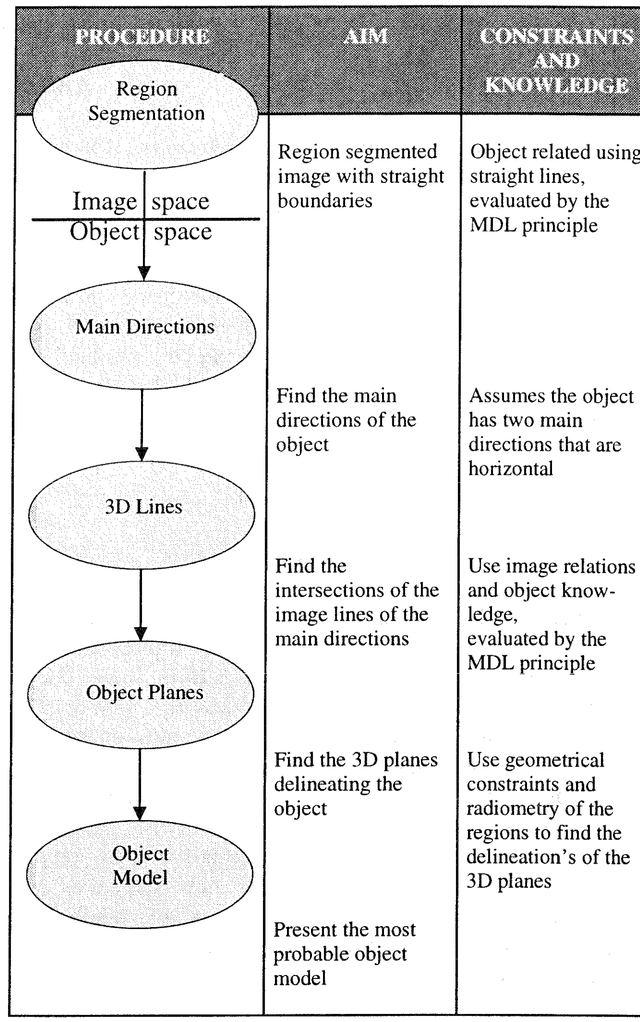

Figure 1. *The main parts of a system for autonomous description of 3D structures.*

3.2 Initial data

Most systems for automated description of buildings use either one image, e.g. (Braun 1994, McGlone *et.al.* 1994, Lin *et.al.* 1995) or two images, e.g. (Jamet *et. al.* 1995, Roux *et.al.* 1994). Our system is specifically designed to handle more than two images without prohibitive increase in search space.

The approximate localisation, but not orientation or shape, of the building is assumed known. The input thus consists of digital image patches, one from each aerial photograph in which the building is imaged. The interior orientation of the camera(s) and the exterior orientation of the images must be known.

We will illustrate our strategy with an example, using a building with a simple geometric shape. The building was imaged in six aerial photographs. One of the six

Figure 3. The segmentation result using Object Related Image Segmentation (ORIS) on six image patches of one building.

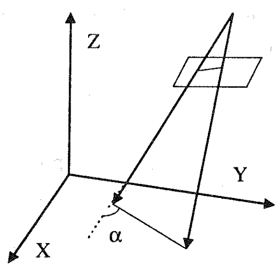

Figure 4. Intersecting the line segment planes with the XY-plane

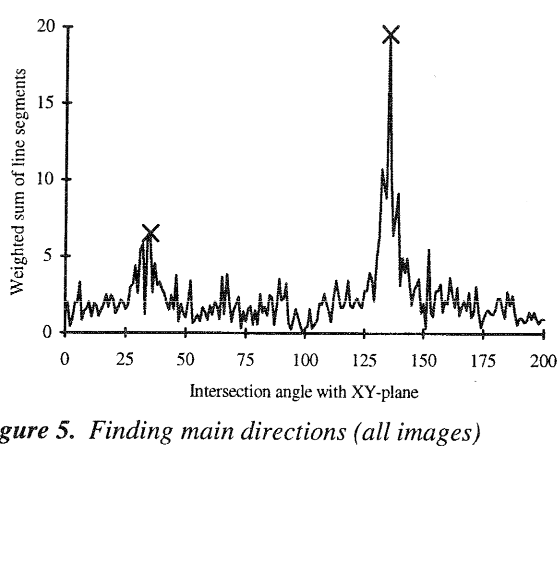

Figure 5. Finding main directions (all images)

3.5 Searching for the main structures

Having found the main directions in object space, *i.e.,* the vanishing directions in the images, a 3D clustering of the line segments that contribute to these directions is carried out. The clustering of intersecting line segments is done for all images simultaneously. The aim of the evaluation of the clusters is to determine the correspondence between image lines (Fig. 6). Geometrical rules like parallelity, perpendicularity and connectivity are incorporated in the clustering algorithm.

In Fig. 7, all line segments belonging to a main direction are projected on a cluster plane and in Fig. 8 only clusters that have at least five intersecting lines. The cluster planes are evaluated using MDL to remove contradictions and to find the most probable solution. The found 3D lines are back-projected on an image in Fig. 9.

extracted image patches is shown in Fig. 2. The following data apply to the example:

Principal distance:	153.19 mm
Approx. flying height:	700 m
Approx. terrain height:	100 m
Patch size:	256 by 256 pixels
Geometric resolution:	30 µm
Radiometric resolution:	256 grey levels

Figure 2 *One image patch out of at least two, typical as input to the system.*

3.3 Region Segmentation

In most related works, e.g. (Roux *et.al.* 1994, Lin *et.al.* 1995), edges or lines are extracted from the image(s) for further analysis. Often, these lines are the only sources of information for solving the task, thus disregarding the original images and their regional information. Lines and regions, extracted by independent algorithms, have been used in combination by e.g. (Jamet *et.al.* 1995). A unified approach for simultaneous extraction of point, line and region areas, has been developed by Förstner (Förstner 1994). The lines are not constrained to connect to each other in the extraction process, but the topology may be determined by creation and analysis of a feature adjacency graph (Fuchs et al, 1995).

We believe that regional information provide helpful clues in automated 3D description in two ways. First, many descriptors can be derived from regions, but not from line segments. Such descriptors are e.g. perimeter, area, average grey level, and variance. Secondly, by the requirements that regions should be closed and non overlapping, the topology between features is known. That the topology is known is manifested in the ability to put queries of the type: which regions neighbour region *x*, which boundaries circumvent region *x*, which two regions are subdivided by boundary *y*, etc. In short, the description of an image by regions and their boundaries admits a compact transfer of most of the information content to higher level processes. This information may be a good help in resolving 3D topology.

Since the objects we are interested in describing mainly consist of linear boundaries, it is recognised that discontinuities should be represented by straight lines.

In consequence, a feature extractor was developed, that combines region and line extraction, by representing the region boundaries by vectors (Wiman 1995). Each region is described by its average grey level (thus assuming a piece-wise constant image model) and noise, assumed to be normal distributed. Each boundary between two regions consists of a set of connected vectors. Given this representation, the feature extractor, called Object Related Image Segmentation (ORIS), uses the Minimum Description Length (MDL) principle (Rissanen 1978) to determine the optimal partitioning of the image. Two actions can be performed by ORIS; (*i*) to simplify a boundary by eliminating vector break points, and (*ii*) to merge regions by eliminating the, possibly simplified, boundaries. The first action may be seen as an algorithm for polygon approximation including radiometry. The second action is similar to traditional region merging. The two actions are iterated until no further action can be made. The segmentation result on all six images used in our example is shown in Fig. 3.

3.4 Searching for the main directions

It is assumed that most of the image lines of the building are horizontal or vertical in object space. Three *main directions* are determined:

- The vertical direction is á priori fully known in the 3D space with a direction vector n_3=[0,0,1].
- Two directions, n_1 and n_2, are known to be horizontal and thus perpendicular to n_3 and also perpendicular to each other. The direction vectors thus have one degree of freedom with n_1 =[x1,y1,0] and n_2=[-y1,x1,0].

The two directions in the XY-plane can be found by forming planes consisting of the line segment and the projection centre and intersecting these planes from one or several images with the XY-plane (Fig. 4). Again, the advantage of accumulating the histogram in object space is that the estimate is based on line segments from all images simultaneously

The accumulated histogram for the six images are shown in Fig. 5. The two peaks are located simultaneously under the condition of being perpendicular. It is assumed that only one object is analysed at the time.

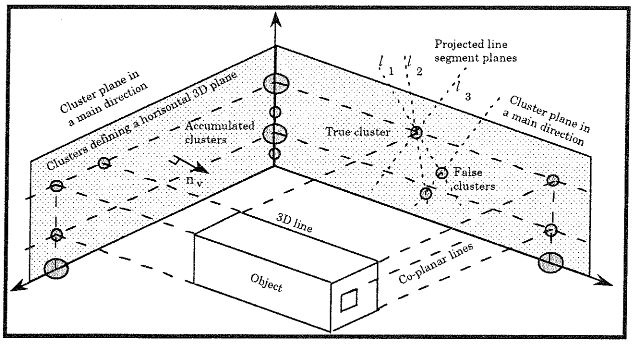

Figure 6. *Clustering on the main direction planes*

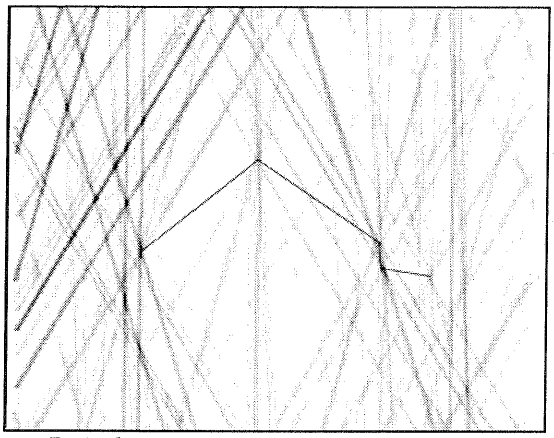

Figure 7. *A clustering with line segments from six images, profile from house as overlay*

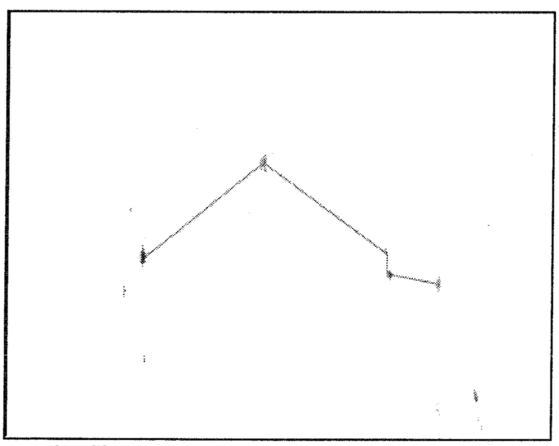

Figure 8. *Clusters with at least five intersecting lines, profile from house as overlay*

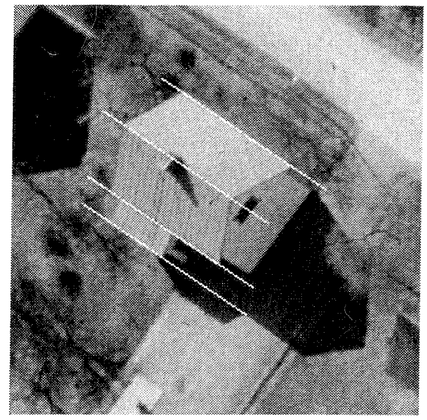

Figure 9. *Found 3D lines projected on one image.*

3.6 Finding 3D Planes

We are now facing a small number of reliable, horizontal 3D lines. The lines are in principle endless, but here truncated to a reasonable object size. The lines may be pair-wise parallel, or, when projected onto the XY-plane, perpendicular. The pairs of lines that are parallel can define a 3D plane. Each of these planes is a hypothesised object plane, regardless if it overlaps other plane hypothesis or not. Each of the plane hypothesis is back-projected to each image, generating a 2D window. The lines of the region boundaries that fall completely within the window are analysed with respect to perpendicularity. In practice, the window is enlarged somewhat (in the example corresponding to appr. 0.5 m) to include lines just nearby the window. Examples of one correct and one false 3D plane hypothesis, projected onto one image and including the lines enclosed by the windows are shown in Fig. 10.

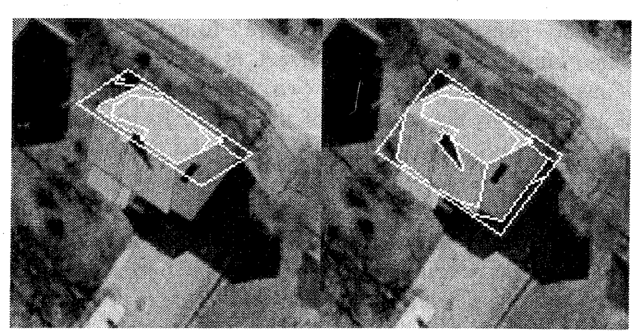

Figure 10. *One correct plane window (left) and one false (right) including the 2D lines from the image that are circumvented by the respective window. The window is somewhat enlarged to include nearby lines. The same procedure is repeated for all plane hypothesis and all images.*

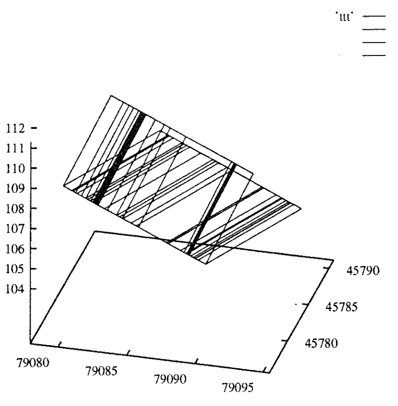

Figure 11. *The tilted, correct plane and the horizontal, false plane from Fig. 10, including the found 2D lines that fit to the corresponding plane and intersect the horizontal lines in right angles. Two structures are clearly visible for the correct plane, whereas the perpendicular lines for the false plane are fewer and more randomly distributed.*

All the 2D lines inside a window are tested if they can (*i*) fit into the 3D plane and (*ii*) intersect the parallel, horizontal lines perpendicularly. If so, the intersection point (due to the constraints only a one-parameter position on one of the 3D lines) is computed.

Fig. 11 shows all the perpendicular lines, from all images, that were found for the correct and the false plane hypothesis, respectively, in Fig. 10. It is noted that less lines were found for the false, horizontal plane, than for the correct, tilted plane. This is because the only real structure that can fit the plane and intersect the horizontal lines perpendicularly are the base lines of the shorter facades, but they are only visible in a few images. For the correct plane more lines were found and, more importantly, many lines intersect at approximately the same position. There are two clear maxima, which correspond to the outlines of the plane. In general, the more perpendicular lines that intersect the two horizontal lines of a plane at approximately the same point, the larger is the probability that there is an actual perpendicular feature causing the lines. One may assume that if the perpendicular lines intersect the horizontal lines randomly the plane is false, whereas if there is at least one pronounced intersection point, the plane may be correct. The same reflections can be made for the other true and false planes.

Radiometry provides further evidence of an actual perpendicular line. The two radiometric criteria used here are (*i*) that the area at one side of the perpendicular line, the one inside the 3D plane, should be radiometrically homogeneous over all images and (*ii*) that the contrast of each contributing 2D line should be high. The first statement assumes that the difference in grey tones between the images is small for the same imaged object. The second statement favours large differences in grey tone between the two sides of a line in one image, and consequently suppresses weak lines, which may occur e.g. inside a roof.

The geometric and radiometric criteria are summarised in one measure. This measure is used first to determine the two largest clusters of intersection points for each plane, secondly to select the best non-overlapping planes. Currently, the best non-overlapping plane hypothesis will always be accepted regardless of how small the measure is. Another restriction is, that there need to be two clear intersections, indicating a 3D rectangle, so that U-shapes can not yet be found. For the best non-conflicting planes the intersection points are determined by averaging the intersection points in that cluster, weighted by the 2D lengths of the lines. The best non-overlapping plane hypothesis and their extensions are shown in Fig. 12.

4. DISCUSSION

A system for finding 3D structures using multiple images has been presented. The system is intended to describe three dimensional buildings using aerial images. Rather than attempting to find the building volumes, the task has been limited to describe only the roofs. Vertical walls and their boundaries are generally only visible in a few images, where the building is far from the nadir point. The ability to find these vertical structures is, we believe, significantly increased by first finding the roof, which in general is visible in all images.

Characteristic for the system is its intense use of object space relations, starting from simple (s.a. there are two main directions) and going to more complex (s.a. two parallel horizontal lines intersected by perpendicular lines at two separate points may form 3D rectangles). There is no image-to-image processing involved; image features are accumulated in a common frame in object space, and analysed in this frame.

The system has been illustrated by a simple example, for which the system works excellently. In spite of rather poor accuracy of the feature extraction, the 3D lines are quite accurate, at least partly explained by the use of multiple images. The chimney and its shadow are bridged over by the use of global search for horizontal lines. The roof of the small addition to the main building structure is in reality *not* connected to the horizontal boundary of the main structure, but intersects the vertical wall somewhat under the main roof. It is however unrealistic to hope for such small deviations to be detected by the system; it is even difficult to interpret for a human. More complex buildings would require additional relations to be defined. For example, it is required that there are at least two salient perpendicular intersection of two horizontal, parallel lines for a plane to be accepted. This omits U-shaped planes, characteristics for houses with additions. Also, the best non-conflicting plane hypothesis is always accepted, regardless of how weak the geometric and radiometric evidence is. This may be overcome by thresholding the measure, that is used for comparing plane hypothesis. Radiometric evidence is used moderately, and should perhaps be used earlier in the process, e.g. by weighting 2D lines contributing to 3D line clusters by their contrast. It is also too restrictive, due to possible occlusion, to require that there should be contributing lines from all images to define a 3D horizontal line. Requiring contributions from less images would reduce the reliability of the lines and increase the risk to introduce false horizontal lines, but may still be necessary to include all relevant 3D lines.

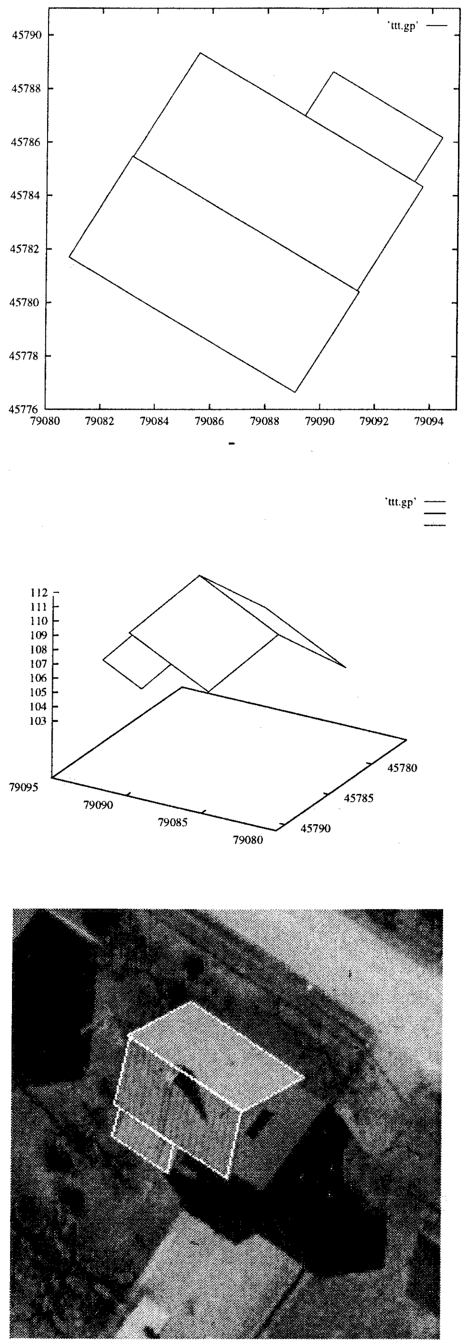

__Figure 12.__ The selected non-conflicting plane hypothesis of the roof. Projected to XY plane (top), 3D view (middle), imposed on one image (bottom).

Although the list of cumbersome restrictions can be made long, there are reasonable solutions to them. The approach seems to be extendible without hitting dead-ends. The system will now be tested on other images.

REFERENCES

Axelsson P. (1994): *Line Correspondence in Aerial Images Using Stochastic Relaxation and Minimum Description Length Criterion,* ISPRS Commission III Symposium, Intl Archives of Photogrammetry and Remote Sensing, Vol 30, part 3, München.

Braun C. (1994): *Interpretation and Correction of Single Line Drawings for the Reconstruction of Objects in Space,* ISPRS Commission III Symposium, IAPRS, Vol. 30, part 3, München.

Fuchs C., Förstner W. (1995): *Polymorphic Grouping for Image Segmentation,* 5th ICCV, Boston, Mass.

Förstner W. (1994): *A Framework for Low Level Feature Extraction,* 4th ECCV, Stockholm, Sweden.

Jamet O., Dissard 0., Airault S. (1995): *Building Extraction from Stereo Pairs of Aerial Images: Accuracy and Productivity Constraints of a Topographic Production Line,* In: Automatic Extraction of Man-Made Objects from Aerial and Space Imagery, Gruen A., *et.al.* (Edt's), Birkhäuser.

Lin C., Huertas A., Nevatia R., (1995): *Detection of Buildings from Monocular Images,* In: Automatic Extraction of Man-Made Objects from Aerial and Space Imagery, Gruen A., *et.al.* (Edt's), Birkhäuser.

McGlone J.C., Shufelt J.A. (1994): *Projective and Object Space Geometry for Monocular Building Extraction,* IEEE Conf. on CVPR, Seattle, Washington.

Rissanen J. (1978): *Modeling by Shortest Data Description,* Automatica, vol. 14.

Roux M., McKeown D.M., (1994): *Feature Matching for Building Extraction from Multiple Views,* IEEE Conf. on CVPR, Seattle, Washington.

Wiman H., (1995): *Object Related Image Segmentation using the Minimum Description Length Principle,* Proc. of 9:th SCIA, Uppsala, Sweden.