PROJECT AMOBE: STRATEGIES, CURRENT STATUS AND FUTURE WORK

Olof Henricsson, Frank Bignone, Wolfram Willuhn, Frank Ade, Olaf Kübler
Emmanuel Baltsavias*, Scott Mason*, Armin Grün*

Communication Technology Laboratory
Swiss Federal Institute of Technology (ETH)
CH-8092 Zurich, Switzerland

* Institute of Geodesy and Photogrammetry
Swiss Federal Institute of Technology (ETH)
CH-8093 Zurich, Switzerland

Commision III, Working Group 2

KEY WORDS: Aerial Image Understanding, Feature Extraction, Building Reconstruction, DEM/DTM, Matching, Colour

ABSTRACT

Automation of Digital Terrain Model Generation and **M**an-Made **Ob**ject **E**xtraction from Aerial Images (AMOBE) is a joint project between the Institute of Geodesy and Photogrammetry (IGP) and the Institute of Communications Technology (Image Science Group) (IKT) at the Swiss Federal Institute of Technology in Zurich. In the project we develop methods and algorithms to detect and reconstruct man-made objects, such as buildings and roads, and to generate Digital Surface Models (DSMs) from high resolution aerial images. Primary attention in AMOBE focuses on high quality reconstruction of buildings as being one of the more predominantly and frequently occurring 3-D man-made objects in high-resolution aerial imagery. In this paper we present our research strategy, current results, and make an outlook onto future work.

1 INTRODUCTION

The reconstruction of houses and other man-made objects in 3-D is currently a very active research area and an issue of high importance to many users of Geographic Information Systems (GIS), including urban planners, architects, and telecommunication and environmental engineers. Manual 3-D processing of aerial images is time consuming and requires the expertise of highly qualified personal and expensive instruments. Therefore, the necessity to interpret, classify and measure aerial images and to integrate the results in GIS is more urgent than ever. It is generally acknowledged that good data is the most valuable and the most needed component, prior to computer hardware, software and user interface.

Methods for computer supported interpretation of aerial imagery have progressed in the wake of Computer Vision and Digital Photogrammetry. The proceedings of the Ascona'95 workshop at Monte Verita, Switzerland gives a good account of the current state-of-the-art [Grün *et al.* 1995]. The objective of the AMOBE project is to develop procedures for extracting quantitative 3-D information of sparsely built-up regions even under difficult terrain conditions. This goal provides complex technical and conceptual challenges and distinguishes the project from existing methods which work on smooth terrain without being able to deal reliably with man-made objects. The applications to Swiss scenery are immediate. Here, we present the strategies, current work, and some ideas for future undertakings within the AMOBE project.

In section 2 we present our main strategies. Section 3 presents the characteristics of the acquired data set. Section 4 deals with digital surface/terrain models and color analysis to provide a means to detect and to provide a coarse description of buildings. In section 5 we present our feature extraction method and show how to relate pairs of contours to each other by similarity in position, orientation, and their photometric and chromatic region attributes. A novel approach to reconstruct complex houses is presented in section 6; it includes segment stereo matching, coplanar grouping and modeling in 3-D. Finally, we present some ideas of future work.

2 STRATEGIES AND GENERAL FRAMEWORK

Although the research topics in the AMOBE project span a large spectrum from Computer Vision to Photogrammetry, attention is focussed on 3-D reconstruction of buildings, and in particular on residential houses. Buildings are the most predominantly and frequently occurring 3-D man-made objects in high resolution aerial images, and their reconstruction requires many components, such as camera models, image processing, matching, texture and color modeling, geometric processing and reasoning, as well as object modeling. The employed imagery is assumed to be digitized photogrammetric color photography. With this aerial imagery primarily building roofs, and not walls, can be reconstructed.

The main features of our strategy for 3-D house reconstruction are illustrated in Fig. 1. The most important feature of our strategy is the mutual interaction of 2-D and 3-D procedures at all levels of processing. This interaction is important since neither 2-D nor 3-D procedures *alone* are sufficient to solve the problems. Three-dimensional information, such as Digital Surface Models and 3-D edges, should therefore be derived as soon as possible. Because 3-D information is available, object modeling can be done in 3-D, right from the beginning. Whenever 3-D features are incomplete or entirely missing, 2-D information should be used to infer the missing information.

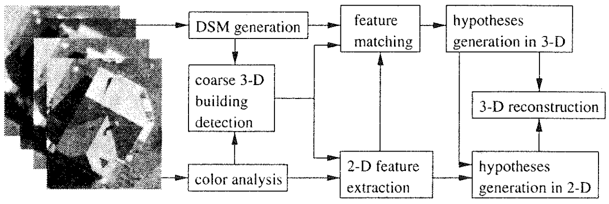
DSM generation
feature matching
hypotheses generation in 3-D
coarse 3-D building detection
3-D reconstruction
color analysis
2-D feature extraction
hypotheses generation in 2-D

Figure 1: Strategy employed in the AMOBE project.

Several roofs of the residential houses in Fig 5A are neither flat nor rectilinear, not even in object space. To reconstruct the roofs of such complex houses, we have developed a procedure that relies on hierarchical hypothesis generation in both

2-D and 3-D. Because low-level feature extraction is error prone, we try to combine as many cues as possible to achieve redundancy. Our general assumption is that the complete roof consists of a set of planar patches that mutually adjoin along their boundary. Our planar primitive can have an arbitrarily complex polygonal boundary, i.e. we do not require rectilinear roof shapes. Considering the detection and discounting of disturbances along the roof boundary, such as chimneys and shadows, the remaining edges should have perceptually uniform color properties. By modeling not only the geometry of the roof, but also the spectral properties along its boundary we can handle complex roof shapes.

In the current approach, the user is only asked to provide a rough location of the houses in one image, the subsequent 3-D reconstruction is fully automatic. The combination of color and DSMs can provide the positions of the houses, as well as a rough 3-D description. This strategy focuses on building reconstruction, however, the concept is general and can be augmented to also include other man-made objects such as roads and bridges.

3 DATA SETS USED IN AMOBE

A data set[1] from Avenches (Switzerland) was acquired for use in the AMOBE project [Mason *et al.* 1994]. The data set consists of a residential and an industrial scene with the following characteristics: 1:5,000 image scale, near-vertical aerial photography, four-way image overlap, color imagery, geometrically accurate film scanning with 15 microns pixel size, precise sensor orientation, and accurate ground truth including a Digital Terrain Model (DTM) and buildings. The manually measured CAD models of the buildings are important to evaluate our results. In Fig. 5A-C we show the residential data set, including the digital surface model and the manually measured CAD models of the houses. The houses shown in the residential scene are representative for Europe and in particular for Switzerland. Since false color infrared images (CIR) were not available for the Avenches data set, we used an additional data set of an urban area with mostly detached buildings for these experiments.

4 USE OF DSMS AND COLOR SEGMENTATION

4.1 Use of Digital Surface Models

Digital surface models are a rich source of information for building detection [Baltsavias *et al.* 1995]:

Building position and separation The approximate position of buildings can be used to guide 2-D feature extraction and grouping, spectral classification and image texture analysis, thereby reducing processing time. Given the approximate position, the DSMs provide means to separate buildings from other objects that have similar low level cues but different DSM characteristics, e.g. separation of buildings from roads and driveways.

Support in matching DSMs support 3-D feature matching, e.g. they provide approximations and they can be used to reduce the number of candidate matches.

Model selection DSMs provide information which allows the inference of 3-D object hypotheses in model-based building reconstruction. Depending on the accuracy and resolution of the DSM, the following information can be provided: approximate 3-D size and shape, distinction between flat and non-flat roofs, distinction between one-peak, ridge, and horizontal roofs, number of major roof planes, and the distinction between I-, T- L-, U-, and X-type buildings.

Ortho-images and ortho-rectified stereo pairs DSMs can be used in the generation of ortho-images and ortho-rectified stereo pairs, whereby the latter can be used to detect DSM errors [Baltsavias *et al.* 1995].

When buildings adjoin each other, which is often the case in dense urban areas, some of the above DSM usages become more difficult or almost impossible.

The extracted DSM must have high accuracy and sufficient density. We have used commercial packages, which employ area correlation for DSM generation at digital photogrammetric workstations in grid mode either in image or object space. Several blunders close to the building boundaries occur, however, the results are still usable. To avoid loss of buildings with these packages, the DSM should have a grid spacing of 0.25 - 0.5 m. Such dense grid spacing is also necessary to distinguish buildings that are close to each other and to avoid strong smoothing of discontinuities. For the same reasons a small patch size should be used in area-based matching. Better DSMs can be derived by use of feature based matching or its combination with area-based matching [Berthod *et al.* 1995], by the use of multi-photo matching with geometric constraints [Grün 1985, Baltsavias 1991], or from airborne laser scanners.

4.2 3-D Blob Detection

Different methods of extracting 3-D blobs, i.e. possible buildings, from a DSM have been investigated. Morphological operators are sensitive to the choice of the structuring element size, particularly in dense urban areas, and have problems when other DSM blobs are situated close to the buildings, or when the terrain is steep and irregular. A subtraction of the DSM from an existing DTM is simple, but DTMs, if they are available, do not usually have sufficient density and accuracy. A sufficient accuracy is essential in order to detect low buildings. Edge detectors extract most of buildings outlines but they do not deliver closed contours. Other structures with a much smaller height than buildings, such as road borders, are also detected.

The most promising method consists of grouping the DSM heights into consecutive bins (height ranges) of a certain size. It corresponds to cutting equidistant slices through the DSM. Thus, the DSM is segmented in relatively few regions that are always closed and easy to extract. The method can be applied hierarchically using different bin sizes, it is simple and fast, and can be applied globally or locally. The maximum and minimum bin sizes are determined from the known height accuracy of the DSM (e.g. 0.5 - 1 m) and the estimated minimum building height in the image (e.g. 3 - 4 m). The hierarchical approach makes use of coarse bins that detect possible buildings, while the fine bins verify the coarse detection, provide information for an approximate building model and separate buildings close to each other. Results of this method are shown in Fig. 2. For details we refer to [Baltsavias *et al.* 1995].

[1] The data set can be acquired by ftp from the authors

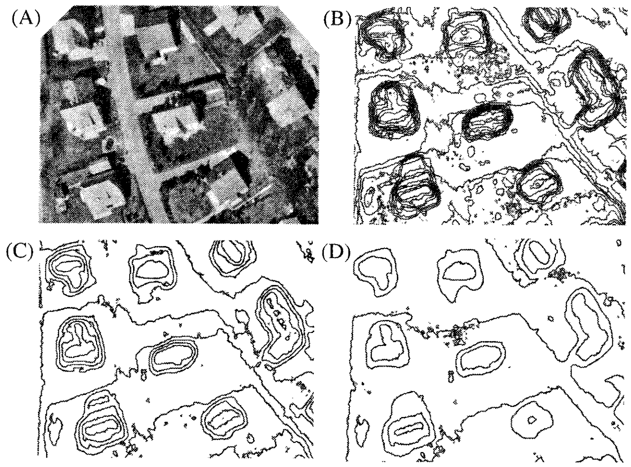

Figure 2: **(A)** ortho-image, **(B-D)** height bins of the Digital Surface Model (DSM) with 1, 2, and 3 meter size (quantization). By decreasing the bin size a better modeling of the buildings is achieved. In addition, gabled roofs, T- and L-shaped buildings, and buildings close to each other can be distinguished.

4.3 Classification of 3-D Blobs

Objects other than buildings will often be detected as blobs, for example trees, bridges/over-passes, transportation means, and big poles. A first elimination of non-building blobs is performed based on the area, height and minimum dimensions of the detected blobs. A further separation can be achieved by using the number and length of extracted straight lines as well as the size and shape of compact homogeneous regions within the projected blobs, the weighted histogram of the local gradient orientation, spectral properties, and context.

Vegetation blobs, in particular trees, are the most prominent non-building blobs that must be detected and eliminated. Apart from using spectral information to separate trees from buildings (see below), we propose a simple procedure which is based on weighted local orientation histograms. A histogram of the local orientations of all edge pixels within the projected blob region is computed. Each entry is weighted with its magnitude. Assuming regularly shaped buildings, the histograms of building blobs will often contain significant peaks 90° apart. Histograms of more complex buildings contain a few additional peaks (usually one or two). On the contrary, histograms of tree blobs are predominantly flat. For details on the approach we refer to [Baltsavias *et al.* 1995].

4.4 Combining Color and False Color Infrared Images with 3-D Blobs

In addition to the above rather simple procedures for blob classification, we have also investigated into the use of color and infrared images together with DSM blobs to separate man-made (MMOs) from natural objects (NOs) [Sibiryakov 1996]. The RGB images are initially transformed into a more suitable color space – the CIE (1976) $L^*a^*b^*$ color space (abbr. CIELAB) [Wyszecki and Stiles 1982]. The CIELAB color space separates the luminant and chromatic components of color and is perceptually uniform. In uniform color spaces, perceptual color differences are computed with Euclidean distances.

In the following analysis we use only the chromatic components a^* and b^*. The lightness component L^* was not used in the classification, because different parts of a roof may have different lightness, however, the same chromatic properties. The CIELAB color space allows us to describe colors more similar to what is perceived by human beings, which is very useful in handling images under non-uniform illumination conditions such as shade, highlight, and strong contrast. A simple classification of the object classes (roads, buildings, vegetation, cars etc.) is not possible using only color images, because the different object classes overlap considerably as can be seen in Fig. 4A. Especially objects with low chroma, such as roads, shadows, trees, brown or grey roofs, and patios, overlap in their chromatic components.

A comparison between color and false color infrared images (CIR) showed, as expected, that a separation between natural and man-made objects is easier with CIR images. Figure 4B shows the main clusters for the CIR image in the a^* and b^* color components. CIR images have essentially three major spectral classes: vegetation, man-made objects and bare soil, and water. Roughly, the man-made objects form a high and well-separated peak in the histogram of the a^* channel, thus a simple thresholding can be used for their detection. With such a simple approach, bare soil mixes with man-made objects. Color images have a larger spectral variability and are more appropriate than CIR images when a larger number of classes must be determined.

An unsupervised classification based on a simple k-mean clustering is used for both the color and CIR images, where k is the number of predefined classes. The clustering is based on minimum distance (Euclidean distance was used). The classification is iterative in a binary tree fashion and it employs 1 - 3 classification steps, see Fig. 3.

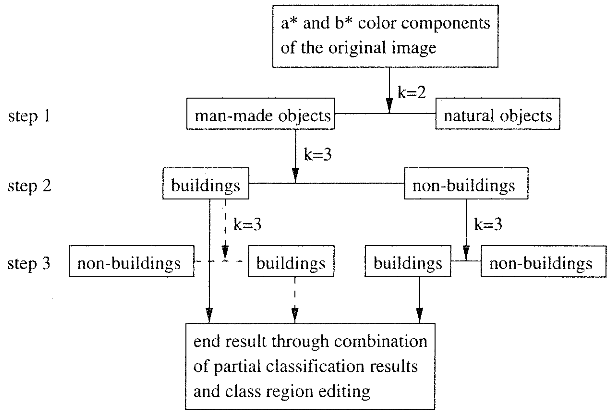
a* and b* color components of the original image
k=2
step 1
man-made objects
natural objects
k=3
step 2
buildings
non-buildings
k=3
k=3
step 3
non-buildings
buildings
buildings
non-buildings
end result through combination of partial classification results and class region editing

Figure 3: An iterative, binary-tree classification scheme using unsupervised k-mean minimum distance clustering, where k is the number of predefined classes. When k=3, the third class is the rejection class. The binary tree can be further densified or reduced. The dashed lines show optional processing steps.

In the first classification step, the number of classes is 2 (NO and MMO). In the second step the MMO image is selected and k=3, i.e. buildings, non-buildings, and a third rejection class. The aim of this step is to separate buildings from other MMO, especially roads, or NO that correlate with MMO like bare soil. This is successful to a large extent but some class mixing does occur, e.g. some buildings are still included in the non-building class. Thus, a third classification step is

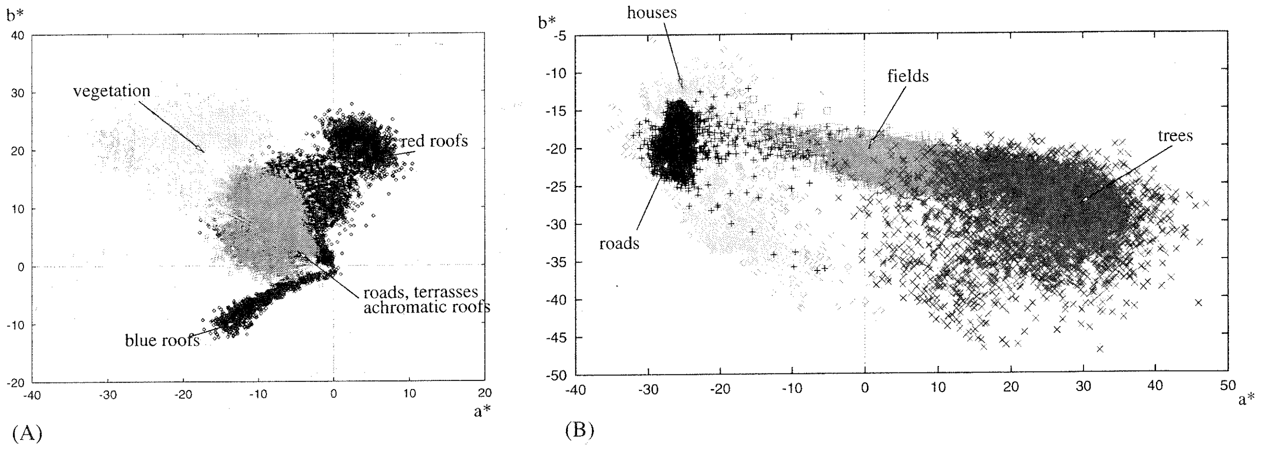

Figure 4: Chromatic clusters of object classes in their (a^*, b^*) color components. **(A)** using a color image, and **(B)** using a false color infrared image. The separation between man-made and natural objects is easier with infrared images.

performed where each of the images for buildings and non-buildings is classified using k=3 as in the second step. Finally, the two building images from the third step are combined, some regions are deleted based on their area, shape and minimum dimensions, and small holes in the remaining regions are filled in.

The MMO class in the first classification step or the building class in the classifications of the second and third steps can be found by a procedure, which is based on the projection of the DSM blobs in the images using the known interior and exterior image orientation. Since the projected blobs might have holes, these are filled in by morphological operations. The MMO or the building class is the one that includes the majority of the projected blob pixels. Other procedures are possible that work also when no DSM is available, such as classification based on color or infrared images, and the characteristics of the edges included in the class regions, such as straightness, length, and orientation. When these additional cues are used, the classification can actually stop after the first step.

As an optional step, a refinement of the detected building outline can be performed. DSM blobs usually do not perfectly outline the building. Therefore, a refinement procedure is applied by using the classification results of the classes MMO and buildings together with the DSM and edges. This optional step is described in [Sibiryakov 1996].

Figure 5A shows the residential scene of the Avenches data set. Figure 5D shows the results of color classification for the building class. Figure 5E shows the result of color classification for the MMO class and the projected DSM blobs, with NOs shown in black (the upper right house has no blob because it was outside the DSM). Figure 5F shows the result of building detection after combining the spectral classification and the DSM blobs and refining the outline of the blobs by the use of edges. It can be noted that edges may introduce some small spurious house elements. With our test images, buildings were always included in the MMO class. Almost all buildings were included in the building class.

The above results demonstrate that an approximate detection of isolated buildings can be performed with practically no human interaction. However, when buildings are connected, human interaction is often required to indicate the outline of the buildings.

5 FEATURE EXTRACTION AND RELATIONS

All intermediate and high level processing in our project needs low-level features, in particular straight contours. In this section, we present methods to generate an attributed contour graph and we show how to relate pairs of straight contours based on similarity in position, orientation, and in photometric and chromatic attributes. The attributed contour graph and the similarity relations form an excellent collection of symbolic data for further processing.

5.1 Edge Detection and Aggregation

Based on the assumption that object boundaries are generally smooth and mostly contrast defined, much effort has been devoted to design suitable edge detectors that reliably detect these 1-D features. The presented work does not require a particular edge detector, however, we believe it is wise to use the best operator available to obtain the best possible results. For this reason, we use the SE energy operator recently presented in [Heitger 1995]. The operator produces a more accurate representation of edges and lines in images of outdoor scenes than traditional edge detectors due to its superior handling of interferences between edges and lines, for example at sharp corners. The edge and line pixels are then linked to produce a contour graph by using the algorithm in [Henricsson and Heitger 1994]. The result in Fig. 6B is a high quality representation of the contours connected to each other at junctions, corner and other important 2-D points.

5.2 Contour and Region Attributes

The contour graph contains only basic information about geometry and connectivity. To increase its usefulness, attributes are assigned to each contour and end-point. The attributes assigned to contours reflect either properties of the contour or region properties on either side. The latter are denoted region attributes and are attached to the generating contour. A region is constructed on both sides of each contour by a translation of the original contour in the direction of its normal. When neighboring contours interfere with the constructed region, a truncation mechanism is applied. For details on the construction of the regions we refer to [Henricsson 1995].

Since each flanking region is assumed to be fairly homogeneous (due to the way it is constructed), the data points contained in each region tend to concentrate in a small region of the color space, however, outliers must also be ac-

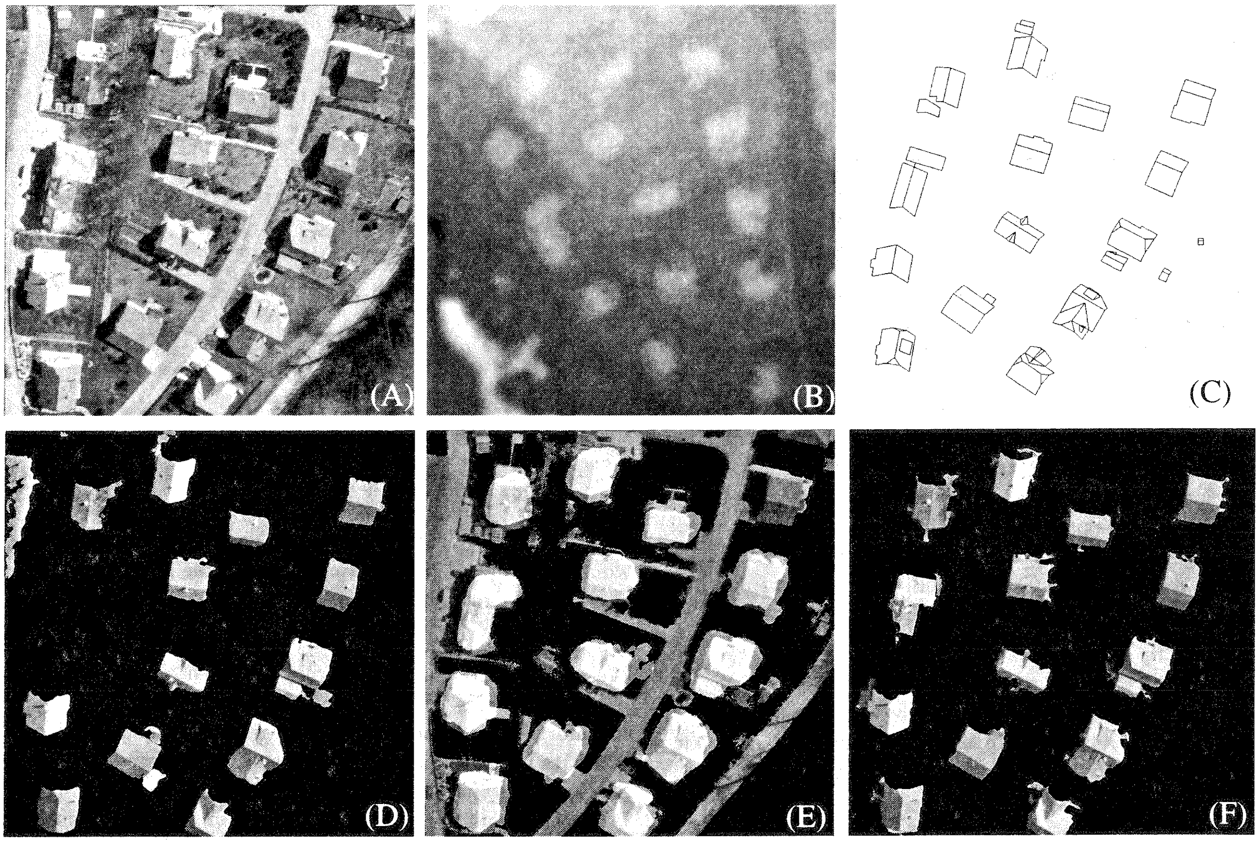

Figure 5: **(A)** the original (RGB) image of the residential scene, **(B)** the Digital Surface Model (DSM), **(C)** the manually measured CAD models of buildings, **(D)** the result of color classification alone. The pixels of the building class are shown (all other pixels are black) **(E)** the result of classification for the MMO class and the projected DSM blobs (in grey), with NO shown in black. The upper right house is not included in the DSM. **(F)** the result of building detection after combining the spectral classification and the DSM blobs and refining the outline of the blobs by the use of edges.

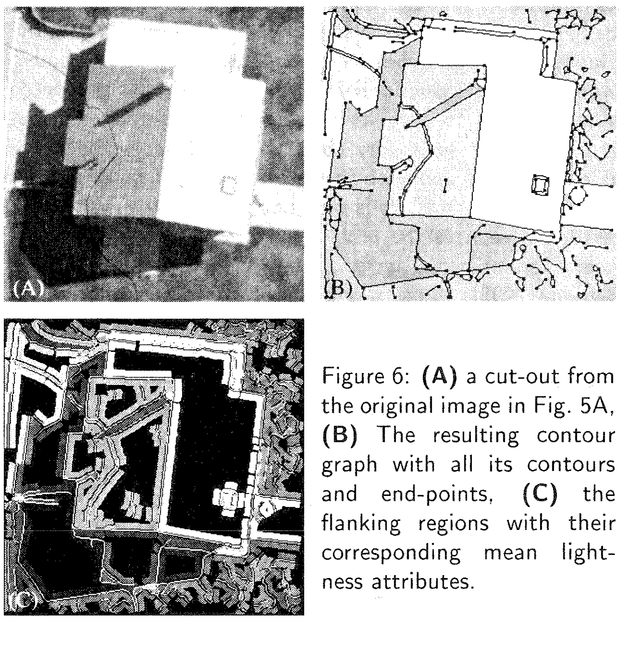

Figure 6: **(A)** a cut-out from the original image in Fig. 5A, **(B)** The resulting contour graph with all its contours and end-points, **(C)** the flanking regions with their corresponding mean lightness attributes.

counted for. The photometric and chromatic attributes are computed for each flanking region using the CIE $L^*a^*b^*$ color space. The photometric region attributes are computed from the lightness component L^* , whereas the chromatic region attributes are derived from the a^* and b^* color components. First, the mean lightness and its standard deviation are estimated by applying the Minimum Volume Ellipsoid (MVE) estimator [Rousseeuw and Leroy 1987] on the L^* data. The inliers in L^* are then used to robustly estimate the mean vector and the scatter matrix for the chromatic components (a^*, b^*). Again, the MVE estimator is used, however, with two variables (a^*, b^*). The estimated scatter matrix of the chromatic cluster is then diagonalized. The chromatic attributes are thereby represented by the mean vector and the two eigenvalues of the scatter matrix. In Fig. 6C we show the mean lightness L^* of each flanking region. The photometric and chromatic region attributes are used to compute similarity relations (next section) and in segment stereo matching (section 6.1).

5.3 Contour Similarity Relations

Although geometric regularity is a major component in the recognition of man-made structures, neglecting other sources of information that corroborate the relatedness among straight contours imposes unnecessary restrictions on the approach. A popular means to relate pairs of straight contours is by geometric primitives, such as parallel, corner, and collinear [Fua and Hanson 1991, Kim and Muller 1995, Lin *et al.* 1995, Henricsson 1995]. Relating contours by defining geometric primitives requires many parameters and specific object models, for example flat and rectilinear roofs. To be able to handle arbitrarily complex roof shapes, we instead propose to form a measure that relates contours based on similarity in position, orientation, and photometric and chromatic properties [Henricsson and Stricker 1995].

Following [Henricsson and Stricker 1995], for each straight contour segment we define two directional contours pointing in opposite directions. Two directional contours form a contour relation with a logically defined interior. For each contour relation we compute four normalized scores based on similarity in position, orientation, and in photometric and chromatic attributes. The final similarity score of a contour relation is the sum of the individual similarity components. A high similarity score proposes that the two contours belong to the same object boundary. A few selection procedures, which are based on local competition on the computed similarity scores, are subsequently applied to yield a small number of similarity relations.

By relaxing the geometrical arrangement of two straight contours, we can handle arbitrarily complex polygonal shapes. These similarity relations are extensively used in coplanar grouping (section 6.2) and to hypothesize the roof boundary (section 6.3).

6 AUTOMATIC HOUSE RECONSTRUCTION

We present a novel approach to reconstruct complex residential houses from sets of aerial images. To solve this problem, we have developed a procedure that relies on hierarchical hypothesis generation, see Fig. 7. The procedure starts with a multi-image coverage of a site, extracts 2-D edges from a source image, computes corresponding photometric and chromatic attributes, and their similarity relationships. Using both geometry and photometry, it then computes the 3-D location of these edges and groups them to planes. In addition, 2-D enclosures are extracted and combined with the 3-D planes to instances of our roof primitive – the 3-D patch. All extracted hypotheses of 3-D patches are ranked according to their geometric quality. Finally, the best set of 3-D patches that are mutually consistent is retained, thus defining the reconstructed house. This procedure has proven powerful enough so that, in contrast to other approaches to generic roof extraction, e.g. [Fua and Hanson 1991, Roux and McKeown 1994, Lin *et al.* 1995, Haala and Hahn 1995, Kim and Muller 1995], we need not assume the roofs to be flat or rectilinear or use a parameterized building model.

Note that, even though geometric regularity is the key to the recognition of man-made structures, imposing constraints that are too tight, such as requiring that edges on a roof form ninety degrees angles, would prevent the detection of many structures that do not satisfy them perfectly. Conversely, constraints that are too loose will lead to combinatorial explosion. Here we avoid both problems by working in 2-D and 3-D, grouping only edges that satisfy loose coplanarity constraints, and weak 2-D geometric and similarity constraints on their photometric and chromatic attributes. None of these constraints is very tight but, because we pool a lot of information from multiple images, we are able to retain only valid object candidates.

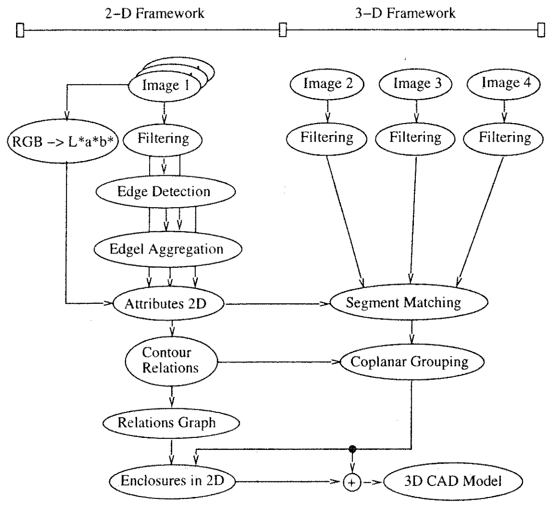

Figure 7: A hierarchical framework, a feed-forward scheme, where several components in the 2-D scheme mutually exchange data and aggregates with the 3-D modules.

We view the contribution of this approach as the ability to robustly combine information derived from edges, photometric and chromatic area properties, geometry and stereo, to generate well organized 3-D data structures describing complex objects while keeping the combinatorics under control. Of particular importance is the tight coupling of 2-D and 3-D analysis. In section 5 we described the 2-D framework, and in the following sections we present the 3-D framework and the combination of 2-D and 3-D processing.

6.1 Segment Stereo Matching

Many methods for edge-based stereo matching rely on extracting straight 2-D edges from images and then matching them. These methods, although fast, they have one drawback: if an edge extracted from one image is occluded or only partially defined in one of the other images, it may not be matched. In outdoor scenes, this happens often, for example when shadows cut edges. Another class of methods [Baltsavias 1991] consists of moving a template along the epipolar line to find correspondences. This can be extended through the introduction of camera models and geometrical constraints to a multi-image (feature/template based) matching technique. Very promising results have been obtained with this approach in close range applications [Grün and Stallmann 1991]. It is much closer to correlation-based stereo and reduces the problem described above. We propose a variant of the latter approach for segment matching [Bignone 1995]. Edges are extracted with the methods in section 5 from *only one* image (the source image) and are matched in the other three images by maximizing an "edginess measure" along the epipolar line. The edginess measure is a function of the gradient in the other images. Geometric and photometric constraints are also used to reduce the number of mismatches. Each matched 3-D segment has a virtual link to its generating 2-D contour, and vice versa.

The photometric constraint consists of computing the photometric region attributes as defined in section 5.2 after a photometric equalization of the images. The photometric consistency means that the photometry in areas that pertains to at least one side of the correspondences should be similar across images. Figure 8A shows all 78 computed 3-D segments of the house in Fig. 6. The details of the algorithm are described in [Bignone 1995].

A more classical approach in stereo matching is under development. This approach simultaneously matches the edges extracted from the four images. To cope with both broken edges and edges with different length, epipolar stripes are defined to first of all relax the epipolar constraint and secondly to reduce the search space. Similar to the above approach, geometric and photometric constraints are used to reduce the number of false matches. The preliminary result of the new stereo matching is shown in Fig. 8B.

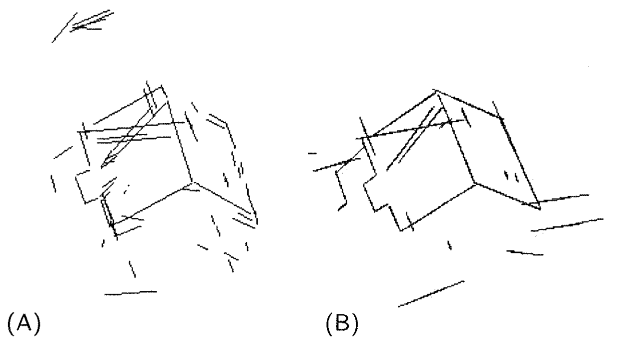

Figure 8: **(A)** the matched 3-D segments (notice the false matches), using the edginess approach, and **(B)** the novel simultaneous stereo matching among all four images. Notice that the new results are more complete than the old ones.

6.2 Coplanar Grouping of 3-D Segments

To group 3-D segments into planes, we propose a simple method that accounts for outliers in the data [Bignone 1995]. The proposed method explicitly uses the similarity relations from section 5.3 to drive the algorithm. This has the advantage that we only extract planes that are somehow related to similar 2-D contours and hence we largely reduce the number of mismatches in the extracted planes. The algorithm proceeds in two steps similar to the procedure in [Stricker and Leonardis 1995]:

Explore: The exploration generates an initial set of hypotheses. Given the similarity relationships of section 5.3 and the 3-D geometry of the segments, planes are fitted to pairs of related contours that are roughly coplanar. The support of those planes are then extended by iteratively including segments that are related to the hypothesis and that are close enough to the plane. After each iteration the plane parameters are re-approximated.

Merge: The exploration produces a set of plane hypotheses. Because all the contours belonging to the same physical plane may not be related in the sense of section 5.3, this plane may give rise to several hypotheses that must be merged. This is done by performing a statistical test on pairs of parallel planar hypotheses to check whether or not they describe the same plane.

For the house in Fig. 8A the exploration instantiated 13 planes and after the merging step only 6 remained. The 2-D contours of the extracted planes are shown in Fig. 9. A plane consists of a number of 3-D segments, most of which are correctly matched and belong to a planar object part. In Fig. 9, plane E is vertical and plane F consists of a correctly matched contour and a false match (the 2-D contour on the ground).

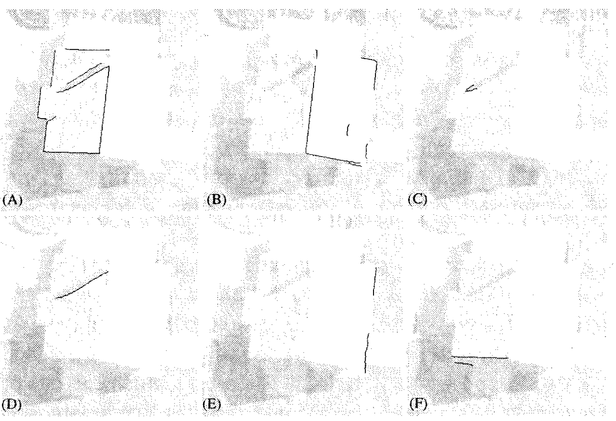

Figure 9: The result of grouping the 3-D segments in Fig. 8A to planes. Plane D consists of two 3-D segments.

As we are interested in the outer boundary of the roofs, we regard those correctly matched 3-D segments that lie inside the roof as disturbances. For example, the shadow contours on plane A and the roof window in plane B, although correctly matched, do not represent 3-D segments of the roof boundary. It is not possible to exclude these disturbing 3-D segments until we have inferred the object boundary of each plane. Some of the planes in Fig. 9 are rejected in the reconstruction of the house, see section 6.4.

6.3 Extract and Select 2-D Enclosures

In the preceding section we described an algorithm that groups 3-D segments into planes. The results in Fig. 9 clearly demonstrate that, in most cases, only a subset of all segments on each plane actually represents the outer boundary of a roof. Furthermore, the planes are often incomplete due to false matches or when the matching algorithm does not find good correspondences for the 2-D contours. The extracted planes themselves are therefore not sufficient to describe the roofs. We therefore need an additional procedure which is capable of inferring the outer boundary of the extracted planes and then rank them according to simple shape criteria [Henricsson and Stricker 1995, Bignone *et al.* 1996].

We propose a graph-based approach similar to [Kim and Muller 1995, Fua and Hanson 1991]. Each similarity relation of section 5.3 defines a node in a relations graph, and compatible nodes represent the graph arcs. A cycle in the graph corresponds to a closed boundary in the image. The strategy consists of grouping related 2-D contours to form 2-D enclosures, thereby using the 2-D contours belonging to the extracted planes to initialize the enclosure finding algorithm. Each 2-D enclosure hypothesizes the boundary for the corresponding plane. The boundaries of the vertical planes are often not entirely visible in single images, hence, we exclude the vertical planes right from the beginning. The tight coupling between the 2-D and 3-D processes plays an important role since we do not need to find *all* possible 2-D enclosures, only those that overlap with non-vertical planes. The major reason for grouping in 2-D instead of in 3-D is that additional and more complete information is available in 2-D. For example, in 2-D *all* straight 2-D contours, their photometric and chromatic attributes and the computed similarity relations are available.

It is not possible to extract the best enclosure among alternatives without considering the neighboring planes and their enclosures. Instead of selecting the best enclosure for each plane we propose to rank the enclosures within each plane according to simple criteria. This is a first important step towards the assembly of 2-D enclosures and planes to form hypotheses of 3-D patches. We assume that each planar part of the roof has a large area, a simple shape, and a large overlap between the contours in the 2-D enclosure and the corresponding 3-D segments of the plane. These three criteria allow us to describe the larger structures of a roof. As we are only interested in ranking the enclosures within each plane, we propose a score for each enclosure, which consists of the product of three relative components: 3-D completeness, relative enclosed area, and relative shape simplicity [Bignone *et al.* 1996].

Figure 10 shows a few extracted 2-D enclosures for the larger planes of the house in Fig. 8A. The algorithm extracted 279 enclosures for the six planes.

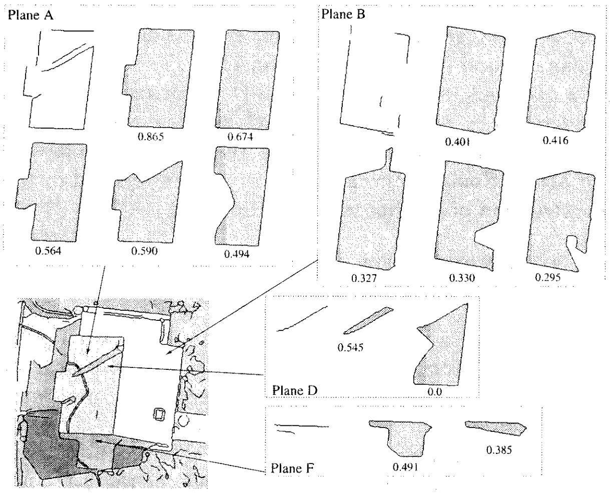

Figure 10: A few representative 2-D enclosures for the planes A, B, D, and F in Fig. 9 with their corresponding scores.

6.4 Assembling Planes and Enclosures to Roofs

Each 2-D enclosure describes a possible boundary description of the corresponding plane. One 2-D enclosure together with one plane form a hypothesis of a 3-D patch. It is reasonable to assume that roofs of residential houses are constructed of adjoining planes. For this reason, only hypotheses of 3-D patches that consistently adjoin with other 3-D patches with respect to the intersection of their planes are retained. In addition, we require that the 2-D contours, that belong to the intersection, are collinear in 2-D. Those 3-D patches that fulfill these constraints are consistent. For example, the 2-D enclosure with the highest score for plane B in Fig. 10, is not consistent and is therefore excluded.

The iterative procedure initially selects a subset of 3-D patches and verifies the total consistency along the boundaries. If one or more 3-D patches do not fulfill this check, they are rejected and new 3-D patches are selected. The first subset of 3-D patches that produce a total consistency among all intersections is the final result. The order of selection is based on the above enclosure score. To obtain the 3-D coordinates of those contours that are contained in the 2-D enclosure but not on the plane, we project their endpoints onto the plane. The result is a complete 3-D boundary for each plane that is likely to describe a roof. Finally, we add artificial vertical walls to the reconstructed roof. The heights of the vertical walls are estimated through the available digital terrain model.

Figure 11C,D show the reconstructed houses in Fig. 11A,B. Notice, that only two planes from Fig. 9 were retained for the final reconstruction. In Fig. 11E,F we superimpose the manually measured CAD model with ours to show the quality of the reconstruction. The accuracy and completeness of the reconstruction will be evaluated in future work.

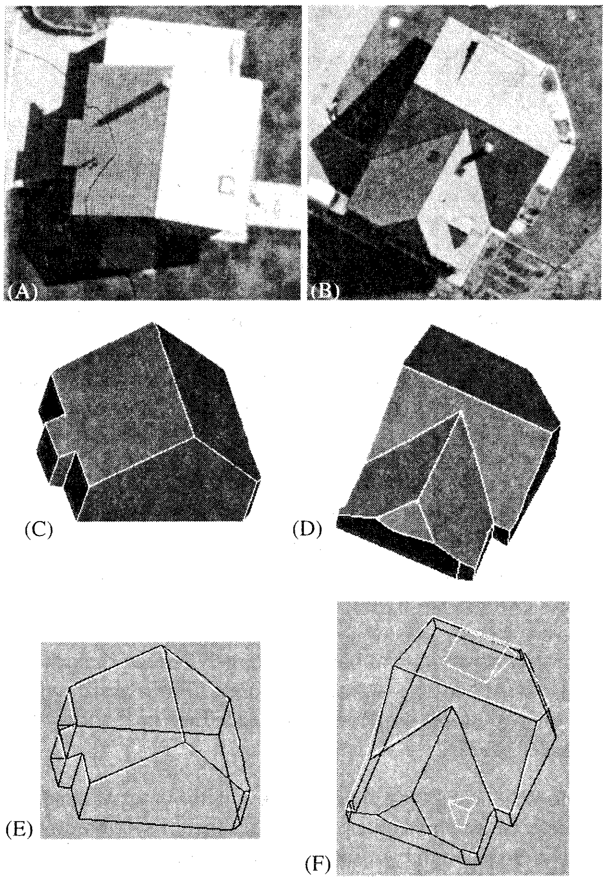

Figure 11: The results of the reconstruction. **(A-B)** the original images, **(C-D)** the reconstructed houses in 3-D and **(E-F)** the manually measured CAD model (white) overlaid on our reconstruction (black).

In Fig. 12 we present our results on the entire residential scene. Eleven of the thirteen roofs are extracted, ten of them with a high degree of accuracy and completeness. The marked house to the right is not complete, since the algorithm fails to extract the two triangular shaped planes, however, the corresponding 2-D enclosures are correctly extracted. The algorithm fails to extract the two upper left houses. The lower of the two is under construction and should not be included in the performance analysis. The upper house is complicated because a bunch of trees cast large shadows on the right roof part. Because of these shadows the algorithm fails to find the corresponding plane, however, the left roof part is correctly reconstructed.

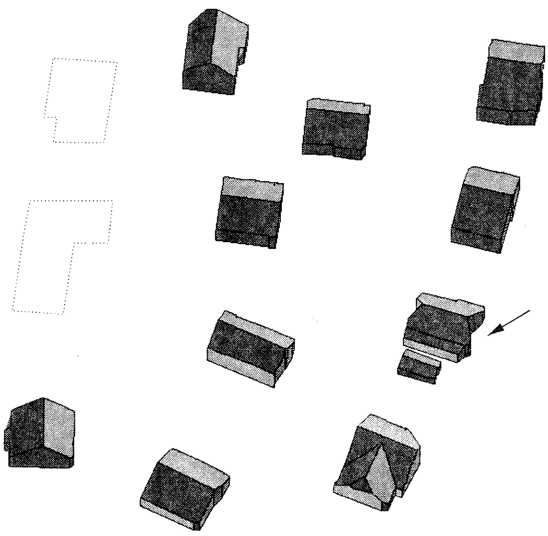

Figure 12: The result of of the 3-D reconstruction on all houses in the scene of Fig. 5A. The artificial vertical walls are added and projected down to the ground. The ground height is estimated through the digital terrain model (DTM). The marked house is not complete, since two triangular patches are missing.

7 RULE-BASED SPATIAL REASONING

In a parallel approach [Willuhn and Ade 1996] we want to incorporate domain-specific knowledge about houses and house roofs into the reconstruction process. We think this step is necessary because, (1) the system should be able to determine the degree of confidence that the reconstructed object is really a house and (2) some peculiarities due to practical or architectural considerations are common in the construction of houses and should be taken into account. Additional constraints, such that decisions take place at all levels of processing, and that previously executed processes may be re-run whenever problems at higher levels occur, imply that we need a system more general than the standard bottom-up. We propose a system that is capable of iteratively activating procedures at different levels and based on a uniform knowledge representation. We have chosen a blackboard architecture with a semantic network as knowledge representation. Due to the variety of possible roof shapes, all knowledge has been coded into rules which have been categorized into the feature, the structure, and the conceptual level. So far only rules at the structure level have been implemented. The generated data from sections 5 and 6.1, i.e. contours, including their attributes and relations, as well as the 3-D contours and the planes are used as initial knowledge in the blackboard.

8 FUTURE WORK

Future work of AMOBE includes not only improvement of each individual component, whenever possible, but also system related and conceptual improvements.

For example, we would like to integrate the operator more actively into the system, especially, for those tasks where the user instantly can provide approximations, or model or contextual knowledge. So far the operator has only been incorporated in the building detection phase. This minimal user interaction works well for the Avenches residential data set, however, in urban scenery fully automatic techniques need to be augmented with operator guidance, at least in the critical phases of the processing.

Up to now the color classification and combination with other cues was performed on only one image. In future investigations all available overlapping images will be used to test the improvement of the classification. Furthermore, our investigations indicate that the building detection is better when more object classes are detected simultaneously. The combination of multiple cues makes such a detection feasible, and a possible extension of our research could be in the detection of all major classes: water, dense forest, separated trees, grass, bare soil, roads and other paved spaces, buildings and shadows. The detection of just trees, buildings and water is important for the reduction of a DSM to a DTM.

The interaction between 2-D and 3-D processing has proven extremely useful, however, its full potential has not yet been investigated. Closely related to the interaction between 2-D and 3-D is the explicit or implicit use of object models. The issues of object modeling has to be investigated further [Mason 1996]. In future work we will validate our algorithms on other data, such as industrial and dense urban scenes. We also plan to improve the data flow by integrating the individual software modules under one joint system.

9 CONCLUSIONS

In this paper we have presented our strategies, the current status of research, and made an outlook onto future work. In the project, we have focused on the 3-D reconstruction of residential houses, as being the most prominent man-made objects in high-resolution aerial images. The approach is highly data-driven, exploits both 2-D and 3-D processing, and reconstructs roofs of houses directly in 3-D. This approach has proven powerful enough so that, in contrast to other approaches of generic roof reconstruction, we can handle more difficult and varying houses.

We have further shown how digital surface models and color classification can be combined to detect buildings and in addition, to provide a coarse description of the buildings. As an alternative approach to house reconstruction, we have also reported on a rule-based system, which is built on a blackboard architecture.

The current status of AMOBE is indeed promising and future undertakings will most certainly profit from the ideas and results presented here.

ACKNOWLEDGEMENTS

We acknowledge the support given to this research by ETH Zurich under project 13-1993-4. The authors would like to acknowledge the work of A. Sibiryakov on color segmentation and blob analysis. We further appreciate the support of P. Fua at SRI International and M. Stricker at IKT, Zurich.

REFERENCES

[Baltsavias *et al.* 1995] E. Baltsavias, S. Mason, and D. Stallman. Use of DTMs/DSMs and Orthoimages to Support Building Extraction. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 199–210, 1995.

[Baltsavias 1991] E. Baltsavias. *Multiphoto Geometrically Constrained Matching*. PhD thesis, Institute of Geodesy and Photogrammetry, ETH Zurich, Mitteilungen 49, 1991.

[Berthod *et al.* 1995] M. Berthod, L. Gabet, G. Giraudon, and J.L. Lotti. High-resolution Stereo for the Detection of Buildings. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 135–144, 1995.

[Bignone *et al.* 1996] F. Bignone, O. Henricsson, P. Fua, and M. Stricker. Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery. In R. Cipolla, editor, *Computer Vision – ECCV'96*, Springer Verlag, Berlin, 1996.

[Bignone 1995] F. Bignone. Segment Stereo Matching and Coplanar Grouping. Technical Report BIWI-TR-165, Institute for Communications Technology, Image Science Lab, ETH, Zurich, Switzerland, 1995.

[Fua and Hanson 1991] P. Fua and A.J. Hanson. An Optimization Framework for Feature Extraction. *Machine Vision and Applications*, 4:59–87, 1991.

[Grün 1985] A. Grün. Adaptive Least Squares Correlation: A Powerful Image Matching Technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, 14(3):175–187, 1985.

[Grün and Stallmann 1991] A. Grün and D. Stallmann. High-accuracy edge-matching with an extension of the MPGC-matching algorithm. In *Proc. of SPIE*, volume 1526, pages 42–55, 1991.

[Grün *et al.* 1995] *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, A. Grün, O. Kübler, and P. Agouris, editors, Birkhäuser Verlag, Basel, 1995.

[Haala and Hahn 1995] N. Haala and M. Hahn. Data fusion for the detection and reconstruction of buildings. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 211–220, 1995.

[Heitger 1995] F. Heitger. Feature Detection using Suppression and Enhancement. Technical Report TR-163, Image Science Lab, ETH-Zurich, Switzerland, 1995.

[Henricsson and Heitger 1994] O. Henricsson and F. Heitger. The Role of Key-Points in Finding Contours. In J.O. Eklundh, editor, *Computer Vision – ECCV'94*, volume II, Springer Verlag, Berlin, pages 371–383, 1994.

[Henricsson and Stricker 1995] O. Henricsson and M. Stricker. Exploiting Photometric and Chromatic Attributes in a Perceptual Organization Framework. In *ACCV'95, Second Asian Conference on Computer Vision*, pages 258–262, 1995.

[Henricsson 1995] O. Henricsson. Inferring Homogeneous Regions from Rich Image Attributes. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 13–22, 1995.

[Kim and Muller 1995] T. Kim and J.P. Muller. Building Extraction and Verification from Spaceborne and Aerial Imagery using Image Understanding Fusion Techniques. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 221–230, 1995.

[Lin *et al.* 1995] C. Lin, A. Huertas, and R. Nevatia. Detection of Buildings from Monocular Images. In A. Grün, O. Kübler, and P. Agouris, editors, *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, Birkhäuser Verlag, Basel, pages 125–134, 1995.

[Mason *et al.* 1994] S. Mason, E. Baltsavias, and D. Stallman. High Precision Photogrammetric Data Set for Building Reconstruction and Terrain Modelling. *Internal Report, Institute of Photogrammetry and Geodesy ETH Zurich*, 1994.

[Mason 1996] S. Mason. 3D Building Reconstruction Using Composites of Surface Primitives: Concept. In *Proc. of 18th ISPRS Congress, Comm. III WG 2*, Vienna, Austria, 1996.

[Rousseeuw and Leroy 1987] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, New York, 1987.

[Roux and McKeown 1994] M. Roux and D.M. McKeown. Feature Matching for Building Extraction from Multiple Views. In *DARPA Image Understanding Workshop*, Monterey, CA, pages 331–349, 1994.

[Sibiryakov 1996] A. Sibiryakov. House Detection from Aerial Color Images. *Internal Report, Institute of Photogrammetry and Geodesy ETH Zurich*, 1996.

[Stricker and Leonardis 1995] M. Stricker and A. Leonardis. ExSel++: A General Framework to Extract Parametric Models. In *Proc. of the 6th Intern. Conf. on Computer Analysis of Images and Patterns, CAIP'95*, Prague, Czech Republic, pages 90-97, 1995.

[Willuhn and Ade 1996] W. Willuhn and F. Ade. Rule-Based Spatial Reasoning for the Reconstruction of Roofs. In ECCV'96 workshop *Conceptual descriptions from images*, Cambride, UK, 1996.

[Wyszecki and Stiles 1982] G. Wyszecki and W.S. Stiles. *Color Science*. John Wiley & Sons, New York, 1982.