Automatic Extraction of Generic House Roofs from High Resolution Aerial Imagery [*]

Frank Bignone, Olof Henricsson, Pascal Fua[+] and Markus Stricker

Communications Technology Laboratory
Swiss Federal Institute of Technology ETH
CH-8092 Zurich, Switzerland

[+]SRI International, Menlo Park, CA 94025, USA

Abstract. We present a technique to extract complex suburban roofs from sets of aerial images. Because we combine 2-D edge information, photometric and chromatic attributes and 3-D information, we can deal with complex houses. Neither do we assume the roofs to be flat or rectilinear nor do we require parameterized building models. From only one image, 2-D edges and their corresponding attributes and relations are extracted. Using a segment stereo matching based on all available images, the 3-D location of these edges are computed. The 3-D segments are then grouped into planes and 2-D enclosures are extracted, thereby allowing to infer adjoining 3-D patches describing roofs of houses. To achieve this, we have developed a hierarchical procedure that effectively pools the information while keeping the combinatorics under control. Of particular importance is the tight coupling of 2-D and 3-D analysis.

1 Introduction

The extraction of instances of 3-D models of buildings and other man-made objects is currently a very active research area and an issue of high importance to many users of geo-information systems, including urban planners, geographers, and architects.

Here, we present an approach to extract complex suburban roofs from sets of aerial images. Such roofs can neither be assumed to be flat nor to have simple rectangular shapes. In fact, their edges may not even form ninety degrees angles. They do tend, however, to lie on planes. This specific problem is a typical example of the general Image Understanding task of extracting instances of generic object classes that are too complex to be handled by purely image-based approaches and for which no specific template exists.

Because low-level methods typically fail to extract all relevant features and often find spurious ones, existing approaches use models to constrain the problem [15]. Traditional approaches rely almost exclusively on the use of edge-based features and their 2-D or 3-D geometry. Although 3-D information alleviates the problem, instantiating the models is combinatorially explosive. This difficulty

---

is typically handled by using very constrained models, such as flat rectilinear roofs or a parameterized building model, to reduce the size of the search space. These models may be appropriate for industrial buildings with flat roofs and perpendicular walls but not for the complicated suburban houses that can be found in scenes such as the one of Fig. 1.

It has been shown, however, that combining photometric and chromatic region attributes with edges leads to vastly improved results over the use of either alone [6, 11]. The houses of Fig. 1 require more flexible models than the standard ones. We define a very generic roof primitive: we take it to be a 3-D patch that is roughly planar and encloses a compact polygonal area with consistent chromatic and luminance attributes. We therefore propose an approach that combines 2-D and 3-D edge geometry with region attributes. This is not easy to implement because the complexity of the approach is likely to increase rapidly with the number of information sources. Furthermore, these sources of information should be as robust as possible but none of them can be expected to be error-free and this must be taken into account by the data-fusion mechanism.

Figure 1 Two of the four registered 1800×1800 images that are part of our residential dataset (Courtesy of Institute of Photogrammetry and Geodesy at ETH Zürich).

To solve this problem, we have developed a procedure that relies on hierarchical hypothesis generation, see Fig. 2. 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 infinite planes. In addition, 2-D enclosures are extracted and combined with the 3-D planes to instances of our roof primitive, that is 3-D patches. 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 are retained, thus defining a scene parse. This procedure has proven powerful enough so that, in contrast to other approaches to generic

roof extraction (e.g. [14, 6, 4, 13, 7, 12]), 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, 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.
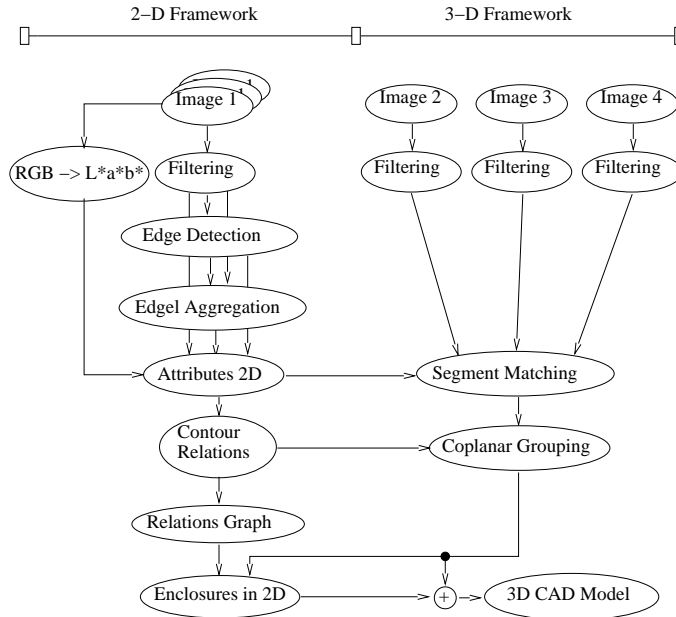

Figure 2 Our hierarchical framework, a feed-forward scheme, where several components in the 2-D scheme mutually exchange data and aggregates with the 3-D modules.

We view the contribution of our 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.

For our experiments, we use a state-of-the-art dataset produced by the Institute of Geodesy and Photogrammetry at ETH Zürich. It consists of a residential and an industrial scene with the following characteristics: 1:5,000 image scale 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 DTM and manually measured building CAD-models. The latter are important to quantitatively evaluate our results.

Our hierarchical parsing procedure is depicted in Fig. 2. Below we describe each of its components: 2-D edge extraction, computation of photometric and chromatic attributes, definition of similarity relationships among 2-D contours, 3-D edge matching and coplanar grouping, extraction of 2-D enclosures, and finally, generation and selection of candidate 3-D object models. Last, we present and discuss our results.

2 Attributed Contours and their Relations

2.1 Edge Detection and Edgel Aggregation

Our approach is based on grouping contour segments. 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 (suppression and enhancement) recently presented in [8]. 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 aggregated to coherent contour segments by using the algorithm described in [10]. The result is a graph representation of contours and vertices, as shown in Fig. 3B. Each contour has geometric attributes such as its coarse shape, that is *straight*, *curved* or *closed*.

Figure 3 (A) a cut-out 350 × 350 from the dataset in Fig.1, (B) The resulting attributed graph with all its contours and vertices, (C) the flanking regions with their corresponding median luminance attributes. The background is black.

2.2 Photometric and Chromatic Contour Attributes

The contour graph contains only basic information about geometry and connectivity. To increase its usefulness, image attributes are assigned to each contour and vertex. The attributes reflect either properties along the actual contour (e.g. integrated gradient magnitude) or region properties on either side, such as chromatic or photometric homogeneity.

Since we are dealing with fairly straight contours the construction of the flanking regions is particularly simple. The flanking region is constructed by a

translation of the original contour in the direction of its normal. We define a flanking region on each side of the contour. When neighboring contours interfere with the constructed region, a truncation mechanism is applied. In Fig. 3C we display all flanking regions. For more details we refer to [9].

To establish robust photometric and chromatic properties of the flanking regions, we need a color model that accurately represents colors under a variety of illumination conditions. We chose to work with HVC color spaces since they separate the luminant and chromatic components of color. The photometric attributes are computed by analyzing the value component, whereas the chromatic attributes are derived from the hue and chroma components. As underlying color space we use the CIE(L*a*b*) color space because of its well based psychophysical foundation; it was created to measure *perceptual* color differences [16].

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. As we deal with images of aerial scenes where disturbances like chimneys, bushes, shadows, or regular roof texture are likely to be within the defined regions, the computation of region properties must take outliers into account. Following the approach in [11] we represent photometric attributes with the median luminance and the interquartile range (IQR), see Fig. 3C. The chromatic region properties are computed analogously from the CIE(a*b*) components and are represented by the center of the chromatic cluster and the corresponding spreads.

2.3   Contour Similarity Relations

Although geometric regularity is a major component in the recognition of manmade structures, neglecting other sources of information that corroborate the relatedness among straight contours imposes unnecessary restrictions on the approach. We propose to form a measure that relates contours based on similarity in position, orientation, and photometric and chromatic properties.

For each straight contour segment we define two directional contours pointing in opposite directions. Two such directional contours form a *contour relation* with a defined logical interior. For each contour relation we compute four scores based on similarity in *luminance*, *chromaticity*, *proximity*, and *orientation* and combine them to a single similarity score by summation.

Three consecutive selection procedures are applied, retaining only the best non-conflicting interpretations. The first selection involves only two contours (resp. four directional contours) and aims at reducing the eight possible interpretations to less or equal to four. The second selection procedure removes shortcuts among three directed contours. The final selection is highly data-driven and aims at reducing the number of contour relations from each directed contour to only include the *locally* best ones. All three selection procedures are based on analysis of the contour similarity scores. Due to lack of space we refer to [11] for more details.

3   Segment Stereo Matching

Many methods for edge-based stereo matching rely on extracting straight 2-D edges from images and then matching them [1]. These methods, although fast and reliable, 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 [2] consists of moving a template along the epipolar line to find correspondences. It is much closer to correlation-based stereo and avoids the problem described above. We propose a variant of the latter approach for segment matching that can cope with noise and ambiguities. Edges are extracted from *only one* image (the source image) and are matched in the other images by maximizing an "edginess measure" along the epipolar line. The source image is the nadir (most top-view) image because it is assumed to contain few (if any) self-occluded roof parts. Geometric and photometric constraints are used to reduce the number of 3-D interpretations of each 2-D edge. We outline this approach below and refer the interested reader to [3] for further details.
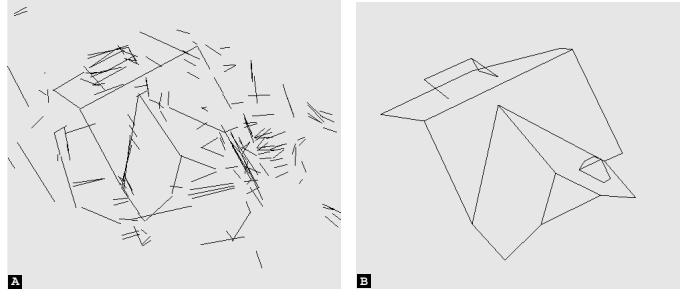

Figure 4 (A) Matched 3-D segments. Notice the false matches.
(B) Manually measured 3-D CAD model.

For a given edge in the source image we want to find the location of its correspondences in the other image. A segment is described by the position of its middle point, its orientation and length. We use the epipolar geometry to constrain the location in the second image so that only 2 parameters are required to describe its counterpart: s_m, the position along the epipolar line, and θ the orientation. The length l, in the other images, is predicted by using (s_m, θ) and the epipolar geometry. For a given s_m and θ, we evaluate its probability of being correct by measuring the *edginess* f. It is a function of the image gradient:

$$f(s_m, \theta) = \sum_{r=-\frac{l}{2}}^{r=\frac{l}{2}} \|G(r)\| \cdot e^{-\frac{(\theta - \theta(r))^2}{2\sigma^2}}$$

where $G(r)$ is the image gradient at r, $\theta(r)$ its orientation. The function f is maximum when the virtual segment lies on a straight edge and decreases quickly

with any translation or rotation. Further, f can be large even if if the edge is only partially visible in the image, that is occluded or broken.

The search for the most likely counterparts for the source edge now reduces to finding the maxima of f by discretizing θ and s_m and performing a 2-D search. In the presence of parallel structures, the edginess typically has several maxima that cannot be distinguished using only two images. However, using more than two images, we can reduce the number of matches and only keep the very best by checking for consistency across image pairs.

We can further reduce the hypothesis set by using the photometric edge attributes of section 2.2 after photometric equalization of the images. We compute the 2-D projections of each candidate 3-D edge into all the images. The image photometry in areas that pertains to at least one side of the 2-D edges should be similar across images. Figure 4 shows all matched 3-D segments as well as the manually measured CAD model for the house in Fig. 3A.

4 Coplanar Grouping of 3-D Segments

To group 3-D segments into infinite planes, we propose a simple method that accounts for outliers in the data. It proceeds in two steps:

- **Explore**: We first find an initial set of hypotheses using a deterministic version of the RANSAC approach [5]: Given the relationships of section 2.3 and the 3-D geometry of the segments, we fit planes to pairs of related contours that are roughly coplanar. We then extend the support of those planes 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**: We now have a set of plane hypotheses. Because all the edges belonging to the same physical plane may not be related in the sense of section 2.3, this plane may give rise to several hypotheses that must be merged. This is done by performing an F-test on pairs of parallel planar hypotheses to check whether or not they describe the same plane.
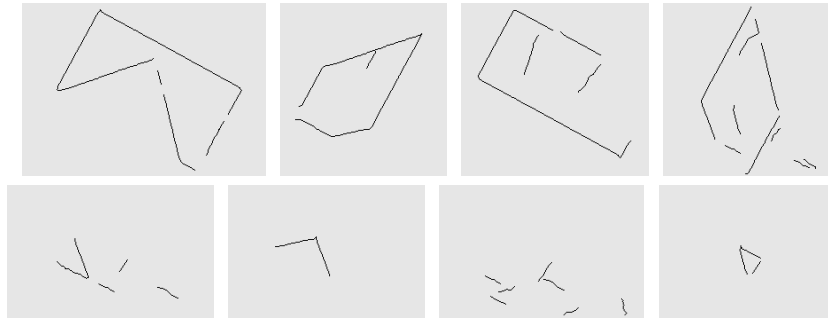

Figure 5 Selection of planes extracted from the 3-D segments of the house in Fig. 3A.

Each plane in Fig. 5 consists of a number of 3-D segments, some of which are correctly matched and do belong to a planar object part. However, quite a few 3-D segments are incorrectly matched and accidentally lie on the plane and other segments, such as the contours on the ground aligned with the roof plane, are correctly matched but do not belong to the object part.

5 Extracting and Selecting 2-D Enclosures

In the preceding section we presented an approach to group 3-D segments into *infinite* planes. However, only a subset of all segments on each plane actually belongs to valid 3-D patches, see Fig. 5. To obtain an ordered list of 2-D contours describing a 3-D patch, we propose to group contours in 2-D where *more complete* data is available and subsequently merge the extracted enclosures with the corresponding planes. The tight coupling between the 2-D and 3-D processes plays an important role; the extracted planes that are *not* vertical initialize the enclosure finding algorithm. We therefore do not need to find *all* possible 2-D enclosures, only those that overlap with non-vertical planes.

We use the edge and region based approach described in [11] since it allows to group contours on other grounds than geometric regularity. The method consists of defining contour similarity relations (section 2.3), which are then used to build a relations graph, in which each cycle defines a 2-D enclosure. At last, all extracted 2-D enclosures are ranked according to simple geometric shape criteria.

5.1 Extracting 2-D Enclosures

Instances of 2-D roof-primitives can be found by grouping related contours to polygonal shaped structures. A computationally attractive approach is to build a relations graph and use it to find these structures [6, 12]. By construction each cycle in the relations graph describe an enclosure. Each contour relation define a node in the graph and two nodes are linked together if they have a compatibly directed contour. We use a standard depth-first search algorithm to find cycles in the *directed* relations graph.

The procedure work as follows: select a not already used node *that belongs to the plane* and find all *valid* cycles in the graph given this start node. Pick the next not already used node *on the same plane* and iterate the procedure until there are no more nodes left. A *valid* cycle is a set of directed contours that have a boundary length not exceeding a large value and that does not form a self-loop; the boundary of the enclosure must be compact.

5.2 Selecting 2-D Enclosures

The above algorithm produces for each plane a set of 2-D enclosure hypotheses. To alleviate the fusion of enclosures and planes, we rank the enclosures within each plane according to simple geometrical shape criteria. We assume that each roof part has a compact and simple polygonal shape. In addition we require a large overlap between the contours in the 2-D enclosure and the corresponding 3-D segments of the plane. We propose the following criteria:

Shape simplicity Shape simplicity is defined as number of straight contours required to represent the enclosure boundary (including missing links). Given an error tolerance, we use a standard polygon approximation algorithm to compute the required number of straight lines. The simpler the description of a 2-D enclosure is, the more likely it is that it will describe a roof part.

Shape compactness Compactness is defined as the squared length of the boundary of the enclosure divided by the enclosed area.

3-D completeness The 3-D completeness is defined as the ratio of the length of the 3-D contours that lie on the enclosure boundary and on the plane, with respect to the total length of the enclosure boundary. This measure will be high whenever a large portion of the extracted 2-D contours have correctly matched 3-D segments that lie on the same infinite plane.

Figure 6 shows a few representative 2-D enclosures for the larger planes of the house in Fig. 3A. Two thresholds are applied, one for shape simplicity (≤ 10) and one for 3-D completeness (≥ 0.4). Together with the 3-D patch consistency test in next section these thresholds preclude highly unlikely hypotheses of 2-D enclosures *before* fusing them with planes to hypotheses of 3-D patches.
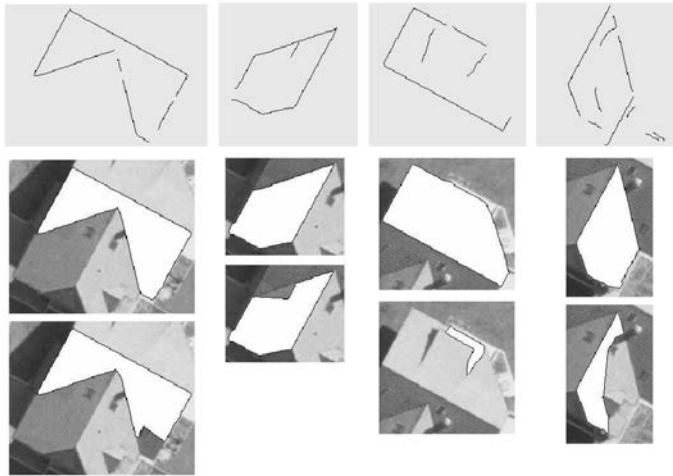

Figure 6 A few representative 2-D enclosures for the larger planes.

6   Finding Coherent 3-D Patches

Each 2-D enclosure describes a possible boundary description of the corresponding 3-D plane. It is reasonable to assume that roofs are usually constructed of adjoining planes. For this reason, only hypotheses of 3-D patches that mutually adjoin with other 3-D patches *along their boundaries* are retained. In addition we require that the 2-D contours belonging to the adjoining boundary of the 3-D patches are collinear in 2-D. Those 3-D patches that fulfill these constraints are consistent.

The iterative procedure initially selects a subset of 3-D patches and verifies the mutual consistency along the boundaries. If one or more 3-D patches do not fulfill this consistency, they are rejected and a new subset of 3-D patches is selected. Moreover, the subset of 3-D patches should be maximally consistent, i.e. have the maximum number of mutually consistent boundaries. The order of selection is initially based on shape simplicity, and in a second step on the product of the normalized compactness and 3-D completeness. 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.

7   Results

We use the presented framework to extract complex roofs of houses in suburban scenes, see Fig. 1. The process is initialized by selecting a rectangular window enclosing the same house in all four images. It has been demonstrated [7] that this initialization procedure can be automated by locating elevation blobs in the digital surface model. After this initialization, the roof is *automatically extracted.*

The roof depicted in Fig. 7A is complex because it consists of several adjoining planar and non-rectangular shapes. The feature extraction finds 171 straight 2-D edges. The segment stereo matching produces 170 3-D segments, and the coplanar grouping extracts 33 infinite planes of which 7 are mutually adjoining and non-vertical. Given these 7 non-vertical planes, the algorithm finds 373 valid 2-D enclosures (resp. 3-D patches). The five 3-D patches in Fig. 7B are finally selected since they maximizes the geometrical shape score and mutual consistency among all 3-D patches. The result is a 3-D CAD model with 3-D segments, 3-D planes and their topology. This procedure yields the main parts of the roof, however, 3-D patches that are not mutually consistent with the final set of 3-D patches, but nevertheless belongs to the house, are not included. One such example is the 3-D patch describing the dormer window in Fig. 7A.
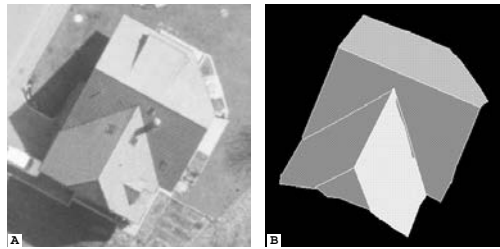

Figure 7 (A) cut-out from the aerial image in Fig.1A, (B) the reconstructed house roof in 3-D.

In Fig. 8 we demonstrate the performance of our approach on the entire scene in Fig. 1. To the automatically extracted CAD models of the roofs we add artificial vertical walls. The height of the vertical walls is estimated through

the available digital terrain model (DTM). Ten of the twelve house roofs are extracted, nine 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. Even manual measuring this house is troublesome. 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.
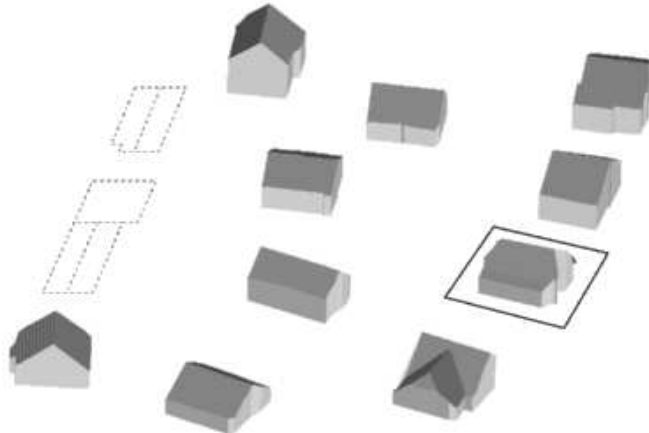

Figure 8 The result of running the algorithm on all houses in the scene of Fig.1. The artificial vertical walls are added and projected down to the ground. The ground height is estimated through the digital terrain model (DTM).

8   Conclusions and Future Work

We have addressed the problem of reconstructing complex objects in real images for which no specific template exists. We have implemented a working framework for the purpose of extracting complicated and generic roofs from sets of aerial images. The framework successfully performs the following tasks: hierarchical integration of multiple data sources, generic shape extraction, and efficient hypothesis generation and selection.

We combine geometry, photometry, chromaticity and stereo information about edges and their flanking regions. We thus make effective use of much of the available 2-D and 3-D information present in several images of a given site. As a result, our procedure is more robust than one that uses only partial information. We further use weak coplanarity constraints together with a generic extraction of 2-D enclosures; a combination that allow us to effectively find instances of 3-D patches. The polygonal shape of these 3-D patches can be arbitrarily complex.

Future work will concentrate on: improvement of each individual module (whenever possible), better exploitation of existing knowledge, e.g. DSMs and

color, better interaction between 2-D and 3-D information, further tests with other data, and improving the selection of 2-D enclosures by also modelling the color homogeneity of the enclosed area with for example MDL encoding, see [6].

Bibliography

[1] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereo vision. In *International Conference on Computer Vision*, June 1987.

[2] E. Baltsavias. *Multiphoto Geometrically Constrained Matching*. PhD thesis, Institute for Geodesy and Photogrammetry, ETH Zurich, December 1991.

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

[4] R.T. Collins, A.R. Hanson, M.R. Riseman, and H. Schultz. Automatic Extraction of Buildings and Terrain from Aerial Images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 169–178. Birkhäuser Verlag,1995.

[5] M.A Fischler and R.C Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications ACM*, 24(6):381–395, 1981.

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

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

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

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

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

[11] 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.

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

[13] C. Lin, A. Huertas, and R. Nevatia. Detection of Buildings from Monocular Images. In *Automatic Extraction of Man-Made Objects from Aerial and Space Images*, pages 125–134. Birkhäuser Verlag, Basel, 1995.

[14] D.M. McKeown, W.A. Harvey, and J. J. McDermott. Rule-Based Interpretation of Aerial Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(5):570–585, 1985.

[15] P. Suetens, P. Fua, and A. Hanson. Computational Strategies for Object Recognition. *ACM Computing Surveys*, 24(1):5–61, 1992.

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