BUILDING RECONSTRUCTION FROM AERIAL IMAGES USING EFFICIENT SEMI-AUTOMATIC BUILDING DETECTION

W. Willuhn, L. van Gool

Computer Vision Lab, Swiss Federal Institute of Technology, 8092 Zurich, Switzerland –
(wolfram, vangool)@vision.ee.ethz.ch

KEY WORDS: Photogrammetry, Building Reconstruction, Building Segmentation, Semi-automation, Aerial Images, Urban Areas

ABSTRACT:

We present a system for the semi-automatic reconstruction of building roofs from aerial images based on a two-step approach. In a first step, a row of buildings is detected, and individual buildings are segmented. After the user identifies the start and end points of the row in one image, region-based and contour-based segmentation algorithms are applied. Both results are combined and fused with 3D data to obtain the 3D bounding boxes of individual buildings. The user can then correct this segmentation by splitting or joining segments or adjusting the width of the building row. The goal is to perform the segmentation with one click or less per building. The second step is the automatic reconstruction of individual building roofs. The system employs a conceptual model based on roof forms, such as gabled or flat roofs, and their constituents as well as a general shape model and is implemented as a rule-based system. It evaluates the input data against these models, builds up a higher-level description of the scene, and hypothesizes and verifies missing parts. With this approach, it is, thus, possible to reconstruct about three quarters of the buildings with high accuracy and detail without further user interaction. Results will be given for the building detection phase which forms the main part of this paper, as well as for the building reconstruction phase.

1. INTRODUCTION

The automatic 3D reconstruction of buildings and entire cities has been a rather popular research topic (Baillard, 2000; Moons, 1998; Scholze, 2003). Yet, the practical use of such methods has remained very limited. The reason is that the level of performance of an automated system needs to be very high, almost perfect, for it to be useful in practice. Otherwise, a human operator will need to go over the results quite carefully, in order to correct for all kinds of errors. The net effect is a marginal gain in productivity, at best. Therefore, it is - at least for the moment - important to supply semi-automatic, on-line solutions. This is the goal of the work described here.

It has been proposed earlier to divide the task of building reconstruction using a two-phased approach: with a detection phase extracting the locations of single buildings and a reconstruction phase of then just a single building. The building detection can take into account the context in which these buildings are located, such as building rows in urban areas or detached houses in suburban areas. The building reconstruction benefits from the smaller images sizes and the constraints this brings with it for 3D reconstruction, as well as the focus on an individual building. Our approach is, thus, to combine a semi-automatic building detection, requiring an acceptable, minimal manual interaction, with a fully-automatic building reconstruction.

2. BUILDING DETECTION

Automatic building detection in dense urban areas (Fradkin, 2001) still requires manual intervention in order to improve the results. We therefore accepted the necessity of some degree of user interaction and build our system around it. The central idea for efficient semi-automatic detection is the concept of a building row. A building row is a number of buildings that are aligned along a street, that share a common general direction, and that are of similar height. This assumption is realistic in urban European areas where many buildings are arranged as city blocks. Such a building row can be easily identified and the segmentation of such a row into individual buildings can be largely automated.

Standard contour extraction in 2D and subsequent line segment reconstruction in 3D is performed beforehand in multiple, overlapping images, and serves as input data for the automatic procedures. The detection starts with the user identifying a building row. The system then perform an automatic section of the row, that is determining its exact orientation, and approximate width and height range, and, most importantly, the start and end parameter of each building along the row. The user has afterwards the chance to verify and, if necessary, correct the segmentation before sending it of for automatic reconstruction.

Figure 1. Image from the Brussels data set

2.1 Manual Initialization

The user determines a building row by clicking once on the first and on the last building of the row in only one image. The manually picked building row line linking the two points selected by the user should be approximately parallel to the main orientation of the building row. The angle of this line will later be adjusted using the 2D contours in the image.

2.2 Pixel-Based Segmentation

The grayscale values of the pixels inside the strips on both sides along this line are analyzed by using a similar approach to region growing, starting with segments of high homogeneity, and extending them in both directions to account for the transition at the border of the segments. For each pixel along the building row line, we determine the average grayscale value of the image pixels within a distance of ten pixels and perpendicular to this line.

We assume nearly homogenous building roofs with only small disturbances caused by chimneys and dormer windows. Thus, a measure for homogeneity can be automatically derived from the image. It is based on the average difference of the newly calculated values of adjacent pixels. Most of the differences of these values should be small, except at the borders from one building to the next.

2.2.1 Determine Homogenous Segments: We decide for each pixel of the line whether it meets the homogeneity criteria by calculating the differences with the values of the pixels which are within a distance of two from it and testing these differences against the homogeneity measure. A segment is made up of an uninterrupted sequence of homogenous pixels. Figure 2 shows the results of the first step. Each segment is characterized by a begin and end pixel of the line and is assigned the average value of its pixels. Non-homogenous pixels will not be associated with any segment.

Figure 2. Initial region-based segmentation

2.2.2 Merge Neighboring Segments: As a result of the first step, homogenous regions may split into two segments due to small disturbances. We therefore test the average values of neighboring segments against the homogeneity measure, and we merge those segments which fulfill the test and are separated by three pixels or less in between them.

2.2.3 Extend Segments: We take these merged segments as seeds for a region growing. Within the limits of the user-picked line, segments are enlarged pixel by pixel. Adjacent pixels are added to the segments given that they keep the segments homogenous and do not create any overlaps. We use a slightly relaxed homogeneity measure, as the goal is to close the gaps between the segments.

2.2.4 Extend Outer Segments: The user has clicked somewhere on the first and last building of the row. In order to determine the correct start and end of these buildings, we therefore have to extend the first and last segment beyond the limits of the user-picked line. This extension is done again pixel by pixel, testing against the homogeneity measure.

The goal of the pixel-based segmentation was to determine possible start and end points for the buildings along the building row. The results for one row are shown in Figure 3. Having too many segments rather than too few will be acceptable due to the merger of the results with the follow-up contour-based segmentation.

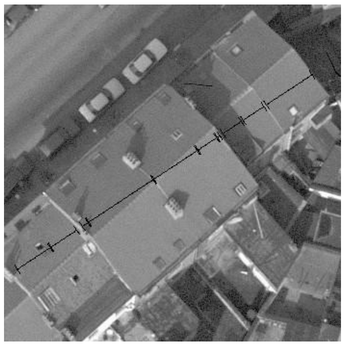

Figure 3. Final region-based segmentation

2.3 Contour-Based Segmentation

2D contours provide a higher positioning accuracy than region growing. Each contour gives us a possible border between two segments. Contours that give rise to similar segment borders will be merged, thus increasing the border's certainty value.

Many contours in the vicinity of the user-picked line belong to the roof ridges and gutters and, thus, are parallel to the building row. We use these contours to perform a robust estimation of the direction of the building row.

2.3.1 Find all appropriate contours: All contours nearby the building row line that are forming a minimum angle with it are candidates for the segmentation. The parameter is determined by the projection of the mid point of the contour onto the building row line. The search is extended beyond both end points of the line. Figure 4 shows all the initial segmentation.

2.3.2 Merge segments: Each contour gave rise to a segmentation hypothesis. We can test now whether two hypotheses may be due to the same building roof border. Such hypotheses are merged. The test is based on our assumption of what the smallest significant structure of a roof is.

Figure 4. Initial contour-based segmentation

2.3.3 Eliminate erroneous segments: Each segment is given a possibility value. This value is determined by the number of contours participating in this segment, their total length, and their distance from the building row line. Segments with low possibility values are discarded.

The results of the contour-based segmentation depicted in Figure 5 demonstrate that they can be usefully combined with the region-based segmentation.

Figure 5. Final contour-based segmentation

2.4 Information Fusion and 3D

Afterwards, the results of both steps will be consolidated. We determine the segmentation with the highest likelihood, where the segments of both approaches coincide and the distribution of the building segments is the most regular. The latter criteria has proven useful to deal with shadows.

Finally, we estimate the width of the buildings as well as their height range using the set of 3D line segments that are close to the line that the user has initially selected. Spurious matches can be eliminated by detecting single outliers in this set. Missed matches have less an effect as we don't look at just an individual building. Determining the bounding boxes of the building roofs in 3D, however, is absolutely necessary in order to position the ROIs that show the same building in the various overlapping images. These bounding boxes are aligned with the direction of the building row (see Figure 6).

2.5 Manual Correction

This whole automatic procedure of segmenting a building row into hypotheses of individual buildings takes less than half a second. The result of this segmentation can, thus, be presented

Figure 6. Results after automatic segmentation

to the user in real time. The user has now the possibility to split or join segments, or to adjust the width of the building row. Even though the user works on only one image, all corrections are translated to the respective changes of the 3D bounding boxes. The requirements to the accuracy of the segmentation are not as high, since this is merely the preprocessing for the later building reconstruction.

2.6 Results of the Segmentation

The scene in Figure 7 depicts the segmentation of three building rows, illustrated by the ROIs determined by the projections of the 3D bounding boxes into one of the images. One row has been segmented with even just the two clicks identifying the building row. As part of the manual corrections, two segments had to be joined in one ambiguous case for the building at the left-hand side and the segments had to be adjusted for the irregularly shaped buildings at the city block corners. This resulted in less than one second per building. The segmentation results are reasonably robust with respect to the position of the two initial clicks identifying the building row.

3. BUILDING RECONSTRUCTION

The second phase is the building reconstruction. We use a system we have developed that is capable of automatically reconstructing individual buildings with high quality and level of detail. A more in-depth description of this system can be found in (Willuhn, 1997).

3.1 Modeling Building Roofs

The system uses a knowledge-based approach with a conceptual model on top of one that is based on shape. The conceptual model takes into account the various roof forms – flat, gabled, or in combinations (Figure 8) – as dictated by architectural and other considerations. A similar approach was used in (Lang, 1999).

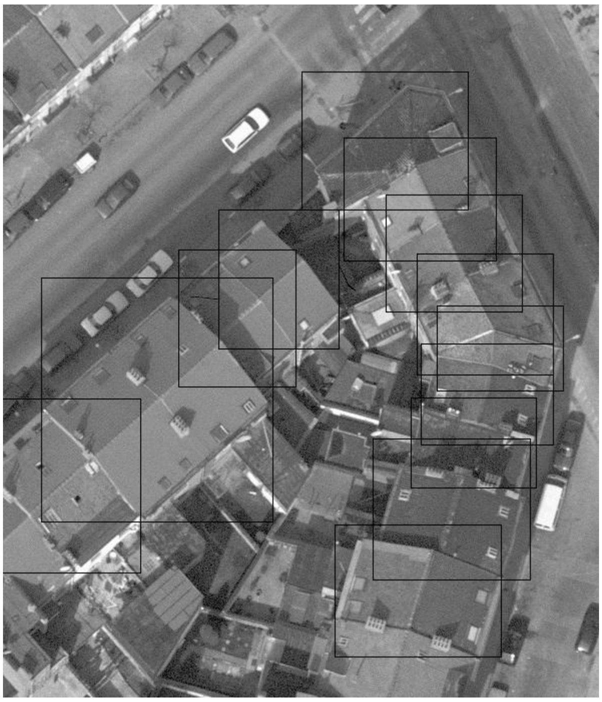

Figure 7. Results after manual correction

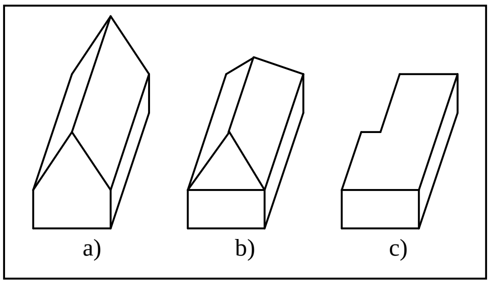

Figure 8. Building roof styles: a) gable, b) hip and c) flat

When developing any kind of model, there is a trade-off between flexibility and robustness. Having rigid models with few parameters gives great robustness. However, few houses have these pure architectures. The reconstruction of the building roofs in our images demand a greater flexibility, while at the same time keeping the system highly robust. For this reason, the general roof styles are decomposed into smaller elements (Figure 9) that can be assembled in a flexible way. We distinguish between conceptual features (e.g. ridge, gutter, roof connection, roof border) and conceptual structures (e.g. roof corner, gable), capturing the knowledge about the architectural elements of building roofs.

The shape model describes a building as a polyhedron. This model comes into play when parts or the whole of the roof do not conform with any of the conceptual features and structures. This model, however, requires a high quality of data in order to ensure that the reconstructed object is indeed the roof that is captured in the image.

3.2 System Implementation

We chose a blackboard (Engelmore, 1988) as the general architecture because it is a flexible framework, allowing us to

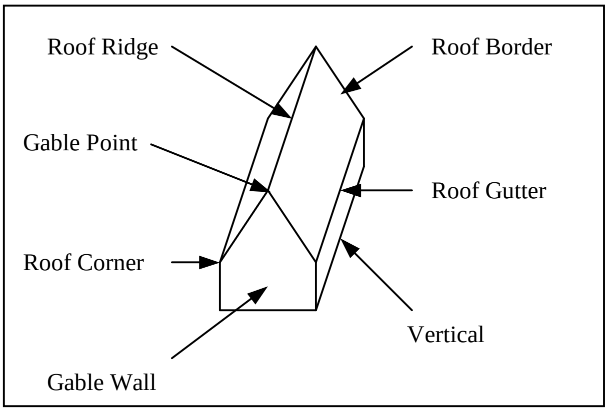

Figure 9. Conceptual feature and structures for a gabled roof

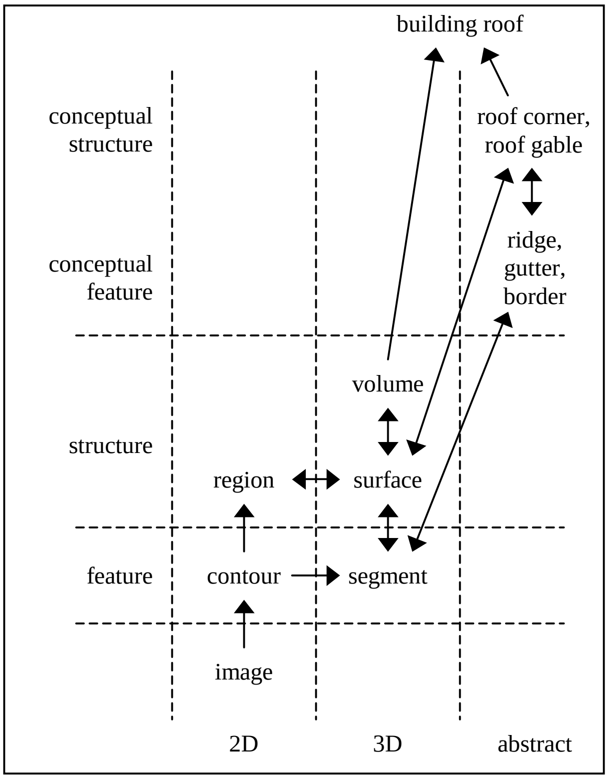

Figure 10. Object reconstruction using conceptual models

freely choose representation and techniques for the knowledge base, the knowledge sources, and the control structures. The data obtained from the images as well as any intermediate and final results are stored in a semantic network. A lot of the information that has proven useful for this approach is not stored in the objects themselves (e.g. contours, segments, roof ridges) but rather in the relationships among them (e.g. a roof ridge and a roof border forming a roof gable.) All knowledge of how to apply our models to the data is coded into rules as part of the knowledge sources. Rules consist of a condition and an action part. When triggered by the blackboard's reasoning control, the rule finds all objects, relations, or attributes from the knowledge base that fulfill a particular condition. It then performs its action part, which can add, remove or change data. Object recognition requires the explicit dealing with evidences. Objects, their attributes and relations, can not determined with absolute certainty. Elements of these categories have therefore a certainty value associated with them. We decided to use possibility theory (Dubois, 1994) for reasoning with evidences, which also comes closest to our goal, that is to determine the more plausible interpretation of the images. The blackboard's control is a standard cycle of tracking the changes of the data caused by the rules, the rating of the rules, considering also rule dependencies, and the execution of the highest-rated rule.

3.3 Reconstruction Process

The input to the system consists of contours in 2D and hypotheses of straight line segments and planar surfaces in 3D. The latter are given as the parameters of the plane made up of a set of coplanar segments. Rules work at various levels of the data as depicted in Figure 10. They

- Check the data against the constraints given by the models,
- Find and describe instances of higher-level elements (upward arrows), and
- Hypothesize about and verify lower-level elements (downward arrows.)

For the conceptual model, rules detect the conceptual features and structures from segments and surfaces in 3D based on the concept's definition, such as a ridge being the line of the junction of two surfaces sloping upwards towards each other. It is important that these conceptual elements can be instantiated without all of their parts being present. A ridge is found with merely a single surface, although with a lower certainty value. Additional rules infer such missing parts and search for evidence in the lower-level data, until the roof is complete. The application of extensive architectural knowledge yields high robustness, although the choice of conceptual features and structures is able to support a variety of roof forms.

A second set of rules reasons about shape in the case that the building does not comply with one of the concepts defined. We assume a building to be of the general shape of a polyhedron, that means a solid figure bounded by planar surfaces. The roof surfaces must therefore be completely enclosed by intersections with either other roof surfaces or the building's vertical walls. Some rules use this knowledge to discard erroneous matches of 3D line segments and hypotheses of planar surfaces. Other rules search for missing segments or surfaces where a roof surface's boundary is not closed. This alternative, shape-based model increases the system's flexibility if the concept-based model fails; although the quality of input data is generally not high enough to reconstruct a building roof based on this model alone.

The rating function of the blackboard control uses information about the data elements that each rule needs for execution as well as those that it changes as a result in order to determine a priority list for invoking the knowledge sources. This particular choice of a rating function implicitly favors rules that check data over those that create (derive or hypothesize) new data, and prioritizes rules that embed more specific knowledge over those with more generic one. This yields the desired order of execution, even though none particular is given, but rather is determined by the input data and the results of each rule. Most rules are used more than once when new lower-level instances have been created and have to be verified and used for the derivation of higher-level information. Therefore, each data set will effectively have a different order in which these rules are invoked.

3.4 Results of the Reconstruction

The system to reconstruct the roofs of single buildings has been used for both suburban and urban scenes (see (Willuhn, 1997) for results on a suburban data set.) Two different systems have been used to provide the input data, demonstrating the independence from the underlying preprocessing. For this data set, the IMPACT system that has been developed by the team at KU Leuven provided the input (Figure 11), consisting of contours in 2D and segments and planar surface patches in 3D, based on ROIs in multiple images for the same single building. The output is a description of the building roof in 3D. The height value of the ground points of the vertical building walls was given as a parameter. It could alternatively be determined from a digital terrain model.

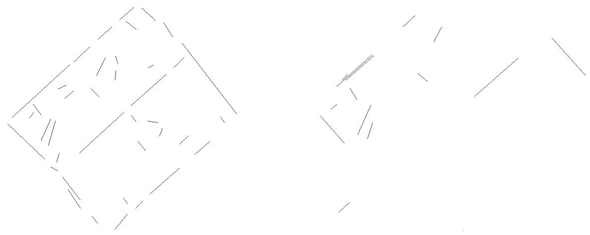

Figure 11. 2D contours of one of the six images and 3D segments from segment stereo matching

The results have been achieved after, in short, finding the two planes that make up this gabled roof, collecting evidences for further segments of the roof borders and roof gutters based on these planes, and choosing the most plausible roof outline. Figure 12 validates the high quality of the obtained roof description.

4. OVERALL RESULTS

We have applied this approach to four scenes at the corners of an intersection in a dense urban part of Brussels, which have previously been used for the European IMPACT project. The data set contains six overlapping color high-resolution images (10cm per pixel) for each scene with internal and external camera parameters. We use the IMPACT system for the purpose of 2D contour extraction and 3D line segment and planar surface reconstruction in both phases.

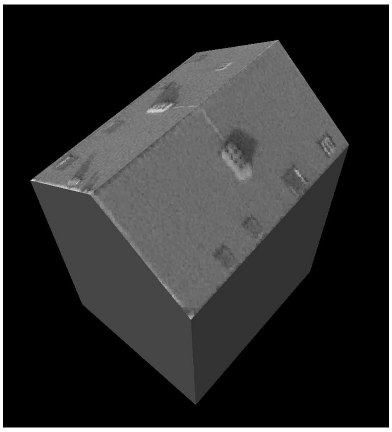

Figure 12. Result of the reconstruction

The user provided input for seven building rows, three for the scene at the bottom of Figure 13 (see also Section 2), two for the scene at the right-hand side, and one for each of the two other scenes. The four scenes, however, do not cover the whole intersection. We achieved our goal of segmenting the images with less than one click per building. Three building rows were segmented with just two clicks, one for the start and one for the end of the row. For the other building rows, the user had to manually join two segments, extend the segments of the irregular corner buildings, and two times adjust the width of the building row.

The automatic building roof reconstruction produced high-quality results for 20 out of the 25 buildings in the scenes. Most of the buildings have gabled roofs, two have flat roofs, and four buildings at the corners have irregularly shaped roofs. The system failed for three corner buildings due to the sole reliance on the shape-based model and insufficient quality of the 3D line segments and planar surfaces, for one flat-roofed building due to strong shadow, and for one building with a gabled roof where occlusion hindered the 3D segment matching.

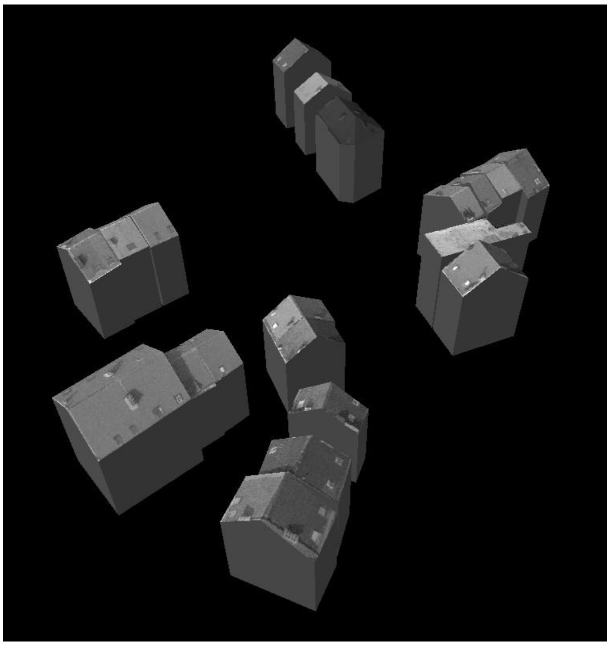

Figure 13. Final results of segmentation and reconstruction

5. CONCLUSIONS

We have presented a system for the semi-automatic reconstruction of building roofs from aerial images. The chosen approach demonstrates the advantages of separating the task into two phases: (1) the efficiency of the segmentation of large aerial images into regions containing single buildings and (2) the quality of an automatic reconstruction that can be achieved by focusing on individual buildings. The combination of a region-based and a contour-based segmentation in 2D and fusion of the results with 3D data has been shown effective in keeping the user interaction at a minimum. The system for roof reconstruction achieves high robustness and flexibility by employing both a concept-based and a shape-based model. It is, thus, possible to reconstruct about three quarters of the buildings with high accuracy and detail with manual interaction of on average one click or less per building.

Several directions for future work are possible. The incorporation of additional conceptual knowledge about irregularly shaped corner buildings will improve the reconstruction of roof forms that still relies on the shape-based model and, therefore, demands a high quality of 3D input data. The improved interaction between both phases – in particular the supply of information about the context of a building row, that is its position and direction, to the building reconstruction – will improve both speed and quality of the reconstruction process. The ROIs are currently aligned with the image axes, but should rather take into account the building row's direction. A third direction is the automatic self-evaluation of the final results in order to support a fully productive system in which the user should be guided to the manual reconstruction of missing or correction of wrongly reconstructed building roofs.

REFERENCES

Baillard, C., Zisserman, A., 2000. A Plane-Sweep Strategy for the 3D Reconstruction of Buildings from Multiple Images. *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXIII, Part B2, pp. 56-62

Dubois, D., Lang, J., Prade, H., 1994. Possibilistic Logic. In Gabbay, D., Hogger, C., Robinson, J., editors. *Handbook of Logic in Artificial Intelligence*, pp. 439-511, Oxford University Press, Oxford

Engelmore, R., Morgan, T., editors, 1988. *Blackboard Systems,* Addison-Wesley Publishing Company

Fradkin, M., Maître, H., Roux, M., 2001. Building Detection from Multiple Aerial Images in Dense Urban Areas. *Computer Vision and Image Understanding 82(3)*, pp. 181-207

Lang, F., 1999. Geometrische und semantische Rekonstruktion von Gebäuden durch Ableitung von 3D-Gebäudeecken. *PhD Thesis, Institut für Photogrammetrie, Universität Bonn*

Moons, T., Frere, D., Vandekerckhove, J., Van Gool, L., 1998. Automatic Modelling and 3D Reconstruction of Urban House Roofs from High Resolution Aerial Imagery. *European Conference on Computer Vision*, Vol. 1, pp. 410-425

Scholze S, 2003. A Probabilistic Approach to Building Roof Reconstruction. *PhD Thesis No. 14932, Computer Vision Laboratory, ETH Zurich*, Zurich, Switzerland

Willuhn, W., 1997. Integrating Models for Object Recognition. *Technical Report 175, Computer Vision Laboratory, ETH Zurich*, Zurich, Switzerland

ACKNOWLEDGEMENTS

We acknowledge the support of Eurosense Belfotop N.V. for providing the data set.