

## AUTOMATIC BUILDING EXTRACTION FROM AERIAL IMAGES

**Wolfgang Faig (Dr.-Ing.) and Thomas Widmer**  
University New Brunswick, Canada  
Department of Geodesy and Geomatics Engineering  
Fredericton, NB, Canada E3B 5A3  
[Wfaig@unb.ca](mailto:Wfaig@unb.ca)  
[Widmer@unb.ca](mailto:Widmer@unb.ca)

Working Group III/3

**KEY WORDS:** 3-D Building modeling, object related method, object representation, bottom-up and top-down approach

### ABSTRACT

This paper describes the initial steps of an ongoing project, which aims to analyze building extraction methods, and their approaches. Automatic building extraction from aerial images uses many approaches from the computer vision technology. The object recognition of man-made features has many difficulties that are discussed and to query. Many steps are involved in the approaches today that lead to loss of accuracy and data. Quality and integrity of data of a semi-automatic building extraction program are evaluated.

## 1 INTRODUCTION

Object recognition with digital imagery is a wide-open field that has yet to be fully discovered, and will lead to new approaches for surveying tasks and demands. Building extraction from aerial images is one of the benefits of image processing. Most of the technologies today are based on the simple image processing. They extend the processing onto generated symbols, that include more specific information about the desired object. The generated symbols are results of classification oriented processes. Both, simple image processing and classification orientated processing are part of the bottom-up approach. The bottom-up approach uses single pixels of the image and combines them to related lines and topological surfaces. Object recognition in bottom-up approaches often faces a limitation, and even very intelligent algorithms are not able to provide a reliable and accurate final solution. Often bottom-up approaches are able to decompose the image into its main components, but are unable to close the object to its final representation. On the other hand, there are top-down approaches. Top-down approaches are using a predefined model, and search for fitting results. Only a limited number of solutions can be reached, because not all reality cases can be covered with given models. Forcing real objects into forms causes generalization and minimization. Simple models and smoothing help to achieve better results, but also loss in accuracy. The combination of bottom-up and top-down methods results in a hybrid approach. The bottom-up approach makes generated symbols available to the top-down approach that intern verifies the information for correctly recognized objects. In the following we want to provide some insight into the bottom-up and top-down approaches. A hybrid approach will be explained using the example of the Building Extraction program ObEx from the University of Bonn, Germany. The possibilities of representing extracted buildings and the quality are studied. Finally a performance characterization of the module is introduced.

## 2 BOTTOM-UP APPROACH

### 2.1 Low-Level methods

Simple image processing algorithms work in most cases without previous knowledge (low-level) and attend to the raw data. They enhance and restore the image and reduce it into basic elements, without discarding important image features, e.g. zero-information-carrying clutter can be removed from the imagery for improvement. Object boundary cleanup reduces the image to a single pixel width. The most familiar operators, such as histogram sliding and stretching, binary contrast enhancement or brightness slicing are single pixel point processes for one image. The geometrical position of a pixel is held fixed, and only a mapping function converts the attributes of the individual pixels. Apply several images for multiple point processing. We are subtracting another image from a reference image, or from the two images, then an average can be calculated. The latter can be used for temporal change detection.

Pixel group processing operates on a group of input pixels surrounding a center pixel. The neighboring pixels enable us to calculate some trends for the area. This information can be a fast change in brightness in a specific direction or a regular plane. It is possible to detect primitives of interest through different filters. In most cases, the primitives of interest are edges or lines. Two major filter types are the low-pass and high-pass filters. Low-pass filters are smoothing the image. High-pass filters accentuate high frequencies in images, while low frequencies remain untouched. With Edge Detection algorithms the image is generalized. Only edges are visible as end products of the process. These outlines can be determined through different methods. All methods use as an indicator the slope of the pixel value changes. Image restoration can eliminate blurred images. If the noise is mostly periodical, or linear, the restoration can be accomplished with good results.

## 2.2 Mid-level methods

We derive a symbolic image description through mid-level methods (Fischler and Boles, 1985). The result of this symbolization is dependent on the algorithm. Every feature extraction algorithm suppresses or neglects data in different ways (Fuchs and Heuel, 1998). Because some methods cannot extract specific features, a combined approach can be utilized, i.e. the integration of region growing and edge detection, as mentioned by Pavladis and Liow.(1990) and Chu and Aggrawal (1993), object related region segmentation (Wiman, 1995), and integrating point-and edge extraction (Henricsson and Heitger (1994). In Fuchs and Heuel (1998) the problem is solved within one framework, the so-called polymorph feature extraction. Points, lines/edges, and regions are the main features to extract, because they are fairly cheap to compute and provide strong visual clues that can help in the recognition process (Marshall, 1997).

In point extraction we have to distinguish between single points, line points, corner points, and junction points. For most tasks the detection of the corners, and junctions is requested. The detection of corner and junction points should satisfy a number of important criteria (Mokhtarian, 1998):

- All the true corners should be detected
- No false corners should be detected
- Corner points should be well localized
- Robust with respect to noise
- Corner detector should be efficient

A common method to detect points is the curvature method [Beaudet, (1978), Dreschler, (1981) and Deriche, (1998)]. This method has been improved and modified in many algorithms, using the derivatives as main components. Another method are points of interest (Moravec, 1977), where an unnormalized local autocorrelation is computed in four directions, and the lowest result is taken as solution. From the extracted points a linking of related points to lines can be executed. For this step the incompleteness is problematic. Some parts of the final line are often undetected and create blanks in the feature; some segments have been classified as line segments, but are noise. To correct these errors, local and global line/edge linkers are used (Marshall, 1997) in most cases.

Mostly gradient operators are used for the line/edge extraction, because they are already used for point extraction, or easily derivable from there. The method of localizing these discontinuities often then becomes one of finding local maxima in the derivative of the signal, or zero-crossings in the second derivative of the signal. The main problem in differential edge detection schemes is noise. The spikes in the derivative from the noise can mask real maxima that indicate edges. Smoothing the image can reduce the effects of noise. Marr (1982), Marr and Hildreth (1980) and later Canny (1986) include smoothing of the image. The Canny edge detector is current standard edge detection scheme widely used around the world is the.

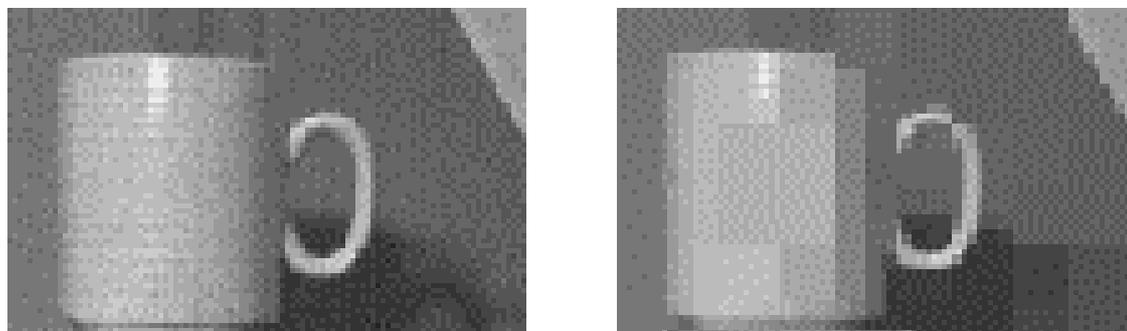


Figure 1: Region growing

At region extraction methods we are mapping individual pixels onto sets of pixels of similar properties (intensity, colour, texture etc.). We assume that pixels of the same region have similar properties. There are several approaches

developed for regional segmentation. Similar to the edge extraction, there are local and global approaches. The local approach is the region growing method (Figure 1). The approach starts from the assumption that all pixels are separate regions, and attempts to merge pixels into regions or groups of similar pixels. The approach has no reference to a more global viewpoint. Pixels are merged if they are homogeneous, i.e. have similar grey level intensity, colour, texture, depth etc. The global approach is the splitting and merging method that attempts to divide an image into uniform regions.

The subdivision is continued until no further splitting is necessary, and all pixels are grouped. Accordingly groups are merged if they are similar, to give larger irregular regions.

### 3 TOP-DOWN APPROACH

The bottom-up approach operated directly on the data without any knowledge about the structure of the data or what the data represents (i.e., higher-level knowledge). It does not necessarily exclude higher-level knowledge, but can operate *with or without* this knowledge. It is up to the implementer to decide what types of high-level knowledge the low-level techniques will use, and how and when this knowledge will be used. The top-down approach uses a model based approach. There are several model based approaches, namely the active contour model (Kass et. al., 1987), the Kalman filter model (Brinkley, 1993), and the Bayesian network model (Thomas et. al., 1989).

For the development of an efficient model based approach in building extraction, the focus has to be placed on important features. Edges, corners and vertices are strong and useful features. The key of the approach is to propose some efficient models associated to each of these features and to efficiently extract and characterize these features directly from the image (Deriche, 1998). Every feature includes a large number of intrinsic parameters (grey level intensities, location, orientation of the line segments/dots). The initialization phase of the process has influence on the computational time, and therefore has to be considered.

### 4 HYBRID APPROACH

#### 4.1 Object related method

The symbolic image description that was achieved through the bottom-up approaches can be used as starting data for the verification of building objects with an object related method. As mentioned, the building object is decomposed into vertices, edges and cells. They are similar to the points, lines/edges, and regions of the bottom-up approach and have to be verified through a reliable comparison.

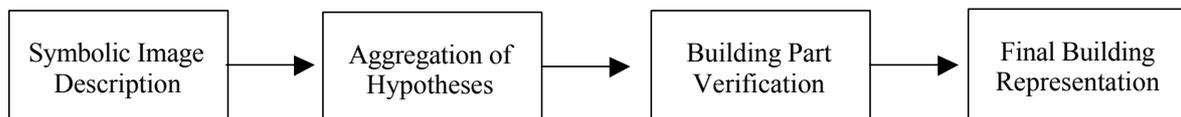


Figure 2: General workflow for building extraction through object related method

Using aerial images for the building extraction, it is assumed that the orientation of the aerial images is known. With hypotheses and verification paradigms a solution can be derived. In general the symbolic image description (see Section 2.2) is used to produce an aggregation of hypotheses for 3D aggregates. These hypotheses have to be verified through the model knowledge. Each hypothesis is semantically interpreted and classified to belong to one or several parts of the object (Brunn et. al., 1996). Identified object parts can be grouped to represent the final object. Completeness is not guaranteed, and has to be accomplished. This can be done by using the given data in a semi-automatic procedure, and to finish the task with a human operator. Linking important parametric information to a CAD modeling tool, an automatic construction of Constructive Solid Geometry structures (CSG) could be possible.

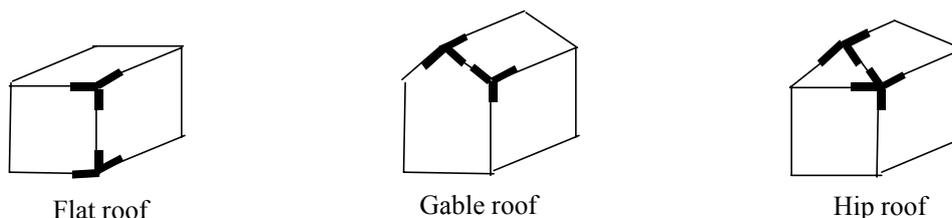


Figure 3: Parametric models

The object related method is a combination of results obtained from the image model in two-dimensions (2D), and the verification at the 3D object model. Starting with the initial step, the decomposition of the image into a symbolic image

description, an interplay takes place between the invariant given results of the image model and the object model. Through a hierarchical advance, the constructed model is increased with semantic information, until a fully completed representation is built.

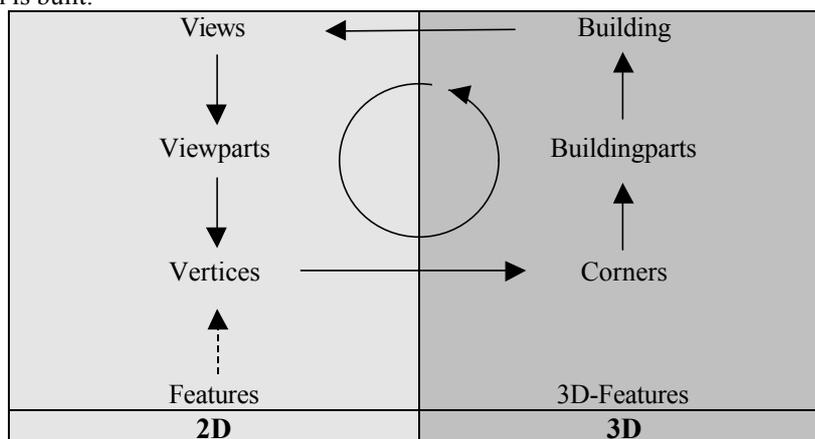


Figure 4: Interplay of image and object model

The production of the hypotheses is driven from the image data to obtain a first 3D description. Starting with the vertices (see Figure 4), a classification is performed, and vertices with high probability of detection in other images are chosen. High stability for the selection of vertices can be improved through multiple overlaps (Lang and Förstner, 1996), and knowledge of the orientation data. The knowledge of the orientation data helps to define geometric restrictions and to choose the best image pairs for occlusion-free representations. This knowledge could be expended to an enhanced level, where additional pre-assumptions can be made for the representation of the object model in image space. Even the classification and the probability could be affected by the estimated parameters and can increase the valuation. To increase the detection the use of epipolar images is recommended. The 3D reconstruction of the vertices helps to make a domain specific interpretation by a 3D corner model, and to decrease the amount of potential hypotheses.

Parameters of building segments that belong to the 3D model corner are stored in a library, to help to build a geometric

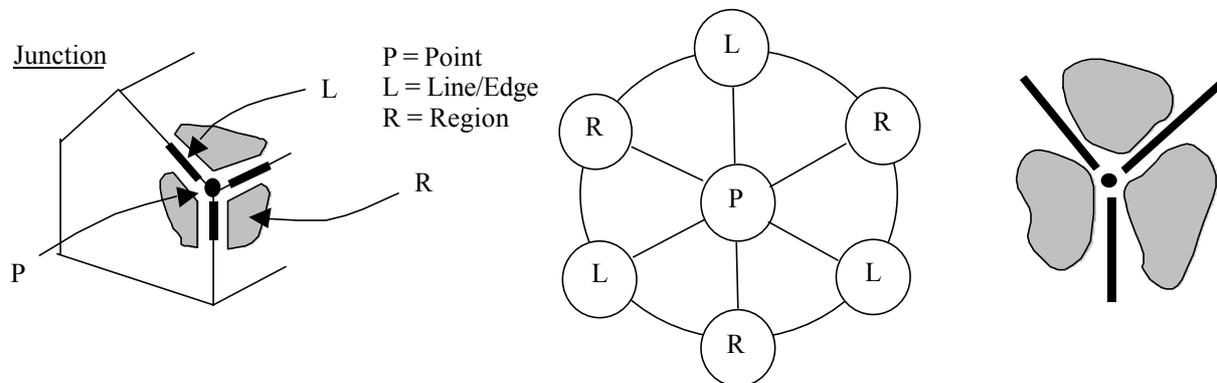


Figure 5: Feature neighborhood (left). Graph of the adjacency relations (center). Expected appearance in the image (right). (from (Fischer et.al., 1998)).

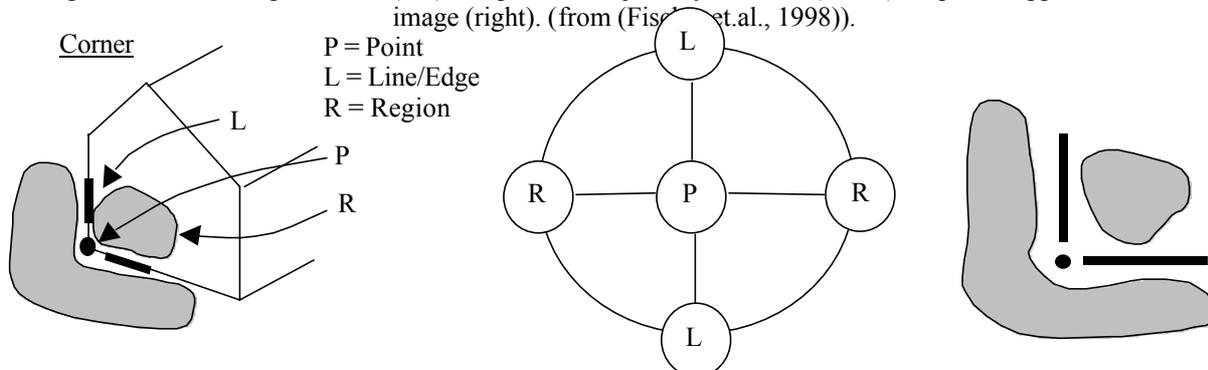


Figure 6: Feature neighborhood (left). Graph of the adjacency relations (center). Expected appearance in the image (right). (from (Fischer et.al., 1998)).

representation of the whole building. The more building segments are given, the higher the possibility to construct a

parameterized building that includes the detected corners. To verify the building segments, they are projected into image space, and compared with the symbolic image description. It is possible that not all building segments may be fixed, and through the application of projective geometry, undetected parts can be discovered.

Using the newly gained information, an iteration can be started, until the building is completely detected, and the building segments are consistent with each other. The iterations are stopped, when no new hypotheses can be generated. Due to the relative stability against partial occlusions, corners are selected as the basic class of feature aggregates for the reconstruction process (Fischer et al., 1998). The interest lays in corner and junction points (see Figure 3). These points have a specific 2D corner model and adjacency relation (see Figures 5 and 6). Detected vertices help for the analysis of following vertices. Heuristic methods aid in classifying correct corners, and by including additional information about corner points. Assuming that all vertices are corners, the vertices are checked for geometric constraints, and their domains. From the corner different lines/edges are expanding with specific attributes. Three-dimensional line orientations of the features can be included for a higher probability of detection. The line orientations can be horizontal, positive oblique, negative oblique, positive vertical, and negative vertical.

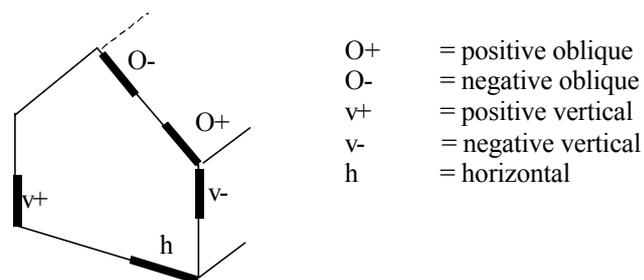


Figure 7: Three-dimensional line orientation

More orientations are possible, but with these 5 major ones, the classification of a corner point in the parametric models used in Figure 3 can be accomplished.

#### 4.2 ObEx – Semiautomatic Object Extraction

This semiautomatic program needs interaction with an operator. Actually, the program is a combination of automatic, semi-automatic, and interactive methods (Müller, 1998). The workflow of ObEx is explained in Figure 8.

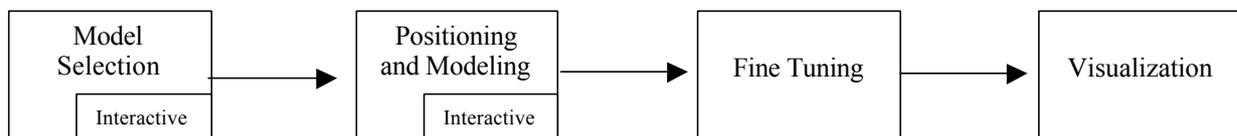


Figure 8: Workflow of semiautomatic program ObEx

The model selection can be expanded with new parameterized models. The positioning and modeling is an interactive step. ObEx uses wireframe models for the different standard models, and projects the models into a single image, computes the expected covered corners, and implements the representation. For this step knowledge of the orientation parameters is needed. The modification possibilities of the wireframes are predefined in an associated spread sheet (Lang and Schickler, 1993). As in every CAD drawing program, different methods for creating an object are possible. ObEx has a two step modification method. The first point selection is the reference, and the second choice in the wireframe is activating one or two parameters that can be changed scale-free. Holding the projective presentation of the wireframe, through all changes with rotation, shift, scaling, and stretching is important for the system. The changes of the wireframe are online, and can be controlled by a 2-button mouse. Differences in the mouse positions are relevant for the magnitude of the parameter change. The method uses a message transfer between the platform-specific Graphical User Interface (GUI) and the extraction system.

To speed up the measuring time in ObEx, automatic fine tuning tools are available. ObEx offers different selective options for automatic measurement tools, which are available in the GUI. They include:

- Ground point measurement tool
- Roof point measurement tool
- and fine adjustment tool

The ground point measurement tool uses a general matching tool of the object-model designed program. It is a point matching tool using a set of 2D-images to get 3D- information of the location. The working range of the tool are ground point measurements (fully to semi-automatic), points of the roof (automatic), and additional implementer specific tasks (Müller, 1997). The semi-automatic option for the ground point measurement is useful for unstable or dense urban areas, where man-made objects are closely built to the referred building, and an automatic location around the building would fail (Figure 9).

This option is not necessary for the roof detection, because the parametric model binds the matching tool to the object. The fine tuning tools assume a coarse adjustment of the wireframe to the image represented object. The fine tuning

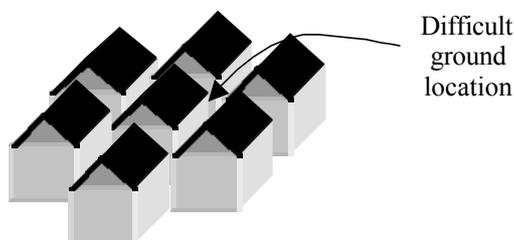


Figure 9: Ground location in high dense urban area

tools are using multiple images, and attribute the different models to each other with a minimum of deviation. Hidden-line algorithms help to distinguish the representation of the object in the image, and epipolar images increase the speed of detection. The allocation of the different models to each other is based on robust estimations (Schickler, 1992), (Lang and Schickler, 1993). Besides the mentioned tools, ObEx uses a random sample method, called RANSAC (RANdom Sample Consensus). It takes a given set of features, most preferably lines/edges and applies a best fit to the used building parameters (Läbe and Gülch, 1998). RANSAC assumes the semi-automatic gross measurement of the parameterized building, with the wireframe (see Figure 8 – Positioning). To complete the model, edge sets are chosen to calculate length, width, and height. The choice of the edges is random, and will be evaluated with an associated weight. A threshold or a number of samples can limit the processing time. The sample with the best calculated weight is chosen to reconstruct the building (see Figure 10).

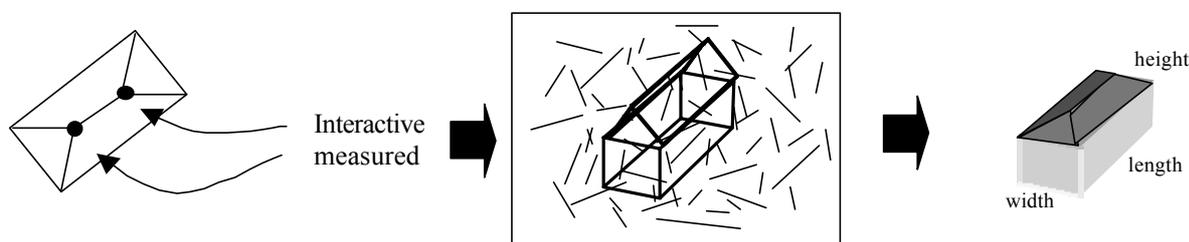


Figure 10: RANSAC – Best weight method

## 5 CONCLUDING REMARKS

The flexibility of the bottom-up approaches and the robust methods of the top-down approaches, are difficult to separate. Trying to automate the process, we tend to build total automatic systems. If we consider the numerous interpretation problems and unknown parameters, we are still far away of a solution without an operator, but the amount of manual steps can be compressed into a few execution-packages. Knowledge of data behavior (probability), time, algorithmic intelligence and accuracy are the key attributes for the quality of an automatic building extraction, and will always face the limit of independence of an operator. The authors are presently utilizing the hybrid method on the practical projects and expect to present the results at the Congress in Amsterdam.

## ACKNOWLEDGEMENTS

The cooperation of the University of Bonn (Germany), and the help of Dr.-Ing. Eberhard Gülch and Mr. Hardo Müller is greatly appreciated.

## REFERENCES

- Beaudet, P.R. (1978). *Rotationally invariant image operators*. Proceedings of the Int. Conference on Pattern Recognition, Kyoto, pages 579-583.
- Brinkley, J.F. (1993). *Hierarchical geometric constraint networks as a representation for spatial structural knowledge*. In 16th Annual Symposium on Computer Applications in Medical Care, New York, NY, pages 140-144.
- Brunn, A., Gülch, E., Lang, F., Förstner, W., (1996). *A Multi-Layer Strategy for 3D Building Acquisition*. Institut für Photogrammetrie, Universität Bonn.
- Canny, J. F. (1986). *A computational approach to edge detection*. IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-698.
- Chu, C.C., Aggrawal, J. (1993). *The Integration of Image Segmentation Maps Using Region and Edge Information*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(12):1241-1252.
- Deriche, R. (1998). [http://www-sop.inria.fr/robotvis/personnel/der/Demo/corner\\_vertex.html](http://www-sop.inria.fr/robotvis/personnel/der/Demo/corner_vertex.html).
- Dreschler, L. (1981). *Ermittlung markanter Punkte auf den Bildern bewegter Objekte und Berechnung einer 3-D-Beschreibung auf dieser Grundlage*. PhD Thesis Universität Hamburg Fachbereich Informatik.
- Fischler, M., Boles, R. (1985). *Perceptual Organization and Curve Partitioning*. IEEE, Image Partitioning and Perceptual Organization, pages 210-215.
- Fischer, A., Kolbe, T.H., Lang, F., Cremers, A.B., Förstner, W., Plümer, L., Steinhage, V. (1998). *Extracting Buildings from Aerial Images using Hierarchical Aggregation in 2D and 3D*. Computer Science Department I/III, Institut für Photogrammetrie, University of Bonn, Institut for Environmental Sciences, University of Vechta.
- Fuchs, C. Heuel, S. (1998). *Chapture 4: Feature Extraction*. Institute for Photogrammetry, University of Bonn.
- Henricsson, O., Heitger, F. (1994). *The Role of Key-Points in Finding Contours*. Computer Vision – ECCV'94, volumell, pages 371-383. Springer Verlag, Berlin.
- Kass, M., Witkin, A.P., Terzopoulos, D. (1987). *Snakes: Active contour models*. International Journal of Computer Vision, pages 321-331, 1987.
- Läbe, T., Gülch, E. (1998). *Robust techniques for estimating parameters of 3D building primitives*. Proceedings of ISPRS Symposium, Comm.II, Cambridge, UK, July 13-17, (5/98, 8p.).
- Lang, F., Förstner, W., (1996). *3D-City Modeling with a Digital One-Eye-Stereo System*. International Archives of Photogrammetry and Remote Sensing, Volume 31, B4, pages 261-266.
- Lang, F., Schickler, W. (1993). *Semiautomatische 3D Gebäudeerfassung aus digitalen Bildern*. Institut für Photogrammetrie, Universität Bonn.
- Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W.H. Freeman and Co.
- Marr, D., Hildreth E. C. (1980). *Theory of edge detection*. Proceedings of the Royal Society, London B, 207:187-217.
- Marshall,David (1997)[http://www.dai.ed.ac.uk/CVonline/LOCAL\\_COPIES/MARSHALL/node24.html](http://www.dai.ed.ac.uk/CVonline/LOCAL_COPIES/MARSHALL/node24.html).
- Mokhtarian, Farzin (1998). <http://www.ee.surrey.ac.uk/Research/VSSP/demos/corners/index.html>.
- Moravec, H.P. (1977). *Toward Automatic Visual Obstacle Avoidance*. Proceedings Int. Joint Conference Artificial Intelligence, Cambridge, MA, pages 584f.
- Müller, H. (1997). *Designing an object-oriented matching tool*. Institut für Photogrammetrie, Universität Bonn.

- Müller, H. (1998). *Object-oriented modeling for the extraction of geometry, texture and reflectance from digital images*. Institut für Photogrammetrie, Universität Bonn.
- Pavlidis, T., Liow, Y.T. (1990). *Integrating Region growing and edge Detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(3):225-233.
- Schickler, W. (1992). *Towards Automation in Photogrammetry – An Example: Automatic Exterior Orientation*. Dorrer E. (Hrsg.): Vorträge der 12. Wiss.-Techn. Jahrestagung der DGPF, Band 1, pages 153-161.
- Thomas, O., Binford, Tod S. Levitt, and Wally B. Mann (1989). *Bayesian inference in model-based machine vision*. In Tod S. Levitt, L.N. Kanal, and J.F Lemmer, editors, *Uncertainty in Artificial In Intelligence*. North Holland.
- Wiman, H. (1995). *Object Related Image Segmentation using the Minimum Description Length Principle*. Proceedings of Scandinavian Conference on Image Analysis, Uppsala, Sweden, pages 541-549.