AUTOMATIC ROAD EXTRACTION IN RURAL AREAS

Albert Baumgartner[1], Carsten Steger[2], Helmut Mayer[1],
Wolfgang Eckstein[2], Heinrich Ebner[1]

1) Lehrstuhl für Photogrammetrie und Fernerkundung
Technische Universität München, D–80290 Munich, Germany
Ph.: +49-89-289-22671, Fax: +49-89-2809573
E-mail: {albert}{helmut}{ebn}@photo.verm.tu-muenchen.de

2) Forschungsgruppe Bildverstehen (FG BV)
Informatik IX, Technische Universität München
Orleansstr. 34, 81667 München, Germany
Ph.: +49-89-48095-122, Fax: +49-89-48095-203
E-mail: {stegerc}{eckstein}@informatik.tu-muenchen.de

KEY WORDS: Road extraction, Multi-Scale, Grouping, Context

ABSTRACT

An approach for the automatic extraction of roads from digital aerial imagery is proposed. It is optimized for rural areas and makes use of versions of an aerial image with different resolutions. Roads are modeled as a network of intersections and links between these intersections, and are found by a grouping process. The context of roads is hierarchically structured into a global and a local level. At first, edges are extracted in the original high resolution image (0.2–0.5 m) and lines in an image of reduced resolution (2 m). Using both resolution levels and explicit knowledge about roads, hypotheses for road segments are generated. They are grouped iteratively into larger segments. In addition to the grouping algorithms, knowledge about the local context, e.g., shadows cast by a tree onto a road segment, is used to bridge gaps. To construct the road network, finally intersections are extracted. Examples and results are given, indicating the potential of the approach.

1 INTRODUCTION

Extensive research has been done on road extraction from aerial and satellite imagery. It is often motivated by the automation of data acquisition and update of Geographic Information Systems (GIS). The most common techniques for road extraction in images with low resolution are the detection and the following of lines. In high resolution imagery, matching of profiles and detection of roadsides, i.e., (anti-) parallel edges, are used. One main criterion to classify extraction schemes is human interaction. In semi-automatic approaches, an operator provides, for example, starting points and starting directions for road following (McKeown and Denlinger, 1988, Vosselman and de Knecht, 1995), or some points along a road are measured and the algorithm finds the road, i.e., a line that connects the points (Neuenschwander et al., 1995, Grün and Li, 1997). By automatic detection of the seed points, semi-automatic schemes can be extended to automatic ones. Automatic approaches have been proposed, e.g., by (Ruskonè et al., 1994, Bordes et al., 1997, Trinder and Wang, 1998). One major finding in recent years is that the characteristic properties of roads described by a model are not the same in different resolution levels (Mayer and Steger, 1998) and in different contexts (Baumgartner et al., 1997, Bordes et al., 1997).

The approach proposed in this paper automatically extracts roads from images with a resolution of 0.2–0.5 m. Although prior GIS data can help to guide the road extraction, old GIS data is not used here. The reasons for this are: First, automatic extraction without priori information shows the potential and deficits of an extraction scheme much better than a GIS-based extraction, because it only relies on the given model and strategy. It therefore can deepen the understanding of the problem. Second, the extraction of new objects is possible only in this manner and is needed for GIS update in any case. Third, to make the system reliable, it is wise to base the decision about an object on new imagery and not on old GIS data.

The proposed approach starts with the generation of hypotheses for roadsides using two resolution levels. Lines are extracted in an image with reduced resolution and edges in the original high resolution image. Using both resolution levels and explicit knowledge about geometric and radiometric properties of roads, road segments are constructed from the hypotheses for roadsides. These road segments are grouped in the following steps, i.e., correct hypotheses are connected and false ones are eliminated. Besides knowledge about roads, additionally knowledge about relations between roads and other objects is employed. The relations are modeled on a *local* and on a *global* level.

The model for road extraction is described in Section 2. Section 3 explains the strategy as well as the basic steps and the characteristic elements of the approach. After an evaluation of the results in Section 4, a short outlook concludes the paper.

2 MODEL

In order to extract roads from an aerial image, it is necessary to have a clear idea about the concept "road." For the proposed approach, the model comprises explicit knowledge about geometry (road width, parallelism of roadsides), radiometry (reflectance properties), topology (network structure), and context (relations with other objects, e.g., buildings or trees). The model described below consists of two parts: The first part (Sect. 2.1) describes char-

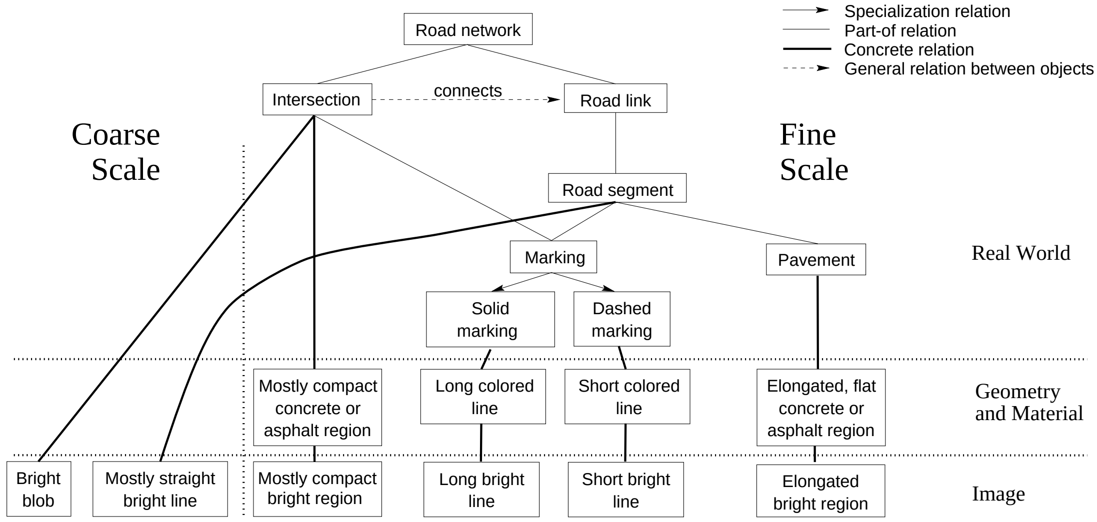


Figure 1: Road model

acteristic properties of roads in the real world and in aerial imagery, and presents a road model derived from these properties. The second part (Sect. 2.2) defines different local contexts and assigns those to the global contexts. In this way, the complex model for the object "road" is split into sub-models which are adapted to specific contexts.

2.1 Roads

A description of roads in the real world can be derived from their function for human beings: Roads are organized as a network connecting all areas inhabited and exploited by human beings. The denser an area is inhabited and the more intensively it is used, the denser the road network is. According to their importance, network components are classified into a hierarchy of different categories with different attributes. Field paths and less important roads follow the natural terrain surface more closely than highways, which serve as fast connections between conurbations. According to the different categories, roads differ with respect to width, minimum curvature radius, and maximum allowed slope. Some important attributes for parts of the road net are the type and state of the road surface material, existence of road markings, sidewalks, and cycle-tracks, or legal instructions, like traffic regulations.

The appearance of roads in digital aerial imagery strongly depends on the sensor's spectral sensitivity and its resolution in object space. The proposed approach is restricted to gray-scale images or to only one channel of a color image, and only scale dependencies are considered. In images with low resolution, i.e., more than 2 m per pixel, roads mainly appear as lines that form a more or less dense network. Contrary to this, in images with a higher resolution, i.e., less than 0.5 m, roads are projected as elongated homogeneous regions with almost constant width. Here the attainable geometric accuracy is better, but background objects like cars, trees, or buildings disturb the road extraction more severely. In a smoothed image — which corresponds to a reduced resolution — lines representing road centerlines can be extracted in a stable manner even in the presence of these background objects. The smoothing eliminates, e.g., vehicles or markings. This can be interpreted as abstraction, i.e., the object road is simplified and its fundamental characteristics are emphasized as shown in (Mayer and Steger, 1998).

From the last paragraph it follows that the fusion of low and high resolution can contribute to improve the reliability of road extraction. Additionally, details like road markings, which can be recognized at a resolution of less than 0.2 m, can be used as further evidence to validate the detected road hypotheses. On one hand, using multiple resolution levels improves the robustness of the road extraction. On the other hand, it results in different features at each resolution level, and the necessity to combine all features of all resolution levels into one road model. The road model condensed from the above findings is illustrated in Figure 1.

This road model describes objects by means of "concepts", and is split into three levels defining different points of view. The *real world* level comprises the objects to be extracted and their relations. On this level the road network consists of intersections and road links that connect intersections. Road links are constructed from road segments. In fine scale, road segments consist of pavement and markings. The concepts of the real world are connected to the concepts of the *geometry and material* level via *concrete* relations, which connect concepts representing the same object on different levels (Tönjes, 1997). The geometry and material level is an intermediate level which represents the 3D-shape of an object as well as its material (Clément et al., 1993). The idea behind this level is that it describes objects independently from sensor characteristics and viewpoint, which is in contrast to the *image* level. Road segments are linked to the "mostly straight bright lines" of the image level in coarse scale. In contrast to this, the pavement as a part of a road segment in fine scale is linked to the "elongated bright region" of the image level via the "elongated, flat concrete or asphalt region."

Whereas the fine scale gives detailed information, the coarse scale adds global information. Because of the abstraction in coarse scale, additional correct hypotheses for roads can be found and sometimes also false ones can be eliminated, while details, like exact width and position, or markings, are integrated from fine scale. In this way the extraction benefits from both scales.

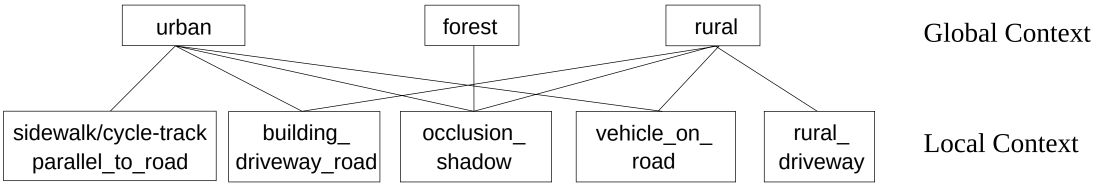


Figure 2: Global and local context

2.2 Context

The road model described above comprises knowledge about radiometric, geometric, and topological characteristics of roads. This model is now extended by context. Background objects, like buildings, trees, or vehicles, can support road extraction (for example, usually there is a road to every building), but also interfere (e.g., a building occludes a part of a road; roofs might look similar to roads). This interaction between road objects and background objects is here modeled *locally* and *globally* (Fig. 2).

In the local context, typical relations between a small number of road and background objects are modeled. Situations, in which background objects make road extraction locally difficult are in rural area, for example, paths to agricultural fields or individual cars. Driveways to buildings cause problems in urban areas. Buildings are mostly parallel to roads. In urban areas sidewalks and cycle tracks are running parallel to roads, potentially hindering or supporting road extraction. For the local context, these situations are described in sketches. E.g., the local context *occlusion_shadow* (Fig. 3) illustrates a situation where a high object occludes a part of a road or casts a shadow on a road, thereby causing the detection of two disconnected road segments. Other local contexts are, e.g., *rural_driveway, building_driveway_road*, or *sidewalk/cycle-track_parallel_to_road*. These basic local contexts can be aggregated into more complex local contexts.

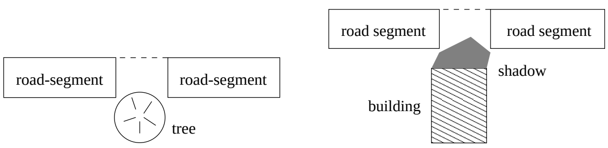


Figure 3: Local context *occlusion_shadow*

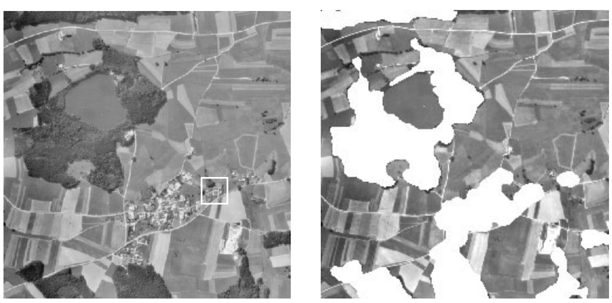
It is not necessary to take every local context into account always. Relations to background objects and their relevance for road extraction depend also on the region where they occur, i.e., on the global context. For instance, roads in urban or suburban areas have a quite different appearance from roads in forest areas or in rural areas. The differences in appearance are partly consequences of different relations between roads and background objects. Therefore, this paper proposes to use different local contexts for different areas, i.e., different global contexts. Here, *urban, forest*, and *rural* contexts are distinguished. The global context is not only relevant for the importance of the local contexts, but also for the extraction of objects. Experience shows that approaches that are suitable for road extraction in rural areas usually cannot be applied in other global contexts without modifications. In forest or urban areas other parameter settings might be necessary or, more likely, even a completely different approach is required. From this, it is clear that the global context enables a more efficient use of the knowledge about roads. In Figure 2, some frequently occurring local contexts are assigned to the global contexts.

3 STRATEGY

In addition to the road and context models, the strategy, i.e., the knowledge about how and when to use which part of the model, is very important for the performance of the approach.

The basic idea of the proposed strategy is to focus the extraction process on those parts of the road network that can be detected most easily and reliably, and that are in addition useful to guide the further extraction. How difficult the extraction of a certain feature is depends strongly on the context in which it is to be extracted. In urban and forest areas knowledge about geometry and radiometry alone is often insufficient because of occlusions and shadows. On the other hand, with a simple model, relying only on attributes of the road itself, good results can be expected for rural areas.

As a consequence of these considerations, road extraction starts in rural areas. Figure 4 shows the result of a texture-based segmentation of *rural* areas in an image with a reduced resolution of about 4 m. The segmentation makes use of the texture filters proposed by (Laws, 1980) and incorporates morphological operations to smooth the boundaries. The pixel size on the ground for this example is about 0.45 m in the high resolution image.



a) b)

Figure 4: a) Aerial image b) *rural areas*

Road hypotheses: According to the road model, apart from the original image also a version of the image with a reduced resolution is used. The lines extracted in the reduced-resolution (about 2 m) image are used to select edges extracted from the original resolution which are candidates for roadsides (cf. Fig. 5 a): The distance between

a) b)

Figure 5: a) Input for the fusion: Lines from coarse scale (black, dashed), edges from fine scale (white) b) Hypotheses for road segments

pairs of edges must be within a certain range, they have to be almost parallel, and the area enclosed by them should be quite homogeneous in the direction of the road. In addition, for each pair of candidates for roadsides, a corresponding line has to exist in the reduced resolution. The selection of edges as roadside candidates is described in detail in (Steger et al., 1995). From these roadsides, road segments are generated (Fig. 5 b). Road segments are represented by the points of their medial axes, and are attributed by the road width.

Grouping: Most of the road segments derived from the fusion of line and edge extraction are not directly connected, and there are also false hypotheses. Therefore, the main task in the next steps is the linking of correct and the elimination of false hypotheses. Initially, the road segments are quite short. The grouping into longer segments, i.e., the closing of gaps and the elimination of false hypotheses, is done according to the "hypothesize and test" paradigm. Hypotheses about which gaps are to be bridged are generated starting with geometric criteria (absolute and relative distance, collinearity, width ratio) and radiometric criteria (mean gray value, standard deviation). Then, the hypothetic road segments are verified in the image. The verification consists of up to three levels: In the first level, radiometric properties of the new segment are compared to the segments to be linked. The geometry of the new segment is defined by the direction at the endpoints of the segments to be linked. If the radiometric attributes do not differ too much, the connection hypothesis is accepted. If not, the verification switches to the second level. Here, a so-called "ribbon snake" is applied to the gradient image, to find an optimum path for the link. In case this verification also fails, a third level is used, in which an explanation by local context is tried to be achieved. The local context is used as last and apparently weakest verification method to explain and close gaps.

According to the above mentioned criteria, hypotheses for connections are generated and verified. This is done iteratively. For every new iteration the maximum length of a gap to be bridged is increased, while the thresholds for other criteria are only slightly relaxed. To avoid hard thresholds for a single criterion in the evaluation of a hypothesis for a connection, all criteria are combined into one value. In parallel to increasing the maximum length of the gaps that are allowed to be bridged, short and unconnected hypotheses for road segments, i.e., hypotheses that are false with a high probability, are eliminated. Figure 6 shows an intermediate result of this grouping process. As can be seen, the mainly collinearity-based strategy sometimes fails, especially for curved segments.

Figure 6: Grouping: Intermediate result

After increasing the threshold for the distance, in the following iterations the constraints for collinearity are relaxed as well. During this phase of the grouping, snakes, (Kass et al., 1988), especially ribbon snakes, become increasingly important. Snakes work according to the principle of energy minimization: The so-called “internal energy” enforces geometric constraints, e.g., length and smoothness of a path. Contrary to this, the so-called “external energy” pushes the snake towards image features. By minimizing internal and external energy simultaneously, image information and geometric properties are fused. In contrast to the conventional snake, the ribbon snake has an additional parameter for the width at each line point. The image features the ribbon snake deals with are anti-parallel edges on both sides of its center line. Using ribbon snakes, road extraction becomes feasible for very fragmented edges and in cases where only one roadside is visible. Bridging a gap between two road segments is performed in two phases: In the first phase, the width of the ribbon is fixed and only the position of its axis is optimized. This is done in analogy to a zip, but starting from both ends, c.f., “ziplock” snake in (Neuenschwander et al., 1995). In the second phase, only the width is optimized, i.e., adapted to the image features. The hypothesis is accepted if the variance of the width is still low after this second step. A more detailed description of this technique is given in (Mayer et al., 1998).

In those cases where the evidence in the image is insufficient to confirm a connection hypothesis, information about the local context of the particular road segment is considered. In other words, a plausible explanation must be given why not enough evidence for a road can be found in the image and the gap can be bridged in spite of this. In this case, especially, the local context *occlusion_shadow* is important. The main part of the information needed for it can be derived from a digital surface model (DSM) and information about when and where the image was taken. With this, shadowed and occluded areas can be predicted and used to explain the gap. For shadows the coarse prediction can be refined in the original image. However, the information about background objects is not required with a high level of detail and accuracy.

Intersections: After the generation of hypotheses for connections and their verification, the road network has to be constructed, i.e., the intersections that link the roads have to be extracted. The generation of hypotheses for intersections is mainly based on geometric calculations: Extracted road segments are extended at their unconnected end points. The length of the extension depends on length and width of the particular segment. If an extension intersects an already existing road segment, a new road segment is constructed, which connects the intersection point with the extended road. The verification of these new road segments is done in the same manner as for the gaps.

Ideally, after this step all road hypotheses are connected, and there is a path between every pair of points on the extracted road network. Usually, such a result cannot be expected. First, due to the limited size of the images some of the nodes will be outside of the image. Second, the results are not error-free. Especially in urban and forest areas only fragments of the network can be expected to be extracted. Because the extraction is reliable only in rural areas, the network characteristics of roads can neither be optimally exploited for the extraction nor for an internal evaluation. However, within a limited scope, it is possible to use topological relations to rate the importance of the roads in the network and to eliminate some of the remaining false hypotheses.

4 EVALUATION OF THE RESULTS

The quality of the results is not the same in different global contexts. For the rural area, for which the approach is optimized, the results are fairly correct and complete (c.f. Fig. 7). The most important condition for an acceptable result is a good contrast between road and background. Based on their width, agricultural paths, which are also extracted, can easily be distinguished from other roads. By integrating knowledge about global and local context, the semantics of individual roads can be determined in more detail. For example, driveways to buildings can be distinguished from paths to fields. Additionally, the existence of road markings can be used as evidence for a correct hypothesis. However, a reliable extraction of markings can only be achieved at resolutions better than 0.3 m.

Figure 7: Extracted roads and intersections

In residential areas there are problems caused by background objects. Roads appear very fragmented in the image and therefore, right from the start, there are fewer correct hypotheses for roads. What is more, the hypotheses cannot be easily grouped because of the many gaps. Since the proposed approach is primarily a bottom-up extraction scheme, the small number and the poor quality of the start hypotheses become the bottleneck for the further grouping in residential areas. The result shown in figure 7 points out the limits of the approach: Most of the roads extracted in the open rural area are ending outside the village. A thorough integration of DSM information,i.e., eliminating wrong hypotheses for roads that lie on the roofs of buildings, which is not implemented by now, is expected to improve the results in this case.

A quantitative evaluation of the results according to the evaluation scheme in (Heipke et al., 1998) has been carried out on a set of test images for which reference data was plotted manually. This evaluation showed that the results for the open rural area are quite reliable (>95%) and also relatively complete (80%–90%). The geometric accuracy for the correctly extracted road axes is about one pixel, i.e., 0.3–0.5 m.

5 DISCUSSION AND OUTLOOK

The proposed approach is well suited for images with a resolution of 0.2 to 0.5 m. A resolution of less than 0.2 m results in a large number of edges and a more inhomogeneous appearance of roads. On the other hand, for a resolution of more than 0.5 m the regions for roads become very narrow and for most roads only line extraction can be used. Whether the resolution of future satellite imagery of about 0.8 m is sufficient for the proposed approach still has to be investigated. For the extraction of roads in urban and forest areas, the extraction in the open rural area provides quite reliable starting points. However, the propagation of the road extraction into these areas requires additional extraction and tracking algorithms and must be based also on other evidence, e.g., grouping of road markings and vehicles.

At this point, the results are not absolutely reliable and complete. Hence, in operational use, a human operator would be needed to edit the results, i.e., to delete wrongly extracted roads and to insert missing parts. Nevertheless, the approach shows that good results can already be achieved based on grouping algorithms. A noticeable improvement seems possible with a more complete integration of context information and global grouping criteria, e.g., as proposed in (Steger et al., 1997). Furthermore, it seems that the results could be improved by using a more detailed modeling of intersections, e.g., as it is addressed in (de Gunst and Vosselman, 1997, Boichis et al., 1998).

ACKNOWLEDGMENTS

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) under grant no. Eb 74/8-3.

REFERENCES

Baumgartner, A., Eckstein, W., Mayer, H., Heipke, C. and Ebner, H., 1997. Context Supported Road Extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag Basel, pp. 299–308.

Boichis, N., Cocquerez, J. and Airault, S., 1998. A top down strategy for simple crossroads automatic extraction. In: International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 2, Cambridge, UK, pp. 19–26.

Bordes, G., Giraudon, G. and Jamet, O., 1997. Road Modeling Based on a Cartographic Database for Aerial Image Interpretation. In: Semantic Modeling for the Aquisition of Topographic Information from Images and Maps, Birkhäuser Verlag Basel, pp. 123–139.

Clément, V., Giraudon, G., Houzelle, S. and Sandakly, F., 1993. Interpretation of Remotely Sensed Images in a Context of Multisensor Fusion Using a Multispecialist Architecture. IEEE Transactions on Geoscience and Remote Sensing 31(4), pp. 779–791.

de Gunst, M. and Vosselman, G., 1997. A Semantic Road Model for Aerial Image Interpretation. In: Semantic Modeling for the Aquisition of Topographic Information from Images and Maps, pp. 107–122.

Grün, A. and Li, H., 1997. Semi-Automatic Linear Feature Extraction by Dynamic Programming and LSB-Snakes. Photogrammetric Engineering & Remote Sensing 63(8), pp. 985–995.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O., 1998. External Evaluation of Automatically Extracted Roads. Photogrammetrie – Fernerkundung – Geoinformation (2/98), pp. 81–94.

Kass, M., Witkin, A. and Terzopoulos, D., 1988. Snakes: Active contour models. International Journal of Computer Vision pp. 321–331.

Laws, K., 1980. Texture Image Segmentation. Ph.D. thesis, Dept. of Engineering, University of Southern California.

Mayer, H. and Steger, C., 1998. Scale-Space Events and Their Link to Abstraction for Road Extraction. ISPRS Journal of Photogrammetry and Remote Sensing 53(2), pp. 62–75.

Mayer, H., Laptev, I. and Baumgartner, A., 1998. Multi-Scale and Snakes for Automatic Road Extraction. In: Fifth European Conference on Computer Vision, pp. 720–733.

McKeown, D. M. and Denlinger, J. L., 1988. Cooperative Methods for Road Tracking in Aerial Imagery. In: Computer Vision and Pattern Recognition, pp. 662–672.

Neuenschwander, W., Fua, P., Székely, G. and Kübler, O., 1995. From Ziplock Snakes to Velcro^{tm} Surfaces. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag Basel, pp. 105–114.

Ruskonè, R., Airault, S. and Jamet, O., 1994. A Road Extraction System Using the Connectivity Properties of the Network. Zeitschrift für Photogrammetrie und Fernerkundung (5/94), pp. 174–180.

Steger, C., Glock, C., Eckstein, W., Mayer, H. and Radig, B., 1995. Model-Based Road Extraction from Images. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag Basel, pp. 275–284.

Steger, C., Mayer, H. and Radig, B., 1997. The Role of Grouping for Road Extraction. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag Basel, pp. 245–256.

Tönjes, R., 1997. 3D Reconstruction of Objects from Aerial Iamges Using a GIS. In: International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 3-2W3, pp. 140–147.

Trinder, J. C. and Wang, Y., 1998. Knowledge-based road interpretation in aerial images. In: International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 4, Stuttgart, Germany, pp. 635–640.

Vosselman, G. and de Knecht, J., 1995. Road Tracing by Profile Matching and Kalman Filtering. In: Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser Verlag Basel, pp. 265–274.