SELF-DIAGNOSIS WITHIN AUTOMATIC ROAD NETWORK EXTRACTION

Stefan Hinz, Christian Wiedemann, Heinrich Ebner

Chair for Photogrammetry and Remote Sensing
Technische Universiteat Muenchen, Arcisstrasse 21, 80290 Muenchen, Germany
{Stefan.Hinz, Christian.Wiedemann, Heinrich.Ebner}@bv.tum.de
www.RemoteSensing-TUM.de

Commission II, IC WG II/IV

KEY WORDS: Road Extraction, Internal Evaluation, Self-Diagnosis

ABSTRACT:

In general, automatic object extraction systems are not to expect to deliver absolutely perfect results and, thus, for meeting predefined application requirements, a human operator has to inspect the automatically obtained results. In order to speed-up the time- and cost-intensive inspection, the system should provide the operator with confidence values characterizing its own performance. In this paper, an approach for self-diagnosis is presented which is part of an existing road extraction system. The design of the self-diagnosis tool is based on considerations from three different perspectives: From an operational point of view, one has to clarify which *representation* of the evaluation result is both convenient and effective for a human operator using the system. The choice of the *evaluation criteria*, on the other hand, is affected by the respective object model and extraction strategy on which the system is founded. Finally, an appropriate *theory and methodology* should be applied for evaluating and combining the criteria.
We exemplify these aspects with the self-diagnosis tool of our road extraction system. Two different types of representation are incorporated to provide the user with the essential information about the extraction quality: A universal and a sectional representation. The criteria used for internal evaluation are derived from the object model. It is special attention payed to employ criteria not used during prior steps of extraction in order to reach a highly unbiased evaluation. Fuzzy-set theory is used as theoretical framework for knowledge representation for both extraction and evaluation.
In order to analyze the quality of the self-diagnosis, we extracted roads in a test series of aerial images and matched the internally evaluated result, i.e., the labelled road sections, to a manually plotted reference. The comparison shows the benefits but also some remaining deficiencies of the self-diagnosis tool.

1. INTRODUCTION

Internal evaluation (self-diagnosis) and external evaluation of the obtained results are of major importance for the relevance of automatic systems for practical applications. Obviously, this statement is also true for automatic image analysis in photogrammetry and remote sensing. However, so far only relatively little work has been carried out in this area. This is mostly due to the moderate results achieved. Only recently, automatic systems for the extraction of objects from imagery reached a state in which a systematic evaluation of the results seems to be meaningful. Both, self-diagnosis and external evaluation yield quantitative results which should be as much as possible independent of a human observer (Foerstner, 1996). The aim of the self-diagnosis is, to determine the geometrical and semantical accuracy of the extracted objects during the extraction process. This information has to be derived from redundancies within the underlying data. Therefore, the self-diagnosis is an integral part of the extraction process. The results of the self-diagnosis will be important if the extraction results are combined with other data, e.g., if they are fused with extraction results stemming from other extraction approaches or if they are used for the update of GIS data. In contrast, an external evaluation determines quality measures referring to actually existing objects. This is usually carried out by a comparison of the extraction results with reference data of superior quality. Therefore, the external evaluation is independent of the extraction approach. The results of the external evaluation can be used for a comparison of different extraction approaches and to detect the strengths and weaknesses of such approaches with respect to practical applications. Recent work on external evaluation can be found, for instance, in (Heipke et al., 1998; Shufelt 1999; Wiedemann and Ebner, 2000).

In this paper, an approach for self-diagnosis is presented which is part of an existing road extraction system developed over the past years. The self-diagnosis focusses on the semantic quality of the extraction rather on its geometric accuracy. We first give a brief description of recent approaches of road extraction and evaluation influencing our work Sect. 2. In Sect. 3, we outline the basic principles of the extraction system followed by the description of the new approach on self-diagnosis (Sect.4). Finally, the results of self-diagnosis are analyzed by means of an external evaluation in Sect. 5.

2. RELATED WORK

It is well-known that besides local road features like width and curvature of a road, which can be detected by rather simple image processing methods, also global and functional properties play an important role for road extraction – especially when using a relatively low ground resolution of 2m and less. An approach which mainly deals with the network characteristics of roads is described in (Vasudevan et al., 1988). After line extraction neighboring and collinear lines are searched for. For each line the locally best neighbor is determined based on the difference in direction and the minimum distance between the end points. A link is inserted which connects the line with its best neighbor. Connected lines form so-called line clusters which represent parts of the road network.

In (Steger et al., 1997) the necessity of global grouping instead of a purely local determination of the continuation of the road is underlined. Each possible link between the endpoints of road segments is regarded as a link hypothesis. A weight is assigned to each link hypothesis which depends on its lengths and direction differences with the adjacent road segments. A weighted graph is constructed from the road segments and the link hypotheses. Seed points are selected on road segments close to the image border. For each pair of seed points the best path is searched through this weighted graph. Only those link hypotheses are accepted which lie on a best path between two seed points.

Fischler et al. (1981) determine a cost array for the entire image from the output of several low-level extraction schemes. The cost array contains a likelihood that a particular pixel belongs to a road. Then, optimal paths between two very likely road pixels are determined by the F* algorithm. The output of this step serves as a road hypothesis, which is deleted from the cost array, and the process is repeated with other likely road pixels to obtain additional road hypotheses. (Fischler, 1994) deals with the problem of extracting single linear structures where a certain percentage of points is missing in a noisy background. This is done by first eliminating most of the noise through an operation similar to a binary rank operator. Then, a neighborhood graph is constructed between pixels having a certain maximum distance, and the diameter path of the minimum spanning tree of this graph is extracted as the single salient line.

The road tracking scheme of (Geman and Jedynak, 1996) uses a simple matching method to search for locally linear structures first. Then, these segments are fed into a complex, recursive decision process in order to remove as much uncertainty as possible about the position of the road. Hence a larger search space is spanned if the road is poorly visible, e.g., due to noise or occlusions. Temporarily, more then one path can be tracked from which finally the path is selected that has the lowest uncertainty about the road position. This can be seen as a form of internal evaluation during the extraction process helping to bridge occlusions and other disturbances.

An interesting approach regarding the role of internal evaluation is employed in the system of (Tupin et al., 1999) for finding consistent interpretations of SAR scenes (Synthetic Aperture RADAR). In a first step, different low level operators with specific strengths are applied to extract image primitives, i.e., cues for roads, rivers, urban/industrial areas, relief characteristics, etc. Since a particular operator may vote for more than one object class (e.g. road *and* river), a so-called focal and non-focal element is defined for each operator (usually the union of real-world object classes). The operator response is transformed into a confidence value characterizing the match with its focal element. Then, all confidence values are combined in an evidence-theoretical framework to assign unique semantics to each primitive attached with a certain probability. Finally, a feature adjacency graph is constructed in which global knowledge about objects (road segments form a network, industrial areas are close to cities, ...) is introduced in form of object adjacency probabilities. Based on the probabilities of objects and their relations the final scene interpretation is formulated as a graph labelling problem that is solved by energy minimization.

3. ROAD EXTRACTION SYSTEM

The road extraction system is designed for extracting roads from imagery with a ground pixel size of about 2 m by 2 m. Due to the limited ground resolution a road model purely based on local characteristics is rather weak. For this reason, the road network is also considered, and regional and global properties are incorporated into the model. Locally, radiometric properties play the major role. The road is modelled as a line. It can have a higher or lower reflectance than the surroundings. Geometry is explicitly introduced on the regional level. Regional characteristics incorporate the assumption that roads are composed of long and straight segments with almost constant width. Globally, roads are described in terms of topology: the road segments form a network, in which all segments are topologically linked to each other.

The extraction strategy is composed of different steps. After the extraction of lines, postprocessing of the lines is performed with three different tasks in mind:

1. Increase the probability that lines either completely correspond to roads or to linear structures not being roads.
2. Fuse lines extracted from different images or channels.
3. Prepare lines for the generation of junctions.

Then a weighted graph is constructed from the lines and the gaps, i.e. potential connections, between them. The weights are derived from local (radiometric) and regional (geometric) criteria by transforming line features like length, curvature, and width into fuzzy values using fuzzy-set theory. Road network generation is carried out by calculation of "optimal paths" between various pairs of points which are assumed to lie on the road network with high probability. By this, global (topological) information is introduced into the extraction process.

A typical result, which can be achieved with our system is illustrated in Fig. 1. Figure 1a) shows the input image with a ground resolution of 2m, and in Fig. 1b) the reference data is plotted. As can be seen from Fig. 1c), most of the roads are extracted correctly (completeness 82%, correctness 94%). However, there are still some roads are missing and some are "false alarms". The geometric accuracy of the road axes yields 0.84 pixel units. In order to test the behavior of the self-diagnosis tool, another extraction result has been generated using intentionally relaxed parameter settings (Fig. 1d). Now a completer road network could be extracted (95%) by debiting, however, the correctness that reaches only 67%. Table 1 summarizes the external evaluation of both extractions.

	Tuned param.	Relaxed param.
Completeness [%]	82.0	95.0
Correctness [%]	94.0	67.4
RMS [pix]	0.84	0.95
Length [pix]	3978.2	6625.5

Table 1: External Evaluation of extractions of Test Scene I

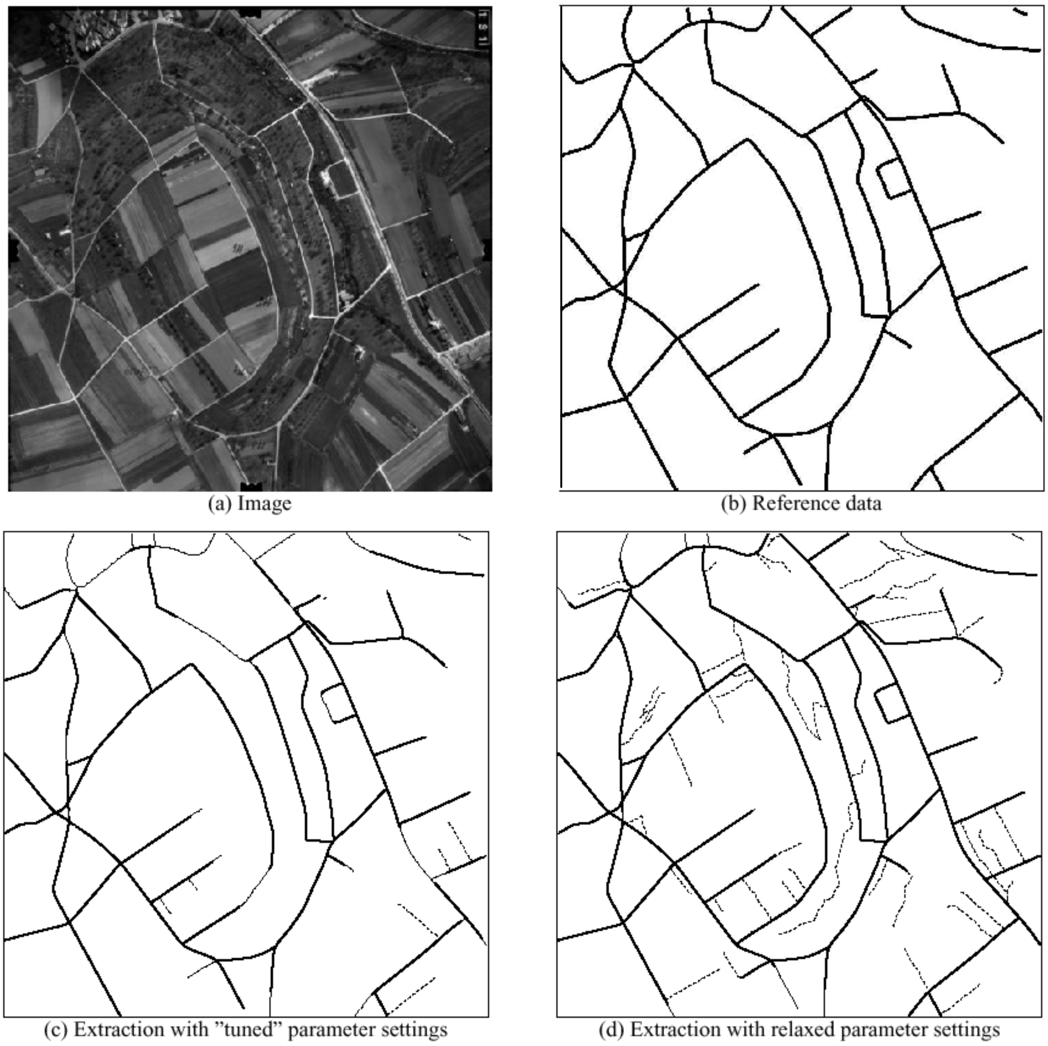


Figure 1: Image, reference, and extraction data of Test scene I: bold lines are correctly extracted roads, thin lines indicate missing roads, dashed lines are "false alarms"

4. SELF-DIAGNOSIS TOOL

The result above may serve as an example that automatic object extraction systems are not to expect to deliver absolutely perfect results. Thus, for meeting predefined application requirements, a human operator has to inspect the automatically obtained results which is naturally time- and cost-intensive. In order to speed-up the inspection, the system should provide the operator with confidence values characterizing its own performance and assisting and guiding the operator during the editing phase.

4.1 General Aspects of Self-Diagnosis

We propose to consider following aspects of a self-diagnosis tool: From an operational point of view, we first have to clarify which *representation* of the evaluation result is both convenient and effective for a human operator using the system. The respective object model and extraction strategy affect which *evaluation criteria* should be chosen. Finally, an appropriate *theory and methodology* should be applied for evaluating and combining the criteria. Although being generally application independent it is desirable to utilize a theory which is in some way consistent with the mathematical foundation of the extraction system.

4.2 Representation

In our case of road extraction, we define two types of representations for illustrating the scores of the self-diagnosis: A *universal representation* which indicates the overall quality of the extraction result. This may help an user in deciding quickly if the whole scene has to be re-processed – be it automatically or manually. In contrast, a *sectional representation* provides the user with more detailed information about the quality of the extracted road network. The road network is split into different road sections (see below) and, for each individual section, the operator gets clues if problems occurred during the extraction process. Consequently, the sectional representation can be used to guide the operator through the network concentrating on those parts of the network that are potentially subject for manual editing which, in the end, helps to shorten the editing phase.

Now, the question arises what kind of information is to present to a human operator by the system. In this context, a frequently mentioned approach is to label the extraction results based upon the so-called traffic-light paradigm (Foerstner, 1996): A green light stands for a result found to be correct, a yellow light indicates that a further investigation / verification is needed, and a red light implies an incorrect result. The self-diagnosis of our system is also based on the traffic-light paradigm, however, we treat the yellow category as a linear transition from red to green. Thus, the operator may inspect the distribution of the evaluation scores inside the yellow category. Especially for the sectional representation, we extend this approach by attaching each category with a small number of attributes that allow a deeper insight which kind of problem occurred during the extraction.

4.3 Evaluation Criteria

The road model described above plays a key role for defining criteria which can evaluate the results internally. In order to guarantee an unbiased evaluation, the criteria should (at least theoretically) be independent of model components that have been already used during the extraction. Regarding the extraction system described above, radiometric and geometric criteria are incorporated on the local and regional level. On the global level, however, only topological criteria have been employed, and so we still can use *global* radiometric and geometric criteria for self-diagnosis. Therefore, we analyze the optimal paths – or large parts of them – found by the global grouping step for their radiometric and geometric characteristics. More specifically, the evaluation strategy consists of two steps:

First, the extracted road network is divided into independent road sections by splitting it at junctions and points with very weighted mean of the fuzzy values used during extraction of this road section.

2. Line support of a road section, i.e., the length ratio between those parts of a road section which originate from line extraction and its total length.
3. Length of a road section.
4. Averaged curvature radius along a road section.
5. Saliency of a road section, i.e., the maximum absolute eigenvalue (see (Steger, 1998) averaged along a road section).

Please note that, even though some of these criteria have been used within the lower levels of road extraction, none of them is employed at the (higher) level of road network generation. Hence, they evaluate roads from a more global point of view and can greatly serve as an independent measure for self-diagnosis. A special treatment is needed for criterion 3: Consider two neighboring junctions connected by a short straight road section. In this case, the road section would get a rather bad rating because of its shortness although being truly a road. Therefore, criterion 3 is omitted whenever a short road section *directly* connects two junctions.

4.4 Evaluation Methodology

As for the extraction process, we utilize fuzzy-set theory as mathematical framework for representing the human knowledge given by the above criteria. Predefined fuzzy functions are used as weighting functions for evaluating road sections regarding the above criteria. The outputs of the different functions are aggregated by the fuzzy-"and" operation which results in a so-called overall fuzzy value between 0 and 1 for each road section. This value specifies how a particular road section satisfies all criteria. In addition to this, we store the output of each fuzzy function as attribute of the sectional representation. From the overall fuzzy value a length-weighted histogram is computed. All road sections yielding a better evaluation than the median of the histogram are assigned to the green category and every road section attached with fuzzy value 0 is assigned to the red category. The remaining road sections are labelled as yellow. Applying the median as threshold implies that the road extraction system is generally able to detect roads with a correctness of at least 50%, which is indeed a reasonable assumption. For the universal representation, we compute the length-weighted mean of all overall fuzzy values from the histogram expressing the total quality of the extraction. Depending on the median also the total quality can attain each traffic-light category. The histogram of the overall fuzzy values is stored as attribute for the universal representation.

4.5 Implementation

For allowing a quick and effective inspection of the results by a human operator, the universal representation is displayed first. It consists of an overview window visualizing the whole scene with the extraction result superimposed, thereby each road section being colored according to its category. In addition, the total quality and its underlying distribution are displayed (see Fig. 2). Whenever a particular road section is sought to be

Figure 2. Global Representation of self-diagnosis. Top: Length-weighted histogram of evaluation scores (left-most bar indicates mean value=*yellow* in this example), thin vertical lines indicate thresholds between categories. Bottom: Overview window (bold=*green*, thin=*yellow*, dotted=*red*)

inspected in more detail, it can be selected via a simple mouse click and the sectional representation is visualized in a separate cutout. Furthermore, the road section's overall fuzzy evaluation and the fuzzy values resulting from each individual criterion are displayed (see Fig. 3). Based on the visualization and the quality information, the operator may decide how to proceed with a road section – whether it should be retained, deleted, or edited.

5. RESULTS AND DISCUSSION

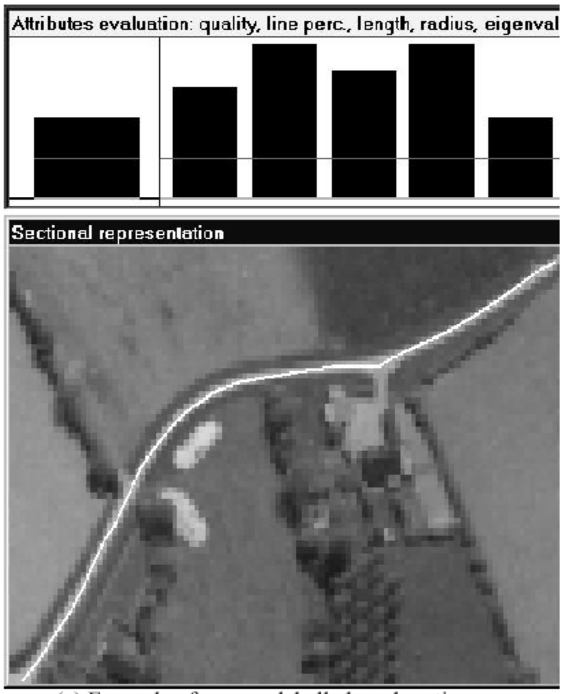
In order to analyze the quality of the self-diagnosis, we extracted roads in a test series of aerial images (see Figs. 2 and 4) and matched the results of self-diagnosis, i.e., the labelled road sections, to a manually plotted reference. The comparison as listed in Table 2 shows that almost every road section of the green category (91% – 98%) and many sections of the yellow category (57% – 76%) are truly roads. The self-diagnosis also detects a high percentage of "false alarms" (83% – 90%). Further interesting aspects of the self-diagnosis tool are captured in Table 3. Here, the result obtained from Test Scene I using "tuned" parameter settings is compared with the union of green- and yellow-labelled road sections resulting from the extraction using "relaxed" parameter settings (see also Sect. 3 and Fig. 1). The self-diagnosis suggests to reject only 0.6% of the "tuned" extraction but 29.5% of the "relaxed" extraction. By means of the external evaluation it turns out that completeness increases by almost 10 percentage points while correctness decreases only by 3 percentage points when using the "relaxed" parameter settings. On one hand, this confirms the ability of the self-diagnosis tool to discern between correct and incorrect results to a great extent. On the other hand, the quality of road extraction gets less dependent on ideal input parameters. This eases the initialization of the extraction significantly and thereby shortens the overall time needed for processing. However, it must be noted that the self-diagnosis is not able to detect all erroneous extractions, as can be seen from the generally lower and quite varying correctness values of the yellow category (see Table 2). Adjusting the weighting functions to enhance this result would not meet the problem's heart, since a system designed for operational use should be insensitive against burdensome parameter tuning as much as possible. Of course, one could also change the thresholds between the three evaluation categories. In an ideal case these thresholds can be derived automatically from the histogram if their justification is statistically apparent. This would be possible, for instance, in case of a "stretched" bimodal histogram. A careful investigation of a variety of results has shown, that the evaluation criteria involved up to now are only partly able to generate such histograms. Hence, our plans for the future are to integrate more knowledge about the road network into the self-diagnosis. More specifically, we plan to integrate the function of particular road sections within the network (not only independent road sections as until now). The evaluation of junctions would be another area of future research.

	Correctness of extraction (*green*)	Correctness of extraction (*yellow*)	Correctness of rejection (*red*)
Test Scene I	97.7%	75.6%	88.6%
Length [pix]: 6625.5	48.7%	21.8%	29.5%
Test Scene II	91.4%	57.1%	90.1%
Length [pix]: 9571.6	48.1%	47.4%	4.5%
Test Scene III	96.9%	64.6%	82.5%
Length [pix]: 12337.1	49.3%	31.1%	19.5%

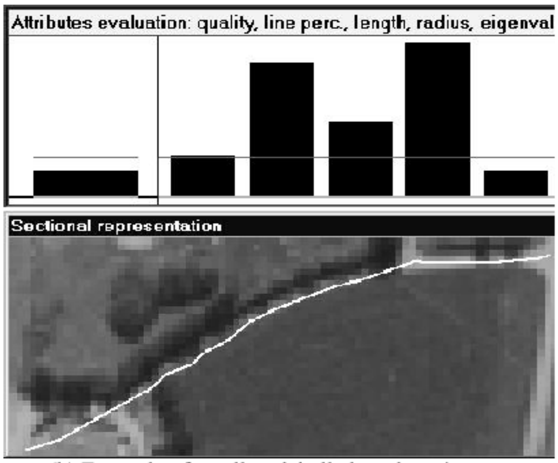
Table 2: Correctness of self-diagnosis

	Tuned parameters	Relaxed parameters (*green + yellow*)
Completeness [%]	82.0	91.8
Correctness [%]	94.0	90.9
RMS [pix]	0.84	0.89
Length (tot.) [pix]	3978.2	4669.7
green [%]	50.4	48.7
yellow [%]	49.0	21.8
red [%]	0.6	29.5

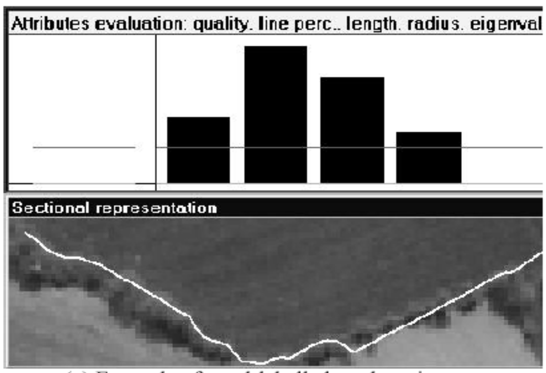
Table 3: Comparison of results achieved with different parameter settings for Test Scene I



(a) Example of a green-labelled road section



(b) Example of a yellow-labelled road section



(c) Example of a red-labelled road section

Figure 3: (a) – (c) Sectional Representation of self-diagnosis. Bars indicate results of road section evaluation ranging [0 ; 1]. Left-most bar is overall fuzzy value. Thin horizontal lines indicate threshold of green category.

a) Test Scene II

b) Result of self-diagnosis for Test Scene II

c) Test Scene III

d) Result of self-diagnosis for Test Scene II

Figure 4. More results of self-diagnosis. Bold = *green*, thin = *yellow*, dashed = *red*.

REFERENCES

Fischler, M. A., 1994. The perception of linear structure: A generic liner. In: *Image Understanding Workshop '94*, Morgan Kaufmann Publishers, San Francisco, CA, pp.1565-1579.

Fischler, M. A., Tenenbaum, J. M. and Wolf, H. C., 1981. Detection of roads and linear structures in low-resolution aerial imagery using a multisource knowledge integration technique. *Computer Graphics and Image Processing* 15, pp. 201-223.

Forstner, W., 1996. 10 pros and cons against performance characterization of vision algorithms. In: H. I. Christrnsen, W. Forstner and C. b. Madsen(eds), *Workshop on Performance Characteristics of Vision Algorithms*, pp. 13-29.

Geman, D. and Jedynak, B.,1996. An active testing model for tracking roads in satellite images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18(1), pp. 1-14.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O., 1998. External evaluation of automatically extracted road axes. *Photogrammetrie – Fernerkundung – Geoinformation (PFG)* 2/1998, pp.81-94.

Shufelt, J., 1999. *Geometric Constraints for Object Detection and Delineation*. Kluwer Academic Publishers, Boston.

Steger, C., 1998. An Unbiased Detector of Curvilinear Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(2), pp. 113-125.

Steger, C., Mayer, H. and Radig B., 1997. The role of grouping for road extraction. In: *A. Gruen, E. Baltsavias and O. Henricsson(eds), Automatic Extraction Of Man-Made Objects from Aerial and Space Images(II)*, Birkhauser verlag, Basel, pp. 245-256.

Tupin, F., Bolch, I. And Maitre, H., 1999. A First Step Toward Automatic Interpretation of SAR Images Using Evidential Fusion of Several Structure Detectors. *IEEE Transactions on Geoscience and Remote Sensing* 37(3), pp. 1327-1343.

Vasudevan, S., Cannon, R. L. and Bezdek,J. C., 1988. Heuristics for intermediate level road finding algorithms.

Computer Vision, Graphics, and Image Processing 44, pp. 175-190.

Wiedemann, C. and Ebner, H., 2000. Automatic completion and evaluation of road networks. In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 33, part B.