Road Extraction from Digital Aerial Imagery

Technical Report

S. Hinz[1], A. Baumgartner[1], C. Steger[2]

1) Chair for Photogrammetry and Remote Sensing

Technische Universität München
Arcisstr. 21, D-80290 München, Germany
Ph.: +49 89 289-23880, Fax: +49 89 2809573
E-mail: {hinz | albert}@photo.verm.tu-muenchen.de

2) MVTec Software GmbH

Neherstr. 1, D-81675 München, Germany
Ph.: +49 89 457695-0, Fax: +49 89 457695-55
E-mail: steger@mvtec.com

September 21, 2000

1 Introduction

In the past, the automation of road extraction from digital imagery has received considerable attention. Research on this issue is often motivated by the increasing importance of geographic information systems (GIS) and the need for data acquisition and update for GIS. However, the influence of this work on automatic road extraction within an operational environment is still quite small. One of the reasons for this might be that fully automatic approaches in many cases do not yield satisfactory results, i.e., the automatically derived results need too much post editing. In contrast, in semi-automatic approaches the optimization of the interaction between operator and computer is the crucial task.

Background objects, like buildings, cars, or trees, often have a strong influence on the appearance of roads. This also has considerable influence on the image features, which are useful for road extraction, and on the extraction strategy. Therefore, it is important to model not only the properties of the roads but also the relations between roads and these background objects. An investigation on the influence of neighboring objects on road extraction has been made in [Bordes et al., 1997].

In this paper, we propose a fully automatic approach for road extraction, which models roads as well as their context. We distinguish in this work three so-called "global contexts": rural, forest, and urban area, cf. [Baumgartner et al., 1997]. The road extraction is based on the "hypothesize and test" paradigm: After the detection of initial hypotheses for road segments, hypotheses for connections are generated and checked. We also make use of the scale-space behavior of roads and combine line extraction and edge extraction. This fusion step is primarily based on local verification of the image content. A global grouping step is then applied to get a topologically sound hypothesis for the road network. In urban areas the road extraction is based on markings. Markings are grouped into lanes and road segments, and knowledge about relations between lanes and vehicles is used to verify the hypotheses for lanes and roads.

The most relevant previous work is discussed in Section 2. The road model in Section 3 describes the most important parts of the model. Section 4 focuses on road extraction in rural areas. In Section 5 an approach for road extraction in urban areas is presented. A short outlook concludes this paper.

2 Previous Work

The existing approaches cover a wide variety of strategies to extract roads automatically from digital aerial or satellite imagery, or at least to automate parts of the manual extraction process. As GIS-driven approaches for road extraction are more useful for verification than for extraction of new roads, we focus in the discussion of previous approaches on those that also aim at extraction of previously unknown roads.

In semi-automatic approaches an operator provides, e.g., starting points and starting directions on the road for a road following algorithm [McKeown and Denlinger, 1988, Vosselman and de Knecht, 1995]. If an operator measures more than one point on the road an algorithm like F* can be applied to find an optimal path, i.e., the road between these points, [Fischler et al., 1981, Merlet and Zerubia, 1996]. If more than one image is used, this can also be done in 3D [Gruen and Li, 1997]. The advantage of the approaches with multiple points is that the path of the road is more constrained, which results in a more reliable handling of critical areas. A similar approach based on so-called "zip-lock" snakes is presented in [Neuenschwander et al., 1995]. By automatic detection of the seed points, semi-automatic schemes can be extended to fully-automatic ones. An automatic approach is described in [Barzohar et al., 1997]. The selection of starting points is based on gray-value histograms. Further assumptions about geometry and radiometry are described in a Markov random field. Road extraction is then performed by dynamic programming. Another fully-automatic approach for the extraction of road networks from digital aerial imagery has also been proposed by [Ruskoné, 1996]: Hypotheses for connections between automatically detected seed points are checked using geometrical constraints. A more comprehensive survey on models and strategies for road extraction is given in [Mayer, 1998].

In the following paragraphs we discuss the most relevant previous work on fully-automatic road extraction. The approaches can be classified according to the resolution of the input images. In the first category, low resolution ($> 1\,\mathrm{m}$ per pixel) aerial or satellite images are used. In the second category, high resolution (20–50 cm per pixel) aerial images are used.

In [Fischler et al., 1981] roads are extracted from low resolution aerial images, each connecting two given points of the road network. Different low level operators for road extraction from low resolution aerial images are classified into two types: Type I operators are assumed to deliver no false extractions, but some roads might not be found. Type II operators may yield false extractions but are assumed to extract all roads completely. In regions where a type I operator has detected a road, the scores of every type II operator is set to a maximum value (zero costs). By this, multiple type II operators are made commensurate. The result of each type II operator is stored in a cost array. Between two given points the best path is calculated for each type II operator using the F* algorithm. The path which yields the lowest so-called "self normalized average cost" per pixel is chosen as the road. In [Merlet and Zerubia, 1996], the F* algorithm is extended to cliques and to neighborhoods larger than one. By means of the cliques, it is possible to introduce contrast information into the calculation of the minimum cost path. The larger neighborhoods allow for the consideration of the curvature of the final path.

An approach that mainly deals with the network character of roads is described in [Vasudevan et al., 1988]. After line extraction from Landsat TM imagery neighboring and collinear lines are searched for. For each line the best neighbor is determined based on the difference in direction and the minimum distance between the end points. Connected lines form so-called "line clusters", which represent parts of the road network. Only local grouping criteria are used, and a line can be part of multiple clusters.

The approach of [Zlotnick and Carnine, 1993] and [McKeown and Denlinger, 1988] is a two-step approach to extract roads from high resolution images. First, road seeds are generated by grouping anti-parallel pairs of edges. The resulting set of incomplete road axes is completed by a pixel-based grouping procedure, and road seeds are finally selected based on length and straightness. These road seeds are input into a road tracking algorithm, which consists of two independent trackers. In the edge tracker, edge points are extracted in a gray value profile perpendicular to the road direction. At least one edge point must be present for a successful tracking. In the profile tracker, an "average" road profile is matched to the current gray value profile perpendicular to the road direction. If the correlation fails, either a change of road surface material or a disturbance on the road is assumed. Limited model knowledge is used to fuse the results of the two trackers, to explain failures of the trackers, and to continue or abort the tracking. An interesting point is that the combination of the two trackers yields better results than each tracker individually [McKeown, 1990].

In [Ruskoné et al., 1994, Airault et al., 1994], an approach for road network extraction on high resolution imagery is presented. In the first stage, a low level road tracker following homogeneous elongated areas is started at automatically extracted seed points. In the second stage, hypotheses for the connection of the extracted road parts are generated and checked based on geometric criteria like distance and direction. The final stage consists of a geometric adjustment of the extracted road network based on snakes. In urban areas, roads are extracted by detecting cars with a neural network classifier, and then collinear cars are grouped into roads [Ruskoné et al., 1996].

The approach presented in [Price, 1999] extracts roads from high resolution images of urban areas. The road network is modeled as a regular grid with roads of approximately constant width. It is assumed that longer portions of the road sides are visible. The grid spacing and orientation is initialized using three points on the grid. The expansion, verification, and refinement of the grid is based on edge and DEM (Digital Elevation Model) information as well as context.

The above approaches show individually promising parts of a road model and extraction strategy. Data from different sources is often useful. For example, in urban areas DEM information helps to remove false road hypotheses. What is missing is the use of different resolutions of the image data, e.g., to eliminate disturbances like cars on the roads [Mayer and Steger, 1998]. Furthermore, context information has proven to be very important. The strategy of road tracking is promising in automatic approaches to bridge gaps in the extracted road hypotheses. Local grouping is also very useful in this case. However, the function of roads is never modeled explicitly, and hence the use of global grouping seems to be an essential step to generate a correct and complete network. Furthermore, all of these knowledge sources need to be integrated into a single system.

3 Modeling Roads

The examples for roads in aerial imagery which are given in Fig. 1 indicate some of the problems for automatic road extraction. Although, in the real world, roads are objects with a smooth and firm surface and quite well defined geometrical properties, e.g., maximum curvature or lower and an upper bound for their width, their appearance in aerial imagery can be quite different. Small objects like cars or markings interfere the homogeneity of the road surface. They can fairly easily be eliminated using the scale-space properties of roads. Larger objects, like buildings or trees, are more difficult to handle. These 3D-objects pose problems due to occlusions and shadows, which cannot be treated using scale-space theory. However, it is clear, that depending on the resolution and on the complexity of the scene, different features and different strategies seem to be useful for automatic road extraction.

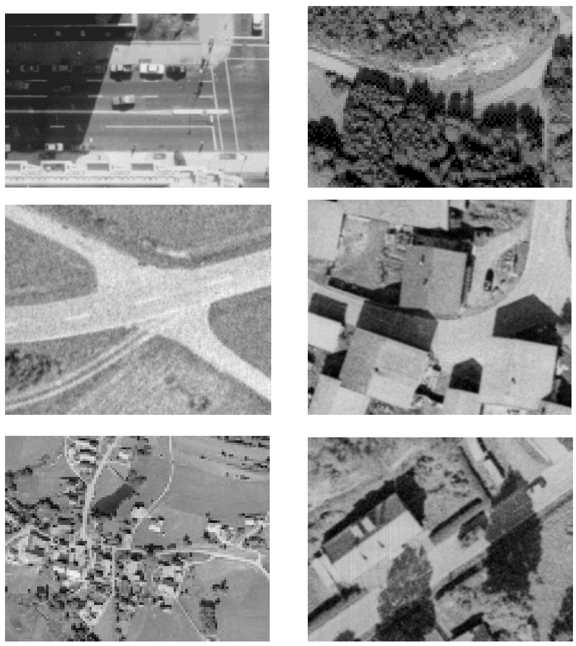


Figure 1: Roads in aerial imagery

After this short analysis of the appearance of roads and the problems which might occur when extracting roads, we now explain the road model on which the extraction is based:

For the proposed approach, the model comprises explicit knowledge about geometry (road width, parallelism of road sides, ...), 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 describes characteristic properties of roads in the real world and in aerial imagery, and represents a road model derived from these properties. The second part 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 that are adapted to specific contexts.

A description of roads in the real world can be derived from their function for human beings: roads are defined as a place, where one may ride, i.e., an open or public passage for vehicles, persons, or animals. They are important for communication and transportation between different places. Therefore, roads are organized as a network. The denser an area is inhabited and the more intensively it is used, the denser the road network is. With respect to their importance, network components are classified into a hierarchy of different categories with different attributes. According to the different categories, roads differ with respect to minimum curvature radius and maximum allowed slope. Some important attributes for parts of the road network are the type and state of the road surface material, existence of road markings, sidewalks, and cycle-tracks, or legal instructions, such as traffic regulations.

The appearance of roads in 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 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. Contrarily 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, cf. Fig. 1.

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 substructure of the road, 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 this makes it necessary to combine all features of all resolution levels into one road model.

3.1 Road Model

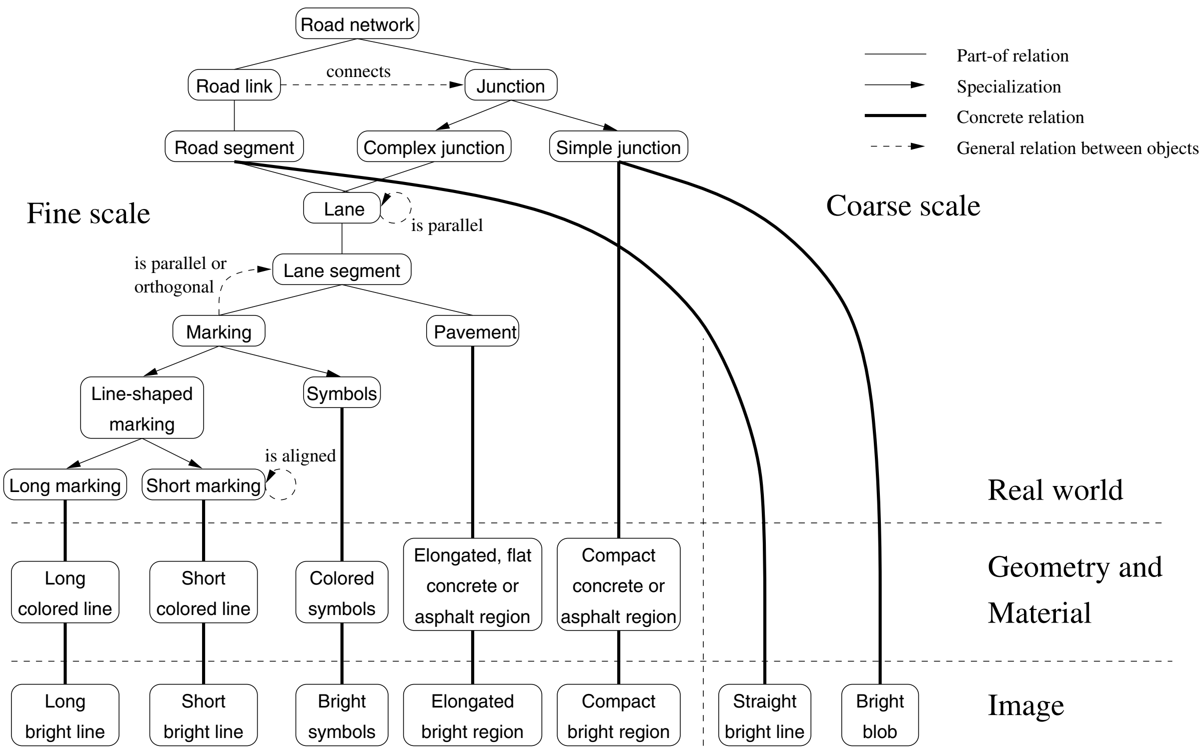


Figure 2: Road model

The road model condensed from the findings above is illustrated in Fig. 2. 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 junctions and road links that connect junctions. Road links are constructed from road segments. In fine scale, road segments are aggregated by lanes, which consist of pavement and markings. For markings there are two specializations: Symbols and line-shaped markings. The concepts of the real world are connected to the concepts of the *geometry and material* level via *concrete* relations [Tönjes, 1997], which connect concepts representing the same object on different levels. 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 in contrast to the *image* level it describes objects independently from sensor characteristics and viewpoint. Road segments are linked to the "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 based on topological criteria, while details, like exact width and position, or markings, from fine scale are integrated. In this way the extraction benefits from both scales.

3.2 Context Model

The road model presented above comprises knowledge about radiometric, geometric, and topological characteristics of roads. This model is extended by knowledge about context: Background objects, like buildings, trees, or vehicles, can support road extraction (e.g., 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 modeled *locally* and *globally*.

With 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 an open rural area, for example, paths to agricultural fields or individual cars. Driveways to buildings are more likely to 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. The local context *occlusion_shadow* illustrates, e.g., a situation where a high object occludes a part of a road or casts a shadow on a road. 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, in which, for example, *occlusion_shadow* and *building_driveway_road segment* interact.

The above mentioned contexts are more specific for rural areas. In addition to Fig. 2, Fig. 3 shows some relations between markings, junctions, lane segments, and vehicles. These are the relations the road extraction in urban area is based on in Sect. 5. Vehicles are aligned with lanes. There are two types of markings: Aligned markings define the borders of the lanes, orthogonal markings define start or end of a lane.

Relations to background objects and their relevance for road extraction depend also on the region where they occur. As mentioned above, roads in urban or suburban areas have a quite different appearance from roads in forest areas or in open rural areas. The differences in appearance are partly consequences of different relations between roads and buildings. In downtown areas, buildings

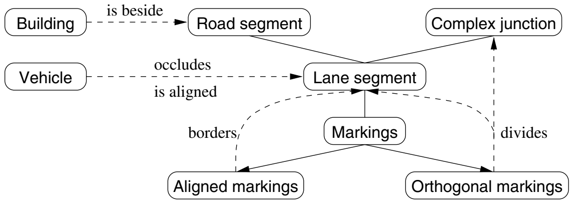


Figure 3: Local context for urban areas

typically are closer and more parallel to roads. 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 Fig. 4, some frequently occurring local contexts are assigned to the global contexts.

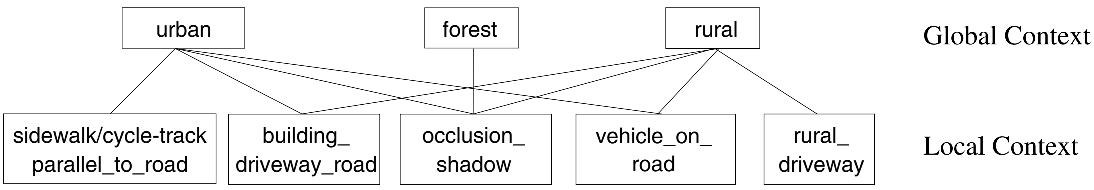


Figure 4: Global and local context

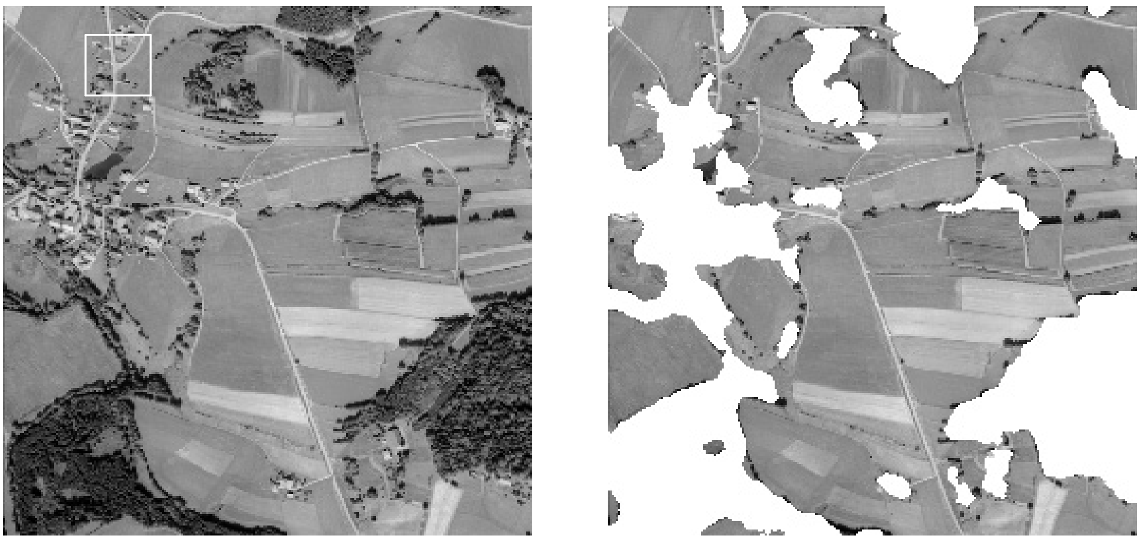
Note, however, that the use of knowledge about local context and the verification of specific relations between local objects will in most cases be possible in high resolution imagery only, because the image features which contribute to the local context are usually not very prominent. Therefore, the local context is more tightly connected with the high resolution, whereas information about global context usually can be derived from images with a resolution $> 2\,\mathrm{m}$ and is useful to guide the road extraction in both scales.

4 Road Extraction in Rural Areas

4.1 Strategy

The knowledge about how and when individual parts of the model can be exploited optimally is condensed into the extraction strategy. The proposed scheme for roads in rural areas consists of two levels: On the first level (Sect. 4.2) knowledge about radiometry, local geometry and local context is used, and road hypotheses are derived by fusion of line and edge extraction, i.e., the scale-space behavior of roads is employed. On the second level (Sect. 4.3) the connectivity of roads, i.e., the use of knowledge about their topology, is enforced using global criteria to derive connection hypotheses instead of purely local criteria as it is done on the first level.

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. According to the "easiest first" principle on the local level salient road segments are extracted first and then connection hypotheses, i.e., the non-salient road segments, between the salient parts of the road network are verified.



a) b)

Figure 5: a) Image at low resolution b) Classification of *open_rural* area

As a consequence of these considerations, road extraction starts in rural areas. Figure 5 shows the result of a texture-based segmentation of *rural* areas in the 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 0.25 m in the high resolution image.

4.2 Local Level

4.2.1 Salient Roads

On the local level we use lines and edges as image features to construct road segments. According to the road model, apart from the original image also a version of the image with a reduced resolution is used. The lines are extracted in the reduced-resolution image (about 2 m) using the differential gemoetric approach of [Steger, 1998a]. They are then used to select edges extracted from the original resolution that are candidates for road sides (cf. Fig. 6 a). In order to reduce the amount of data lines and edges are approximated by polygons. Here, the term "edge" is used for an individual segment of an edge polygon. Edges that are candidates for road sides must fulfill the following criteria:

a) b)

Figure 6: a) Input to the fusion process: hypotheses for road axes (dashed, black), hypotheses for road sides (solid, white) and other edges (dotted, white). b) Final road side hypotheses.

- The distance between pairs of edges must be within a certain range. Minimum and maximum distance depend on the classes of roads to be extracted.
- The edges have to be almost parallel, i.e., there is an overlap and the differences in the direction of the edges is low. For long edges a smaller direction difference is tolerated because the direction is the better defined the longer an edge is.
- The area enclosed by a pair of parallel edges should be quite homogeneous in the direction of the road. Markings perpendicular to the road axis or tire tracks can cause inhomogeneities.
- In addition, for each pair of candidates for road sides, a corresponding line has to exist in the reduced resolution.

The selection of edges as road side candidates and the fusion of line and edge extraction is described in detail in [Steger et al., 1995]. The fusion of lines from low resolution and edges from high resolution has proven to be very useful in order to get more reliable results. This advantage is also confirmed by the results of [Trinder and Wang, 1998] who use a quite similar approach to fuse low and high resolution imagery for road extraction.

Figure 7: Road segments

From these road sides, road segments are constructed (Fig. 7). The road segments consist of quadrilaterals which are generated from parallel road side candidates. Quadrilaterals sharing points with neighboring quadrilaterals are connected. The geometry of the road segments is represented by the points of their medial axes, attributed by the road width. These road segments are the semantic objects which are used as input for the extraction of the non-salient parts of the road network.

The computational effort for the construction of road segments depends on the number of involved lines and edges. In order to reduce this effort, hypotheses for road sides and road segments are generated locally, i.e., by working on small, overlapping image patches. In a second step, the hypotheses for road segments are collected from all patches, conflicting road segment hypotheses caused by overlapping patches are examined, and only the best hypotheses are kept (Fig. 8).

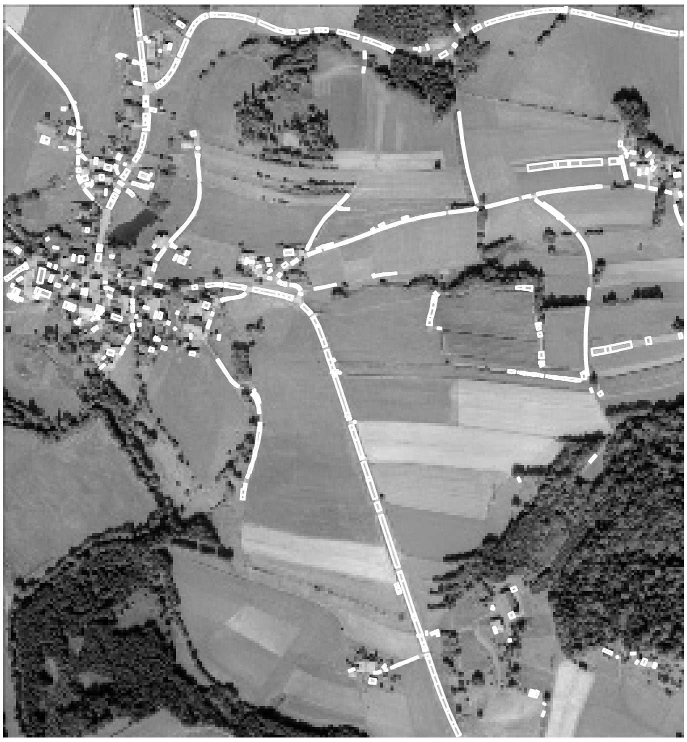


Figure 8: Road segments

4.2.2 Non-salient Roads

In the previous step only a small part of the knowledge about roads was explored. The extracted road segments are clearly visible in the image, and — based on local criteria — they have a high probability for being roads. Non-salient roads are those parts of the road network that could not be extracted due to a lack of suitable image features. For the extraction of non-salient road segments, additional knowledge about roads must be applied or the parameters of the algorithms used before have to be adapted. We assume that non-salient roads correspond to gaps between salient road segments. Therefore, the extraction of non-salient roads is equivalent to the problem of linking salient roads extracted in Sect. 4.2.1. In addition to the extraction of non-salient road segments, we want to get rid of incorrect hypotheses for salient road segments.

Most of the road segments derived from the fusion of line and edge extraction are not directly connected and they are quite short. The linking of correct and the elimination of false hypotheses is achieved by grouping the salient road segments into longer segments. The grouping is done according to the "hypothesize and test" paradigm. Hypotheses about which gaps should be bridged are generated starting with geometric criteria (absolute and relative distance, collinearity, width ratio) and radiometric criteria (mean gray value, standard deviation). Because information about only two road segments is involved, we call this local grouping, in contrast to the grouping in Sect. 4.3 which additionally uses global criteria. 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. If this verification fails too, 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. This mainly collinearity-based strategy sometimes fails, especially for curved segments.

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 increas-

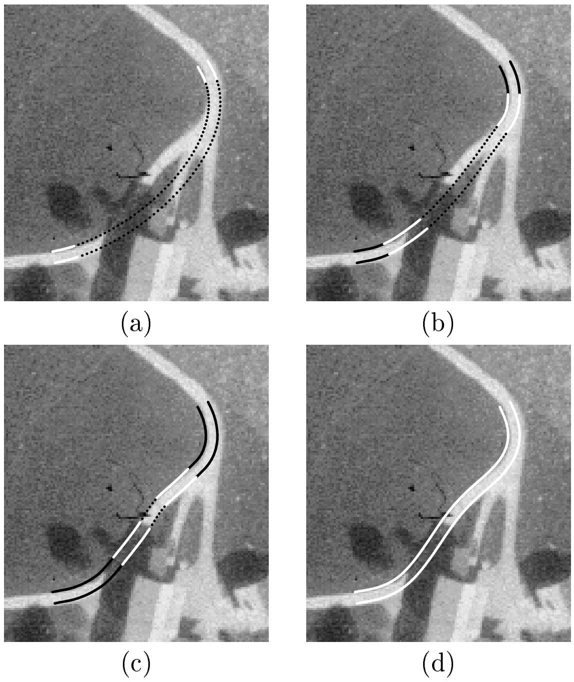


Figure 9: Optimization steps of a "zip-lock" ribbon. (a)-(c) Dotted lines indicate the passive part of the ribbon. White parts are currently optimized. Black ends indicate the result of the optimization so far. (d) Final result

ingly 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. In contrast 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. As an extension to the conventional snake, the ribbon snake has an additional parameter for the width at each line point. The image features the ribbon snake is attracted to 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 road side 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-lock, but starting from both ends, c.f., "zip-lock" snake in [Neuenschwander et al., 1995]. The zip-lock behavior of the ribbon is achieved by splitting the ribbon in active and passive parts, and only the active parts are optimized. Figure 9 illustrates how this zip-lock ribbon is applied.

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

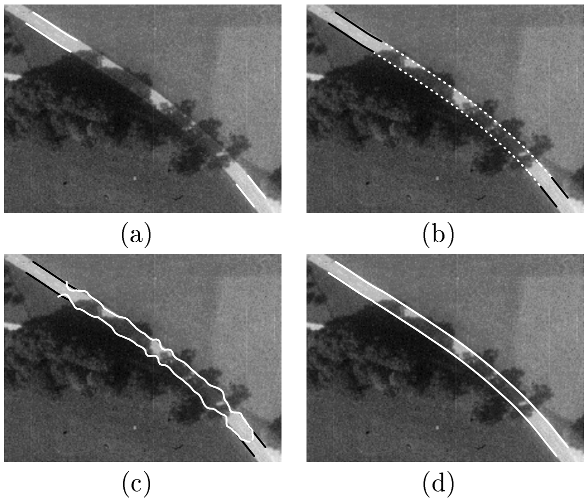


Figure 10: Extraction of *non-salient* roads. (a) Selection of initial hypothesis (b) Optimal path; (c) Verification by optimization of width (d) Selection of hypothesis with constant width

this second step. Figure 10 shows that this is possible even if the road sides do not correspond to strong edges in the image. A more detailed description of this technique is given in [Mayer et al., 1998], where ribbon snakes are initialized by lines extracted from low resolution. With this, the automatic extraction of salient and non-salient roads can be completely based on the snake technique. The results of the completely snake based approach are quite similar to the grouping based approach described in this paper, see [Heipke et al., 1998]. However, for the proposed way to bridge gaps between salient road segments it makes no difference, whether the road segments were constructed by a grouping process or by any other method.

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 is allowed to 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 DEM and information about when and where the image was taken [Eckstein and Steger, 1996]. 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.

Figure 11 displays the road hypotheses derived from the road segments shown in Fig. 8.

Figure 11: Road hypotheses

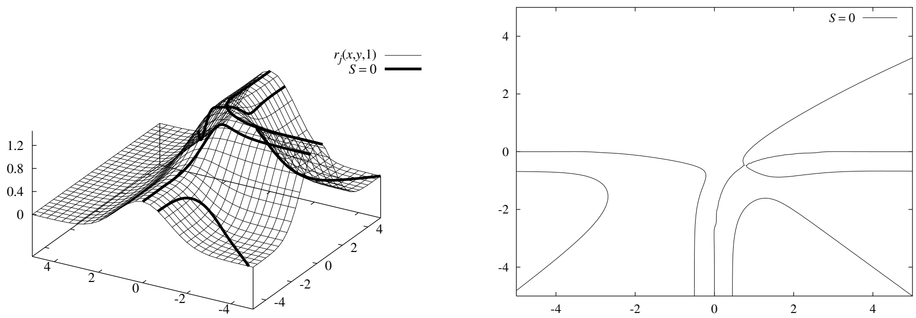
4.2.3 Road Junctions

After the generation of hypotheses for connections and their verification, the road network must be constructed, i.e., the junctions that link the roads must be extracted.

For lines in the reduced resolution image, the construction of missing junctions is based on an analytical scale spaces analysis. In order ot keep consistency, this is carried out in the same way as for the line detector mentioned in Sect. 4.2.1 (details can be found in [Steger, 1998a]).

A suitable functional model for a T-junction — a junction where three lines of different width and different contrast meet — is given by:

$$f_j(x,y) = \begin{cases} h_1, & x \geq 0 \wedge |y| \leq w_1 \\ h_2, & x < 0 \wedge |y| \leq w_2 \\ h_3, & |x| \leq w_3 \wedge y < -w_1 \\ 0, & \text{otherwise} \ . \end{cases} \tag{1}$$



(a) Extracted lines

(b) Lines in the (x, y)-plane

Figure 12: Extracted lines by the De Saint-Venant condition for a T-junction with lines of equal width and different contrasts of $h_1 = h_3 = 2$ and $h_2 = 1$: (a) mapped onto the response $r_j(x, y, 1)$, and (b) in the (x, y)-plane.

The scale-space description of this junction is given by r_j and its partial derivatives of first and second order where

$$\begin{aligned} r_j(x,y,\sigma) \quad = \quad & h_1(\phi_\sigma(y+w_1) - \phi_\sigma(y-w_1))\phi_\sigma(x) + \\ & h_2(\phi_\sigma(y+w_2) - \phi_\sigma(y-w_2))(1-\phi_\sigma(x)) + \\ & h_3(\phi_\sigma(x+w_3) - \phi_\sigma(x-w_3))(1-\phi_\sigma(y+w_1)) \end{aligned} \tag{2}$$

and $\phi_\sigma(x)$ is the integral of the Gaussian kernel with standard deviation σ.

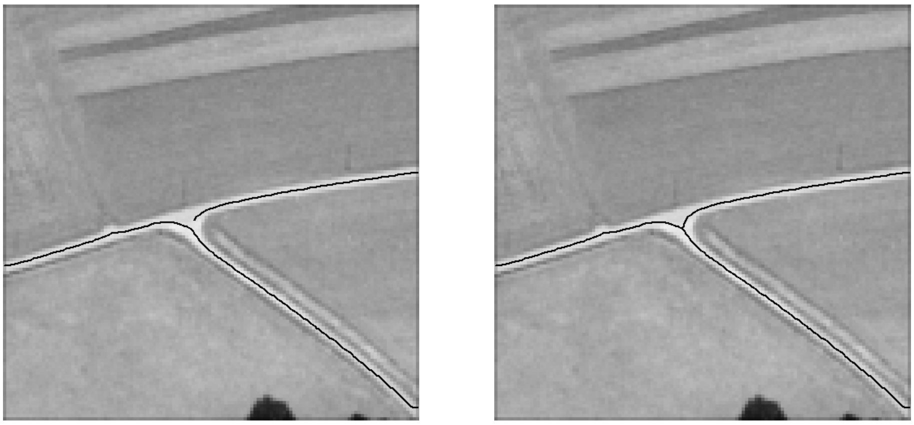
In order to analyze the scale-space behavior of the line positions leading to the junction, the zero crossings of the first directional derivative in the direction of the maximum second directional derivative must be obtained. To analyze the behavior of lines in the 2D case, we can use De Saint-Venant's condition $S = 0$ [Koenderink and van Doorn, 1994]. This condition is given by

$$S = \frac{\nabla r_j^T H \nabla r_j^\perp}{\nabla r_j^T \nabla r_j} = \frac{r_x r_y (r_{xx} - r_{yy}) - (r_x^2 - r_y^2) r_{xy}}{r_x^2 + r_y^2} = 0 \ , \tag{3}$$

with r_x,r_y,r_{xx},r_{xy}, and r_{yy} beeing the partional derivatives of r_j.

This condition will extract a superset of the lines the algorithm of [Steger, 1998a] extracts because — besides extracting the maxima of the second derivative perpendicular to the line direction — it will also extract the points where there is a restricted extremum in the direction of minimum second directional derivative, and because it does not discern between restricted maxima and minima. Nevertheless, because the extracted lines form a superset of the lines returned by the proposed algorithm, the reasons for missed junctions can be studied.

Figure 12 shows a junction with $h_1 = h_3 = 2$ and $h_2 = 1$, i.e., the left line on the y-axis is darker than the other two lines. As can be seen, the line on the left part of the y-axis again changes to a restricted maximum in

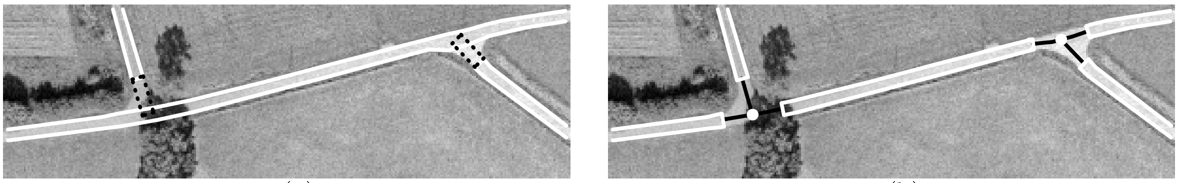


(a) Extracted lines (b) Junction completed

Figure 13: Example of junction reconstruction

the direction of the minimum second directional derivative. However, as this line approaches the other two lines it rapidly turns away from the junction, and ends up running parallel to the line on the y-axis. Therefore, allowing lines to be restricted extrema in the direction of the minimum second directional derivative is definitely a wrong approach. What is striking, though, is that the direction of the line on the y-axis at the point where it changes its type points exactly at the junction. More examples with different configurations seem to indicate that this is true in general [Steger, 1998b]. Therefore, a good algorithm to extract the missed junctions is to search for other lines along the last extracted line direction. As can be seen from the examples, the lines not continuing to the junction will in general be the lines with the smaller contrast. Hence, while following the line, the gray value in the image should increase monotonically for bright lines and decrease monotonically for dark lines, respectively. Since it is desirable to avoid using the image intensity directly, this condition can be reformulated to the condition that the dot product of the gradient of the image and the line direction must always be positive for bright lines and always negative for dark lines. Of course, the length of the line along which the search is performed should not be arbitrarily long. A good restriction is to use the a search length of 2.5σ . Figure 13 a) shows a junction of a rural road and the extracted lines without junction completion. In Fig. 13 b) the junction is reconstructed with the proposed method.

The generation of hypotheses for junctions in the high resolution image 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 (Fig. 14). The verification of these new road segments is done in the same manner as for the gaps.



(a) (b)

Figure 14: a) Road segments (white) and extensions (black, dotted) b) Road segments and final junctions

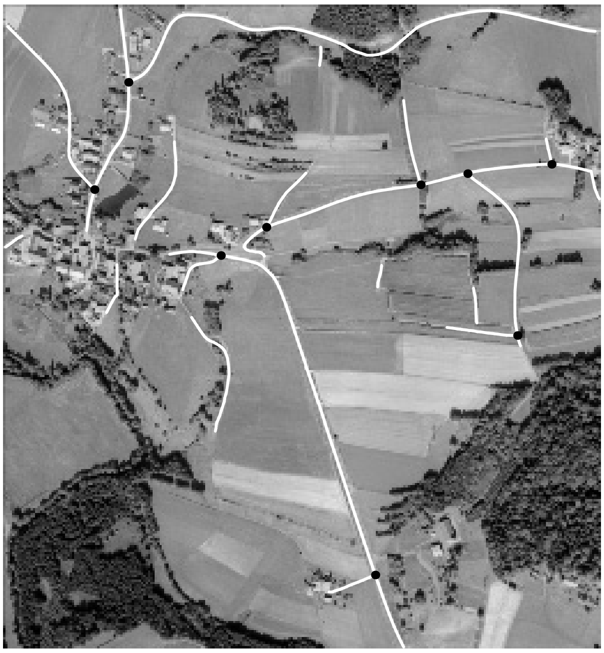


Figure 15: Road axes and junctions

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 (Fig. 15). 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 primarily based on local information and is reliable only in rural areas, the network characteristics of roads are not optimally exploited. 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.

Up to now primarily local criteria were applied to construct a road network. In Sect. 4.3 we make use of global topological properties of roads to further improve the results with respect to completeness and connectivity.

4.3 Global Level

The intrinsic function of roads is to connect different "important places," even if they are far away from each other. Hence, roads form a (hierarchical) network that is mostly optimized to provide an economic and convenient way for reaching different places. Because of this property, searching for the globally best connection between such places is an essential step for road extraction. Moreover, since there usually exists only one good connection between two "important places" [Ruskoné et al., 1994] (at least in open and rural terrain) we restrict the search to the best connection between two places.

In addition to exploring the global road properties, the scheme should provide reliable road extraction results by itself. In some cases, images with a resolution of less than 0.5 m needed for an extraction strategy like the one described in Sect. 4.2 might not be available. Thus, the extraction should not exclusively rely on the results of the local level. Instead, a flexible strategy is desirable which is able to integrate and process different sources. Hence, the scheme outlined in this paragraph is basically designed for the extraction of roads from low resolution imagery that, for instance, comprises aerial images which are down sampled to a pixel size of 1–4 m or satellite imagery with a ground resolution of 6–18 m, e.g., as in [Wiedemann and Hinz, 1999], MOMS-2P imagery. Additionally, other information sources can be integrated in the early stage of processing. In the remainder of this section, such an integration of another source is illustrated by combining the results of the previous section, i.e., the road segments after local grouping, with the proposed algorithm for global grouping. By doing so, we are able to exploit both the local shape and reflectance features and the global network characteristics of roads.

With the above mentioned range of resolutions, it is obvious that, when using low resolution imagery only, both road model and extraction strategy are only able to capture the basic shape and reflectance features that roads exhibit in the real world. In contrast to these local properties, the global topology of roads, i.e., the network characteristics, play the major role. However, since the image domain only covers a part of the whole road network, different sub-networks, which are not necessarily connected, may occur in the image.

Having these requirements in mind, the following extraction strategy has been developed: First, bright lines are extracted and processed in order to build up road segments. Optionally, the road segments can be fused with the results of the local level. Then, a weighted graph is constructed from the road segments and from possible connection hypotheses between them. Thereafter, pairs of "important places" are selected and the optimal path between each pair is calculated. The extracted road network results from the combination of all optimal paths computed through the graph. In the following, a detailed description of each step is given.

4.3.1 Low Level Processing

Line extraction is performed using the approach described in [Steger, 1998a]. It is based on differential geometry and captures the local radiometric road characteristics. To initialize the procedure, a few, semantically meaningful parameters have to be determined: The maximum width of the lines to be extracted can be chosen according to the road width scaled to the image. The two hystereses values that control the process of linking individual line pixels into pixel chains can be derived from the expected gray value contrast between roads and their surroundings. It can be decided whether bright or dark lines should be extracted. The result of the line extraction is a set of pixel chains and junction points with sub-pixel precision. Additionally, local line attributes like width, direction, and contrast are obtained at each line pixel. Due to the exclusive use of local road characteristics (for low resolution images) the result is not complete and contains false alarms, i.e., some roads are not extracted and some extracted lines are not roads.

During the next step, road segments are constructed from the extracted lines. This has two different goals:

- Reduce the probability that a road segment represents partly a road and partly another linear structure, i.e., a feature should either completely correspond to a road or to none at all.
- Obtain the attribute values describing the quality of a road segment with respect to the model.

The attribute values of road segments are used to include additional evidence about the presence of roads into the grouping process at a later stage. Therefore, it is necessary to ensure that the lines to which the attributes belong either completely correspond to roads or to linear structures not being roads, i.e., lines have to be split at the point where they might cross the road side. A careful analysis of the behavior of several line attributes has shown that the most significant feature for a change in the line semantics ("road"/"not road") is high curvature. Hence, lines are split at points where the direction difference between two consecutive polygon points on the line exceeds a given threshold. Please note that this procedure does not exclude any part of a line. A line might be erroneously split, i.e., the line is split although it completely belongs to a road (for instance in mountainous regions where roads often are highly curved). However, there is still a high probability that the split parts are joined again if they are found to be part of the paths computed during road network generation.

Each resulting line defines a road segment. Assuming the image resolution is known, the following attribute values are calculated to obtain an extended description of each road segment:

- length of a road segment,
- straightness of a road segment, i.e., the standard deviation of its direction,

- width of a road segment, i.e., the mean width of the extracted line,
- width constancy of a road segment, i.e., standard deviation,
- reflectance constancy of a road segment, i.e., standard deviation of the intensity values along the segment,
- flatness of a road segment, i.e., the mean of the absolute gradients along the segment.

4.3.2 Fusion

In order to make use of road data originating from different sources, we developed a method for fusing the road segments with every kind of linear data. Here, we combine the center axes of the road segments constructed and verified in the high resolution image with the road segments obtained from the low resolution image, i.e., after the fusion both types of segments are contained in one set of linear road data. Since the road segments achieved from the low resolution image are less constrained, a more complete network might be extracted. Segments, or parts of segments, which lie within a buffer with a suitable chosen width (e.g., 3 m), are candidates for unification. In addition, if two candidates have a direction difference less than, e.g., 15°, they are unified. Otherwise, they are checked for an intersection.

4.3.3 Graph Representation

From the resulting set of road segments, an initial unweighted graph is constructed. The road segments define the edges of the graph, and their end points represent the set of vertices. In case of junctions, i.e., if two or more road segments end in the same point, only one vertex of the appropriate degree is inserted in order to preserve topology.

The attribute values of the road segments are used to weight the graph by associating every edge with a single weight. This is done by defining linear fuzzy functions ranging from 0 to 1 [Mendel, 1995] for transforming the attribute values into fuzzy values. These fuzzy values are aggregated by the fuzzy *and*

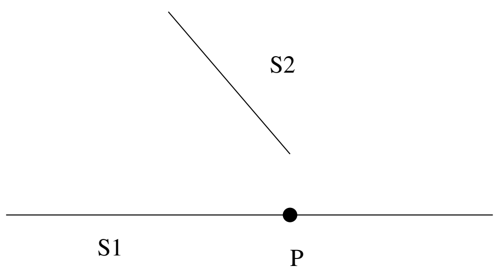
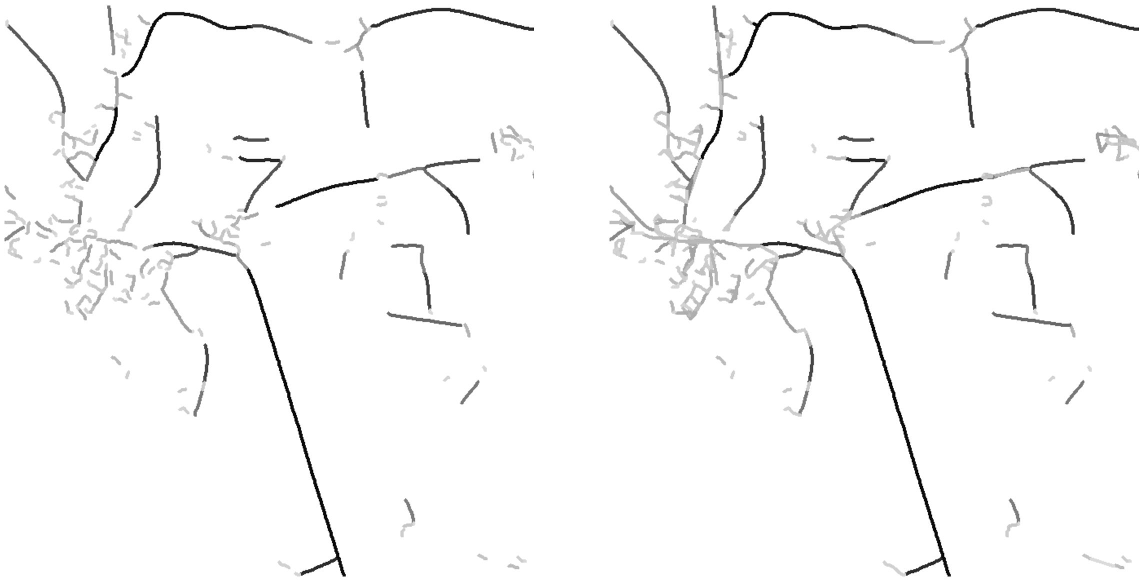


Figure 16: Candidate for a junction

operation into one overall fuzzy value for each edge, i.e., an overall fuzzy value 1 stands for a road segment that matches the road model perfectly and 0 means that the road segment should not be considered anymore.

The following phase of processing is addressed to prepare the graph for detection of possibly missing road junctions at a later stage. Due to deviations from the road model, it might happen that some of the road junctions, especially the larger ones, are not detected. Junctions are, however, an essential topological part of the road network. Hence, it should (at least) be possible to form connection hypotheses in situations where a junction might be present. For this reason, the edges of the graph are split at points which can be regarded as a priori candidates for junctions, and a new vertex of degree 2 is inserted. See, e.g., point P in Fig. 16 that lies on segment S1 closest to the end of segment S2.

A reliable decision whether a junction candidate truly represents a road junction is not possible at this stage of processing. It could be caused by blunder, e.g., by other linear structures close to a road like certain kinds of vegetation. However, the splitting of a road segment affects its properties (e.g., its length), which leads to an incorrect evaluation in such cases. Therefore, the fuzzy value of a split edge is inherited, i.e., the weights in the graph are not changed by inserting a new vertex.



(a) (b)

Figure 17: Weighted graph of road segments based on lines (a), with inserted connection hypotheses (b)

The resulting weighted graph is used to generate and evaluate connection hypotheses (Fig. 17). The following criteria are introduced to measure the quality of a hypothesis:

- the direction difference between adjacent road segments, where either collinearity (within a road) or orthogonality (at a T-junction) are preferred,
- the absolute length of a connection,
- the relative length of a connection (compared to the length of the lower rated of the adjacent road segments),
- the mean gradient along the connection,
- an additional constraint which avoids that a connection hypothesis is evaluated higher than its adjacent road segments.

As above for road segments, linear fuzzy functions are defined to obtain individual fuzzy values for each criterion, which are then aggregated into an overall fuzzy value by the fuzzy *and* operation. A special case is the evaluation of the direction difference between two road segments. In order to either prefer the continuation of a road or to support a possible road junction, a fuzzy function with two peaks is defined (e.g., at 10° and 85°), one supporting the collinearity of two segments and one supporting their connectivity with respect to a T-junction. See Fig. 18 for an illustration. Since a connection hypothesis can represent only one of these grouping principles, but not both at the same time, the proposed examination of the direction difference can be understood as classification of the connection hypotheses in road connections and junction connections, whereby each connection is associated with a fuzzy value. Depending on this classification one may chose different parameter settings of some of the other fuzzy functions, e.g., for evaluating the absolute distance.

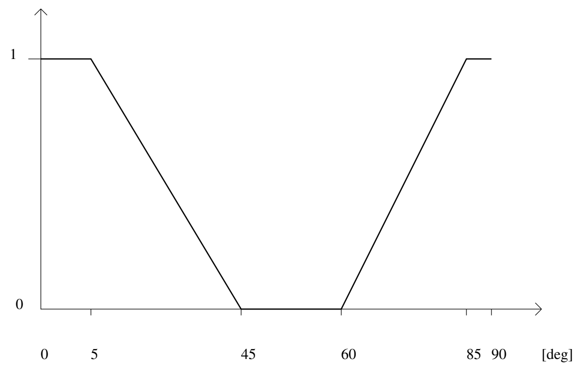


Figure 18: Fuzzy function for evaluating direction difference

4.3.4 Road Network Generation

The extraction of the road network relies on the selection of “important places” (seeds), which are then connected by the optimal path through the network. Such places are usually buildings, industrial areas and other sites of interest. Since this approach exclusively deals with roads without considering additional objects, we define “important places” as road segments that represent portions of the road network with high probability. An indication for the probability of a road segment being truly a road is its fuzzy value. Hence, all road segments yielding a high evaluation are chosen as seeds for road network generation. Please note that this threshold can be derived from semantically meaningful and reasonable parameters using the same fuzzy functions as before, e.g., by demanding that a seed must have a given minimum length.

It might happen that several disconnected sub-networks instead of one complete road network are visible in the image. Thus, the strategy for road network extraction should be able to extract disconnected road networks but still should incorporate the function of roads connecting places far away from each other. To this end, only those pairs of seeds are considered for path calculation which guarantee that the length of the resulting path exceeds some suitably chosen threshold, e.g., 1 km. A lower bound of the respective path length can be achieved without executing path calculation by summing the length of both road segments and the minimal distance between their endpoints (see Fig. 19). With this strategy, larger isolated parts of the road network can be detected without losing the postulated globality of the proposed grouping algorithm.

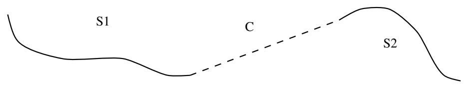


Figure 19: Minimum path length: $l_{min} = l_{S1} + l_{S2} + l_C$

Now, the basic idea is to find the shortest paths in the weighted graph with suitably chosen distances. Therefore, the weights in the graph should reflect the true distances in the object, but depending on how a road segment or a connection hypothesis is evaluated, the distance between two vertices should be increased to make it harder to bridge obviously bad links. If a link is considered perfect, i.e., has the fuzzy value 1, the true distance between the vertices is used. If the link has the fuzzy value 0, its weight must be ∞. Therefore, the following formula is used to construct the weighted graph:

$$w_{i,j} = \begin{cases} l_{i,j}/r_{i,j} & \text{if vertices } i \text{ and } j \text{ are connected} \\ & \text{by a road segment of length } l_{i,j} \\ & \text{and } r_{i,j} > 0 \\ d_{i,j}/r_{i,j} & \text{if } i \text{ and } j \text{ are not connected in} \\ & \text{the original graph, but } r_{i,j} > 0 \\ & (d_{i,j} \text{ is the Euclidean distance}) \\ \infty & \text{otherwise (no edge in the graph)} \end{cases}$$

where $w_{i,j}$ is the weight of the edge between the vertices i and j.

The final step is the computation of the respective optimal path between each seed pair. This is carried out by using the Dijkstra-Algorithm [Knuth, 1994]. The combination of all detected paths defines the extracted road network.

It should be noted that the resulting network is inhomogeneous with respect to the geometric accuracy since parts of the network originate from purely geometry-based gap bridging without considering the radiometric content in between. This implies a final verification of the bridged gaps using similar criteria as in Sect. 4.2.

4.3.5 Algorithmic Aspects

Grouping in general is exponential in the number of features to be grouped. In our case, only pairs of vertices are evaluated ($O(n^2)$). The Dijkstra-algorithm has a complexity $O(n \log n)$. Therefore, in the worst case, the presented approach has an overall complexity $O(n^3 \log n)$. Nevertheless, restrictions on the search space (without affecting the final result) should be included whenever possible. In our implementation the following points help to minimize the runtime:

- Use of range search if only local relations between features are relevant (as in the early stages of processing). Line features are described by their bounding box plus a buffer covering the area where a possible grouping candidate has to be searched for.
- Ordering the criteria for local evaluation of the connection hypotheses such that the most significant criteria are calculated first (e.g., the distances). The calculation can be aborted if the overall score becomes too low for further consideration of a particular hypothesis.
- Partitioning the graph in its connected components before selecting seeds and computing best paths. Only those seed pairs are selected which are contained in the same connected component. By doing so, it is guaranteed that only successful path calculations are carried out. Of course, the efficiency of this strategy depends on the number of connected components but the computational effort of calculating the connected components is negligible compared to the possibly drastical reduction of seed pairs, especially if the algorithm has to cope with a highly fragmented segmentation result that cannot be linked completely.
- Ordering the list of seed pairs such that the presumably longest paths are computed first. Every seed pair that lies on one single path already found can be removed before calculating the next path.

4.4 Evaluation

Internal self-diagnosis and external evaluation of the obtained results are essential for any automatic system. In the long run these factors are of major importance for the introduction of the system into practical applications. Both, internal self-diagnosis and external evaluation should yield quantitative measures which make different results commensurable. Here, we deal with the external evaluation of the automatically extracted road data by means of comparing them to manually plotted linear road axes used as reference data. In the following the evaluation procedure is briefly described. More details can be found in [Heipke et al., 1998].

The comparison is carried out by matching the extracted data to the reference data using the so-called "buffer method," in which every portion of one network within a given distance (buffer width) from the other is considered as matched. For the evaluation of the road extraction results a number of quality measures is defined based on the matching results. Two questions can be answered by the quality measures: (1) how complete is the extracted road network, and (2) how correct is the extracted network? The completeness indicates how much is missing in the network, whereas the correctness is related to the probability of an extracted linear piece to be indeed a road.

Completeness is defined as the percentage of the reference data that is explained by the extracted data, i.e., the percentage of the reference data which lies within the buffer around the extracted data:

$$\mathit{completeness} = \frac{\mathit{length\ of\ matched\ reference}}{\mathit{length\ of\ reference}}$$

The correctness represents the percentage of correctly extracted road data, i.e., the percentage of the extracted data that lies within the buffer around the reference network:

$$\mathit{correctness} = \frac{\mathit{length\ of\ matched\ extraction}}{\mathit{length\ of\ extraction}}$$

In addition, the geometric accuracy of the extraction is assessed. It is expressed as the RMS difference between the matched extracted and the matched reference data.

The geometric accuracy of the final result in Fig. 20 is about 0.54 m, whereas the accuracy of the results of the previous steps is about 0.37 m. The accuracy decreases since all during the global grouping additionally detected roads result either from line extraction in low resolution (here 2 m) or from connection hypotheses between the lines which are inserted as straight segments without considering the radiometry in between.

In contrast to the RMS value, both the completeness and the correctness have increased from 83% to 89% and from 90% to 95%, respectively, because of the global grouping step. However, the main improvement caused by the global grouping, which is captured by none of the used quality measures, is the topological connectivity of the extracted network. As illustrated in Fig. 20, now all road sections are topologically linked to each other.

The high quality of this result proves that the approach extracts most of the roads in rural areas, but it fails in some parts of the residential areas. Although,

(a) Final result after global grouping (b) Evaluated result

Figure 20: Result and evaluation. Bold: Correct extraction. Thin: False extraction. Dashed: Missing extraction.

compared to the local level, the global grouping step has improved the general network connectivity, some of the streets in the village are not detected at all. One obvious reason for this are fragmented or missing parallel structures defining the road sides. Such structures, however, are the basic features for constructing road segments, and by this way they are the most essential part of the approach. In other words, for a successful and reliable road extraction in suburban and, especially, urban areas, the focus must be turned to another part of the model and the use of different extraction strategies is required. For example, it can be expected that a thorough integration of DEM information, which is not implemented by now, might improve the results, i.e., wrong hypotheses for roads that lie on the roofs of buildings can be detected and eliminated more easily.

A quantitative evaluation of the results for rural areas according to the above described evaluation scheme has been carried out on a set of test images. This evaluation has shown 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 Road Extraction in Urban Areas

In this section, the most obvious differences that roads exhibit in rural and urban areas are exemplified. Based on this, an extraction strategy is presented that exploits parts of the road model not yet used above (see Sect. 3). Finally, preliminary results for the extraction of separate lanes are shown.

5.1 Appearance of Roads

The discussion of the results in Sect. 4.4 indicates that the appearance of roads in the context of urban areas is generally different from their appearance in rural terrain.

Figure 21 visualizes two examples of urban roads. It is obvious that these images exhibit a more complex content than scenes showing rural areas since the number of different objects and their heterogeneity is much bigger. Generally, this implies that more details of the road and context model must be exploited for road extraction. In dense urban areas, for instance, some of the roads comprise several lanes that are linked by complex road crossings. What is more, by the increase of the number of objects the complexity of their relations grows, too. In the left image, for example, some parts of the streets are occluded by vehicles, especially at the road sides. Hence in this particular case, a road is mainly defined by groups of (parking) cars but not by parallel road sides or homogeneous surface. A similar relation is the occurrence of shadows cast by high buildings. A road generally appears bright in open areas, but in the case of shadows two problems for the extraction arise: (1) the surface is darkened

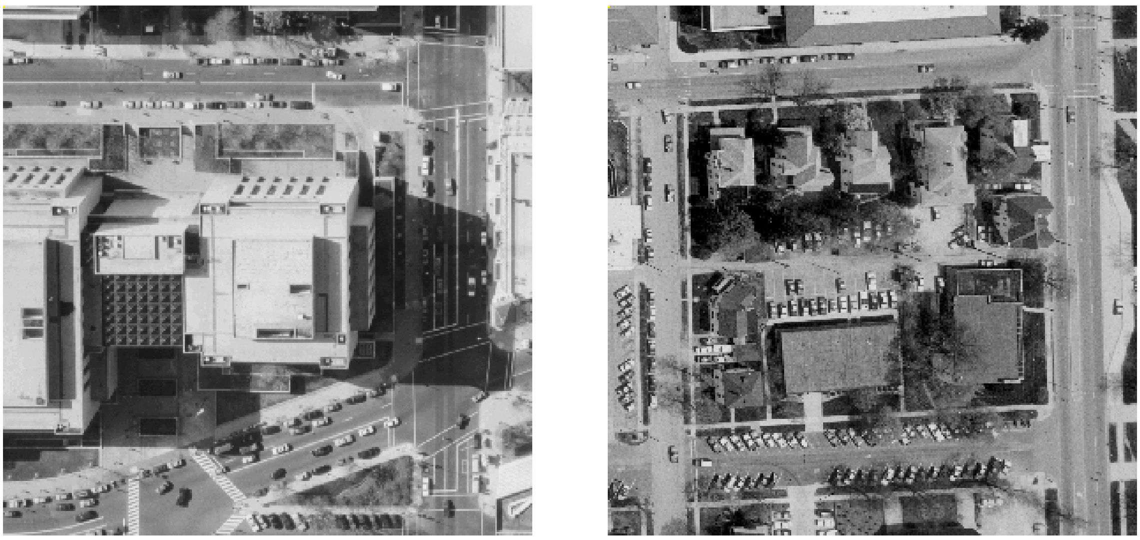


Figure 21: Examples of roads in urban areas

significantly, and (2) at the margin of the shadow regions, strong gray value edges in almost any direction may occur on the road disturbing the usually homogeneous reflectance.

The right image shows a different kind of problem: The roof of the rectangular building in the center of the image could be wrongly identified as a parking lot because its shape and reflectance properties match the ones of a road-like object almost perfectly. Only the combination with height data as given by a DEM or, as in this case, implicitly given by a corresponding shadow region provides enough information for avoiding this misdetection.

What follows is, that on one hand those features of a road ought to be selected on which the influence of the above mentioned phenomena is minimal. On the other hand, it is very important to consider context objects, in particular different kinds of vehicles, in order to explain abnormal changes in the appearance of a road.

Besides an (at least partially) homogeneous surface and more or less densely arranged vehicles, one obvious feature of roads in urban areas are road markings. To make use of them, we model roads and complex junctions as a combination of several lanes consisting of one or more lane segments. Dashed or solid linear markings define the border of a lane segment. The interior of a lane segment should either exhibit the typical homogeneous reflectance of the pavement, or a vehicle that occludes the pavement has to be detected. The influence of high objects is considered twice: first, roads are allowed to be dark (shadow), and second, they cannot lie on locally high regions (buildings/dense vegetation).

5.2 Road Extraction Based on Markings

From these components of the model, the strategy for road extraction based on markings is derived. Preconditions for a successful extraction are, of course, (1) markings must be painted on the roads, and (2) they must be detectable in the image. Condition (1) is in fact fulfilled for many roads, especially for the larger ones in built-up areas. Condition (2) depends on the image recording circumstances, i.e., especially the resolution. Furthermore, it also depends on objects that might occlude road markings, e.g., large cars or trucks. Fortunately, as can be seen from Fig. 21, if the viewing angle is not too oblique, lanes are generally wide enough so that markings are visible even if cars are next to them.

The extraction starts with the segmentation of areas of interest based on height information, e.g., as given by a DEM. Then, faint bright lines are extracted and iteratively connected to marking groups that represent the lane sides. On both sides of every group of markings a lane segment is hypothesized. Lane segments are verified by different criteria using geometric, radiometric, and context knowledge.

5.2.1 Preprocessing

In the first step, areas of interest are segmented using the context of roads: most buildings are higher than the road surface. Therefore, the parts that correspond to locally high regions in a DEM are removed from the image. In this example, the imagery has been downsampled from approximately 0.25 m to 3 m. The segmentation procedure compares a smoothed version of the DEM with the original DEM and removes regions where the height difference between both data exceeds a threshold. Both parameters, the size of the smoothing mask and the minimum height difference, can be derived from the expected size and height of the buildings. Figure 22 shows the downsampled image, the DEM image, and the segmented image.

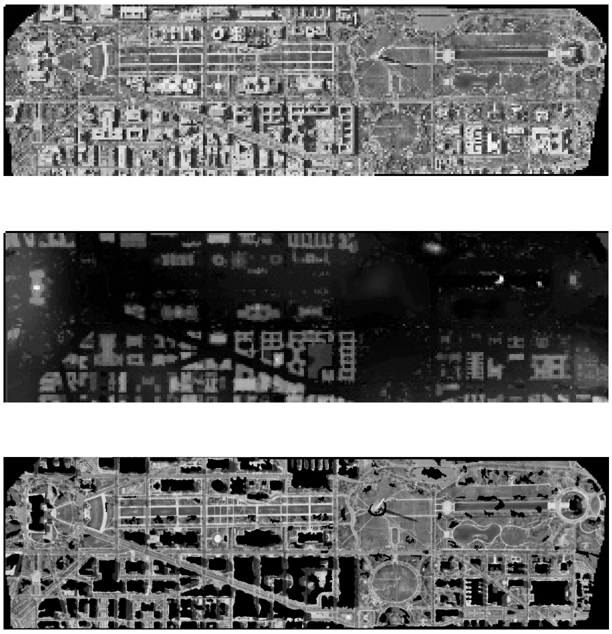
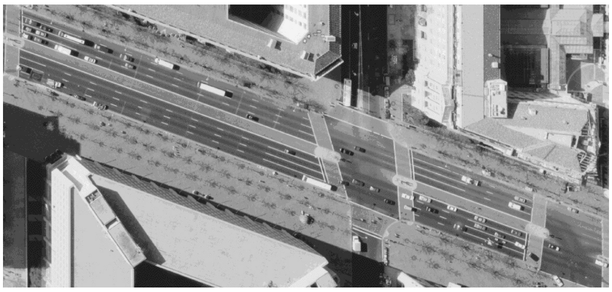


Figure 22: Segmentation of Areas of Interest

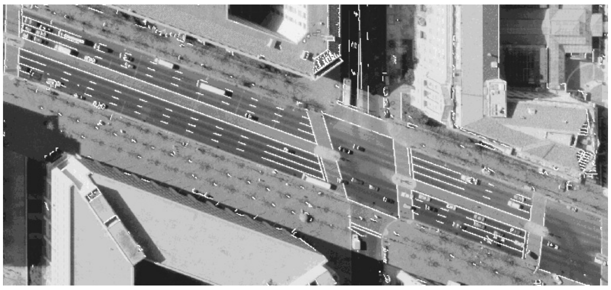
The segmentation results are then transformed to the original image resolution, after which the image is partitioned into small patches (Fig. 23a). Please note that, based on a DEM, a variety of segmentation techniques could be used in order to limit the search space, e.g., gray value morphology. Furthermore, a combination with other road extraction approaches, e.g., [Price, 1999], where approximate roads positions can be derived easily, is possible at this stage.

5.2.2 Extraction and Grouping

During the next step, lines are extracted and grouped. Fig. 23b) shows the line extraction result which is obtained, by analogy to the previous sections, using the approach of [Steger, 1998a]. Thereafter, lines are grouped according to perceptual principles: absolute and relative proximity of lines as well as their continuation. Basically, the algorithm performs in a very similar way like the one outlined in Sect. 4.3. Only the selection of seeds has been changed: from the lines and possible connection hypotheses, a weighted graph is constructed. In contrast to Sect. 4.3, the optimal path between every pair of vertices with degree 1 is calculated. By doing so, all possible groups of lines that show rather good continuation are detected. Thereafter, all paths are combined by means of deleting identical parts of different paths and splitting paths at intersections. The resulting set of unique and topologically consistent paths serves as input for the next iteration. A new graph with new connection hypotheses is constructed and the path calculation is carried out again. This procedure is repeated until no new connections are found. Figure 24a) visualizes the achieved result which represents the finally extracted groups of markings.



(a) Image patch

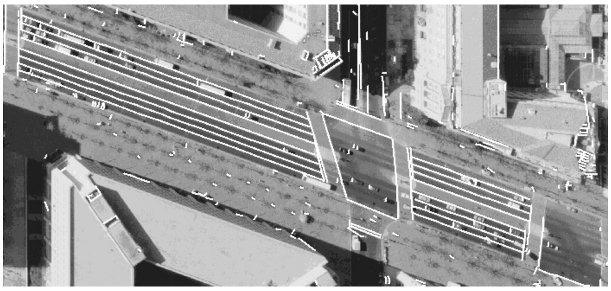


(b) Extracted lines

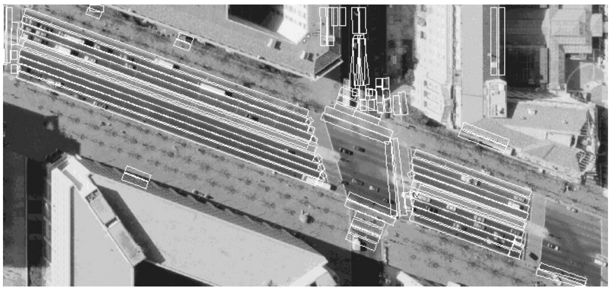
Figure 23: Original image patch and extracted lines

5.2.3 Generation of Hypotheses

The strategy for generating hypotheses is intentionally designed to be very liberal. This has the following two reasons: (1) Markings usually appear as very faint lines. Thus, the line extraction may miss some of them. Such a failure, however, can be often compensated by the iterative grouping procedure de-



(a) Grouped lines



(b) Hypothesized lanes

Figure 24: Grouped lines and hypothesized lanes

scribed above. (2) Due to occluding objects like big trucks or trees, markings are more reliable to extract in the center of a road than on its sides. Therefore, two lane segments are hypothesized, one on each side of a detected group of markings.

In order to construct lane segments from the markings, generic knowledge about the geometry of lanes is used. Lanes have some lower bounds for length and curvature. Therefore, after polygon approximation and splitting the groups of markings at sharp bends, short polygon segments are deleted (e.g. $< 5\,\mathrm{m}$). Additionally, lanes have in general a certain constant width, e.g., $3\,\mathrm{m}$. Hence, lane segments are constructed as rectangular regions on each side of a group of markings (see Fig. 24b).

5.2.4 Verification

Since the lane segments are hypothesized in a liberal manner, a sophisticated verification is needed in order to discriminate good hypotheses from bad hypotheses. To this end, not only the geometric and radiometric properties of lane segments are considered but also knowledge about their context is included. Here, the following criteria are used to collect evidence for the presence of a lane segment:

- The surface of a lane segment should be homogeneous in the direction of a lane. In regions where this criterion is not fulfilled, a car must be present. Figure 25 visualizes the extracted homogeneous regions.

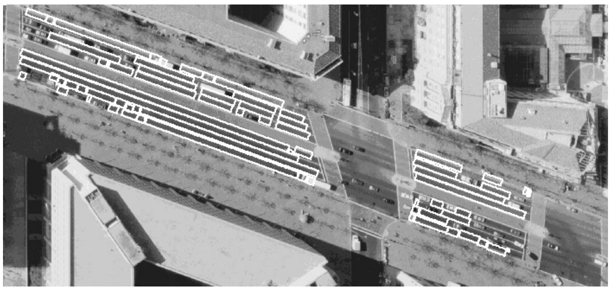


Figure 25: Homogeneous regions inside lanes

- In case of parallel marking groups or lane segments there is a high evidence for the presence of a road (see Fig. 26).

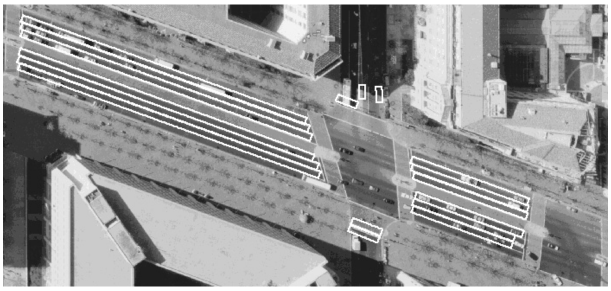


Figure 26: Parallel lanes

- Additional markings at the margin of a lane segment are searched for by using lower thresholds than in the previous steps. A lane segment is rated depending on the percentage of dashed or solid markings (see Fig. 27).

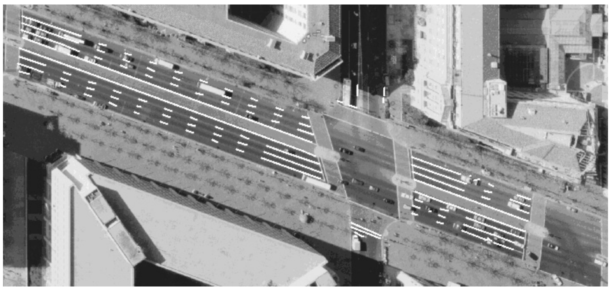


Figure 27: Additionally extracted markings

- Parallel edges or lines — possibly highly fragmented — at the margin of a hypothesized lane segment support a hypothesis, too. As mentioned above, there is a low probability that markings can be detected at the sides of urban roads. However, as can be seen from Fig. 28, in some cases small pieces of markings, curbstones, and other parallel structures that can support a hypothesis might be found.

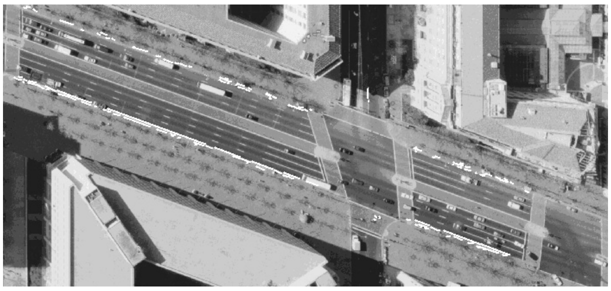


Figure 28: Extracted line and edge support

- Finally, orthogonal lines at the ends of lane segments are extracted (see Fig. 29). These lines are interpreted as cross walks or stop-lines at traffic lights. Such information is on one hand useful to get information about the end of a lane, and on the other hand, it provides a strong cue for a junction.

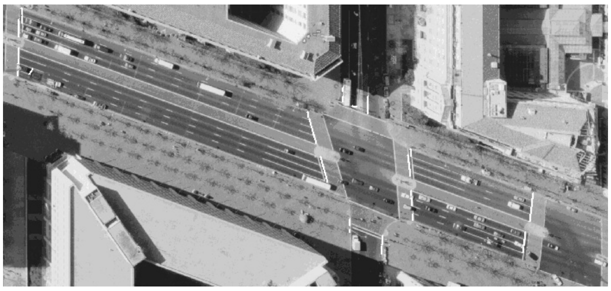


Figure 29: Extracted orthogonal lines at lane ends

Please note that a hypothesis is not ultimately rejected at this stage of processing. A reliable decision if a lane segment belongs to a road or not is only possible when considering additional features, e.g., the connectivity of different lane segments and the global network topology. As mentioned, the results presented above should be regarded as intermediate steps of a more complex strategy. Further implementations are to come.

6 Conclusions

The proposed approach for rural areas is suited for images with a resolution of 0.2 to 0.5 m. However, the results are not 100% 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. By means of global grouping criteria, the knowledge about the topological properties of roads is incorporated, and we are able to overcome some deficiencies of purely local grouping. We showed that a noticeable improvement concerning the connectivity of the resulting road network is possible with an integration of global grouping criteria in the last step. A full integration of local and global grouping should be even better, at least theoretically. A problem is that the new connections are not verified with the image data. This is the main reason for the loss of geometric accuracy after global grouping. Furthermore, it seems that the results could be improved by using a more detailed modeling of junctions, e.g., as it is proposed in [de Gunst and Vosselman, 1997, Boichis et al., 1998].

For road extraction in urban areas markings are the most important features. DEM information has proven to be very useful to restrict the search space for the extraction of markings. Compared with the approach for rural areas, the extraction uses more knowledge about substructure of roads (markings, lanes) and relations between vehicles and lanes. The preliminary results of this approach for road extraction in urban areas are encouraging for our future work on this topic.

What is missing in our road extraction scheme is the link between urban (in our example: downtown) and rural areas. The roads extracted in rural or urban areas could be used as starting points for road hypotheses in suburban areas. However, the problem in suburban areas is that it is not easy to decide which parts of roads and which image features one should focus on. A consequence would be to employ a suitably extended road and context model and to employ a more flexible extraction strategy.

Acknowledgments

The work presented in this paper was carried out in cooperation of the Chair of Photogrammetry and Remote Sensing and the Department of Computer Science (Image Analysis Group) at TU München during the past six years. Funding by the Deutsche Forschungsgemeinschaft (DFG) under grant no. Eb 74/8 is gratefully acknowledged.

Some parts of the used imagery were provided by members of the Institute for Robotics and Intelligent Systems at USC, Los Angeles/Ca. The program for the external evaluation of road networks was provided by Christian Wiedemann. Their cooperation is also gratefully acknowledged.

References

[Airault et al., 1994] Airault, S., Ruskoné, R. and Jamet, O., "Road detection from aerial images: a cooperation between local and global methods". In: J. Desachy (ed.), *Image and Signal Processing for Remote Sensing*, *Proc. SPIE 2315*, pp. 508–518, 1994.

[Barzohar et al., 1997] Barzohar, M., Cohen, M., Ziskind, I. and Cooper, D. B., "Robust Method for Automatic Aerial Detection of Occluded Roads Based on Multihypothesis Generalized Kalman Filter". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. 32, Part 3-2W3, pp. 1–7, 1997.

[Baumgartner et al., 1997] Baumgartner, A., Eckstein, W., Mayer, H., Heipke, C. and Ebner, H., "Context-Supported Road Extraction". In: [Gruen et al., 1997], pp. 299–308, 1997.

[Boichis et al., 1998] Boichis, N., Cocquerez, J. and Airault, S., "A top down strategy for simple crossroads automatic extraction". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, Part 2, pp. 19–26, 1998.

[Bordes et al., 1997] Bordes, G., Giraudon, G. and Jamet, O., "Road modeling based on a cartographic database for aerial image interpretation". In: [Gruen et al., 1997], pp. 123–139, 1997.

[Clément et al., 1993] Clément, V., Giraudon, G., Houzelle, S. and Sandakly, F., "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, 1993.

[de Gunst and Vosselman, 1997] de Gunst, M. and Vosselman, G., "A semantic road model for aerial image interpretation". In: W. Förstner and L. Plümer (eds), *Semantic Modeling for the Acquisition of Topographic Information from Images and Maps*, pp. 108–122, 1997.

[Eckstein and Steger, 1996] Eckstein, W. and Steger, C., "Fusion of digital terrain models and texture for object extraction". In: *Proceedings of the Second International Airborne Remote Sensing Conference and Exhibition*, Vol. III, Environmental Research Institute of Michigan, pp. 1–10, 1996.

[Fischler et al., 1981] Fischler, M. A., Tenenbaum, J. M. and Wolf, H. C., "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, 1981.

[Gruen and Li, 1997] Gruen, A. and Li, H., "Linear feature extraction with LSB-snakes". In: [Gruen et al., 1997], pp. 287–298, 1997.

[Gruen et al., 1997] Gruen, A., Baltsavias, E. and Henricsson, O. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II).* Birkhäuser Verlag, Basel, 1997.

[Gruen et al., 1995] Gruen, A., Kuebler, O. and Agouris, P. (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images.* Birkhäuser Verlag, Basel, 1995.

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

[Kass et al., 1988] Kass, M., Witkin, A. and Terzopoulos, D., "Snakes: Active contour models". In: *International Journal of Computer Vision*, pp. 321–331, 1988.

[Knuth, 1994] Knuth, D. E., *The Stanford GraphBase: A Platform for Combinatorial Computing.* Addison-Wesley Publishing Company, Reading, MA, 1994.

[Koenderink and van Doorn, 1994] Koenderink, J. J. and van Doorn, A. J., "Two-plus-one-dimensional differential geometry". Pattern Recognition Letters 15(5), pp. 439–443, 1994.

[Laws, 1980] Laws, K., *Texture Image Segmentation.* PhD, Department of Engineering, University of Southern California, 1980.

[Mayer, 1998] Mayer, H., *Automatische Objektextraktion aus digitalen Luftbildern.* Habilitation, Deutsche Geodätische Kommission (C) 494, München, 1998.

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

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

[McKeown, 1990] McKeown, D. M., "Toward Automatic Cartographic Feature Extraction". In: *Mapping and Spatial Modelling for Navigation*, *Nato ASI Series*, Vol. F65, Springer-Verlag, pp. 149–180, 1990.

[McKeown and Denlinger, 1988] McKeown, D. M. and Denlinger, J. L., "Cooperative methods for road tracking in aerial imagery". In: *Computer Vision and Pattern Recognition*, pp. 662–672, 1988.

[Mendel, 1995] Mendel, J. M., "Fuzzy logic systems for engineering: A tutorial". Proceedings of the IEEE 83(3), pp. 345–377, 1995.

[Merlet and Zerubia, 1996] Merlet, N. and Zerubia, J., "New prospects in line detection by dynamic programming". IEEE Transactions on Pattern Analysis and Machine Intelligence 18(4), pp. 426–431, 1996.

[Neuenschwander et al., 1995] Neuenschwander, W., Fua, P., Székely, G. and Kübler, O., "From ziplock snakes to velcro surfaces". In: [Gruen et al., 1995], pp. 105–114, 1995.

[Price, 1999] Price, K., "Road Grid Extraction and Verification". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, part 3-2W5, International Society for Photogrammetry and Remote Sensing, 1999.

[Ruskoné, 1996] Ruskoné, R., *Road Network Automatic Extraction by Local Context Interpretation: application to the production of cartographic data.* PhD thesis, Université Marne-La-Valleé, 1996.

[Ruskoné et al., 1994] Ruskoné, R., Airault, S. and Jamet, O., "Road network interpretation: A topological hypothesis driven system". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXX, part 3/2, pp. 711–717, 1994.

[Ruskoné et al., 1996] Ruskoné, R., Guiges, L., Airault, S. and Jamet, O., "Vehicle detection on aerial images: A structural approach". In: *13th International Conference on Pattern Recognition*, Vol. III, pp. 900–903, 1996.

[Steger, 1998a] Steger, C., "An Unbiased Detector of Curvilinear Structures". IEEE Transactions on Pattern Analysis and Machine Intelligence 20(2), pp. 113–125, 1998a.

[Steger, 1998b] Steger, C., *Unbiased Extraction of Curvilinear Structures from 2D and 3D Images*. Dissertation, Fakultät für Informatik, Technische Universität München, 1998b. Herbert Utz Verlag, München, ISBN 3-89675-346-0.

[Steger et al., 1995] Steger, C., Glock, C., Eckstein, W., Mayer, H. and Radig, B., "Model-based road extraction from images". In: [Gruen et al., 1995], pp. 275–284, 1995.

[Tönjes, 1997] Tönjes, R., "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, 1997.

[Trinder and Wang, 1998] Trinder, J. and Wang, Y., "Knowledge-Based Road Interpretation in Aerial Images". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, part 4, pp. 635–640, 1998.

[Vasudevan et al., 1988] Vasudevan, S., Cannon, R. L. and Bezdek, J. C., "Heuristics for intermediate level road finding algorithms". Computer Vision, Graphics, and Image Processing 44, pp. 175–190, 1988.

[Vosselman and de Knecht, 1995] Vosselman, G. and de Knecht, J., "Road tracing by profile matching and kalman filtering". In: [Gruen et al., 1995], pp. 265–274, 1995.

[Wiedemann and Hinz, 1999] Wiedemann, C. and Hinz, S., "Automatic Extraction and Evaluation of Road Networks from Satellite Imagery". In: *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, part 3-2W5, International Society for Photogrammetry and Remote Sensing, pp. 95–100, 1999.

[Zlotnick and Carnine, 1993] Zlotnick, A. and Carnine, P., "Finding road seeds in aerial images". Computer Vision, Graphics, and Image Processing: Image Understanding 57(2), pp. 243–260, 1993.