Road extraction focussing on urban areas*

S. Hinz, A. Baumgartner
Chair for Photogrammetry and Remote Sensing, Technische Universität München, Munich, Germany

H. Mayer
Institute for Photogrammetry and Cartography, Bundeswehr University Munich, Munich, Germany

C. Wiedemann, H. Ebner
Chair for Photogrammetry and Remote Sensing, Technische Universität München, Munich, Germany

ABSTRACT: In this paper, we present work on automatic road extraction from aerial imagery. After briefly reviewing our system for road extraction in rural areas, we focus on our current research on road extraction in urban areas. In order to deal with the high complexity of this type of scenes, we integrate detailed knowledge about roads and their context using explicitly formulated models. The road model includes, for instance, small sub-structures such as markings but also knowledge about the global network characteristics of roads, while the context model describes relations between roads and other objects, e.g., buildings casting shadows on the road or cars occluding parts of a lane. Most of the context information is gained from the image and a given Digital Surface Model at the beginning of the extraction, which allows us to automatically adapt specific parts of the road model and extraction strategy to the contextual situation. This makes it possible to extract roads even if their appearance is heavily affected by other objects. We illustrate intermediate steps of the extraction by various examples and, based on an external evaluation of the final results, we discuss the advantages but also the remaining deficiencies of the approach.

1 INTRODUCTION

Due to the need for efficient acquisition and update of data for Geographic Information Systems (GIS), the automatic extraction of man-made structures from aerial imagery became an important research issue more than two decades ago. Especially the reconstruction of buildings and the extraction of roads has received considerable attention. For roads, besides many user- or map-guided approaches, also numerous automatic approaches have been developed. Most of these efforts are directed towards the extraction of roads in rural areas. Approaches designed to process satellite or low resolution aerial images generally describe roads as curvilinear structures (Gruen & Li 1997, Heller et al. 1998, Wiedemann & Hinz 1999) while those using large scale imagery (i.e., a ground resolution less than 1 m) model roads mostly as relatively homogeneous areas satisfying certain shape and size constraints (Ruskoné 1996, Trinder & Wang 1998, Mayer et al. 1998, Harvey 1999, Zhang & Baltsavias 1999, Baumgartner et al. 1999).

Compared to the relatively high number of research groups focussing their activities on rural areas, only few groups work on the automatic extraction of roads in urban environments. Here, the road network is often modeled as a combination of regular grids with an approximately constant mesh size, i.e., the size of a single building block. Moissinac et al. (1995), for instance, incorporate this dualism of building blocks and urban roads to achieve a consistent scene interpretation for features extracted from maps and images. Faber & Förstner (2000), in contrast, rely purely on directional information of lines extracted from scanned maps or low resolution images for segmenting regions showing similar grid orientation. Price (1999) combines multiple high resolution images and a Digital Surface Model (DSM) to extract the urban road grid in complex,

* Preprint, to appear in: *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III)*. Balkema Publ.

Figure 1. Left: Aerial image from the Zurich Hoengg dataset. Right: Texture based segmentation of *open rural area.*

though stereotypical, residential areas. After manual initialization of two intersecting road segments defining the first mesh, the grid is iteratively expanded by hypothesizing and verifying new meshes. During final verification, Price (2000) exploits the contextual knowledge that high objects such as buildings or trees may define the sides of urban roads. Thus, few consecutive road segments are simultaneously adjusted by moving them to local minima of the DSM, i.e., the extracted roads are constrained to run along ”valleys” in the DSM.

Throughout all the different approaches, some issues have proved to be of great importance: By integrating a *detailed road and context model* one can capture the varying appearance of roads and the influence of background objects such as trees, buildings, and cars in complex scenes (Ruskoné 1996, Baumgartner et al. 1999). The *fusion of different scales* helps to eliminate isolated disturbances on the road while the fundamental structures are emphasized (Mayer & Steger 1998). This can be supported by considering the function of roads connecting different sites and thereby forming a fairly dense and sometimes even regular network. Hence, exploiting the *network characteristics* adds global information and, thus, the selection of the correct hypotheses becomes easier (Heller et al. 1998, Wiedemann & Ebner 2000).

In the remainder of this paper, we give a short review on our approach on road extraction in rural areas which has been tested on various images and reached a rather matured state (Section 2). Then, in Section 3 we turn to our present work on road extraction in urban areas and present the basic components of our road model. Illustrated by examples using the Zurich Hoengg dataset, we outline the extraction strategy in Section 4. In Section 5, a numerical evaluation of the results currently achievable with our system is given followed by a discussion of the advantages and remaining deficiencies of the proposed approach. We conclude the paper with an outlook on future work.

2 ROAD EXTRACTION IN RURAL AREAS

Usually, in rural areas there are only a few background objects, e.g., single buildings or trees, that influence the appearance of roads in aerial imagery. Therefore, we do not need to make extensive use of contextual information, i.e., of knowledge about relations between roads and other objects, and base the main part of the extraction on attributes of the object road itself. Our approach to road extraction in rural areas employs a model that comprises the scale-space behavior of roads and describes the road network on different levels. We combine extraction of lines in small scale

Figure 2. Result of road extraction for the upper part of Figure 1.

and edges in large scale to generate hypotheses for segments of the road network and integrate local and global grouping to construct the network. The texture based segmentation of *open rural areas* (cf. Fig. 1) helps to define the regions where the approach is supposed to deliver reliable results.

We developed a road extraction scheme for rural areas that consists of three different modules with specific strengths. The first module employs multiple scales as well as the local grouping of lines and edges to reliably extract most parts of the road network. Contextual information is used to verify connection hypotheses on a local level (Baumgartner et al. 1999). The second module is able to fuse linear structures from various sources and constructs a weighted graph. Network characteristics are exploited by selecting pairs of seed points within this graph and by searching for shortest paths between these pairs. Compared to the first module, the second module is characterized by global grouping (Wiedemann & Hinz 1999). The third module completes the network based on an analysis of path lengths within and between connected components of the network. Its most important contribution is the extraction of roads which improve the topology of the network (Wiedemann & Ebner 2000). An evaluation of the results on several test scenes shows that the system benefits from the integration of different types of knowledge within the road extraction scheme (Hinz et al. 2000). Figure 2 displays the result achieved by our system for the upper part of Figure 1.

3 MODELLING ROADS AND CONTEXT IN URBAN AREAS

Road extraction in urban areas is in particular motivated by the high demand for accurate, detailed, and up-to-date information for applications related with urban planning as, e.g., traffic flow analysis and simulation, estimation of air and noise pollution, street maintenance, etc. The survey on 3D City models by the European Organization for Experimental Photogrammetric Research (OEEPE) confirmed this demand showing that about 85% of the participants have greatest interest in information about road networks (Fuchs et al. 1998). However, many of the developed road extraction approaches would presumably fail when applied to images taken over densely built-up areas, though they show good results in rural areas. A reason for this is that the incorporated models and strategies usually can not cope with situations typical for urban areas. The most important of them are:

1) The variability of objects belonging to the same class is usually bigger. For instance, buildings in downtown areas are typically much more complicated than more or less isolated houses in rural areas.

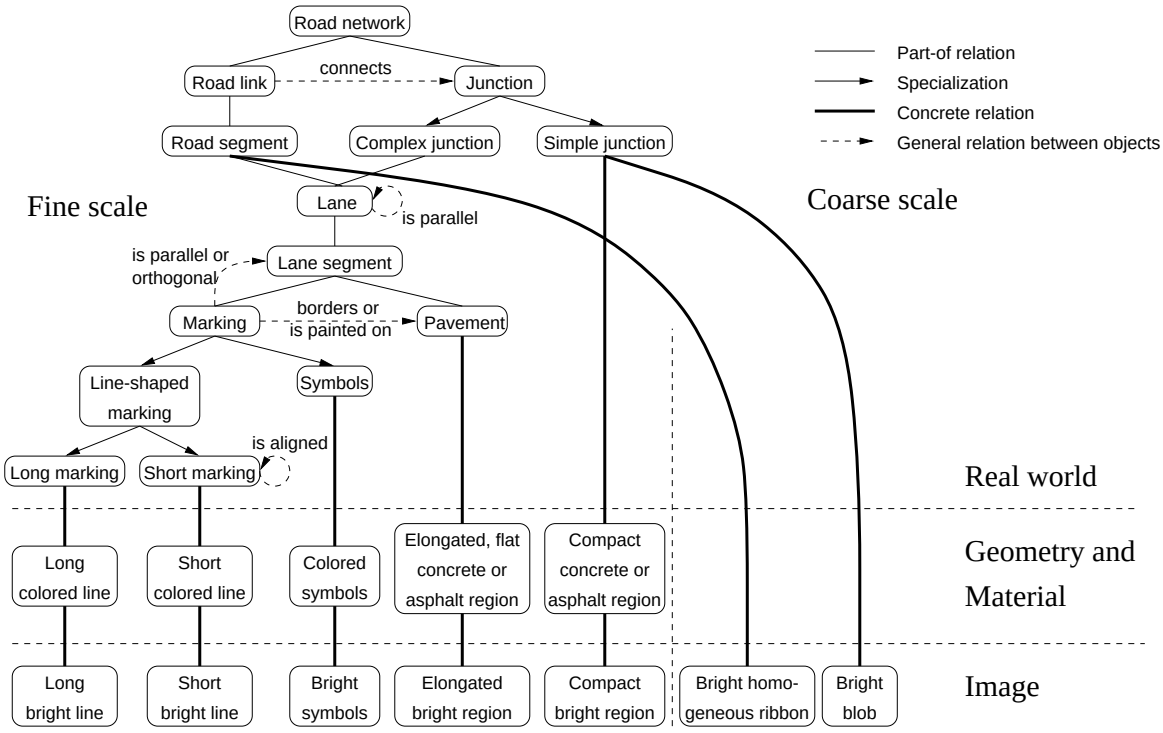


Figure 3. Road model.

2) The appearance of an object is often influenced by its neighboring objects, and thus, it deviates from the underlying object model. For example, buildings casting shadows on a road lead to strong shadow edges across the usually bright and homogeneous road. Such disturbances happen less frequently in rural areas.

Hence, beside using a sophisticated extraction strategy, detailed object and context modelling plays a key role for road extraction in urban areas.

3.1 *Road model*

The road model illustrated in Figure 3 compiles knowledge about radiometric, geometric, and topological characteristics of urban roads in form of a hierarchical semantic network. The model represents the *standard case*, i.e., the appearance of roads is not affected by relations to other objects. It 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. On this level the road network consists of junctions and road links connecting junctions. Road links are constructed from road segments. In fine scale, road segments and complex junctions are aggregatings of lanes, which in turn 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 of sensor characteristics and viewpoint. Road segments are linked to the bright homogeneous ribbons of the image level in coarse scale. In contrast to this, the pavement as a part of a lane 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 (Mayer & Steger 1998), 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 of the lanes and markings, are integrated from fine scale. In this way the extraction benefits from both scales.

3.2 *Context model*

The road model presented above is extended by knowledge about context: So-called context objects, i.e., background objects such as buildings, trees, or vehicles, may hinder road extraction if they are not modelled appropriately but they can substantially support the extraction if they are part of the road model. Vehicles, for instance, occlude the homogeneous pavement and, thus, interfere the extraction. On the other hand, if the extraction system is able to detect a vehicle automatically, a strong hint for a road has been found. External GIS data can also be regarded as context object. Experience from our work on road extraction in rural areas (Baumgartner et al. 1999) shows that modeling this interaction between road objects and context objects on a local as well as on a global level is a strong aid for guiding the extraction. The difficult task of image interpretation is split into smaller sub-problems which can be solved efficiently by using specific models and extraction strategies.

Global context:
The motivation for employing global context stems from the observation that it is possible to find semantically meaningful image regions – so-called *context regions* – where roads show typical prominent features and where certain relations between roads and background objects have a similar importance. Consequently, the relevance of different components of the road model and the importance of different *context relations* (described below) must be adapted to the respective context region. In urban areas, for instance, markings separating individual lanes are often a remarkable feature and have therefore a high relevance for the extraction, while they are usually less important in rural areas. Similarly, relations between vehicles and roads are more important in urban areas since traffic is usually much denser inside of settlements than in rural areas. As in our previous work, we distinguish *urban*, *forest*, and *rural* context regions (an example of the texture-based segmentation is given in Figure 1 b).

Local context:
We model the local context with so-called *context relations*, i.e., certain relations between a small number of road and context objects. Typical context relations in the rural context region can be found in Baumgartner et al. (1997). In the following we turn our focus on context relations in urban areas (see Fig. 4).

Almost every building in the real world is connected to the road network. The denser the settlement, the closer the buildings move to the road and the more parallel is their outline to the road sides. Therefore, this context relation is especially useful for the extraction in downtown areas, where, in extreme cases, roads and junctions are purely defined by the building outlines. Vice-versa, buildings or other high objects standing close to the road potentially occlude larger parts of it or cast shadows on it. Hence, a context relation ”occlusion”, gives rise to the selection of another image providing a better view on this particular part of the scene, whereas a context relation ”shadow” can tell an extraction algorithm to choose modified parameter settings, e.g., for the extraction of road markings. Both context relations imply that roads lie lower than the surrounding objects. Consequently, there is no need to search for roads on locally high objects.
For some settlements, road axes might be available digitally. Such kind of information can be integrated in a very consistent way by using a context relation that models parallelism and closeness between an extracted piece of road and the mapped road axis (Moissinac et al. (1995)). With such a context relation, cues are provided where a road *might* be present. Nevertheless, the extraction has to prove *independently* whether the road truly exists in the image.
Vehicles are related to a road, or more specifically, to a lane segment by means of occluding the lane’s pavement. However, since vehicles drive or stand aligned with a lane – at least in most cases – we can directly use a detected vehicle or vehicle convoy for road extraction: in particular, we treat a detected vehicle as lane segment. In so doing, we need not to take care of moving vehicles if we want to fuse results achieved from images taken at different times.

In addition to relations between road and context objects, we also consider relations between the object and its sub-structures, i.e., objects which can help to define the semantics of an object, but which can only be detected at a very fine scale in relation to the object extent. They are

exemplified here by orthogonal markings which relate the end of a lane to a junction. Figure 4 summarizes the relations between road objects, context objects, and sub-structures by using the concepts ”Lane segment” and ”Junction” as the basic entities of a road network. Note, however, that the exploitation of specific context relations will in most cases be possible in high resolution imagery only (< 0.15 m), because sub-structures and other 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 > 0.5 m and is useful to guide the road extraction in both scales.

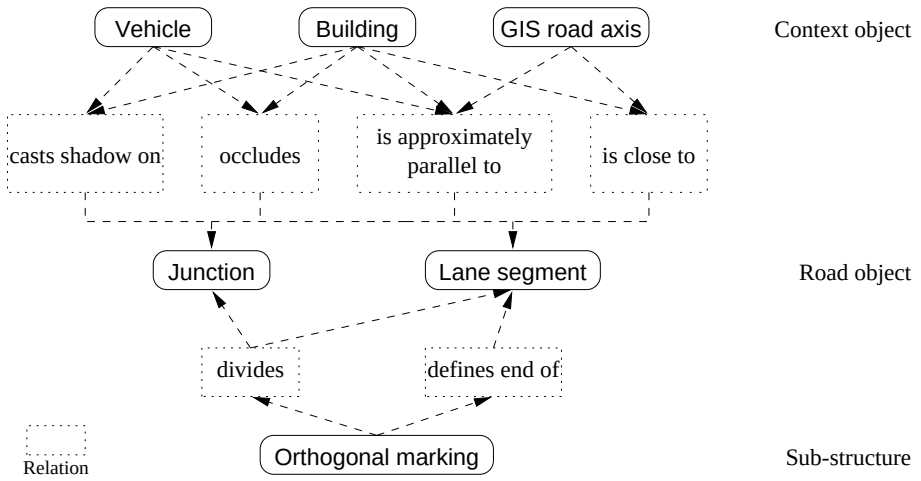


Figure 4. Context relations in urban areas.

4 EXTRACTION STRATEGY

In a very general sense, the extraction strategy comprises knowledge about how and when certain parts of the road and context model are optimally exploited. Since the road model incorporates 3D information as well as small sub-structures to a considerable extent, the extraction relies on one side on aerial imagery consisting of overlapping gray scale images with a fairly high resolution (< 15 cm) and on the other side on an quite accurate DSM. In contrast to other approaches, we neither use orthophotos for extraction nor we extract completely in 3D as, e.g., Grün & Li (1997) do, by using multiple images simultaneously. The latter procedure is conceptually elegant, but it inherently implies matching procedures throughout the extraction process, which become burdensome for higher scene complexities. Furthermore, we want to avoid feature extraction in orthophotos – despite of using accurate DSM information. Inaccuracies of the DSM due to erroneous height measurements, filtering, resampling, or moving objects remain in orthophotos and, for instance, could disturb in particular collinear properties of image structures like road markings. Hence, during extraction, we separate the processing of image and height information to a certain extent. The overall flow of our extraction is illustrated in Figure 5. It consists of 3 levels: (1) *Context-based data analysis*, (2) *Extraction of salient road segments*, and (3) *Road network completion.*

Context-based data analysis:
We start the extraction by exploiting contextual knowledge to make it available for all following processing stages. After segmenting the image into urban, rural, and forest areas, we continue in the urban area with the context relation between roads and buildings. To take into account the ”parallel to” and ”close to” relations, we extract coarse building outlines, i.e., bright blobs, from the DSM followed by the detection of valleys, i.e., dark wide lines, between them.

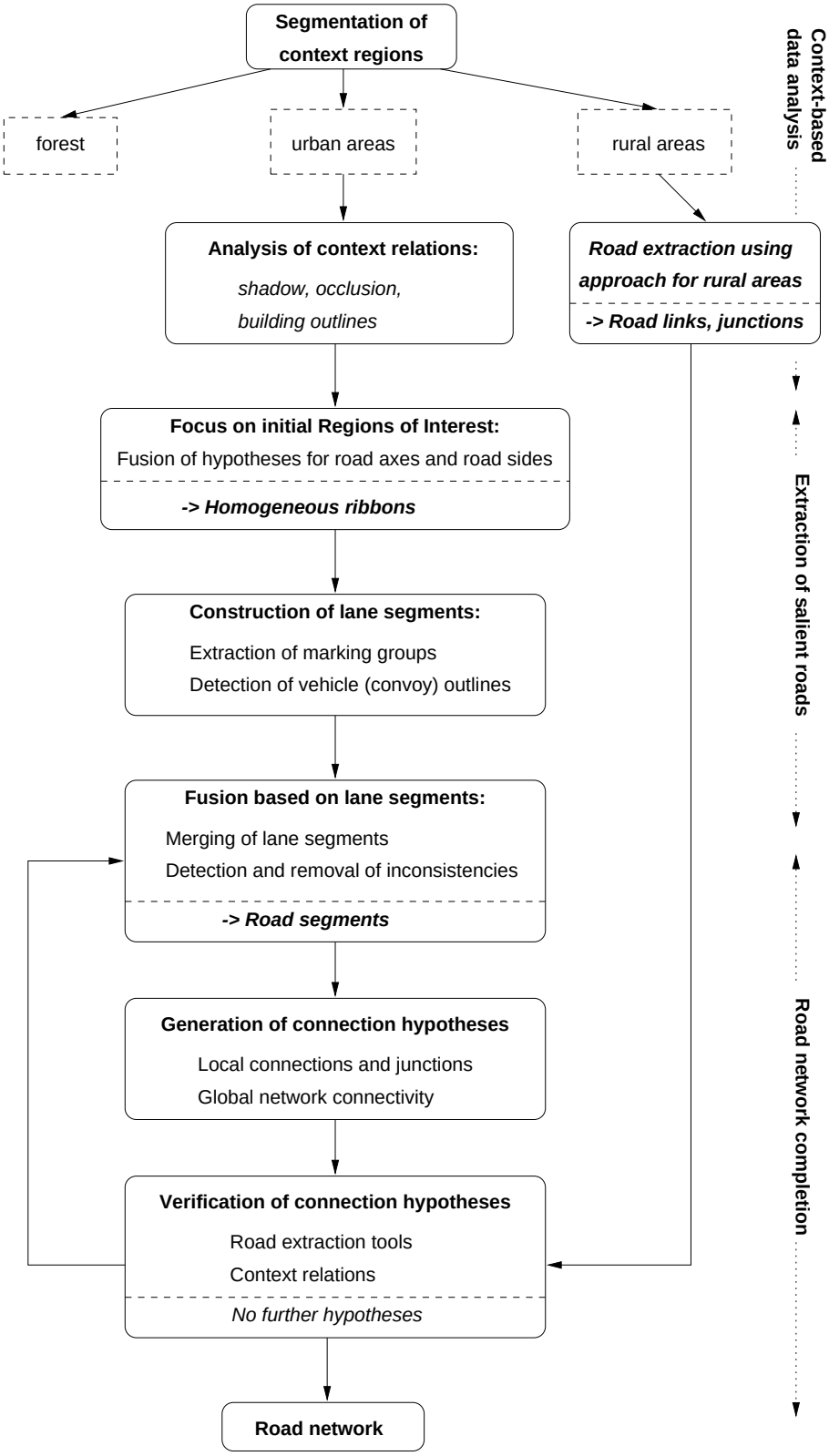
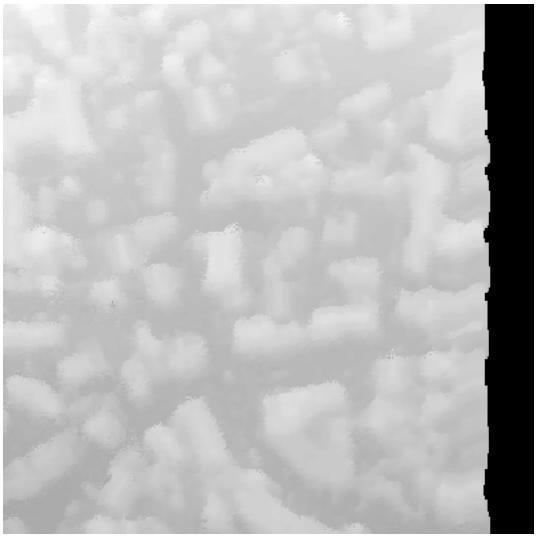


Figure 5. Extraction strategy for urban areas.



(a) DSM

(b) Building outlines and valleys derived from DSM

(c) Shadow regions derived from DSM

(d) Refined shadow regions

Figure 6. Analysis of context relations.

This gives us a rough idea about the position and direction of a potential road as well as about regions where no road is expected. Then we explore the ”shadow” relation. Shadow regions are detected in a similar way as proposed by Eckstein & Steger (1996) by combining height and image information with knowledge about the sun angles derived from image orientation parameters and image capture time. The results of these steps are presented in Figure 6.

Extraction of salient roads:
For deriving initial Regions of Interest (RoI) we fuse the DSM-valleys with markings, i.e., thin bright lines with symmetric contrast, and potential road sides, i.e., image edges extracted from a high resolution image, as well as dark ribbons extracted from a low resolution gradient image. This operation returns homogeneous ribbons with partial edge support and optionally markings inside (see Fig. 7 a). Within these regions we start a grouping procedure which iteratively connects consecutive markings and constructs lane segments from parallel marking groups. This operation also expands in regions outside the RoI. The resulting lane segments are validated by checking their interior for gray value homogeneity in direction of the lane.

(a) Homogeneous ribbons

(b) Detail: verified lanes (white), remaining (black)

(c) Detected car

(d) Connected road segments

Figure 7. Steps for extracting salient roads.

However, cars driving on the road can cause severe disturbances of the homogeneity which results usually in a gap in the extraction (see Fig. 7 b). Therefore, the system calls a module for automatic vehicle detection (Hinz 2001) trying to find a reason for the gap (Fig. 7 c). The gaps are closed whenever a car has successfully been detected. Finally, the verified lanes segments are connected to construct lanes, and parallel and collinear lanes are aggregated to set up road segments. The axes of the resulting road segments are displayed in Figure 7 d.

Road network completion:
Once the road segments have been constructed, the network is iteratively completed by generating and verifying connection hypotheses. By doing so, road segments are linked and road junctions are reconstructed. The verification can be carried out by using road extraction tools such as ribbon snakes and homogeneity tracking but also by exploiting the context relations described above. Up to now, we only integrated modules for checking homogeneity in junction and shadow regions into this loop (see *road network completion* in Fig. 5). Hence, the bridging of such regions is currently not very accurate. However, the modules applied during the extraction of salient roads have to be

(a) Image Part I (b) Image Part II

Figure 8. Extracted road networks.

modified only slightly and will be included in the near future. Figure 8 a shows the successful verification of the connection leading through the shadow region in the central part of the image.

5 RESULTS AND DISCUSSION

Figure 8 shows the final result of road extraction in two parts of the Zurich Hoengg dataset. The results have been evaluated by matching the extracted road axes to manually plotted reference data (Wiedemann & Ebner 2000). As can be seen, major parts of the road networks have been extracted (white lines indicate extracted road axes). Expressed in numerical values, we achieve a completeness of almost 70 % and a correctness of about 95 %. The system is able to detect shadowed road sections or road sections with rather dense traffic. However, it must be noted that some of the axes' underlying lane segments have been missed. This is most evident at the complex road junctions in both image parts, where only spurious features for the construction of lanes could be extracted. Thus, not enough evidence was given to accept connections between the individual branches of the junction. Another obvious failure can be seen at the right branch of the junction in the central part of Image Part I (Fig. 8 a). The tram and trucks in the center of the road have been missed since our vehicle detection module is only able to extract vehicles similar to passenger cars. Thus, this particular road axis has been shifted to the lower part of the road where the implemented parts of the model fit much better.

In summary, the results indicate that the presented system extracts roads even in complex environments. An obvious deficiency exists in form of the missing detection capability for vehicle types as busses and trucks. However, the main bottleneck of our system is the (still) weak model for complex junctions. Hence, one of our next steps will be directed towards the modeling and reliable detection of road junctions. As indicated in Section 4, another extension of our system is the incorporation of multiple overlapping images. Also for multiple images, we plan to treat the processing steps up to the generation of lanes purely as 2D-problem. The results for each image are then projected on the DSM and fused there to achieve a consistent dataset. Then, new connections are hypothesized and, again, verified in each image separately.

REFERENCES

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

Baumgartner, A., Steger, C., Mayer, H., Eckstein, W. & Ebner, H. 1999. Automatic Road Extraction Based on Multi-Scale, Grouping, and Context. *Photogrammetric Engineering and Remote Sensing* 65(7): 777–785.

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

Eckstein, W. & Steger, C. 1996. Fusion of digital terrain models and texture for object extraction. *Proc. Second International Airborne Remote Sensing Conference and Exhibition, vol. III, 1–10.* Environmental Research Institute of Michigan.

Faber, A. & Förstner, W. 2000. Detection of Dominant Orthogonal Structures in Small Scale Imagery. *International Archives of Photogrammetry and Remote Sensing*, vol. 33, part B3(1), 274–281.

Fuchs, C., Gülch, E. & Förstner, W. 1998. OEEPE Survey on 3D-City Models. *OEEPE Official Publication, No. 35*.

Gruen, A. & Li, H. 1997. Linear feature extraction with LSB-snakes. In A. Gruen, E. Baltsavias & O. Henricsson (eds), *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II)*, 287–298. Basel: Birkhäuser Verlag.

Harvey, W. A. 1999. Performance Evaluation for Road Extraction. *Proc. ISPRS Workshop on 3D Geospatial Data Production: Meeting Application Requirements, 175–184, Paris, 7-9 April 1999.*

Heller, A., Fischler, M., Bolles, R. & Connolly, C. 1998. An Integrated Feasibility Demonstration for Automatic Population of Spatial Databases. *Proc. Image Understanding Workshop.*

Hinz, S. 2001. Vehicle Detection in Aerial Imagery based on Generic Features, Grouping, and Context. *Technical Report PF–2001–02*, Chair for Photogrammetry and Remote Sensing, Technische Universität München.

Hinz, S., Wiedemann, C. & Baumgartner, A. 2000. A Scheme for Road Extraction in Rural Areas and its Evaluation. *Proc. 5th IEEE Workshop on Applications of Computer Vision, 134–139.*

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

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

Moissinac, H., Maître, H. & Bloch, I. 1995. Graph based urban scene analysis using symbolic data. In D. M. McKeown Jr. & I. Dowman (eds), *SPIE Conf. on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, SPIE Proc. vol. 2486, 93–104.*

Price, K. 1999. Road Grid Extraction and Verification. *International Archives of Photogrammetry and Remote Sensing*, vol. 32, part 3-2W5, 101–106.

Ruskoné, R. 1996. *Road Network Automatic Extraction by Local Context Interpretation: application to the production of cartographic data.* Ph.D. Thesis, Université Marne-La-Valleé.

Tönjes, R. 1997. 3D Reconstruction of Objects from Aerial Images Using a GIS. *International Archives of Photogrammetry and Remote Sensing*, vol. 32, part 3-2W3, 140–147.

Trinder, J. & Wang, Y. 1998. Knowledge-Based Road Interpretation in Aerial Images. *International Archives of Photogrammetry and Remote Sensing*, vol. 32, part 4, 635–640.

Wiedemann, C. & Ebner, H. 2000. Automatic completion and evaluation of road networks. *International Archives of Photogrammetry and Remote Sensing*, vol. 33, part B3(2), 979–986.

Wiedemann, C. & Hinz, S. 1999. Automatic Extraction and Evaluation of Road Networks from Satellite Imagery. *International Archives of Photogrammetry and Remote Sensing*, vol. 32, part 3-2W5, 95–100.

Zhang, C. & Baltsavias, E. 1999. Road Network Detection by Mathematical Morphology. *Proc. ISPRS Workshop on 3D Geospatial Data Production: Meeting Application Requirements, 185–200, Paris, 7-9 April 1999.*