AUTOMATIC ROAD EXTRACTION BASED ON MULTI-SCALE MODELING, CONTEXT, AND SNAKES

Helmut Mayer[1], Ivan Laptev[2], Albert Baumgartner[1], Carsten Steger[3]

[1]Chair for Photogrammetry and Remote Sensing
Technische Universität München, D-80290 Munich, Germany
[2] Computational Vision and Active Perception Laboratory
NADA, KTH, S-10044 Stockholm, Sweden
[3]Forschungsgruppe Bildverstehen, Informatik IX
Technische Universität München, D-80290 Munich, Germany

E-mail: {helmut|albert|ivan}@photo.verm.tu-muenchen.de; stegerc@informatik.tu-muenchen.de
URL: http://www.photo.verm.tu-muenchen.de

used to decide which segments should be verified first and, what is more, also which algorithm should be used for extraction.

This paper follows the work presented in (Baumgartner et al., 1997b), which essentially uses context information and multiple scales, based on the experience that distinct characteristics of roads can be detected best at different scales and in different contexts. Basically, multiple scales are used to model detailed information of fine scale, like the markings, as well as abstract information of coarse scale, like the road network. When fusing multiple scales, on the one hand, the abstract information is used to focus the extraction of the details. This avoids getting lost in the plethora of features in an image. On the other hand, details like markings can give very reliable evidence for a road.

The multi-scale modeling is complemented by context information which is divided into local context sketches, like *occlusion_shadow*, which is modeling a tree casting a shadow on the road, and global context regions, i.e., *open_rural*, *suburb_urban*, and *forest* areas, which comprise the whole image. The context sketches are related to the context regions: The context sketch *occlusion_shadow* for instance belongs to the context region *suburb_urban*. By this means the complex model for the object *road* is split into more specific sub-models which are adapted to the contextual environment. The sub-models emphasize certain characteristics of the objects and therefore can be regarded as specialized models. An advantage of context regions is that they can be used to focus the extraction: Road extraction in *open_rural* areas is easier and much more robust than in *suburb_urban* or in the *forest*. Thus, from the scale-space behavior and the context of the road a strategy for the extraction of the road can be deduced.

Up to now, snakes were mainly used for semi-automatic extraction. In the scope of the work presented in this paper they were found to show many advantages compared to the grouping scheme used in previous work for automatic road extraction (Baumgartner et al., 1997b). More specifically, the results of the coarse scale line extraction were used as the approximate centers for so-called ribbon-snakes used to extract the road in fine scale. Especially the ability to bridge gaps resulting from shadows cast on the road by trees or buildings (context sketch *occlusion_shadow*) is a new and very important feature of this approach. The main advantage of the snake for this application is that due to its geometrical stabilization, it can make use of the little information in the shadowed or short visible parts.

The paper proceeds as follows. In Section 2 the scale-space behavior of roads is analyzed, and a model for the extraction of roads is condensed from it. Section 3 defines the context sketches and assigns them different context regions. The model derived from scale-space behavior and the context of the road is complemented with a strategy for road extraction in Section 4. The snake-based approach is presented in Section 5. After giving some basics for snakes, the so-called ribbon snakes are introduced. After distinguishing *salient* and *non-salient roads*, results for road extraction are shown. The paper concludes with an outlook in Section 6.

2 SCALE-SPACE BEHAVIOR

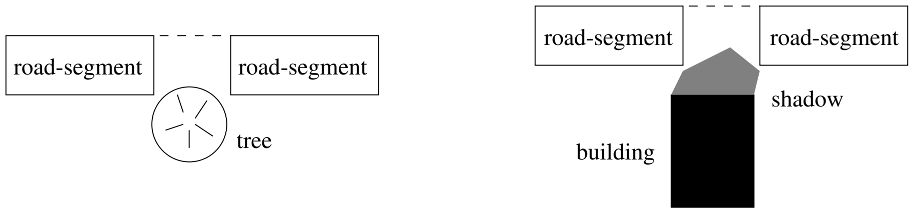
The appearance of roads in digital imagery depends on the sensor's spectral sensitivity and its resolution, i.e., inherent scale in object space. The remainder of this paper is restricted to grey-scale images, and only scale dependencies are considered. Images with various scale exhibit different characteristics of roads. In images with coarse scale, i.e., more than 2 m per pixel, roads mainly appear as lines establishing a more or less dense network. Opposed to this, in images with a finer scale, i.e., less than 0.50 m, roads are depicted as elongated homogeneous areas with more or less parallel borders and almost constant width.

In a smoothed image, i.e., a coarser scale, lines representing road axes can be extracted in a stable manner even in the presence of background objects like trees, buildings, or cars. Smoothing an image is hereby closely linked to the concept of "scale-space" for which (Lindeberg, 1994) gives a good introduction. Figure 1(a) displays a bar shaped bright line (= road) with a bright disturbance on the right side (= bright car on the right lane) and its behavior in scale-space for the line extraction model presented in (Steger, 1996). It is intuitively clear that only one line should be detected for all levels of smoothing, i.e., scales σ, and this is indeed the case. Figure 1(a) displays the line and edge positions mapped onto the smoothed profiles, while Figure 1(b) compares them to the corresponding positions of an undisturbed profile (ideal position). For small σ the extracted line position will be the one of the bright object, while for large σ it will correspond to the center axis of the line.

The outcome is that just by increasing the scale σ one can eliminate the car from the road. It can also be seen that the two edges corresponding to the bright object will vanish along with the flat inflection points on the undisturbed part of the line. As (Mayer and Steger, 1996) have shown, the appropriate scale for line extraction can be computed if the width of the line (=road) and of the disturbance (=car) as well as the contrast between background, line, and disturbance is given. Seen from a symbolical point of view, in the finer scale the substructure of the road (the car on the road or also objects like markings) has been eliminated. This can be interpreted as the *abstraction*, i.e., the increase of the level of simplification and emphasis of the road. Abstraction is achieved simply by changing the scale of the object.

From the last paragraph follows that fusion of coarse and fine scale results can contribute to improve the reliability of the road hypotheses. Additionally, details like road markings, which can be recognized at a resolution of less than 0.25 m, can be used as evidence to corroborate the detected road hypotheses. On the one hand, using multiple scales improves the robustness of road extraction. On the other hand, it results in the necessity to use different features at each scale, and to simultaneously combine all features of all scales into one road model. The semantic network in Figure 2 illustrates a simplified road model condensed from this (for a more complete model refer to (Baumgartner et al., 1997a)).

The model is split into three levels, defining different points of view. The *real world* level consists of the objects and their relations on a natural language level. In fine scale the road-segment is constructed of road-parts which in turn comprise the pavement and the markings. The objects in the *real world* level are connected to the objects in the *geometry and material* level by means of the *concrete* relation which connects *concepts* describing the same object on different levels, i.e., from different points of view. The *geometry and material* level is an intermediate level which



a) b)

Figure 3: Context sketch *occlusion_shadow*

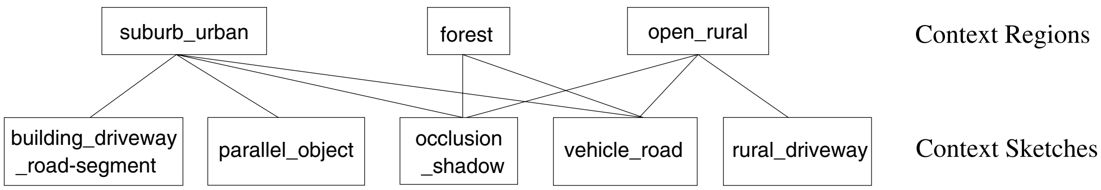


Figure 4: Context regions and context sketches for roads

4 STRATEGY

From the scale-space behavior of the roads and the model for their context the following strategy for their extraction can be deduced. It is based on the principle "hypothesize and test", and focuses on single objects ("local feature focus" (Grimson, 1990)) by postulating that the focus should consider:

- Object is easy to extract
- Object can be extracted with high confidence
- Object has a big importance for the overall extraction

The resulting strategy is (Baumgartner et al., 1997a):

1. Start extraction in *open_rural* area
2. Extract hypotheses for roads in coarse scale
3. Extract roads in fine scale; verify by means of markings or cars
4. Expansion of the road-network by a complex interaction of
   (a) Closing gaps based on local context
   (b) Propagation of the network in other global contexts

For the expansion of the network in 4. (Steger et al., 1997) show an approach which uses the information of the whole network to find out which parts should be connected.

In the next Section a snake-based approach is presented which not only gives good results for the extraction of roads but also gives a means to bridge the gaps caused by small occlusions or by shadows cast on the roads.

5 SNAKES

This Section is based on results from (Laptev, 1997), where details of the approach can be found.

5.1 Basics of Snakes

The concept "snake", also called "active contour model" was originally introduced in (Kass et al., 1987). It combines internal smoothness constraints like bending of a curve with image forces like the gradient. This idea can be represented as a sum of its energies

$$E(\vec{v}) = E_{img}(\vec{v}) + E_{int}(\vec{v}) + E_{ext}(\vec{v}), \quad (1)$$

where E_{int} represents the *internal energy*, E_{img} the *image energy* and E_{ext} the *external forces*. The position of the snake where all these forces compensate each other corresponds to the local minimum of the snake's total energy E. Thus, the problem of the optimization of the snake's position is equivalent to the minimization of its energy.

The image energy of the snake can be defined as:

$$E_{img}(\vec{v}) = -\int_0^1 P(\vec{v}(s,t))ds, \quad (2)$$

where $P(\vec{v}(s,t))$ is a function with high values corresponding to the features of interest. When attracting the snake to edges in images, $P(\vec{v}(s,t))$ is usually taken equal to the magnitude of the image gradient, that is

$$P(\vec{v}(s,t)) = |\nabla \ \ (\vec{v}(s,t))|, \quad (3)$$

where $\ (\vec{v}(s,t))$ is the raw image or – more often – the image convolved with the Gaussian kernel. The convolution with Gaussian kernel smoothes the image and removes disturbances which prevent the snake from moving toward the positions with lower image energy corresponding to the more salient image features.

The internal energy makes it possible to introduce geometric constraints on the shape of the snake. It can be defined as

$$E_{int}(\vec{v}) = \frac{1}{2}\int_0^1 \alpha(s)\left|\frac{\partial\vec{v}(s,t)}{\partial s}\right|^2 + \beta(s)\left|\frac{\partial^2\vec{v}(s,t)}{\partial s^2}\right|^2 ds, \quad (4)$$

where $\alpha(s)$ and $\beta(s)$ are arbitrary functions that control the snake's tension and rigidity. The constraint on tension is introduced by the first order term and makes the snake act like a membrane. The rigidity is constrained by the second order term and makes the snake act like a thin plate.

In order to find the optimal position for the snake, its energy has to be minimized. According to the variational calculus this must be a solution to the *Euler-Lagrange* differential equation of motion. When choosing a particular deformation energy the differential equation controling the motion of the snake becomes linear and can be separated. This has the advantage of solving one optimization step in linear time. For the actual implementation the equations have to be discretized. For details of this refer to (Laptev, 1997).

5.2 Ribbon Snakes

The goal of this paper is to extract roads, i.e., linear features with significant width. They can be modeled by ribbons whose sides correspond to the features' boundaries. Using ribbon snakes, linear features can be extracted by optimizing the position and the width of the ribbon. In order to represent ribbon snakes, the parametric curve $\vec{v}(s,t)$ can be augmented by the third component $w(s,t)$ (Fua and Leclerc, 1990):

$$\vec{v}(s,t) = (x(s,t), y(s,t), w(s,t)), \qquad (0 \leq s \leq 1), \tag{5}$$

Such representation implies that each slice of the ribbon snake $\vec{v}(s_0,t_0)$ is characterized by its width $2w(s_0,t_0)$ and the location of its center $(x(s_0,t_0), y(s_0,t_0))$. All center points compose the centerline of the ribbon (cf. Figure 5 (a)).

In order to perform the optimization of the ribbon snake, the forces which act on it have to be defined. The advantage of the ribbon's representation in equation (5) is that the expression for the snake's internal energy E_{int} can be directly used for ribbon snakes. Doing so, the width of ribbons will be constrained by tension and rigidity in the same way as the two coordinate components. The internal forces which act on the ribbon snake will on the one hand constrain ribbon's centerline to be a smooth curve. On the other hand, they will control the distance between the ribbon's sides, forcing the sides to be parallel.

In contrast to the original snakes, the image information for ribbon snakes has to be taken into account not at the center of the curve $(x(s,t), y(s,t))$, but at the ribbon's left and right sides. As shown in Figure 5 (a), for each slice of the ribbon $\vec{v}(s_0,t_0)$ there exist two points $\vec{v}_L(s_0,t_0)$

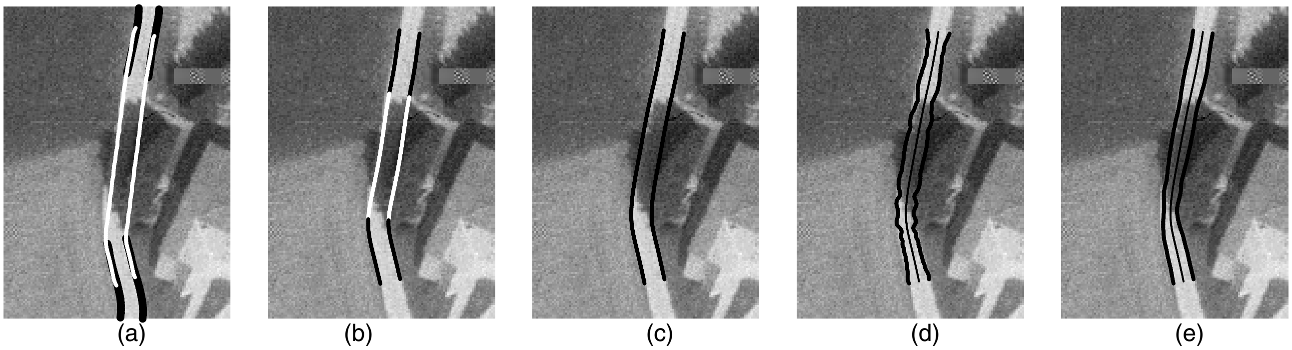
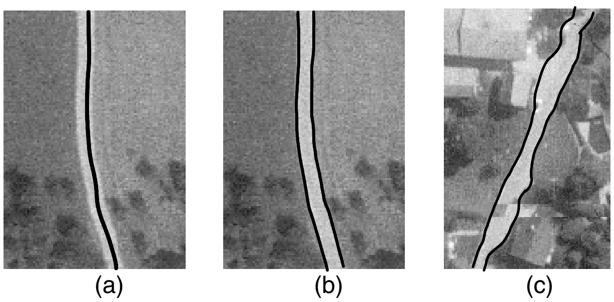
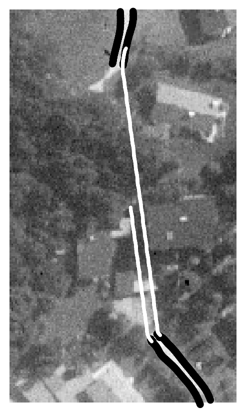


Figure 7: Optimization of ribbon snakes at a **correctly hypothesized** *non-salient* road. (a)-(c) Optimization of ziplock ribbon with constant width and fixed end points. White lines indicate passive part. (d)+(e) Optimization of the ribbon's width with fixed centerline



(a)

Figure 6: Optimization of ribbon snakes initialized at the centerlines of hypothesized roads. (a) centerlines (b) constant width ⇒ *salient road* (c) no constant width ⇒ *non-salient road* or no road at all



accepted as a road.

When all *salient roads* have been extracted, in most cases smaller or larger parts of the road-network are still missing. To find further parts of the network so-called *non-salient* roads are introduced. They are extracted as follows: The starting points are always two ends of *salient roads* for which a connection is hypothesized (cf. point number 4.(a) in Section 4). The hypotheses can be generated by the sophisticated algorithm of (Steger et al., 1997), but in many cases a connection of nearby ends of *salient roads* works as well. The two ends are connected by a ribbon snake and are optimized using the ziplock principle (Neuenschwander et al., 1995). This idea prevents that in case the two endpoints of a snake are known and one optimizes the whole snake at once the snake is stuck to disturbances in between. This is done by optimizing at first only the parts close to the ends and then propagating this information from both sides until the whole snake is optimized. For locating *non-salient roads* this helps but is not enough, and another strategy was found to be of major importance (Laptev, 1997): with the ziplock snake only the center of the road is optimized. The width is taken to be constant and equal to the average of the width of the two *salient roads*. Then in a second step the centers are fixed and only the width is optimized. *Non-salient roads* can again be distinguished from other objects by the constancy of the width (cf. Figures 7 and 8). Whereas in Figure 7 the width is relatively constant, and therefore the hypotheses of a *non-salient road* is accepted, in Figure 8 it is not.

In Figure 9 the result for a larger image combining the extraction of *salient* and *non-salient roads* is given. The approach is able to bridge not only short gaps caused by sin-

r

r

Figure 9: *Salient* and *non-salient roads*

Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Birkhäuser Verlag, Basel, Switzerland, pp. 299–310.

Baumgartner, A., Steger, C., Mayer, H. and Eckstein, W., 1997b. Multi-Resolution, Semantic Objects, and Context for Road Extraction. In: Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, Birkhäuser Verlag, Basel, Switzerland, pp. 140–156.

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

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

Eckstein, W. and Steger, C., 1996. Fusion of Digital Terrain Models and Texture for Object Extraction. In: Second International Airborne Remote Sensing Conference and Exhibition, Vol. III, pp. 1–10.

Fua, P. and Leclerc, Y., 1990. Model Driven Edge Detection. Machine Vision and Applications 3, pp. 45–56.

Grimson, E., 1990. Object Recognition by Computer. MIT Press, Cambridge, USA.

Grün, A. and Li, H., 1996. Linear Feature Extraction with LSB-Snakes from Multiple Images. In: International Archives of Photogrammetry and Remote Sensing, Vol. (31) B3/III, pp. 266–272.

Kass, M., Witkin, A. and Terzopoulos, D., 1987. Snakes: Active Contour Models. International Journal of Computer Vision 1(4), pp. 321–331.

Laptev, I., 1997. Road Extraction Based on Snakes and Sophisticated Line Extraction. Master's thesis, Computational Vision and Active Perception Lab (CVAP), Royal Institute of Technology, Stockholm, Sweden.

Lindeberg, T., 1994. Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, Boston, USA.

Matsuyama, T. and Hwang, V.-S., 1990. SIGMA - A Knowledge-Based Aerial Image Understanding System. Plenum Press, New York, USA.

Mayer, H. and Steger, C., 1996. A New Approach for Line Extraction and its Integration in a Multi-Scale, Multi-Abstraction-Level Road Extraction System. In: IAPR TC-7 Workshop: Mapping Buildings, Roads and other Man-Made Structures from Images, Oldenbourg Verlag, Vienna, Austria, pp. 331–348.

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

Merlet, N. and Zerubia, J., 1996. New Prospects in Line Detection by Dynamic Programmming. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(4), pp. 426–431.

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

Plietker, B., 1994. Semiautomatic Revision of Street Objects in ATKIS Database DLM 25/1. In: International Archives of Photogrammetry and Remote Sensing, Vol. (30) 4, pp. 311–317.

Ruskoné, R., 1996. Road Network Automatic Extraction by Local Context Interpretation: Application to the Production of Cartographic Data. Thèse de doctorat, Université de Marne-La-Vallée, Noisy-le-Grand, France.

Ruskoné, R., Guigues, L., Airault, S. and Jamet, O., 1996. Vehicle Detection on Aerial Images: A Structural Approach. In: 13th International Conference on Pattern Recognition, Vol. III, pp. 900–903.

Steger, C., 1996. Extracting Curvilinear Structures: A Differential Geometric Approach. In: Fourth European Conference on Computer Vision, Vol. I, pp. 630–641.

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

Tönjes, R., 1996. Knowledge Based Modelling of Landscapes. In: International Archives of Photogrammetry and Remote Sensing, Vol. (31) B3/III, pp. 868–873.

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

Wiedemann, C. and Mayer, H., 1996. Automatic Verification of Roads in Digital Images Using Profiles. In: Mustererkennung 1996, Springer-Verlag, Berlin, Germany, pp. 609–618.

Zlotnick, A. and Carnine, P., 1993. Finding Road Seeds in Aerial Images. Computer Vision, Graphics, and Image Processing: Image Understanding 57, pp. 243–260.