Automatic Road Extraction Based on Intersection Detection in Suburban Areas

Go Koutaki† and Keiichi Uchimura
Graduate School of Science and Technology, Kumamoto University, Kurokami, Kumamoto, JAPAN

Zhencheng Hu
Faculty of Engineering, Kumamoto University, Kurokami, Kumamoto, JAPAN

We propose an intersection model and strategy for automatic road extraction from aerial imagery. The proposed approach is able to detect typical intersections such as crossroads, T-junctions and Y-junctions based on matching the model to the image features. Compared to the traditional morphological methods, for example the combination with thinning and 8-neighbor pattern matching, our approach is less sensitive to noise and holes, and less like to produce a false match. The road network is constructed by connecting the detected intersections. The connecting hypothesis is generated and validated using the road tracking method and the road shape including the width is refined using ribbon snakes. We show the feasibility of our approach by presenting results for a suburban area, and evaluating them in comparison to the existing road map.

Journal of Imaging Science and Technology 49: 163–169 (2005)

Introduction

Shortening the cycles of updating map information is required, given especially the shift to digital maps from analog mapping. A digital topographic database is an essential part of the Geographic Information System (GIS). Especially, building the outline or road arc data is demanded in the field of Intelligent Transport System (ITS) such as car navigation, forwarding agencies, city planning, etc. Major existing digital maps are currently digitized from the topographical 1:25,000 paper-based maps. Since these contain some errors, for instance, missing road arcs or geometric spatial errors (the maximum about 17.5 m), it may be not sufficient for the specific field requiring more detailed data and higher scale. Aerial and satellite imagery are important data sources for acquiring topographic objects with high accuracy over a wide range. Since the manual extraction from the images requires enormous cost and time, automation should be employed for updating map information.

An early road extraction approach has focused on low resolution aerial images. Fischler et al.[1] proposed road detectors considering local and global criteria. The road tracing step exploits a local criterion calculated by low level processing. Steger[2] presented the method of line extraction based on differential geometry. For each pixel in the image convoluted with the Gaussian kernel, the image profile along the principal direction is examined. Line points for which the first and second derivatives of the profile vanish and exhibit a minimum respectively, are detected and connected. Mayer et al.[3] proposed the multi-resolution approach. In a coarse scale image, the lines are extracted using the line extraction algorithm of Steger. In a fine scale version, salient roads with a low variance of width are verified by ribbon snakes[4] initialized by the results of the line detection in the coarse scale analysis. Non-salient roads that include a shadow, car, or have no roadside, are bridged with so called Ziplock Snakes.[5] Wiedemann[6] developed a graph structure for the road network. A road network is accordingly constructed by calculation of the optimal path between road segments for which there is high evidence. Weight of the graph is evaluated with a fuzzy value that incorporates some criteria, i.e., length of segments or gap, straightness and curvature.

The above works have concentrated on open rural areas. The multi-resolution approach mainly depend on the result of line detection at low resolution. Most roads are distinct from background objects (mostly fields) and have the roadsides generally clearly presented. Therefore, because the line detector or ribbon snakes extract a lot of salient roads very successfully, the construction step of road network works well, even if some gaps derived from shadows or trees exist. However, roadsides in suburban areas are absent because of houses, shadows or bushes. Additionally, some house

Original manuscript received September 24, 2004

†Corresponding Author: G. Koutaki, koutaki@navi.cs.kumamoto-u.ac.jp

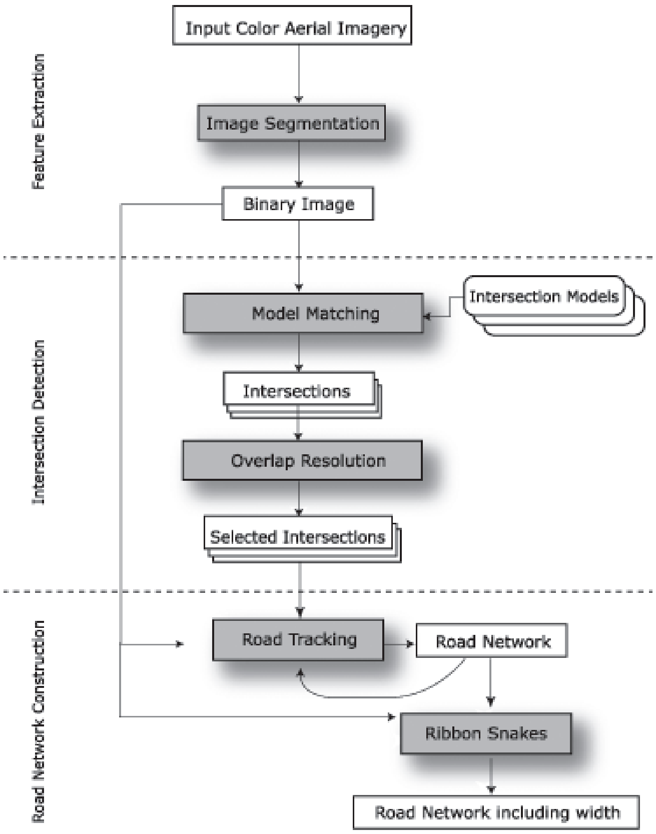


Figure 1. Proposed system

roofs that have parallel edges will be extracted as roadsides, and these will cause false connections on the road network. Price[7] attempted to detect roads in suburban and urban areas. The road network was modeled as a regular street grid pattern topology, and individual roads were approximated with constant width and length. The initial grid was given by a human operator and propagated across the scene. However the grid description only deals with intersection with mostly right angles and regular intervals. It is difficult to extract curved roads.

Our approach proposes to detect various intersections aligned with non-regular intervals such as crossroads, T-junctions and Y-junctions (three-forked roads) automatically. They are extracted by a model matching method like template matching, and redundant results are reduced using overlapped resolution. The road network is constructed by connecting the detected intersections.

A hypothesis of connecting is generated and validated using the road tracking method and the road shape including the width is refined using ribbon snakes. We show the feasibility of our approach by presenting results for a suburban area, and evaluating them in comparison to the existing road map.

Data Sources and Imaging System

Input aerial imagery is a digital ortho-image provided by Digital Earth Technology. The registration of ortho-images is performed using a 50 m mesh digital elevationmodel (DEM). Table I shows the image and camera specifications. Our approach is summarized in Fig. 1. The developed system consists of three parts

TABLE I. Image Specification

Channels	3 bands (RGB)
Scan resolution	25 μm
Camera focal length	300 mm
Flying height	About 3000 m
Image scale	About 1/20,000
Ground resolution	About 50 cm/pixel
Depths	256

which are feature extraction, intersection detection, and road network construction. Each part contains some image analysis tools. The basic idea is to connect between each intersection for road network construction. Because it is rare to isolate a road, most roads will be extracted. In contrast to various forms of representation for roads, i.e., curved road, klothoid curve, and roads with changing width, the intersections have similar simple shape. For instance, the roads entering intersection are mostly straight because of traffic safety regulations. We consider the geometric intersection model, such as three or four straight roads entering a certain point. The road intersections that fit the models are extracted. Certain features are extracted from images, and intersections and roads are extracted using the features. Those processes depend on “Road seed binary image”. In the following sections, we illustrate each step in detail.

Feature Extraction

Image segmentation is one of the fundamental steps in low-level image processing and pattern recognition. The purpose of the segmentation is to partition an image into connected regions of similar colors or features, so we can deal with meaningful regions. There are a lot of methods for image segmentation in image analysis. We apply ISODATA (Iterative Self-Organizing Data Analysis Technique) clustering method for color aerial imagery following Zhang.[8] With this approach, an optimal number of spectral clusters is automatically determined by iteratively applying split and merge operations. The original RGB images contain redundancy (the covariance is very high in each channel). The selection of input data should efficiently provide useful information and reduce the clustering cost by reduced dimensionality. We transformed the RGB color spaces into three different color spaces, i.e., luminous, greenness in RGB space, as $(G - R)/(G + R)$, and saturation from HIS color space. The greenness allows enhancing the vegetation, and the saturation is able to emphasize the shadows and house roofs with vivid color.

Figure 2(b) shows the result of the classification for input image (Fig. 2(a)). Then, the road class is selected by an operator. If a reference road arc exists, it is possible to select the road class automatically by taking the statistics of classfied pixels around the reference roads. A distribution of road classes may contain errors of non-road pixels, and they should be cut out as noise. Therefore, for each pixels of road class, we calculate the squared Mahalanobis distance d^2,

$$d^2 = \left(p - \mu_{road}\right)^t \Sigma_{road}^{-1}\left(p - \mu_{road}\right) \tag{1}$$

where, μ_{road} and Σ_{road}^{-1} road are mean vector and a covariance matrix of the road cluster respectively. If d^2

(a) an input image (b) classified 5 classes (c) road seed

Figure 2. Result of the image classfication

is more than a threshold k then the pixel is removed from the road cluster. Finally, we divide the classes into either road class or background class. We call the pixels of the road class "road seed". Figure 2(c) shows road seed corresponding to Fig. 2(a).

Intersection Detection

A road seed is a binary image, where a pixel with value of "1" denotes a probably road like ob ject. Since an extraction error arises from some pixels where building roofs or soil have similar spectral response to roads, general morphological operators, for example combinations with closing, thinning and 8-neighbor pattern matching, will not work well because of high sensitivity to noise. Therefore, stronger constraints and knowledge about intersections are required. We consider intersections to be of the following three types:

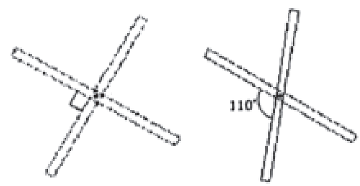
- Crossroads, which represent the intersection of two road portions.
- Three-forked roads have three road segments. Each branch has a different direction
- T-Intersection consists of one straight road and a connected branch.

We provide intersection models as templates shown in Fig. 3; they consist of two or three elongate rectangles with different or same widths with respect to each other. Each branch has 4 m, 7 m, or 13 m widths corresponding to one lane, two lanes, and three lane roads, respectively with constant length L = 30 m. Figure 4 shows models with combinations of the widths. We represent the all models as a set $\mathcal{M}$, For $M \in \mathcal{M}$, $M^{\theta}_{(x,y)}$ denotes that center of intersection is (x,y) and direction of 1st branch is $\theta (0 \le \theta < 2\pi)$.

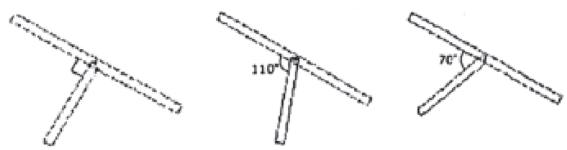
Consider matching the above models to the road seed and calculating matching values between the models and the road seed. The model is rotated and positioned over the binary image. The matching measure is defined following,

$$D\left(M^{\theta}_{(x,y)}\right) \equiv \begin{cases} \mu(S) - \mu(B) & \text{if } \min\limits_{n=1,2,\dots N} \mu(S_n) > k_1 \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

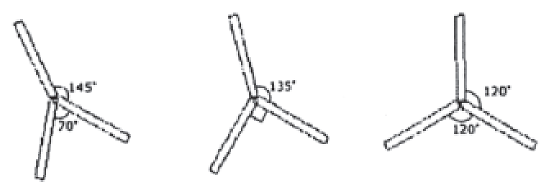
Here, S and B are inner and outer regions of the model. S_n denotes a region inside nth branch of intersection and $S = S_1 \cup S_2 \cup \cdots \cup S_N$. The constraint of Eq. (2)



(a) Crossroads



(b) T-Intersection



(c) Three-forked road

Figure 3. Types of intersections

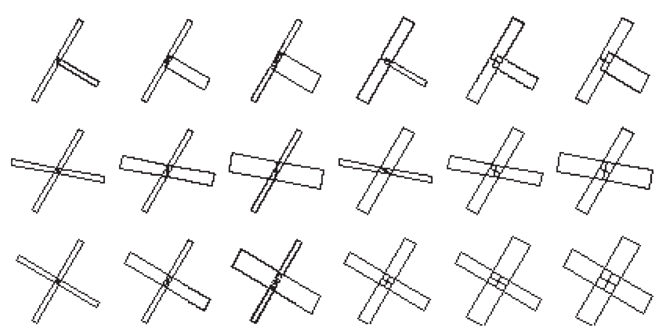


Figure 4. Combinations of some widths for crossroads.

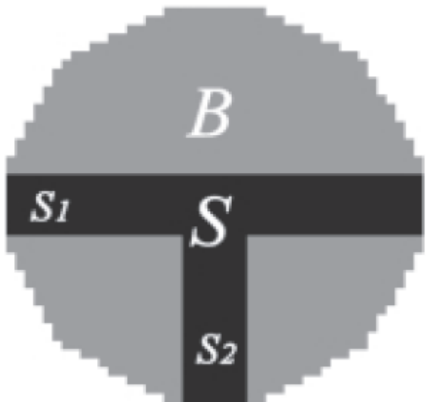
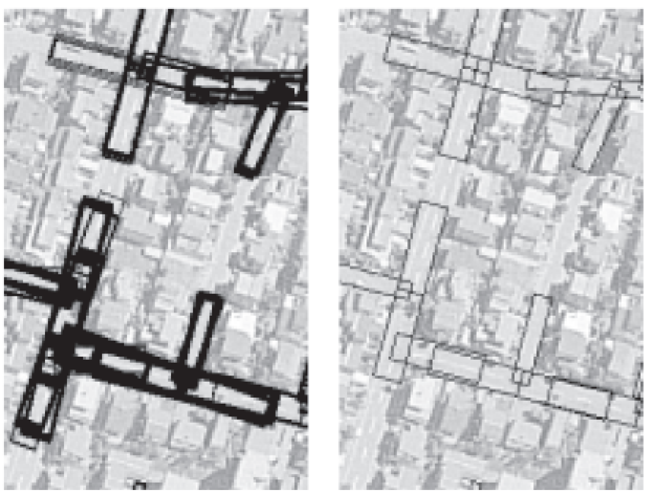


Figure 5. Inner and outer region of model



(a) Extracted intersections by thresholding (b) After overlap resolution

Figure 6. Result of Intersection detection

requires that the whole branch should have values more than threshold k_1; $\mu(\cdot)$ is defined as,

$$\mu(R) = \frac{1}{|R|}\sum_{(x,y)\in R} I(x,y). \tag{3}$$

where $I\,(x,y) = \{1, 0\}$ is the binary image of the road seed and $|R|$ denotes the number of pixels in region R. The background region B is defined subtracted from S in a circular region C (radius L) is shown Fig. 5. The matching algorithm is as follows. For each position (x,y) in input image, the maximum matching measure is calculated following,

$$D(x,y) = \max_{M\in\mathcal{M},0\le\theta<2\pi} D\left(M^{\theta}_{(x,y)}\right). \tag{4}$$

$$M^{\max}_{(x,y)} = \arg\max_{M\in\mathcal{M},0\le\theta<2\pi} D\left(M^{\theta}_{(x,y)}\right). \tag{5}$$

If $D(x,y) > k_2$, $M^{\max}_{(x,y)}$ is extracted as an intersection and put in the intersection cue P, it is possible to use "Coarse-To-Fine" matching. First, the algorithm tests the matching not by whole pixels but over the interval q_1. If $D(x,y) > k_3$ in the pixels in the neighborhood of (x,y), the matching is performed by the interval q_2. We use q_1 corresponding to 2 m, and $q_2 = 1$. Typically, a number of pixels around the center of intersection will have a high matching value, thus many intersections will be detected. Thus P contains redundant intersections. Overlap resolution is needed to remove redundant results. First, we group adjoining intersections that have similar shape and direction. Then, the intersection with the highest evaluation value in the group is represented. Overlap resolution of intersections is shown in Fig. 6.

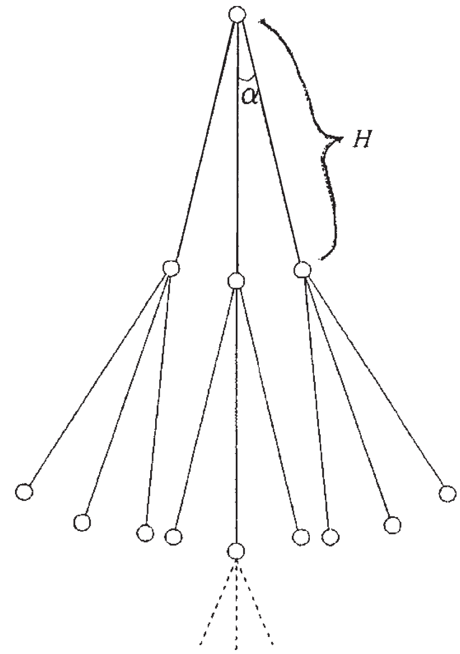


Figure 7. Road structure for tracking

Road Network Construction

The road network is constructed by connecting branches of each intersection. A connection hypothesis is performed according to directions and distance between two branches. A road tracking method[9,10] is available for this hypothesis. A structure of curve-linear road is modeled as a ternary tree (Fig. 7). We set H = 30 m and α = 5°. A starting point and direction of the tracking are given by the center point and direction at each branch of the intersection. A portion of the path is evaluated by matching an elongate rectangle template to the road seed as follow equation,

$$E(a) \equiv \mu(A_{in}) - \mu(A_{out}), a \in \mathcal{A} \tag{6}$$

where $\mathcal{A}$ is a set edge of the tree; A_{in} and A_{out} are inside and outside regions respectively around the edge, a, as shown in Fig. 8. The size of this template is the same as the branch of intersection. A road line is extended from the starting point in an iterative way by calculating maximum cost.

Using this tracking approach, a road network is generated according to the following rules.

[**R1**] if the road tracker from a branch can reach another intersection, the result is inserted as road arc in network

[**R2**] while tracking, if the tracker meets with a previously extracted road arc,

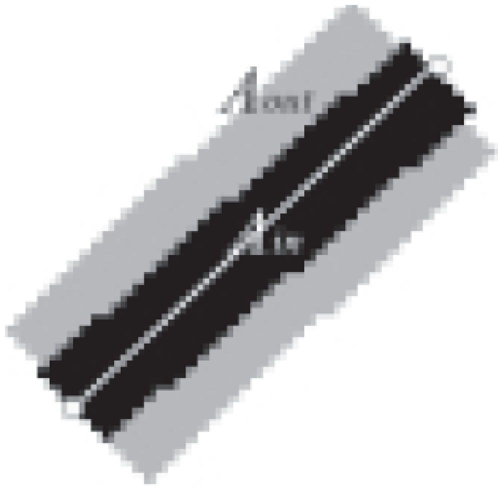


Figure 8. Line template for tracking

[R3] even if the tracker does not fulfill the conditions above, it is possible to insert a road arc in the case of 30 m or more of tracking.

Extraction of Road Shape Including Width

To refine the road shape containing the width, we implement the ribbon snakes[4] method that is a kind of active contour model.[11] The ribbon snake is initialized using the road arc derived in the road network construction step. We represent ribbon snakes as parametric curves according to the following equation.

$$v(s) = (x(s), y(s), w(s)) \tag{7}$$

where s is proportional to arc length, $x(s)$ and $y(s)$ are center coordinates of ribbon, and $w(s)$ is the curve's width. The total energy of the ribbon snakes can be de fined as:

$$E_{total} = \int E_{\text{int}}\big(v(s)\big) + E_{img}\big(v(s)\big)ds, \tag{8}$$

$$E_{\text{int}}\big(v(s)\big) = \alpha\left|\frac{\partial}{\partial_s}v(s)\right|^2 + \beta\left|\frac{\partial^2}{\partial_{s^2}}v(s)\right|^2. \tag{9}$$

Here, α and β are control parameters for the ribbon's rigidness and smoothness. Typically, E_{img} is defined by intensity or gradient of image. However, in order to embed road seed in the image energy, we define total energy as follows,

$$E_{total} = \int E_{\text{int}}\big(v(s)\big)d_s + E_{img}\big(R_{in}, R_{out}\big). \tag{10}$$

R_{in} and R_{out} are regions corresponding to inner and outer tracks of the ribbon (see Fig. 9). In order to perform the numerical optimization of the energy, Eq. (10) is discretized. E_{int} is calculated in finite difference. R_{in} and R_{out} is approximated as sequential rectangles,

$$\left\{R_{in}^1, R_{in}^2, \ldots, R_{in}^N\right\} \text{ and } \left\{R_{out}^1, R_{out}^2, \ldots, R_{out}^N\right\}$$

as shown in Fig. 9. Discretized version of E_{img} is then defined as follows,

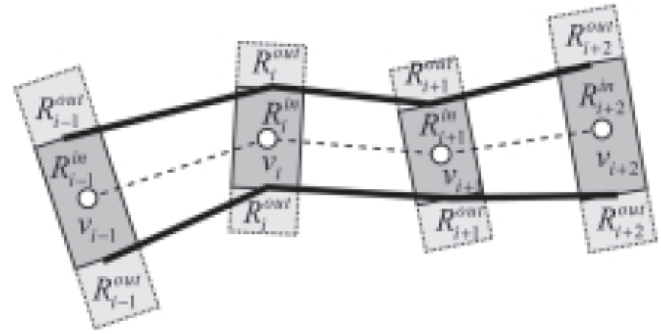


Figure 9. Region of ribbon for image energy

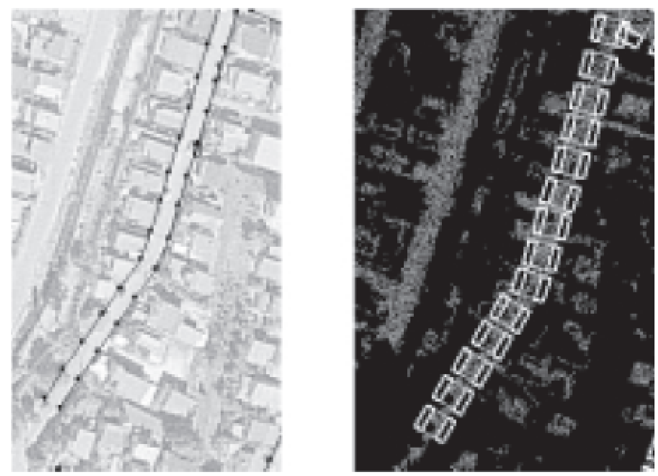


Figure 10. Result of ribbon and aspect in road seed

$$E_{img}\big(R_{in}, R_{out}\big) \approx \lambda\sum_i\Big(\mu\big(R_{in}^i\big) - \mu\big(R_{out}^i\big)\Big) \tag{11}$$

where λ is the control parameter of image energy. It is possible to optimize the total energy using dynamic programming.[12] Figure 10 shows an example of ribbon snakes based on road seed.

Results

Our approach has been tested in one rural area and two residential suburban areas. The specifications of the input image have already been stated. Size of input images is 1200×1200 pixels corresponding to 600×600 m. We used $k_1 = 0.3$, $k_2 = 0.25$, $k_3 = 0.125$, $q_1 = 4$, $q_2 = 1$, and as ribbon parameters, $\alpha = 1.0$, $\beta = 2.0$, $\lambda = 3.0$ are used. For the reasons of comparison, the test was performed with the multi-scale snakes (MS) method.[13] Figures 11 through 13 show the results of road network extraction using MS, and Figs. 14 and 15 show the results for our approach. Here, white ribbons show extracted roadsides and white dotted lines show the roads that our system could not extract.

The results have been evaluated by matching the extracted road axes to Digital Map 2500 that the Geographic Survey Institute Japan published (1/2500 scale and maximum errors is 1.75 m). The buffer method[14] is used for the quality measures. Note that we compared in respect to center lines since the reference

TABLE II. Results of MS

	Rural	Suburban 1	Suburban 2
Completeness [2%]	50.2	43.1	42.9
Correctness [%]	98.4	57.9	69.7
Quality [%]	49.4	25.0	29.9
Time [min]	20.2	40.5	46.6

TABLE III. Results of Proposed Approach

	Rural	Suburban 1	Suburban 2
Completeness [2%]	39.0	72.8	93.2
Correctness [%]	90.9	95.5	95.7
Quality [%]	35.5	70.0	89.2
Time [min]	14.4	15.2	16.5

Figure 11. Rural (MS)

Figure 12. Suburban 1 (MS)

data did not contain width. Buffer size used is 15 pixels. Tables II and III show the quality of the results, which were obtained on a PC with Pentium III processor, 2.8 GHz, and 1024 MB of RAM.

In rural areas, MS can extract road network correctly since line extraction in coarse scale works well. However in the roads adjoining soil, line and edge extraction failed because of low contrast. For the same reason, our approach could not detect the roads either. In suburban areas, MS failed considerably since line and parallel edges could not be extracted because of house roofs, junctions or bushes. Further, some building roofs are identified as road segments, which results, unfortunately, in extracting false road connections. On the other hand, our approach could extract intersections correctly and not extract buildings as road even in suburban areas. As a result, we could construct the road network with high correctness.

Because the proposed method uses only features of the color distribution of the road's surface, the roads including white lane lines and the roads where many cars exist could not be extracted. However, a lot of roads that do not have lane marks could be completed. Existing 1/25,000 scale Digital Map is likely not to contain them because of their small width. Thus, our system is useful to identify those of small width. Thus, our system is useful to identify those small width roads.

Conclusion

In this research, we first proposed a method to extract from input aerial images the intersections that have typical shape, such as T-intersections, crossroads, and three-forked junctions. They are detected using model matching, whose evaluation depends on the road seed that is selected, using ISODATA clustering. Redundant results of intersection extraction are reduced by overlap resolution. Road networks are constructed by connecting the detected intersections. A road tracking method is performed for the connection. The starting point and direction of tracking are given by the center position of the detected intersections and the orientation of each branch of the intersections. The

Figure 13. Suburban 2 (MS)

Figure 14. Rural (proposed approach)

Figure 15. Suburban 1 (proposed approach)

road shape containing the width is reconstructed using ribbon snakes that depend on the road seed. Experiments are conducted using aerial imagery with a ground resolution of 50 centimeters. Experimental results show that the method is valid in extracting roads in a suburban residential area, especially where many typical intersections exist.

Figure 16. Suburban 2 (proposed approach)

References

1. M. Fischler, J. Tenebaum and H. Wolf, Detection of roads and linear structures in low resolution aerial images using multi-source knowledge integration techniques, in *DARPAN79*, Morgan Kaufmann, Menlo Park, CA, 1979, pp. 87–100.
2. C. Steger, An unbiased detector of curvilinear structures, *IEEE Trans. Pattern Analysis Machine Intell.* **20**(2), 113–125 (1998).
3. H. Mayer, I. Laptev, A. Baumgartner, and C. Steger, Automatic road extraction based on multi-scale modeling, context, and snakes, *Int'l. Arch. Photogram. Remote Sensing* **32,** 106–113 (1997).
4. P. Fua and Y. Leclerc, Model driven edge detection, *Computer Vision* **3,** 45–56 (1990).
5. W. Neuenschwander, *Elastic Deformable Contour and Surface Models for 2-d and 3-d Image Segmentation*, Hartung-Grove Verlag, Konstant, Germany, 1996.
6. C. Wiedemann and S. Hinz, Automatic extraction and evaluation of road networks from satellite imagery, *Int'l. Arch. Photogram. Remote Sensing* **32,** 95–100 (1999).
7. K. Price, Road grid extraction and verification, in *USC Computer Vision*, Univ. Southern California, Los Angeles, CA, 1999.
8. C. Zhang, E. Baltsavias and A. Gruen, Knowledge-based image analysis for 3d road reconstruction, *Proc. 21st Asian Conference on Remote Sensing*, vol. 1, Online: Asian Association of Remote Sensing, Taipei, Taiwan, 2000, pp. 100–105.
9. H. Zhao, J. Kumagai, M. Nakagawa, and R. Shibasaki, Semi-automatic road extraction from high-resolution satellite image, *Proc. International Archives of Photogrammetry and Remote Sensing,* vol. 34, ISPRS, Graz, Austria, 2002, pp. 406–411.
10. D. Geman and B. Jedynak, An active testing model for tracking roads in satellite images, *IEEE Trans. Pattern Analysis Machine Intel.* **18,** 1–14 (1996).
11. M. Kass, A. Witkin and D. Terzopoulos, Snakes: Active contour models, computer vision, *Computer Vision* **1,** 321–331 (1988).
12. A. Amini, T. Weymouth and R. Jain, Using dynamic programming for solving variational problems in vision, *IEEE Trans. Pattern Analysis Machine Intell.* **12**(9), 855–867 (1990).
13. I. Laptev, *Road extraction based on snakes and sophisticated line extraction*, M.Sc. Thesis, Computational Vision and Active Perception Lab (CVAP), Royal Institute of Technology, Stockholm, Sweden, 1997.
14. C. Heipke, H. Mayer and C. Wiedemann, *Empirical evaluation of automatically extracted road axes*, IEEE Computer Society Press, Los Alamitos, CA, 1998, pp. 172–187.