KNOWLEDGE-BASED IMAGE ANALYSIS FOR 3D ROAD RECONSTRUCTION

Chunsun ZHANG, Emmanuel BALTSAVIAS, Armin GRUEN
Institute of Geodesy and Photogrammetry
ETH-Hoenggerberg, CH-8093 Zurich, Switzerland
Tel.: +41-1-6332931, Fax: +41-1-6331101
E-mail: {chunsun, manos, agruen}@geod.baug.ethz.ch

KEY WORDS: Road reconstruction, Context, Knowledge base, Spatial reasoning

ABSTRACT: The extraction of road networks from aerial images is one of the current challenges in digital photogrammetry and computer vision. In this paper, we present our system for 3D road network reconstruction from aerial images using knowledge-based image analysis. In contrast to other approaches, the developed system integrates knowledge processing of color image data and information from digital spatial databases, extracts and fuses multiple object cues, takes into account context information, employs existing knowledge, rules and models, and treats each road subclass accordingly. The key of the system is the use of knowledge as much as possible to increase success rate and reliability of the results, working in 2D images and 3D object space, and use of 2D and 3D interaction when needed. Another advantage of the developed system is that it can correctly and reliably handle problematic areas caused by shadows and occlusions. This work is part of a project to improve and update the 1:25,000 vector maps of Switzerland.

1. INTRODUCTION

The extraction of roads from digital images has drawn considerable attention lately. The existing approaches cover a wide variety of strategies, using different resolution aerial or satellite images. Overviews can be found in Gruen et al. (1995, 1997) and Foerstner and Pluemer (1997). Semi-automatic schemes require human interaction to provide interactively some information to control the extraction. Roads are then extracted by profile matching (Airault et al., 1996, Vosselman and de Gunst, 1997), cooperative algorithms (McKeown et al., 1988), and dynamic programming or LSB-Snakes (Gruen and Li, 1997). Automatic methods usually extract reliable hypotheses for road segments through edge and line detection and then establish connections between road segments to form road networks (Wang and Trinder, 2000). Contextual information is taken into account to guide the extraction of roads (Ruskone, 1996). Roads can be detected in multi resolution images (Baumgartner and Hinz, 2000). The existing approaches show individually that the use of road models and varying strategies for different types of scenes are promising. However, all the methods are based on relatively simplistic road models, and most of them make only insufficient use of a priori information, thus they are very sensitive to disturbances like cars, shadows or occlusions, and do not always provide good quality results. Furthermore, most approaches work in single 2D images, thus neglecting valuable information inherent in 3D processing.

In this paper, we present a knowledge-based system for automatic extraction of 3D roads from stereo aerial images which integrates knowledge processing of colour image data and existing digital spatial databases. The information of the existing road database provides a rough model of the scene. Color aerial images give the current situation of the scene, but are complex to analyze without the aid of other auxiliary data. Therefore, the information provided by the existing spatial database can help the understanding of the scene, while the images provide real data useful for improving the old road database and updating it. The system combines different
input data that provide complementary, but also redundant information about road existence, therefore, it can account for problematic areas caused by occlusions and shadows, and the success rate and the reliability of the extraction results are increased. The system has been developed within the project ATOMI (Automated reconstruction of Topographic Objects from aerial images using vectorized Map Information), in cooperation with the Swiss Federal Office of Topography (L+T), with aims to improve road centerlines from digitized 1:25,000 topographic maps by fitting them to the real landscape, improving the planimetric accuracy to 1 m and providing height information with 1–2 m accuracy. The details of ATOMI can be found in Eidenbenz et al. (2000). We currently use 1:16,000 scale color imagery, with 30 cm focal length, and 60% forward overlap, scanned with 14 microns at a Zeiss SCAI. The other input data include: a nationwide DTM (DHM25) with 25 m grid spacing and accuracy of 2-3/5-7 m in lowland/Alps, the vectorised map data (VEC25) of 1:25,000 scale, and the raster map with its 6 different layers. The VEC25 data have a RMS error of ca. 5-7.5 m and a maximum one of ca. 12.5 m, including generalization effects. They are topologically correct, but due to their partly automated extraction from maps, some errors exist. In addition, DSM data in the working area was generated from stereo images using MATCH-T of Inpho with 2 m grid spacing. This paper is organized as following. We describe the general strategy in section 2, while details for the extraction of various cues are presented in section 3. Section 4 explains cue combination for road reconstruction. In section 5 results are presented, together with the quality evaluation by comparing the results with manually measured data. The paper is concluded in section 6 with a discussion and outlook.

2. GENERAL STRATEGY FOR 3D ROAD RECONSTRUCTION

The developed system makes full use of available information about the scene and contains a set of image analysis tools. The management of different information and the selection of image analysis tools are controlled by a knowledge-based system. The general strategy is shown in Fig. 1.

![Fig. 1. Strategy of road network extraction in ATOMI.](image-url)
The initial knowledge base is established by the information extracted from the existing spatial data and road design rules. This information is formed in object-oriented multiple object layers, i.e. roads are divided into various subclasses according to road type, land cover and terrain relief. It provides a global description of road network topology, and the local geometry for a road segment. Therefore, we avoid developing a general road model; instead a specific model can be derived for each road segment. This model provides the initial 2D location of a road in the scene, as well as road attributes, such as road class, presence of road marks, and possible geometry (width, length, horizontal and vertical curvature, land cover and so on). A road segment is processed with an appropriate method corresponding to its model, certain features and cues are extracted from images, and roads are derived by a proper combination of cues. The knowledge base is then automatically updated and refined using information gained from previous extraction of roads. The processing proceeds from the easiest subclasses to the most difficult ones. Since neither 2D nor 3D procedures alone are sufficient to solve the problem of road extraction, we make the transition from 2D image space to 3D object space as early as possible, and extract the road network with the mutual interaction between features of these spaces. More details of the general strategy can be found in Zhang and Baltsavias (2000).

3. METHODS FOR CUE EXTRACTION

When a road segment from VEC25 is selected, the system focuses on the image regions around the road. The regions are defined using the position of the road segment and the maximal error of VEC25. Then a set of image processing tools is activated to extract features and cues. 3D straight line generation is a crucial component of our procedure. We are interested in 3D straight lines because the correct road sides are among them. The 3D information of straight lines is determined from the correspondences of line segments between stereo images. With color images, a multispectral image classification method is implemented to find road regions. We also exploit additional cues such as road marks to support road extraction.

3.1. 3D STRAIGHT LINE GENERATION

Due to the complexity of aerial images, different view angles and occlusions, straight line matching for 3D line generation is a difficult task in computer vision and photogrammetry. We developed a structural matching method that exploits rich line attributes and line geometrical structure information (Fig. 2). The rich line attributes include the geometrical description of the line and the photometrical information in the regions right beside the line (flanking regions). The epipolar constraint is applied to reduce the search space. The similarity measure for a line pair is first computed by comparing the line attributes. The similarity measure is used as prior information in structural matching. The locally consistent matching is achieved through structural matching with probability relaxation.

Fig. 2. Flow-chart of straight line matching.
The input images are first filtered with a Wallis filter for contrast enhancement and radiometric equalization (Baltsavias, 1991). The Canny operator is employed to extract edges and straight line fitting is applied based on the procedures developed in a previous project (AMOBE, see Henricsson (1996)). For each straight line segment, we compute the position, length, orientation, and photometric and chromatic robust statistics in the left and right flanking regions. The photometric and chromatic properties are estimated from the “L”, “a” and “b” channels after an RGB to Lab color space conversion and include the median and the scatter matrix.

With known orientation parameters, the epipolar constraint can be employed to reduce the search space. The two end points of a line segment in one image generates two epipolar lines in the other image. With the approximated height information derived from DHM25 or DSM data, an epipolar band of limited length is defined (Baltsavias, 1991). Fig. 3 illustrates this idea. Therefore, a search region is determined in the right image for each segment in the left image. Any line included in this band (even partially) is a possible candidate, if it intersects the two epipolar lines $ep$ and $eq$ (through the two line endpoints in the left image) within this band. For example in Fig. 3, lines $i$, $j$, $k$ are accepted and will be compared with line $pq$ in the left image for similarity measurement, while line $r$ is rejected because it intersects $eq$ outside $q_1$, $q_2$. The size/height of this search band is proportional to edge length and orientation difference to the epipolar lines. The comparison with each candidate edge is then made only in the common overlap length, i.e. ignoring length differences and shifts between edge segments.

![Fig. 3. The epipolar band $p_1p_2q_2q_1$ defines the search space for line $pq$.](image)

For each pair of lines that satisfies the epipolar constraints above, their rich attributes are used to compute a similarity score. Therefore, the similarity score is a weighted combination of various criteria. The detailed computation can be found in Zhang and Baltsavias (2000).

After performing the similarity measurement computation, we construct a matching pool and attach a similarity score to each candidate line pair. However, one still has problems to determine the best matches. The difficulty comes first from how to decide on a threshold and how to treat the case of occlusions or multiple solutions existence. In addition, matching using a very local comparison of line attributes does not necessarily give results consistent in a local neighbourhood. For this reason, structural matching receives more and more attention in computer vision and photogrammetry (Vosselman, 1992; Haralick and Shapiro, 1993; Christmas et al., 1995; Wilson and Hancock, 1997). We developed a method to conduct structural matching for line correspondence using probability relaxation.

The image line structure is described as a graph, where the nodes of the graph represent straight lines, and the links between lines the relations. To find a correspondence, both individual lines and the graphs should be matched. We represent the straight lines in the left image as a set $L$, $L=\{li\}$, $i=1, 2, \ldots n$, the straight lines in the right image as a set $R$, $R=\{rj\}$, $j=1, 2, \ldots m$. The
mapping from the left description to the right one is represented as $T$. Assuming the right type of mapping $T$, we seek the probability that line $l_i$ matches $r_j$, i.e. the matching problem becomes the computation of $P\{ l_i = r_j \mid T \}$.

$$P\{ l_i = r_j \mid T \} = \frac{P\{ k = r_j \} Q(k = r_j)}{\sum_{n=1}^{m} P\{ k = r_n \} Q(k = r_n)}$$

where

$$Q(k = r_j) = \prod_{k=1}^{n} \sum_{k=1}^{m} P(T(l_i, r_j; l_h, r_k) \mid l_i = r_j, l_h = r_k) P(l_h = r_k)$$

$P\{ T(l_i, r_j; l_h, r_k) \mid l_i = r_j, l_h = r_k \}$ is called compatibility function. The evaluation is derived from the geometrical relation measurement between $(l_i, l_h)$ and $(r_j, r_k)$.

The final 3D position of each edge pixel is computed from the original edge pixels and not the fitted straight lines. This is done in the overlap length of the corresponding edges. A 3D straight line is then fitted to the 3D edge pixels. Fig. 4 shows an example of line extraction and matching. The dataset is extracted from a stereo pair of the ATOMI project, the extracted lines are presented in white, and matched lines are shown in yellow.

![Fig. 4. Straight line matching.](image)

### 3.2. IMAGE CLASSIFICATION FOR ROAD REGION DETECTION

An unsupervised classification method, ISODATA (Jain and Dubes, 1988), is applied to the images to separate road regions from other objects. The algorithm automatically classifies the selected image data into desired clusters. It recursively generates a new partition by assigning
each pattern to its closest cluster center and merges and splits existing clusters or removes small or outlier clusters. The success of image classification also depends on the data used. As we are concerned with road surfaces and shadows (especially shadows on road surfaces), our purpose is to separate them from other objects in the image, and we do not pay much attention to separate other objects, for example, we do not try to separate trees and grass. The original RGB color image is transformed into different color spaces, and 3 bands are selected for image classification, they are:

- a* Band from Lab color space;
- a band calculated with R and G bands in RGB space as \( \frac{(G-R)}{(G+R)} \);
- S band from HSI color space.

With this classification, we avoid using any hard thresholds for image segmentation, and 5 classes are obtained. They correspond to road regions, green objects, shadow areas, dark roofs and red roofs. An example of classification is shown in Fig. 5. The above 5 classes are shown in white, green, black, blue and pink respectively.

3.3. DSM AND DTM ANALYSIS

As described in section 3.2, the DTM or DSM is used in straight line matching to reduce search space. We also exploit the DTM and DSM to support road extraction. DSMs have recently found many applications in digital photogrammetry, such as orthophoto generation and building extraction (Baltsavias et al. 1995). We use the DSM and DTM to verify if a 3D straight line is on the ground. Because a DSM ideally models the man-made objects as well as the terrain, subtracting the DTM from DSM results in the extraction of above ground objects, including buildings and trees. Thus, the above ground objects and ground objects are separated (see Fig. 6, ground objects are shown in blue). This information is used in our system to reason if a region is on the ground. Further, it can also be used to compensate the missing information in classification data during the spatial reasoning process. An example is shown in Fig. 7. When the system focuses on the rectangle area (left), it finds that the area is composed of a road region and shadow from classification data (middle). However, the result from subtracting DTM from DSM (right) shows that both the shadow and the road regions are on the ground, and they have the same height. Thus, the system gains a certain confidence that the rectangle area belongs to a road.

![Fig. 5. Example of classification.](image1)

![Fig. 6. Subtracting the DTM from DSM.](image2)
3.4. ROAD MARK AND ROAD ZEBRA EXTRACTION

Road marks and zebra crossings are good indications of the existence of roads. They are generally found on main roads and roads in urban areas. Both of them have distinct color (usually white or yellow). In high resolution images, such as the ones used in our project, road marks are white thin lines with a certain width. The zebra crossings are shown as yellow strips. As far as road extraction is concerned, road marks give the road direction and even the road centerline, while the zebra crossings define the local road width. Thus, they can be used to guide the road extraction process or verify the extraction results. In addition, in many cases the correct road centerlines can be even derived directly from present road marks and/or zebra crossings. This is especially useful when the roadsides are occluded or not well-defined, such as in cities or city centers. In the following, we describe our developed methods to extract 3D road marks, and the zebra information: the zebra center, the short axis (local road direction), and the long axis (local road width). Note that with the existing knowledge such as DTM, VEC25, and derived information from image processing like DSM and image classification data, the processing can be focused on the road surface within the VEC25 error buffer.

Since road marks are white, the image is first segmented using color information. The road marks are then extracted using an image line model and the geometrical description of each road mark is obtained. The shape of a image line can be presented as a second order polynomial (Haralick et al. 1983, Busch 1994); it is fitted to the grey values \( G(x, y) \) as a function of the pixel’s row and column coordinates \((x, y)\). The line local direction is determined when the second directional derivative of \( G(x, y) \) becomes maximal. This can be done by computing a Hessian matrix:

\[
\begin{pmatrix}
\frac{\partial^2 G}{\partial x^2} & \frac{\partial^2 G}{\partial x \partial y} \\
\frac{\partial^2 G}{\partial x \partial y} & \frac{\partial^2 G}{\partial y^2}
\end{pmatrix}
\]

Then, the line local direction is given by

\[
\alpha = \frac{1}{2} \arctan \left( \frac{\frac{\partial^2 G}{\partial x \partial y}}{\frac{\partial^2 G}{\partial y^2} - \frac{\partial^2 G}{\partial x^2}} \right)
\]
In order to find the precise location of the line point, we get the profile in the direction perpendicular to \( \alpha \), and the profile is described as a parabola \( f(r) = a + br + cr^2 \). When \( f'(r) = 0 \), we obtain the position of the line point by \( r_0 = \frac{b}{2c} \); \( b \) and \( c \) are determined by

\[
\begin{align*}
    b &= G_{\alpha x} = \frac{\partial G}{\partial x} \sin \alpha + \frac{\partial G}{\partial y} \cos \alpha \\
    c &= \frac{1}{2} G_{\alpha x} = \left( \frac{\partial^2 G}{\partial x^2} \sin^2 \alpha + \frac{\partial^2 G}{\partial x \partial y} \sin \alpha \cos \alpha + \frac{\partial^2 G}{\partial y^2} \cos^2 \alpha \right) / 2
\end{align*}
\]

where \( G_{\alpha x}, G_{\alpha y} \) are first and second directional derivatives.

With the detected line points, those with similar direction and second directional derivative are linked. Straight lines are obtained by least squares fitting. The 3D lines are generated by our developed structural matching method. The 3D lines are then evaluated using knowledge. Only those on the ground (as defined by DSM-DTM), belonging to road region (as determined by the classification) and in the buffer defined by VEC25 are kept as detected road marks.

Zebra crossings are composed of several thin lines. Using color information, the image is first segmented. Morphological closing is applied to bridge the gaps between lines. We then obtain several clusters by connected labeling. Only the clusters with a certain size are kept, while the small ones are discarded. Then, the shape of the cluster is analyzed. The rectangle-like clusters are selected as zebra crossings. The center, the short and long axes of the detected zebra crossings are computed using spatial moments.

Fig. 8 is an example of road mark and zebra crossing extraction. The extracted road marks are shown in red lines superimposed on the image. The center, the short and long axes of the detected zebra crossings are also presented.

**Fig. 8. Example for road mark and zebra crossing extraction.**

4. KNOWLEDGE-BASED ROAD RECONSTRUCTION: CUE COMBINATION

With the information from existing spatial data and image processing, the knowledge base is established according to the general strategy. Note that one of the important characteristics of the built knowledge base is that all information in it is spatially related, and relations between 2D edges and their corresponding 3D straight lines are kept. The system then extracts roads by finding 3D parallel lines that belong to a road and link them in sequence. In case of shadows, occlusions caused by trees and buildings, spatial reasoning is applied using the knowledge base. Thus, also the cases when only one or no side is visible can be handled by the system. The main
procedures are shown in Fig. 9. The key is the use of knowledge and image context as much as possible, working in 2D images and 3D object spaces, use of 2D and 3D interaction when needed, and reasoning the problematic areas. The details of the implementation can be found in Zhang (2000).

The system checks extracted lines to find 3D parallel lines. Only lines located in the buffer defined by VEC25, having a similar orientation to VEC25 segments and a certain slope are further processed. Since roads are on the ground, lines above ground are removed by checking with the DHM25. By checking with the image classification results, a relation with the road region (in, outside, at the border) is attached to each line. Two lines are considered as parallel if they have similar orientation in 3D space. The lines of a pair must overlap in direction along the lines, and the distance between them must be within a certain range. The minimum and maximum distances depend on the road class defined in VEC25. The found 3D parallel lines are projected onto the images and evaluated using multiple knowledge. The region between the projected lines must belong to the class road as determined by the image classification. If road marks are presented on this road, the extracted road marks are used to confirm that the line pair corresponds to correct road sides.

Fig. 9. Flow-chart of image analysis procedures for road extraction.
For each found 3D parallels lines, the system tries to extend them as much as possible using 3D and 2D line information. And for each extension, spatial reasoning (see below) is applied to guarantee that the extension area is road region and extended lines are road sides. An example is shown in Fig. 10. We use upper case to represent 3D line segments, and lower case for 2D line segments. Suppose that $ef$ is a 2D straight line, and $eg$ is occluded in the other image. Then, we can only generate a 3D straight line for segment $gf$, which is shown as $AB$ in the figure. $CD$ is another 3D straight line parallel to $AB$. Region 1 is the overlap area of $AB$ and $CD$, and the system finds that this area belongs to road. Thus, the overlap area is extended to $C'$ as long as region 2 is also road.

![Fig. 10. Extension of parallel lines.](image)

Once the system finds with a certain confidence that $AB$ is a correct road side, region 3 will be checked even if there is no 3D information for segment $eg$. If region 3 is also road, the overlap area is further extended to $C''$. The 3D information of point $e$ is obtained by the intersection of image ray passing through $e$, and the 3D line $AB$. This procedure is conducted for 3D segments $AB$ and $CD$, and their corresponding 2D straight lines in both images, until there is no possible extension found in 3D and 2D any more.

The system also checks each individual 3D straight line, if this line does not belong to any 3D parallel pair. When one of the sides of the line is road, the system hypothesizes its opposite side using the width from already found 3D parallel lines. Again the hypothesized area is checked using accumulated knowledge. Compared with the visible 3D parallel lines, the system assigns a low reliability to the hypothesized parallel. In case the single visible line is close to a line of a found 3D parallel pair and has similar orientation, its reliability is increased.

All found parallel lines are considered as Possible Road Sides that are Parallel (PRSP). They compose a weighted graph. The nodes of the graph are PRSPs, the arcs of the graph are the relations between PRSPs. Note that in occlusion areas, the arcs also represent the missing parts of a road between a pair of PRSPs. The width of two PRSPs should be similar. If there is no gap between two PRSPs, i.e. one PRSP shares points with another, and the linking angles between them in 3D space comply with VEC25, they are connected directly. In case of an existing gap, the gap area is checked. This is called spatial reasoning in our work. If the gap is not too long, and

- within the gap is a road region, or
- within the gap is a shadow, or shadow mixed with road region, or
- the gap is caused by tree occlusion (determined from the image classification results and the data of DSM minus DTM), or
- within the gap is terrain as determined by the DSM, or
- road marks are extracted within the gap

and the connecting angles between PRSPs and gap comply with VEC25, a link is made for the two PRSPs.

The road is then found by searching the graph using the best-first method. The method maximizes a merit function. The function is defined as a weighted summation of parallel
measurement for PRSP, image information measurement for PRSP and gap, width similarity measurement between PRSPs, and geometrical linking measurement between PRSPs.

For main roads, on which the system knows that road marks are present, the system also extracts roads using detected road marks and zebra crossings. The road marks are linked using a similar method as described in the previous paragraph. This procedure increases the effectiveness and reliability of our system. In complex areas, such as in city centers, the road sides are generally occluded very much, and sometimes it is impossible to identify them. However, the road centerlines are successfully extracted by the system using road marks. In rural and suburban areas, the extracted road using road marks is used by the system to verify the extraction results using 3D parallel lines.

5. RESULTS

The described system is implemented as a standalone software package with a graphic user interface running on SGI platforms. Fig. 11 shows a road image in a rural area where the road is occluded by tree shadows. The outdated road from VEC25 is presented in yellow, the extracted straight lines and 3D parallel lines are shown in white and orange respectively. The red line is the extracted road centerline. In Fig. 12 we show the extracted road in a suburban area. Fig. 13 is an example where the road is occluded by buildings and shadows. Fig. 14 shows a complex scene where the roadsides are not well defined, but the correct road centerline can be reliably extracted through road mark extraction.

In order to evaluate the extraction results, a method is developed to compare the extracted road with reference data. The reference data is measured by L+T in a stereoplotter. The method matches the two datasets and computes the coordinate differences. The comparison results for the Fig. 12 and Fig. 14, where reference data were available, are listed in Table 1.

Fig. 11. Road extraction in a rural area with occlusions.        Fig. 12. Road extraction in a suburban area.
6. DISCUSSION AND CONCLUSION

We presented a knowledge-based image analysis system for road extraction from stereo aerial color images. The system has several advantages over other approaches. It uses existing knowledge, image context, rules and models to restrict the search space, treats each road subclass differently, checks the plausibility of multiple possible hypotheses, therefore provides reliable results. The system contains a set of image processing tools to extract various cues about road existence, and fuses multiple cues and existing information sources. This fusion provides not only complementary information, but also redundant one to account for errors and incomplete results. Working on stereo images, the system makes an early transition from 2D image space to 3D object space. The road hypothesis is generated directly in 3D object space. This not only enables us to apply more geometric criteria to create hypotheses, but also largely reduces the search space, and speeds up the process. The hypotheses are evaluated in images.
using accumulated knowledge information. Whenever 3D features are incomplete or entirely missing, 2D information from stereo images is used to infer the missing features. By incorporating multiple knowledge, the problematic areas caused by shadows, occlusions etc. can be handled. We also present in this paper the results of road extraction in different landscapes and quantitative analysis using accurate reference data. The comparison of the reconstructed roads with such data showed that the results fulfill the specified accuracy requirements. Our future work will concentrate on road extraction in cities and city centers. Another important issue is the reliability indicator of the extraction results. A method will be developed to quantify reliability using accumulated knowledge information.

ACKNOWLEDGEMENTS

We acknowledge the financial support of this work and of the project ATOMI by the Swiss Federal Office of Topography, Bern.

REFERENCES


