Extracting straight road structure in urban environments using IKONOS satellite imagery

Donna Haverkamp
Space Imaging
Department of Research and Product Development
12076 Grant Street
Thornton, Colorado 80241
E-mail: dhaverkamp@spaceimaging.com

Abstract. We discuss a fully automatic technique for extracting roads in urban environments. The method has its bases in a vegetation mask derived from multispectral IKONOS data and in texture derived from panchromatic IKONOS data. These two techniques together are used to distinguish road pixels. We then move from individual pixels to an object-based representation that allows reasoning on a higher level. Recognition of individual segments and intersections and the relationships among them are used to determine underlying road structure and to then logically hypothesize the existence of additional road network components. We show results on an image of San Diego, California. The object-based processing component may be adapted to utilize other basis techniques as well, and could be used to build a road network in any scene having a straight-line structured topology. [DOI: 10.1117/1.1496785]

Subject terms: automated processing; image processing; object-based processing; IKONOS; high-resolution imagery; satellite imaging.

Paper REM-07 received Feb. 26, 2002; revised manuscript received Mar. 8, 2002; accepted for publication Apr. 10, 2002.

1 Introduction

With the advent of the satellite IKONOS comes the opportunity to utilize high-resolution, commercial satellite data for the purposes of large-scale feature extraction and analysis. The relative spectral response curves for the IKONOS bands are shown in Fig. 1. With its 1-m resolution panchromatic band and its highly coregistered 4-m resolution multispectral bands, IKONOS offers a unique opportunity to combine panchromatic and multispectral data to derive a greater amount of information for use in automated feature extraction methods. For additional detail concerning the properties of IKONOS satellite imagery, refer to Ref. 1.

The extraction of roads and, subsequently, road networks from imagery is a difficult and complex task. Roads can be either dark or bright in comparison to their background, either narrow or wide in size, exhibiting either high or low contrast, highly variable surfaces or very smooth surfaces. We realize that even with the additional information offered by the multiple bands of IKONOS, there will still be no single method that will be successful in extracting all types of roads from this imagery. Instead, we believe that a number of techniques developed for different classes of roads will lead to a many-branched solution for road extraction that will be effective for a wide range of road types. Our recent focus within this overall solution framework has been on the extraction of primarily linear streets in urban environments. These streets are characterized by recognizable spatial relationships between neighboring intersections, and our method takes advantage of these relationships to build a vectorized road network.

2 Background

Fully automatic, highly accurate extraction of road networks has long been an elusive goal. As a compromise to full automation, many procedures begin with two or more user-selected road points and track the path of the road among these points;[2,3] others require starting vectors and track the roads from there;[4–6] and yet others use a series of user-supplied seed points and derive roads from snake-based approaches[7,8] or dynamic programming methods.[9] On the fully automated end of the scale are methods that derive such seed points automatically.[10,11] Other automated methods use information from low-level processing to isolate linear structures,[12] or they combine the results of different low-level processes at different resolutions to extract the roads.[13,14] There are also artificial intelligence-based approaches that use similar low-level processes followed by symbolic or meta-level reasoning about the extracted objects.[15–17]

Our approach is similar to some of the latter processes in that we follow low-level, pixel-based techniques with higher-level reasoning to extract networks of roads.

3 Extracting Roads from Imagery of Urban Environments

As stated before, our recent focus has been in the extraction of roads from urban environments. As a basis for this extraction, we utilize low-level pixel techniques that are not specific to urban environments and which can be applied to data with either a well-structured underlying network or a seemingly random one.

When it comes to identifying road pixels, we can find them both by looking for them directly or, conversely, by ruling out their existence. Each of our basis techniques utilizes one of these two approaches. We first compute a vegetation mask from the multispectral data. This tells us where there are no roads, and offers the benefit both of reduced false positives and reduced computational time in

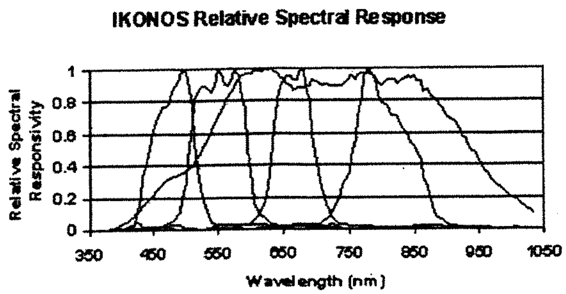


Fig. 1 IKONOS spectral response curves.

the subsequent (and more compute-intensive) road-pixel-finding technique. We apply this vegetation mask to the panchromatic data and look specifically for road pixels in the remainder of the image.

Given a set of pixels for the image that is considered to be "road," we can group them into objects that correspond to real-world road components such as road segments and intersections. From these components, we determine a local road network pattern (if one exists) and use it along with the existing road network components to hypothesize and verify the presence of missing portions of the road network.

3.1 *Vegetation Masking*

Masking the vegetation in an image prior to further processing is important for the reduction of both false alarms and processing time. The red and near-infrared bands are instrumental in the identification of vegetation in imagery. Our procedure for deriving the vegetation mask uses the spectral average, the difference between the red and green bands, and the ratio of the near-infrared to the spectral average as discriminators for vegetation. All three of these measures are taken at the pixel level and compared to a series of corresponding thresholds to determine whether or not a pixel is composed of vegetation. The binary mask resulting from this operation is resampled using bilinear interpolation to match the resolution of the panchromatic image. This enables direct correspondence between the mask and the panchromatic pixels in subsequent processing.

3.2 *Panchromatic Texture Measurement*

A road should be relatively straight when considered at any local point, as well as being smooth and displaying a low variance along the trajectory of the road. In response to these characteristics, we developed a matched filter to determine the existence and nature of a potential road pixel.

In this approach, a rectangular filter is extended from and rotated a full 360 deg about each potential road pixel of the image (those pixels that remain after vegetation masking). At discrete intervals about the pixel, the variance over the rectangular window is calculated. For most areas, a filter size of 3×20 pixels and 16 to 18 discrete points is appropriate. The size of this rectangular filter may be adjusted based on observation of the input image, as can the number of intervals for which to calculate the variance. When the values of the variances at the discrete locations are graphed in a linear fashion, locations at which the filter lay along a section of road will be identifiable as low points or "valleys" in the texture graph (see Fig. 2 for illustration). If these valleys are significant enough in terms of depth and width, we assume that the pixel under scrutiny is part of a road. For each of these road pixels, we save two pieces of information. One is the number of valleys in its texture mapping. This indicates the nature of the pixel: whether it is part of a two-directional road segment or a three- or four-directional road intersection. The other piece of information is the set of discrete locations at which these valleys occur; this gives directional information useful for the grouping of road pixels into objects for higher-level reasoning.

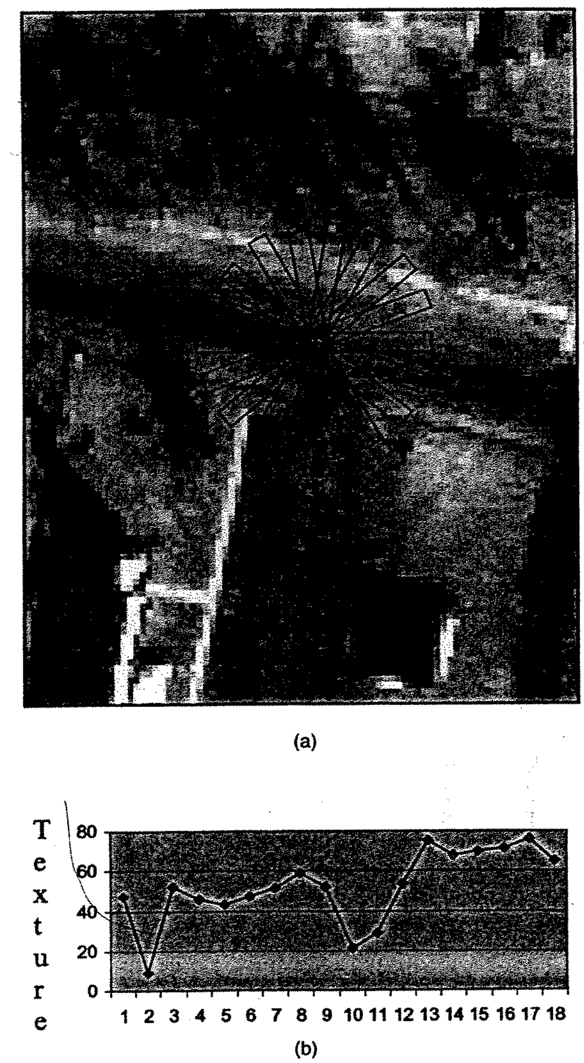
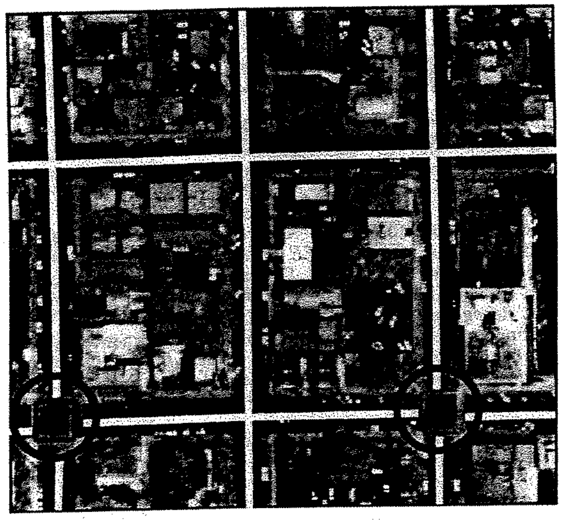


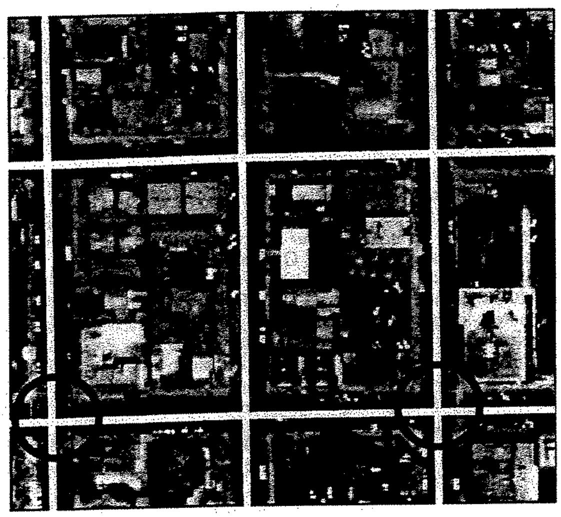
Fig. 2 Panchromatic texture measurement. The illustration in (a) shows how the texture filter is applied to the image, while (b) shows a graph of the corresponding texture values, taken at 18 discrete locations over 360 deg about the pixel.

3.3 *Recognizing and Utilizing Underlying Road Network Structure*

Many cities, particularly in the United States, contain road structures that follow a set, repetitive, and sometimes even recursive pattern. In any given local area, this pattern can be defined by the spatial relationship between a single in-



(a)



(b)

Fig. 3 Knowledge concerning the underlying road network structure enables us to identify missing components; (a) shows how undetected road pixels affect the initial road extraction, while (b) illustrates that the underlying road network structure can be used to hypothesize and verify missing information.

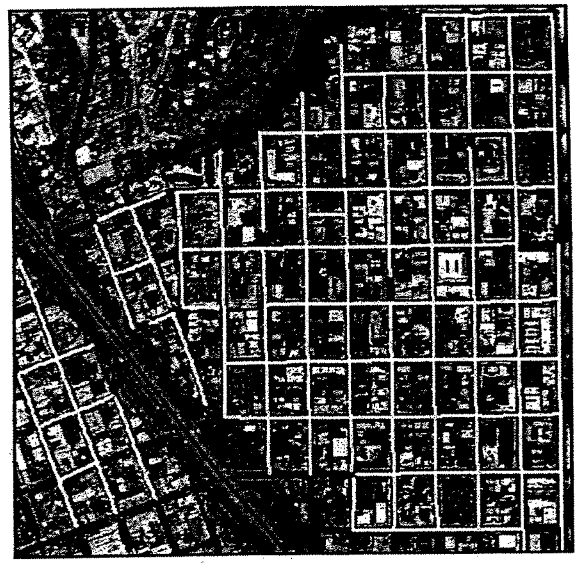


Fig. 4 Example results: a panchromatic image of San Diego, California, with the extracted road network overlaid on the data in white. Image size is approximately 1100×1100 pixels.

tersection and its n neighbors, if the relationship of that intersection to its neighbors is characteristic of the local neighborhood.

To get to the point where we can take advantage of these spatial relationships, we must first derive and recognize them. The product of the low-level pixel processing is a determination of the nature of the local road component (two-way segment versus three- or four-way intersection) and the direction of the road at that pixel. Given these two pieces of information, we can group pixels together into objects that potentially correspond to intersections and the road segments between those intersections. As objects, we can analyze them to see which ones will be useful for extracting the underlying road network. Here, the geometric characteristics of the objects become important. For example, the only objects that we can be certain are road segments are those which are straight and which are fairly compact in structure. We use size, eccentricity, and length to determine which two-directional (road segment) objects will contribute to our estimate of the underlying road network structure. Likewise, to be sure that an intersection is part of our road network, there must be a number of incoming/outgoing road segments arranged in some spatial pattern about it. Once we have established these "recognizable" components, we can use them to determine the underlying road network structure.

To derive the road network structure, the spatial relationships between neighboring intersection objects are used. Through utilization of the connecting road segment objects, we can derive the following spatial description for an intersection A:

has-neighbor A N_1	has-neighbor A N_2 ...	has-neighbor A N_x
distance d_1	distance d_2	distance d_x
angle θ_1	angle θ_2	angle θ_x

where intersection N_x represents a neighbor of intersection A, d_x is the Euclidean distance from A to N_x, θ_x is the angle measured between the line from A to N_x and the horizontal, and $x=\{1,2,3,4\}$ for a valid intersection object. If there are enough intersection objects in a local area with a spatial description similar to A, then that spatial description becomes the estimate of the underlying road network structure.

The recognition of these spatial relationships not only allows us to derive a good straight-line approximation of the road centerlines in the image, it also benefits us by providing a framework within which to propose hypotheses concerning the existence of missing road network structure. If we have identified the spatial pattern of the underlying road network, we can select intersection objects that fit this spatial pattern and then hypothesize the existence of "missing" neighbors. As an example, consider Fig. 3, a zoom showing a small section of our results. In Fig. 3(a), notice how our network is interrupted in the vicinity of two of the road intersection areas. Because of the crosswalks at those intersections, our variance-based texture measure detects no road pixels. However, because there is enough surrounding structure to identify the underlying road network, the system is able to hypothesize the existence of intersections in those areas and verify them using the patterns of incoming road segments as proof, adding those intersections to the symbolic road network representation, as shown in Fig. 3(b).

4 Results

Current efforts are under way to characterize performance in a number of urban environments, using an array of examples from not only the United States, but from nations and continents on various parts of the globe. We show the results of our road extraction on a piece of imagery taken of the San Diego, California area (see Fig. 4). As you can see, this road extraction technique works quite well in this class of urban environment. For this particular scene, 73 intersections were initially identified, and 14 more were hypothesized and verified using the road network estimate and existing objects. The entire process ran in approximately 1 min for the 1100×1100 scene shown in Fig. 3. In comparison to hand-digitized roads of the same area used as ground truth, the automated method extracted 82.9% of the streets with an rms error of 1.0 pixel and introduced one false alarm. Much of the missing structure is in the lower left corner of the image, where there are many cars. This high concentration of vehicles is detrimental to the texture measure used in low-level processing, and its effects naturally carry over to the object-level processing. The addition of complementary basis techniques to the system in the future will assist in overcoming problems such as this.

5 Conclusions and Recommendations

The road extraction process explained here works well in urban areas exhibiting a straight-line, gridded road network structure. Current efforts are underway to test the method on a wide variety of images to determine precisely the class or classes of road for which it works well, to determine what portion(s) of our many-branched solution it will satisfy.

There are certain weaknesses in the process due to the low-level basis technique used to classify the pixels as road or nonroad. That technique requires predetermination of road width, and it is tuned for the detection of roads having a specific level of contrast and a low along-road variance. Other basis techniques must be integrated to perform the pixel classification to make the higher-level algorithm suitable for extracting road networks for a wider range of road classes.

References

1. G. Dial, L. Gibson, and R. Poulsen, "IKONOS satellite imagery and its use in automated road extraction," from *Automatic Extraction of Man-Made Objects from Aerial and Space Images (III), Proc. Centro Stefano Franscini Ascona*, E. Baltsavias, A. Gruen, and L. Gool, Eds., pp. 357–367, Swets and Zeitlinger, Lisse, Netherlands (2001).
2. A. Gruen, P. Agouris, and H. Li, "Linear feature extraction with dynamic programming and globally enforced least squares matching," from *Automatic Extraction of Man-Made Objects from Aerial and Space Images, Proc. Centro Stefano Franscini Ascona*, A. Gruen, O. Kuebler, and P. Agouris, Eds., pp. 83–94, Birkhauser Verlag, Basel, Switzerland (1995).
3. J. Hu, B. Sakoda, and T. Pavlidis, "Interactive road finding for aerial images," *IEEE Workshop on Applications of Computer Vision*, pp. 56–63 (1992).
4. G. Vosselman and J. de Knecht, "Road tracing by profile matching and Kalman filtering," *Automatic Extraction of Man-Made Objects from Aerial and Space Images, Proc. Centro Stefano Franscini Ascona*, A. Gruen, O. Kuebler, and P. Agouris, Eds., pp. 265–274, Birkhauser Verlag, Basel, Switzerland (1995).
5. D. Geman and B. Jedynak, "An active testing model for tracking roads in satellite images," *IEEE TPAMI* **18**(1), 1–14 (1996).
6. D. M. McKeown, Jr. and J. L. Denlinger, "Cooperative methods for road tracking in aerial imagery," Proc. *IEEE Conf. Computer Vision Patt. Recog.*, pp. 662–672, Ann Arbor, Michigan (1988).
7. J. Trinder and H. Li, "Semi-automatic feature extraction by snakes," *Automatic Extraction of Man-Made Objects from Aerial and Space Images, Proc. Centro Stefano Franscini Ascona*, A. Gruen, O. Kuebler, and P. Agouris, Eds., pp. 95–104, Birkhauser Verlag, Basel, Switzerland (1995).
8. A. Gruen and H. Li, "Linear feature extraction with 3DLSB-snakes," *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Proc. Centro Stefano Franscini Ascona*, A. Gruen, E. P. Baltsavias, and O. Henricsson, Eds., pp. 287–298, Birkhauser Verlag, Basel, Switzerland (1997).
9. L. Iverson, "Dynamic programming delineation," *ARPA Image Understanding Workshop*, pp. 951–956 (May 1997).
10. M. Barzohar, M. Cohen, I. Ziskind, and D. B. Cooper, "Fast, robust tracking of curvy partially occluded roads in clutter in aerial images," *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Proc. Centro Stefano Franscini Ascona*, A. Gruen, E. P. Baltsavias, and O. Henricsson, Eds., pp. 277–286, Birkhauser Verlag, Basel, Switzerland (1997).
11. Z. Aviad and P. D. Carnine, Jr., "Road finding for road-network extraction," *Proc. IEEE Conf. Computer Vision Patt. Recog.*, pp. 814–819 (1988).
12. C. Steger, C. Glock, W. Eckstein, H. Mayer, and B. Radig, "Model-based road extraction from images," *Automatic Extraction of Man-Made Objects from Aerial and Space Images, Proc. Centro Stefano Franscini Ascona*, A. Gruen, O. Kuebler, and P. Agouris, Eds., pp. 275–284, Birkhauser Verlag (1995).
13. A. Baumgartner, W. Eckstein, H. Mayer, C. Heipke, and H. Ebner, "Context-supported road extraction," *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Proc. Centro Stefano Franscini Ascona*, A. Gruen, E. P. Baltsavias, and O. Henricsson, Eds., pp. 299–308, Birkhauser Verlag, Basel, Switzerland (1997).
14. A. Baumgartner, C. Steger, H. Mayer, W. Eckstein, and H. Ebner, "Automatic road extraction based on multi-scale, grouping, and context," *Photogramm. Eng. Remote Sens.* **65**(7), 777–785 (1999).
15. J. C. Trinder, Y. Wang, A. Sowmya, and M. Palhang, "Artificial intelligence in 3-D feature extraction," from *Automatic Extraction of Man-Made Objects from Aerial and Space Images (II), Proc. Centro Stefano Franscini Ascona*, A. Gruen, E. P. Baltsavias, and O. Henricsson, Eds., pp. 257–266, Birkhauser Verlag, Basel, Switzerland (1997).
16. R. Tönjes and S. Growe, "Knowledge based road extraction from multisensor imagery," *ISPRS Symp. Object Recog. Scene Classification from Multispectral and Multisensor Pixels*, 6–10 July 1998, Columbus, OH (1998).
17. F. Wang and R. Newkirk, "A knowledge-based system for highway network extraction," *IEEE TGARS* **26**(5), 525–531 (1988).

Donna Haverkamp received the PhD degree in electrical engineering from the University of Kansas, Lawrence, in 1997. During her graduate studies, she was supported by a NASA Fellowship and worked with imagery of polar areas for applications in climate change. After graduation, she took a position with Harris Corporation in Melbourne, Florida. At Harris, she participated in research toward the development of 3-D site models and later served as principal investigator on a contract under DARPA's Dynamic DataBase (DDB) program. She is currently with the Research and Product Development Department of Space Imaging in Thornton, Colorado, working with applications and research in image enhancement and feature extraction technologies. Her research interests are in image processing, computer vision, and artificial intelligence. She is a member of Tau Beta Pi, Eta Kappa Nu, IEEE Computer Society, and AAAI.