Automated Road Extraction from High Resolution Multispectral Imagery

Pete Doucette, Peggy Agouris, and Anthony Stefanidis

Abstract

This work presents a novel methodology for fully automated road centerline extraction that exploits spectral content from high resolution multispectral images. Preliminary detection of candidate road centerline components is performed with Anti-parallel-edge Centerline Extraction (ACE). *This is followed by constructing a road vector topology with a fuzzy grouping model that links nodes from a self-organized mapping of the ACE components. Following topology construction, a* Self-Supervised Road Classification (SSRC) *feedback loop is implemented to automate the process of training sample selection and refinement for a road class, as well as deriving practical spectral definitions for non-road classes. SSRC demonstrates a potential to provide dramatic improvement in road extraction results by exploiting spectral content. Road centerline extraction results are presented for three 1 m color-infrared suburban scenes which show significant improvement following SSRC.*

Introduction

The extraction of geospatial features from remotely sensed imagery remains the primary means by which to create or update geospatial databases. Presently, most feature extraction is done manually by humans in typical photogrammetric production settings. The growing availability of high-resolution imagery drives the user demand for correspondingly high-resolution feature data. The anticipated volume of imagery collected by next generation space-borne sensors is expected to overwhelm the ability of human analysts to perform feature extraction in a timely manner. Hence the motivation for automated feature extraction algorithms.

Road networks represent a vital component of geospatial data sets in high demand, and thus contribute significantly to extraction labor costs. Efforts to automate (to some extent) the task of road extraction have been the subject of considerable research activity over the last two decades. When extracting roads from high-resolution imagery (e.g., GSD $\leq$ 1.0 m/pixel), automated extraction techniques have traditionally relied on edge and textural analysis applied to single layer, e.g., panchromatic images. Since multispectral imagery has only recently become widely available at high spatial resolutions, modeling spectral content had traditionally received limited consideration for road extraction algorithms. Given the trend of technological advances in multispectral sensor design, the availability of space-borne multispectral imagery on the order of 1.0 m/pixel is presumably inevitable. At this spatial resolution, spectral content takes on new meaning in terms of detection potential for road features. The objective of this work is to develop a methodology that can meaningfully exploit spectral content toward fully automated road extraction.

Approach

In this paper, we present an approach toward fully automated road extraction from high resolution multispectral imagery. It is a bottom-up approach that integrates image gradient analysis, road topology analysis, and spectral classification. The overview of the approach is illustrated in Figure 1. The first step consists of identifying candidate road centerline pixels by performing anti-parallel edge centerline extraction (ACE) on selected layers from a multispectral image. In the second step, a regional node-based representation of the ACE pixels is generated through self-organized road mapping (SORM). The third step consists of road topology construction by linking the SORM nodes according to a fuzzy grouping model. The fourth step is the refinement feedback loop. If refinement is opted for, nodes from the constructed road vector layer are used to select road training samples. Spectral training samples for roads are derived from SORM nodes, and for non-road classes, from unsupervised classification. Supervised classification is then performed to separate roads from non-roads. The road class is then cleaned with morphological filtering prior to being cycled through the process again. The refinement feedback loop represents our concept of *self-supervised road classification* (SSRC). All algorithms are implemented in MATLAB, and use the image processing and fuzzy logic toolboxes.

A summary of the problem domain conditions we assume for this approach are as follows:

- *Multispectral imagery* with a ground sample distance (GSD) no larger than about 2 meters per pixel. We experiment with USGS color-infrared orthoimages at 1 meter per pixel.
- *No ancillary reference data*, such as digital terrain models (DTM), digital raster graphics (DRG), or vector data, are used. Although ancillary reference data often does exist in practice, we are motivated to approach the problem as a *vision* process, i.e., limited to what can be derived from the scene at hand, which humans perform fluently. This strategy is employed to reveal the capabilities and limitations of our proposed approach. It also mitigates the need to address issues related to the quality and registration accuracy of ancillary reference data on the extraction algorithm's effectiveness.

Pete Doucette is with the Pacific Northwest National Laboratory, 1529 W. Sequim Bay Road, Sequim, WA 98382 (peter.doucette@pnl.gov).

Peggy Agouris and Anthony Stefanidis are with the Department of Spatial Information Science & Engineering, National Center for Geographic Information and Analysis, University of Maine, 348 Boardman Hall, Orono, ME 04469-5711.
PNNL-SA-41582

Photogrammetric Engineering & Remote Sensing
Vol. 70, No. 12, December 2004, pp. 1405–1416.

0099-1112/04/7012–1405/$3.00/0

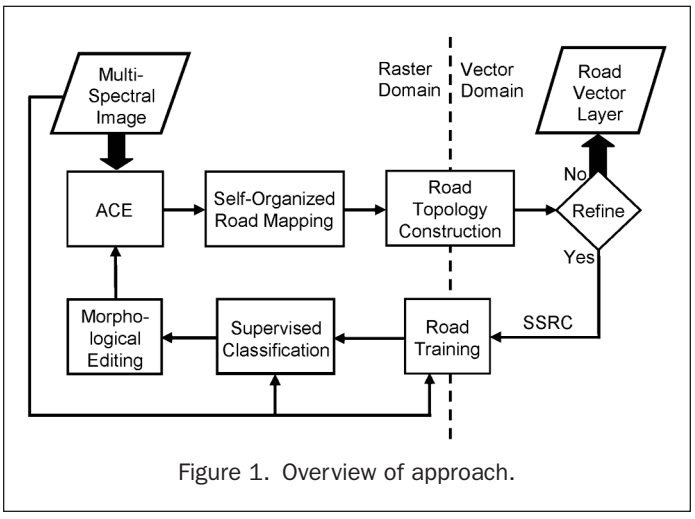


Figure 1. Overview of approach.

- *Suburban-to-urban scene content* is considered in this paper. This is not a constraint, but rather a conscious effort to examine increasingly complex scene structure compared to that found in rural scenes.
- *Automated* algorithm operation that is data driven, i.e., no requirement for interactive human input derived from the scene.
- *Performance evaluation* of automated extraction is measured with respect to manually derived ground truth.

The remainder of the paper proceeds as follows: We provide a review of related road extraction algorithms, and an overview of algorithm performance evaluation followed by the ACE technique, and results are derived from color-infrared imagery. Road topology construction that is based on self-organization and a fuzzy weighted graph model is described, and finally we present our approach of self-supervised road classification, and provide experimental results for three suburban scenes.

Background

Roads in remotely sensed images are among the most easily recognizable features for the human vision system to identify and over a wide range of image resolutions. Despite significant advancements in computer vision research in recent years, few road extraction algorithms in the literature offer reliability levels suitable for general use in practice. Road extraction researchers are increasing the complexity of road detection models by incorporating more high-level, or "cognitive" processing, which is only in its infancy. In this section we provide an overview of some of the core issues involved with automating road extraction.

The Road Extraction Strategy

Road extraction strategies are often categorized according to the degree of automation, i.e., semi- to fully automated. The objective of semi-automated methods is to assist the human operator on-the-fly. This strategy is designed to use interactive user-provided information typically in the form of seed points, widths, and directions, and allow manual editing of the extraction results on the fly. Since operation generally relies upon sustained interaction with a human, the algorithm must execute in real-time. Although real-time execution clearly sets limits on computational complexity, and thus, the capability of an algorithm, Moore's Law mitigates this problem to some extent over time. Road *trackers* (Barzohar, *et al.*, 1997; Mckeown & Denlinger, 1988; Vosselman & Knecht, 1995) and *snakes* (Gruen & Li, 1997; Trinder & Li, 1995) exemplify state-of-the-art techniques used for semi-automated road extraction.

Generally, a fully automated extraction strategy is intended to extract roads from a scene without the need of sustained interactive operation from a human. A frequently used approach is a top-down, or "update" strategy, in which coarse or outdated road data is used to guide a *revision* extraction, e.g., Agouris, *et al.*, 2001b; Bordes, *et al.*, 1997; and Zhang, *et al.*, 2001. The dominant factor that influences the extent of success of an automatic update is the registration accuracy of the reference data relative to the source image for extraction. Given current worldwide database holdings of road network data, an update strategy offers a practical approach to revise and refine existing data. However, the extraction of *new* road features requires a bottom-up approach in lieu of reference data. A bottom-up strategy is motivated by treating the extraction problem fundamentally as a *vision* process, e.g., modeling aspects of perceptual organization, which is characteristic of human vision. The bottom-up approach usually begins with low-level detection that generates initial hypotheses for candidate road components, followed by mid-level grouping of components, and concludes with high-level reasoning for road network completion.

In bottom-up approaches, *scale space* analysis (commonly used for automated DEM extraction) can be used to define multi-resolution descriptions of roads in a scale-based hierarchical detection process (Baumgartner, *et al.*, 1999). Using scale space to enhance computer vision algorithms draws inspiration from the biological vision processes of *early* and *late* vision. For instance, in early or *preattentive* vision, the vision task consists of scanning an entire scene at low resolution to generate initial hypotheses for objects of interest. Then, in late or *attentive* vision, high resolution analysis is directed toward scene objects of interest found during early vision. Vision systems in biological organisms are very efficient at this process.

Road Model Considerations

The road detection process is normally based on a predefined model description of a road. A road is modeled according to problem domain parameters, such as, image spatial resolution, scene content, road geometry, radiometry, and texture. When the road model is over-constrained with respect to certain parameters, poor generalization is often the result, which is typified by the cliché, "every image has its algorithm". Conversely, when road model parameters are under-constrained, higher false alarm rates result. The standard approach to deal with this issue is to implement a user-controlled *dial-setting* capability for road model parameters to accommodate variable scene content and road geometry.

Image spatial resolution usually establishes the extent of complexity of the road model. For example, when image GSD is greater than about 5 meters per pixel, roads are basically lines. When image GSD is between 1 and 5 meters per pixel, roads become elongated regions bounded by parallel edges, and texture becomes a definable parameter to some extent. But, when image GSD is less than 1 meter per pixel, traffic markings, vehicular traffic, shoulders, sidewalks, and curbs become model-able objects that serve as contextual cues for roads.

Cultural scene content determines modeling aspects of road geometry, and contextual cues such as road markings and vehicular traffic. For instance, in rural scenes roads generally manifest as long curvilinear segments with well-defined edges, and connect in irregular networks. By contrast, roads in urban scenes tend to be linear segments that connect in dense regular grid networks, and are subject to occlusions from a wide variety of sources. Designing a road model for urban scene content is generally a more complicated process than for rural scene content.

A simple, but effective road model for high resolution images that is prevalent in the literature is the anti-parallel, or

"apar," edge detection technique (Nevatia & Ramesh, 1980; Zlotnick & Carnine, 1993). Implementation is straight forward, and application is particularly effective with rural scene content. However, the apar model is often inadequate for road extraction when scene content becomes more urban, and/or occlusions become numerous. Nonetheless, the apar model serves as a good starting point upon which to build. In this paper, we propose an approach that builds upon the apar model for road extraction from suburban to urban scene content.

Exploiting Spectral Content for Road Extraction

Because high spatial resolution multispectral imagery has only recently become widely available (e.g., using remote sensing systems such as QuickBird and Ikonos) modeling spectral content has received somewhat limited consideration for road extraction in the literature. Given the trend of technological advances in multispectral sensor design, it is anticipated that the spatial resolution of multispectral imagery will steadily improve. As it does, the exploitation potential of spectral content is enhanced for automated road extraction in particular. We summarize three approaches that could be used to exploit spectral content for automating road extraction.

1. Discrete layer gradient analysis (DLGA) applies edge detection techniques to individual spectral layers in a multispectral stack, and merges the results. As the number of image layers increases, DLGA becomes inefficient. The main limitation of DLGA is that it deals with each image layer as a separate unit, which does not exploit spectral content explicitly.

2. Unsupervised classification (*clustering*) assumes that road spectra are separable from non-road spectra using a minimum distance metric. Then the road class(es) would need to be identified as such. The key parameter is the number of spectral clusters K into which a scene is classified. When K is too small, roads become lumped spectrally with non-road features. When K is too large, roads become further divided into subclusters. What is needed is a procedure that automatically validates an adequate size of K for roads. One approach to road cluster validation is to start with an empirically derived value for K, and for each spectral cluster, examine its spatial properties (in raster space) for topology consistent with road networks (Agouris, *et al.*, 2001a). If insufficient evidence is found among the K spectral clusters for road topology, then the process is repeated for different values of K until sufficient evidence is found. Limitations of clustering methods include sensitivity to initial conditions, use of only first order spectral statistics, and computational expense.

3. Supervised classification requires the selection of training samples for explicitly defined feature classes within a given scene. Although we are mainly interested in roads, multiple and comprehensive class definitions are generally required for a given scene, and training samples need to be completely representative for each class. In practice, this process requires substantial human input. One approach to automatically select and label training samples is to use existing reference data as the basis for training site selection (Walter & Fritsch, 1998). To ensure the accurate selection of training samples, reference data training requires adequate registration accuracy of the reference data. If reference data does not exist, or is not sufficiently accurate, training samples must be selected through other means. In this paper we present a methodology for "self-supervised" road classification, in that the selection of spectral training samples for road and non-road classes is driven by an algorithm rather than a human. The training procedure is based on an automated spatial analysis of a scene from which to derive spectral samples for road and non-road classes.

Performance Evaluation

In recent years, more attention has been given to developing standards for performance evaluation of computer vision algorithms. The *usage cost* ultimately determines the practical utility of an algorithm in a production environment. Timeline analysis is the basic strategy used to assess the usage cost of an automated extraction algorithm. The idea is to compare the level of effort expended between algorithm-assisted extraction and manual extraction. An algorithm's usage cost includes algorithm initialization, execution time, and manual editing of its output. Operator biases also need to be taken into account.

Delineation accuracy is measured by comparing algorithm output against manually derived ground truth. Human biases introduced in manual delineation can be accounted for to some extent by using a buffered version of the output upon which to base the comparison between human and algorithm. Algorithm output pixels are compared against ground truth and tagged into one of the four categories:

- *true positives* (*TP*): correctly tagged as a feature of interest.
- *true negatives* (*TN*): correctly tagged as background.
- *false positives* (*FP*): incorrectly tagged as a feature of interest.
- *false negatives* (*FN*): incorrectly tagged as background.

Using these tags, standard interpretative metrics used in the literature (Mckeown, *et al.*, 2000; Harvey, 1999; Wiedemann, *et al.*, 1998) include,

$$correctness = \frac{TP}{TP + FP} \tag{1}$$

$$completeness = \frac{TP}{TP + FN} \tag{2}$$

$$quality = \frac{TP}{TP + FP + FN} \tag{3}$$

The *correctness* is a measure between 0 and 1 that indicates the detection accuracy rate relative to ground truth. It is associated with *commission error* in that *correctness* + *commissiom_error* = 1. The *completeness* is also a measure between 0 and 1 that is associated with *omission error* such that *completeness* + *omission_error* = 1. Completeness and correctness are complementary metrics that are only meaningful when interpreted concurrently. The *quality* provides a normalized measure between correctness and completeness. Its value is between 0 and 1, and can never be larger than either completeness or correctness. These three metrics will serve as the basis for evaluating the performance of the road extraction approach presented in this paper.

Road Detection from Anti-Parallel Edges

Our approach to automate road extraction begins with gradient analysis. The gradient signature for roads is demonstrated in Figure 2. Horizontal profiles for intensity and gradient (first derivative) are shown for the road in Figure 2a. When a road is darker than its background (as in this case), the intensity profile takes on the well shape as shown in Figure 2b. The gradient magnitude profile shown in Figure 2c was created by convolving the image with a vertical 3×3 Sobel kernel, which is most sensitive to vertical edges. The minimum and maximum gradient magnitude peaks indicate the locations of edges on either side of the road. The gradient orientations (i.e., for low to high intensity values) for either side of the road are shown with the white arrows in Figure 2a. Because the orientations are directed away from each other, we refer to these as *repelling gradients*. The description of "anti"-parallel edges is motivated by the anti (i.e., opposing) gradient magnitude peaks and orientations for parallel roadsides. If the road were brighter than its background, then the intensity and gradient profiles would be inverted (top to bottom) relative to those shown in Figure 2b and c, and the gradient orientations would be *attracting*. When using anti-parallel edge detection

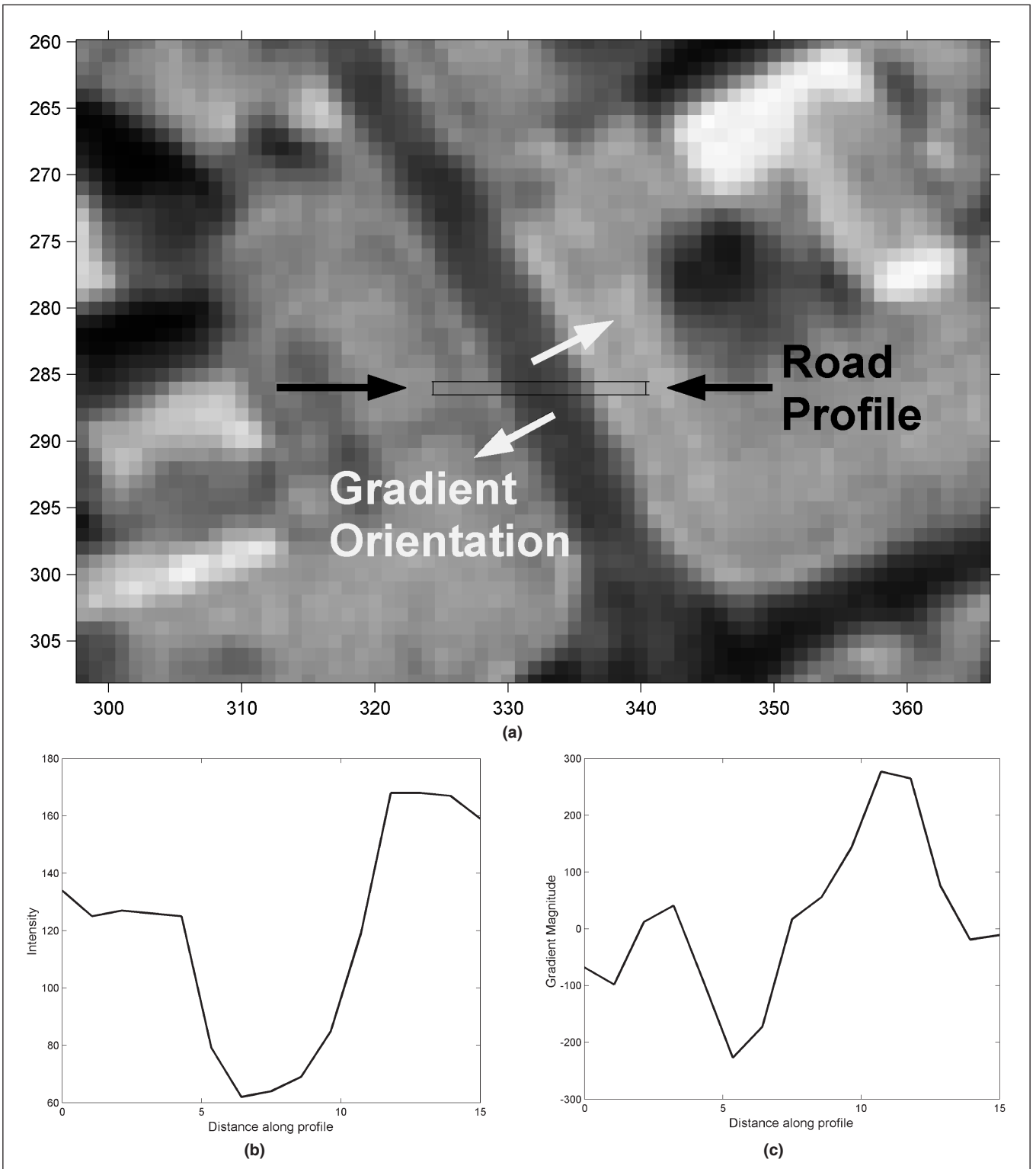


Figure 2. Gradient signature for roads. (a) repelling gradients for a dark road against a lighter background; (b) image intensity profile; (c) gradient intensity profile from a 3 × 3 vertical Sobel filter.

to find roads, both attracting and/or repelling gradients will need to be considered.

To extract candidate road centerline pixels from single layer images, we developed the Anti-parallel edge Centerline Extractor (ACE) algorithm (Agouris, *et al.*, 2004). ACE first computes a Canny edge image for which threshold and smoothing parameters are derived empirically. Then, a gradient orientation image is computed from 3 × 3 Sobel operators according to the relations,

$$G_x = (z_1 + 2z_2 + z_3) - (z_7 + 2z_8 + z_9), \quad \text{and}$$
$$G_y = (z_1 + 2z_4 + z_7) - (z_3 + 2z_6 + z_9). \tag{4}$$

$$\varphi_{(z_5)} = \tan^{-1}\left(\frac{G_x}{G_y}\right). \tag{5}$$

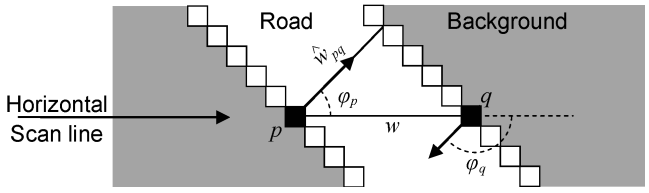


Figure 3. ACE horizontal scan line.

In Equation 4, z_1 through z_9 represent gray values in a 3×3 image region ordered from left to right and top to bottom. In Equation 5, the gradient orientation angle φ is computed for the center pixel (located at z_5).

ACE then scans the Canny edge image in horizontal and vertical directions in search of successive edge pixels p and q that satisfy distance and gradient orientation criteria. Figure 3 demonstrates an ACE horizontal scan line that first finds an edge pixel p, and next finds an edge pixel q on the same row. If pixels p and q satisfy distance and gradient orientation criteria, a centerline pixel is tagged at the midpoint between p and q. The horizontal and vertical scan results are merged and thinned into an output of candidate road centerline pixels.

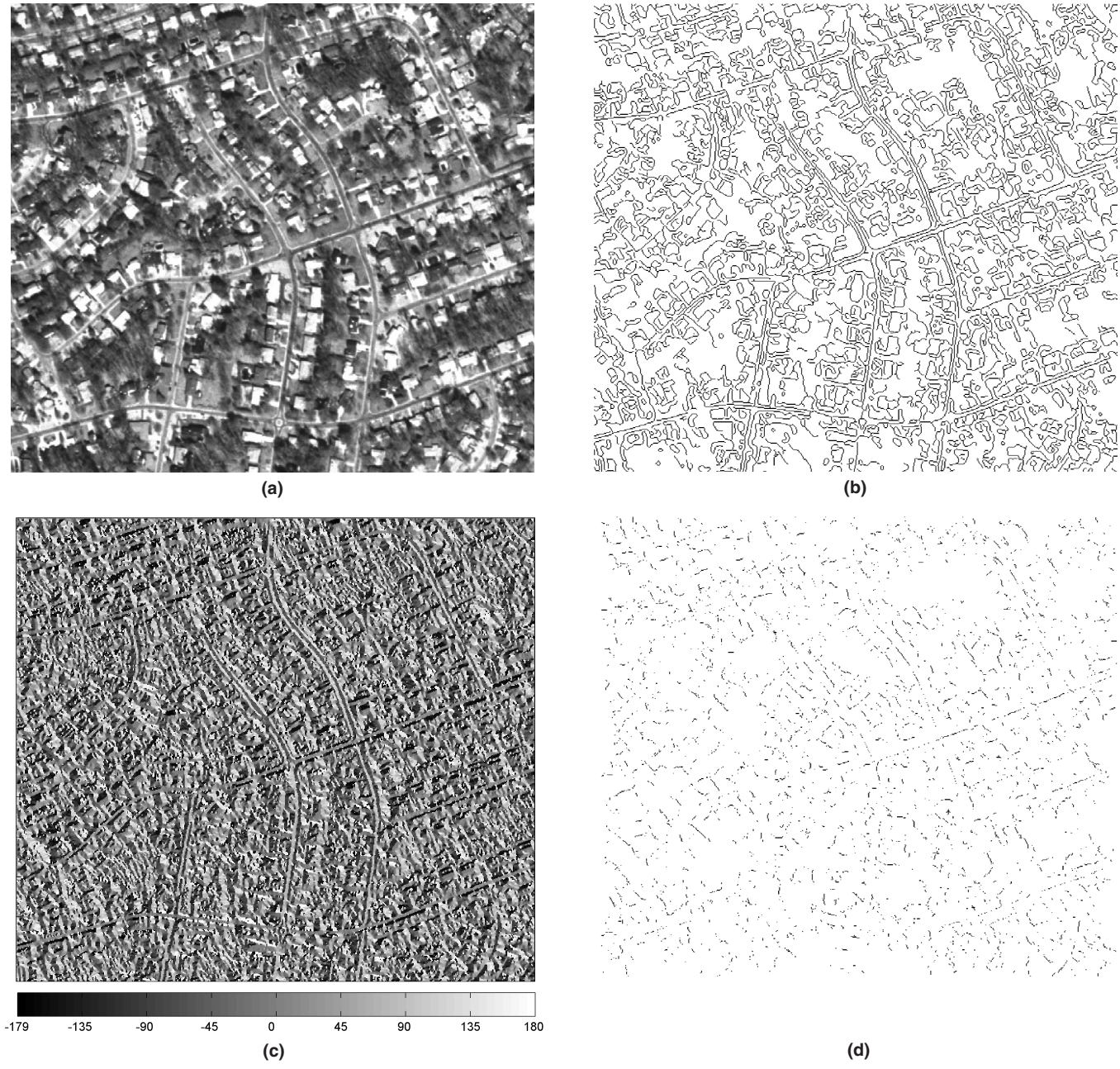


Figure 4. ACE processing. (a) input scene dcfm05 (single-layer with GSD of 1 m/pixel); (b) Canny edge results; (c) Sobel gradient orientation image (units are degrees); (d) candidate road centerline components from ACE.

The basic input parameters for ACE are (1) minimum feature width w_{min}, (2) maximum feature width w_{max}, (3) maximum deflection angle between gradient orientations, α_{max}, and (4) minimum number of pixels per connected component cc_{min}. The estimated perpendicular feature width $\hat{w}_{pq}$ shown in figure 3 for two edge pixels p and q on a scan line is calculated according to,

$$\hat{w}_{pq} = w \cdot \cos(\varphi_p) \qquad \text{(horizontal scan line)} \tag{6}$$

$$\hat{w}_{pq} = w \cdot \sin(\varphi_p) \qquad \text{(vertical scan line)} \tag{7}$$

where $w = |column(p) - column(q)|$ for a horizontal scan line, and φ_p and φ_q are the gradient orientation angles. The deflection angle α between gradient orientations at pixels p and q is determined as,

$$\alpha = |180^\circ - |\varphi_p - \varphi_q||. \tag{8}$$

Figure 4 shows the ACE process for scene dcfm05 (Figure 4a). The scene was clipped from a DOQ image that contained suburban streets. The ACE output in Figure 4d is generated using input parameters of w_{min} = 5 m, w_{max} = 15 m, $\alpha_{max} = 45^\circ$, and cc_{min} = 1 pixel. The extraction of road centerline pixels using ACE is effective to the extent that, (1) road sides are described by anti-parallel edges according to predefined criteria, and (2) anti-parallel edges are exclusive to roads. It is evident from the ACE output in Figure 4d that these two assumptions can easily break down. Nonetheless, ACE serves as a practical starting point from which to detect roads. Specifically, ACE represents the low-level processing stage in our approach, which is followed by a high-level linking stage.

Edge Detection in Multi-Layer Images

The output from ACE may be improved by exploiting multispectral image content. Discrete layer gradient analysis (DLGA) is a straightforward approach in which edge detection is applied to individual image layers, and the results merged. Merging can be performed after the computation of the gradient magnitudes, or after applying the gradient threshold. Figure 5 depicts the process of merging after gradient thresholding, which is how it is implemented in our investigation. The merging process consists of determining the union of all layers, and thinning the results. When using ACE, DLGA merging is applied to ACE output from each layer.

Performance evaluation of ACE with DLGA merging applied to color-IR DOQ scenes (3 layers) revealed that the quality measure of the merged results were generally not optimal. Table 1 shows experimental results for scene dcfm05. As expected, the DLGA merged output has a higher completeness compared to the *best layer* result (66.9 percent versus 59.7 percent), but at the expense of lower correctness (i.e., more false positives), leading to lower road detection quality (0.141 versus 0.188).

In an effort to enhance the spectral contrast of roads against background features, we conducted DLGA experiments on the principal component analysis (PCA) layers. Again, the road detection quality of merged results was sub-optimal compared to the best single layer (0.130 versus 0.231). We concluded that the best single PC layer (PC2) generally provided a more efficient means by which to exploit spectral content when applying ACE road detection to these scenes compared to DLGA merging.

TABLE 1. PERFORMANCE EVALUATION OF ACE FOR SCENE dcfm05

Procedure	Correctness	Completeness	Quality
DLGA (merge)	0.152	0.669	0.141
DLGA (optimal)	0.216	0.597	0.188
PCA (merge)	0.135	0.777	0.130
PC2 (cc_{min} = 1)	0.267	0.631	0.231
PC2 (cc_{min} = 3)	0.371	0.564	0.288

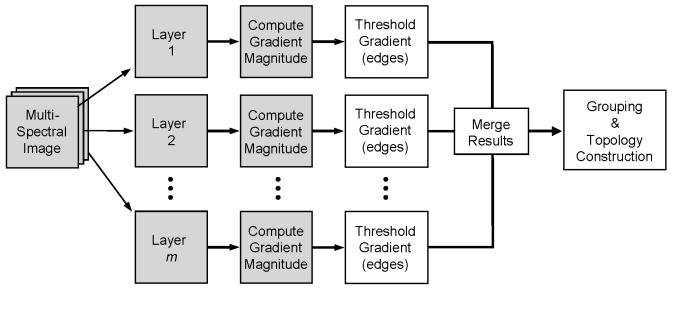
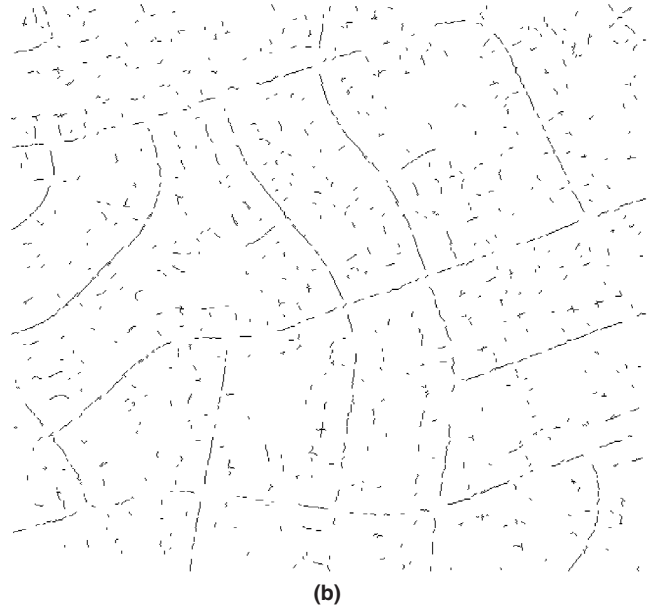


Figure 5. DLGA merging following gradient thresholding (gray: gray level images; white: binary images).

(a)



(b)

Figure 6. ACE processing multiple layers. (a) PC2 input layer; (b) candidate road centerline components from ACE.

Our performance strategy at this low-level stage of processing was to maximize the quality metric in particular. In general, a minimum connected component size ($cc_{\min}$) slightly larger than 1 pixel was most effective at filtering out smaller components that more likely contributed to false positive detections. For example, with $cc_{\min} = 3$ pixels for PC2, the quality increased by 0.057 as shown in Table 1. Figure 6 shows the ACE results for PC2 with $cc_{\min} = 3$ pixels. Note that the roads are brighter than the background in PC2, thus necessitating the detection of attracting gradients.

Topology Construction

Following detection of candidate road centerline pixels using ACE, the next step consists of constructing a coherent road network topology. Centerline component grouping and linking is performed on a node-based representation of the ACE pixels. This node representation is generated from a *self-organized mapping* of the ACE pixels. Then nodes are linked with a rule-based fuzzy grouping model to construct topologies that are consistent with road networks.

Self-Organized Road Mapping

We apply the concept of self-organized road mapping (SORM) to partition ACE pixels into node-based regions (Doucette, *et al.*, 2001). Our SORM technique is an adaptation of the K-means and self-organizing map (SOM) (Kohonen, 2001) algorithms. The procedure is as follows:

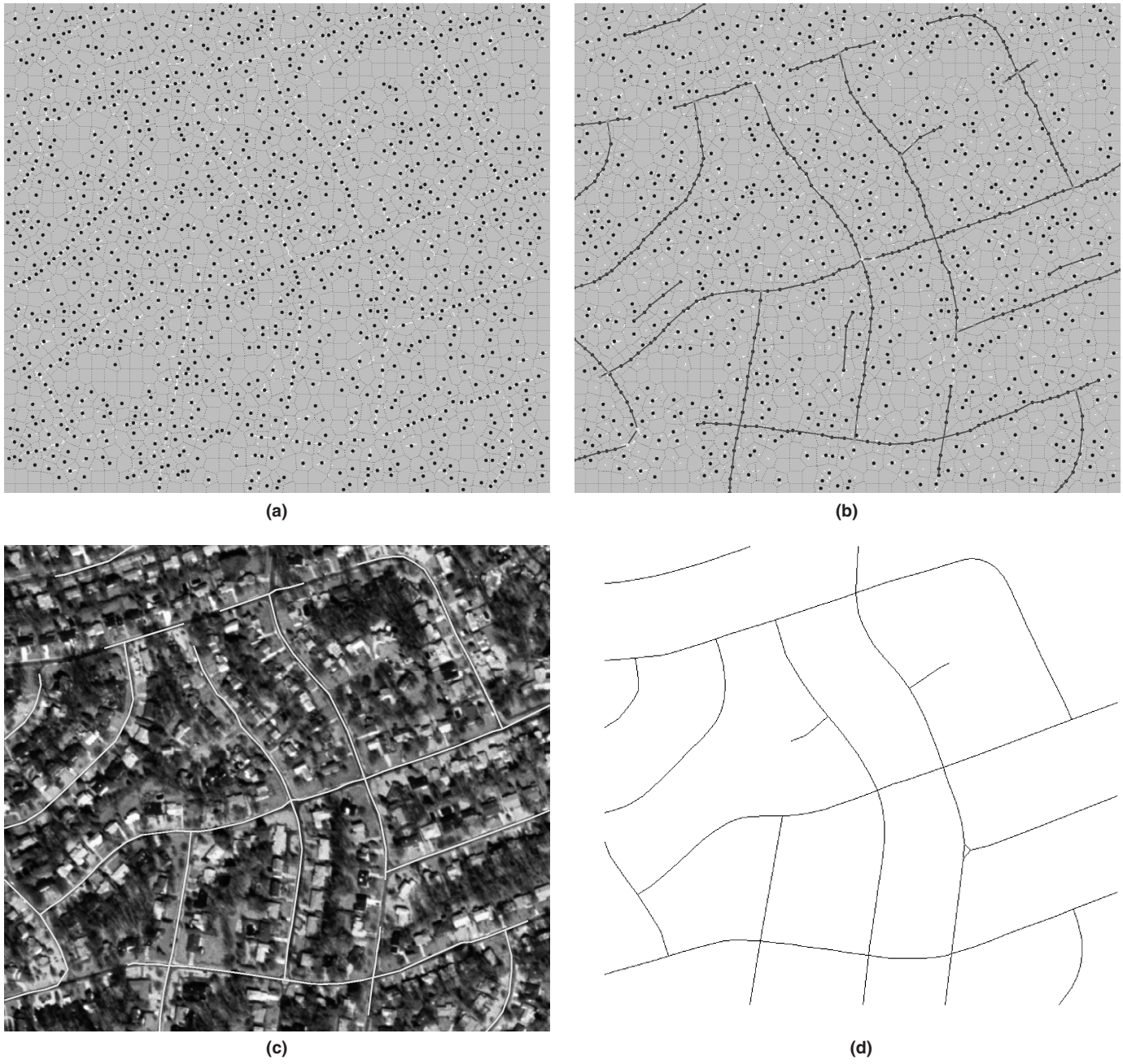


Figure 7. Topology construction for scene dcfm05. (a) node convergence pattern from SORM; (b) results of component grouping; (c) final extracted road vectors. (d) ground truth.

Step 1. (*Initialization*) Initialize K nodes (cluster centers), $\mathbf{c}_1$, $\mathbf{c}_2, \ldots, \mathbf{c}_K$, in the ACE output space. The nodes are arranged as a regular grid, with node spacing defined by β_{grid}.

Step 2. (*Determine sample-node associations*) At the nth iteration, determine which node each sample vector $\mathbf{x}$ (x, y coordinate for a candidate road centerline pixel from ACE) is closest to according to,

$$\mathbf{x} \in S_j(n) \quad \text{if} \quad \|\mathbf{x} - \mathbf{c}_j(n)\| < \|\mathrm{x} - \mathbf{c}_i(n)\| \quad i = 1, 2, \ldots, K, \quad \text{and} \quad i \neq j. \tag{9}$$

In Equation 9, $S_j(n)$ represents the set of samples whose closest node is $\mathbf{c}_j(n)$. Ties are resolved by the order rule. If $S_j(n)$ is less than a minimum threshold number of samples, then node $\mathbf{c}_j(n)$ is deleted.

Step 3. (*Update node positions*) Compute the median of the samples associated with each node from Step 2, and let each median represent the updated location of the node according to,

$$\mathbf{c}_j(n+1) = \left\{ \begin{array}{l} \text{sort}(\mathbf{x})_{(N_j+1)/2}, \text{ if } N_j \text{ is odd} \\ \frac{1}{2}\left(\text{sort}(\mathbf{x})_{N_j/2} + \text{sort}(\mathbf{x})_{(N_j/2)+1}\right), \text{ if } N_j \text{ is even} \end{array} \right\}, \quad j = 1, 2, \ldots, K. \tag{10}$$

In Equation 10, N_j is the number of samples in $S_j(n)$, and sort($\mathbf{x}$) represents the sorted samples for a given node, where $\mathbf{x} \in S_j(n)$. The median statistic is used because it is less sensitive to clutter than the mean.

Step 4. (*Node merge check*) The Euclidean distance between any two nodes i and j, is defined by d_{ij}. A minimum distance allowance between nodes is defined by $\beta_{\min}$. If $d_{ij} \leq \beta_{\min}$, then merge nodes i and j. If node merging occurs, then return to Step 2.

Step 5. (*Convergence check*) Iterate from Step 2 until all node positions remain unchanged such that,

$$\mathbf{c}_j(n+1) = \mathbf{c}_j(n), \qquad j = 1, 2, \ldots, K. \tag{11}$$

Using the ACE output for PC2 ($cc_{\min} = 3$) from scene dcfm05, Figure 7a shows the node convergence pattern from SORM for $\beta_{\text{grid}} = 15$ m and $\beta_{\min} = 5$ m. The black dots are nodes, and the polygon surrounding each node represents its voronoi region. Polygons that are devoid of nodes indicate deleted nodes. The white pixels (input samples) are ACE output. Each node-sample set combination represents a candidate road centerline *component*.

Grouping of Centerline Components

Component orientation statistics are analyzed to find evidence of topology that is consistent with roads and road networks. The topological analysis is performed with a fuzzy Gestalt grouping model (Doucette, 2002). This process incorporates a weighted graph technique to link nodes into progressively larger components. The challenge is to draw inferences from topological structure to complete road networks, while avoiding imposing topological structure where it should not exist. The vector links between nodes in Figure 7b are the results of component grouping. The final grouping results are captured as a vector layer, which is overlaid on the original scene in Figure 7c. Comparing the grouping results to the ground truth in Figure 7d, the gaps and false positive segments are evident.

Table 2 shows the performance evaluation results for scene dcfm05 when ACE is followed by topology construction (TC). To perform the evaluation, both ground truth and TC vector layers are converted to raster representations (1 m per pixel) to accommodate a pixel by pixel comparison. Comparing Table 2 with Table 1 shows a substantial improvement with using TC, e.g., for the case of PC2 ($cc_{\min} = 3$), extraction quality jumps from 0.288 to 0.672. The RMSE for the node positions with respect to ground truth is included in Table 2.

TABLE 2. PERFORMANCE EVALUATION OF ACE WITH TOPOLOGY CONSTRUCTION FOR SCENE dcfm05

Procedure	Correctness	Completeness	Quality	RMSE (m)
DLGA (merge) + TC	0.654	0.346	0.292	13.10
DLGA (optimal) + TC	0.861	0.569	0.521	11.20
PCA (merge) + TC	0.837	0.730	0.639	7.10
PC2 ($cc_{\min} = 3$) + TC	0.892	0.731	0.672	1.69

Self-Supervised Road Classification

The objective of self-supervised road classification (SSRC) is to automate the process of training sample selection and refinement for a *road* class, and deriving a practical spectral definition for a *non-road* class. We are therefore faced with a two-class problem. For the road class, nodes from the road topology construction serve as source of training sample sites. As illustrated in Figure 8, training samples are gathered from regions surrounding each node, which in this case is a 3×3 neighborhood. Additional training sites could be captured by increasing the neighborhood size, as well including additional nodes through vector densification. Spectral statistics (mean and variance) for the road class are derived directly from these training samples. Rather than attempting to derive spectral statistics for a single non-road class (which would be spectrally broad), they are derived from multiple non-road candidate classes found from unsupervised classification. This strategy provides for a multi-class spectral distribution of the non-road candidate classes that can be used more effectively in the subsequent step of supervised classification.

With road and non-road candidate class statistics in hand, a supervised classification algorithm is applied to the scene. The outcome for the road class is weighted slightly higher than all other classes in order for it to supersede any non-road

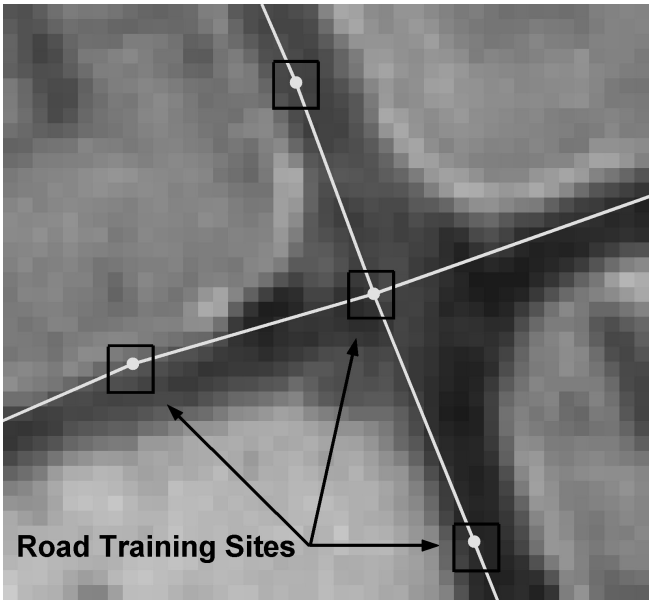


Figure 8. Training samples gathered from 3×3 neighborhoods around each node.

class candidates that may be spectrally similar to roads. Figure 9 demonstrates the process for scene dcfm05. Figure 9a shows the results from unsupervised classification for four non-road candidate classes (the use of four was empirically derived). Figure 9b shows the results from a Bayesian maximum likelihood classification that uses spectral statistics derived from the road and non-road candidate classes. (A radial basis function (RBF) neural network classifier was also tested, and provided similar results). Figure 9c shows the final two-class output for roads and non-roads. The relative weighting of the road class in the supervised classification controls the extent of error, e.g., heavier road class weighting contributes to fewer false negatives and more false positives. Morphological filtering is applied to the binary road class image to fill gaps (using closing) and eliminate noise (using minimum component size filtering).

With the binary road classification image in hand (Figure 9c), we can repeat the ACE and topology construction process. This constitutes the self-supervised road classification (SSRC) feedback loop (Figure 1), which can be repeated as needed. Table 3 shows the significant increase in extraction performance through two iterations of SSRC. For scene dcfm05, the road extraction quality goes from 0.672 (prior to SSRC), to 0.832 on the first iteration, and 0.868 on the second. This trend demonstrates an automated refinement process for road training sample selection. We observed that

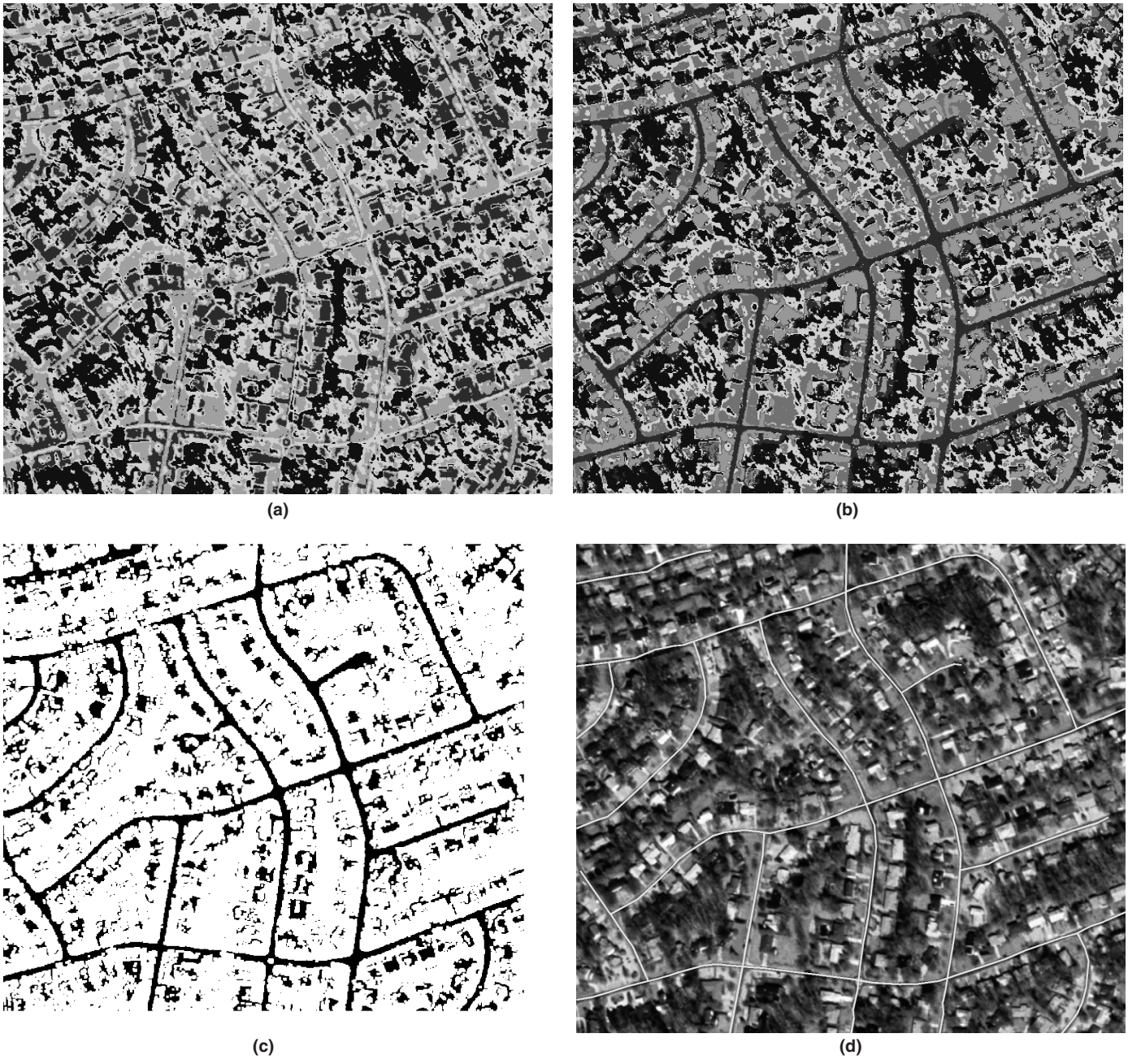


Figure 9. SSRC process for scene dcfm05. (a) unsupervised classification of four classes; (b) Bayesian maximum likelihood classification; (c) two-class output for roads (black) and non-roads (white); (d) road extraction results following second iteration of SSRC.

TABLE 3. PERFORMANCE EVALUATION OF THREE SUBURBAN SCENES WITH SSRC

Scene	Procedure	Correctness	Completeness	Quality	RMSE	Samples
dcfm05	PC2 (cc_{min} = 3) + TC	0.892	0.731	0.672	1.69 m	n/a
dcfm05	SSRC (1st iteration)	0.957	0.864	0.832	1.13	4710
dcfm05	SSRC (2nd iteration)	0.940	0.919	0.868	2.07	5385
dcfm01	PC2 (cc_{min} = 3) + TC	0.925	0.082	0.082	1.39	n/a
dcfm01	SSRC (1st iteration)	0.915	0.373	0.361	1.68	515
dcfm01	SSRC (2nd iteration)	0.935	0.632	0.605	1.83	2375
dcfm04	PC2 (cc_{min} = 3) + TC	0.367	0.049	0.045	21.60	n/a
dcfm04	SSRC (1st iteration)	0.578	0.237	0.202	12.90	750
dcfm04	SSRC (2nd iteration)	0.602	0.388	0.309	16.00	2090

extraction performance generally did not increase much beyond two SSRC iterations, and slight decreases were possible with additional iterations. Figure 9d shows the road extraction results for scene dcfm05 following a second SSRC iteration, and vector smoothing using standard generalization (also reflected in the performance evaluation numbers). The *samples* column in Table 3 indicates how many road training samples are automatically selected and used in SSRC. As to be expected, the number of road samples selected increases with a higher completeness percentage from the previous iteration.

Road extraction results from SSRC for two additional 1 m color-infrared suburban scenes (dcfm01 and dcfm04) are provided in Table 3 and Figure 10. These scenes contain more complex content than dcfm05, which is reflected by the lower extraction results. Nonetheless, SSRC provides dramatic improvement over ACE + TC alone. For scene dcfm01 (Figure 10a), where road occlusion from trees is substantial, extraction quality goes from 0.082 to 0.605 through two iterations of SSRC.

Experiments with these and similar scenes indicated that the selection of a sufficient number of correct road training samples was crucial for SSRC to be effective. SSRC was able to accommodate incorrectly selected road training samples only to a certain extent. To facilitate this constraint, we adopted a strategy in which extraction correctness (fewer false positives) was favored over completeness earlier in the process, and then gradually shifted to favoring extraction quality for the latter iteration(s). This *correctness-to-quality* approach was incorporated into the topology construction stage. Future work will consider other multispectral and multi-sensor images.

Summary

This paper presented a methodology that exploits spectral content for fully automated road centerline extraction. Preliminary detection of road centerline pixel candidates was performed with Anti-parallel-edge Centerline Extraction (ACE). This was followed by constructing topologies that are consistent with road networks. A road vector topology was constructed with a fuzzy grouping model that links nodes from a self-organized road mapping (SORM) of the ACE pixels. Following topology construction, a self-supervised road classification (SSRC) methodology was implemented to automate the process of training sample selection and refinement for a road class, as well deriving practical spectral definitions for non-road classes. SSRC was incorporated into the process as an iterative feedback loop, which demonstrated a potential to provide dramatic improvement by exploiting spectral content. Road centerline extraction results were presented for three 1 m color-infrared suburban scenes. Significant improvements in extraction results were observed following SSRC, particularly when dealing with substantial road occlusion from trees. The selection of a sufficient number of correct road training samples was crucial for SSRC to be effective. To facilitate this constraint, a *correctness-to-quality* strategy was adopted in which extraction correctness was favored over completeness initially, and then gradually shifted to favoring extraction quality. Future work will consider other multispectral and multi-sensor images, and refinement of the training sample selection process.

Acknowledgments

This work was supported by the National Science Foundation through grant ITR-0121269, and the National Geospatial-Intelligence Agency through NMA501-03-BAA-0002.

References

Agouris, P., P. Doucette, and A. Stefanidis, 2004. Automation and Digital Photogrammetric Workstations (Chapter 12), *Manual of Photogrammetry* (Fifth Edition), (C. McGlone, M. Mikhail, and J. Bethel, editors), American Society for Photogrammetry and Remote Sensing.

Agouris, P., P. Doucette, and A. Stefanidis, 2001a. Spatiospectral Cluster Analysis of Elongated Image Regions, *International Conference on Image Processing* (*ICIP*), pp. 789–792, IEEE Signal Processing Society.

Agouris, P., A. Stefanidis, and S. Gyftakis, 2001b. Differential Snakes for Change Detection in Road Segments, *Photogrammetric Engineering & Remote Sensing*, 67(12):1391–1400.

Barzohar, M., M. Cohen, I. Ziskind, and D. Cooper, 1997. 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*) (A. Gruen, E. Baltsavias, and O. Henricsson, editors), pp. 277–286, Birkhauser Verlag.

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

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

Doucette, P., 2002. *Automated Road Extraction from Aerial Imagery by Self-Organization*, Ph.D. Dissertation, Department of Spatial Information Science and Engineering, University of Maine.

Doucette, P., P. Agouris, A. Stefanidis, and M. Musavi, 2001. Self-Organized Clustering for Road Extraction in Classified Imagery, *ISPRS Journal of Photogrammetry and Remote Sensing*, 55(5–6): 347–358.

Gruen, A., and H. Li, 1997. Semi-Automatic Linear Feature Extraction by Dynamic Programming and LSB-Snakes, *Photogrammetric Engineering & Remote Sensing*, 63(8):985–995.

Harvey, W., 1999. Performance Evaluation for Road Extraction, *Bulletin de la Societé Francaise de Photogrammetrie et Télèdetection*, No. 153(1999–1):79–87.

Figure 10. SSRC results for scene dcfm01 (a) input; (b) output following second iteration; (c) ground truth, and for scene dcfm04 (d) input; (e) output following second iteration; (f) ground truth.

Kohonen, T., 2001. *Self-Organizing Maps*, Springer-Verlag, Boston, Massachusetts.

Mckeown, D., T. Bulwinkle, S. Cochran, W. Harvey, C. McGlone, and J. Shufelt, 2000. Performance Evaluation for Automatic Feature Extraction, *International Archives of Photogrammetry and Remote Sensing*, Vol. 33(B2), pp. 379–394.

Mckeown, D., and J. Denlinger, 1988. Cooperative Methods for Road Tracking in Aerial Imagery, *IEEE Proceedings of Computer Vision and Pattern Recognition*, pp. 662–672, Ann Arbor, Michigan.

Nevatia, R., and B. Ramesh, 1980. Linear Feature Extraction and Description, *Computer Vision, Graphics, and Image Processing*, 13, 257–269.

Trinder, J., and H. Li, 1995. Semi-Automatic Feature Extraction by Snakes, *Automatic Extraction of Man-Made Objects from Aerial and Space Images* (A. Gruen, O. Kuebler, and P. Agouris, editors), pp. 95–104, Birkhauser, Basel.

Vosselman, G., and J. Knecht, 1995. Road Tracing by Profile Matching and Kalman Filtering, *Automatic Extraction of Man-Made*

Objects from Aerial and Space Images (A. Gruen, O. Kuebler, and P. Agouris, editors), pp. 265–274, Birkhauser, Basel.

Walter, V., and D. Fritsch, 1998. Automatic Verification of GIS Data Using High Resolution Multispectral Data, *International Archives of Photogrammetry and Remote Sensing*, Vol. 32(3), pp. 485–489.

Wiedemann, C., C. Heipke, H. Mayer, and O. Jamet, 1998. Empirical Evaluation of Automatically Extracted Road Axes, *Empirical Evaluation Methods in Computer Vision* (K. Bowyer, and P. Phillips, editors), pp. 172–187, IEEE Computer Society Press.

Zhang, C., E. Baltsavias, and A. Gruen, 2001. Knowledge-Based Image Analysis for 3D Road Reconstruction, *Asian Journal of Geoinformatics*, 1(4):3–14.

Zlotnick, A., and P. Carnine, 1993. Finding Road Seeds in Aerial Images, *Computer Vision, Graphics, and Image Processing*, 57(2): 243–260.