Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Delineation and geometric modeling of road networks

Charalambos Poullis[*], Suya You

Computer Graphics and Immersive Technologies Lab, Integrated Media Systems Center, University of Southern California, United States

ARTICLE INFO

Article history:
Received 7 April 2008
Received in revised form
2 October 2009
Accepted 6 October 2009
Available online 14 November 2009

Keywords:
Road extraction
Network delineation
Road detection
Road modeling

ABSTRACT

In this work we present a novel vision-based system for automatic detection and extraction of complex road networks from various sensor resources such as aerial photographs, satellite images, and LiDAR. Uniquely, the proposed system is an integrated solution that merges the power of perceptual grouping theory (Gabor filtering, tensor voting) and optimized segmentation techniques (global optimization using graph-cuts) into a unified framework to address the challenging problems of geospatial feature detection and classification.

Firstly, the local precision of the Gabor filters is combined with the global context of the tensor voting to produce accurate classification of the geospatial features. In addition, the tensorial representation used for the encoding of the data eliminates the need for any thresholds, therefore removing any data dependencies.

Secondly, a novel orientation-based segmentation is presented which incorporates the classification of the perceptual grouping, and results in segmentations with better defined boundaries and continuous linear segments.

Finally, a set of gaussian-based filters are applied to automatically extract centerline information (magnitude, width and orientation). This information is then used for creating road segments and transforming them to their polygonal representations.

1. Introduction

Recent technological advancements have caused a significant increase in the amount of remote sensor data and of their uses in various applications. Efficient and inexpensive techniques in the area of data acquisition have popularized the use of remote sensor data and led to their widespread availability. However, the interpretation and analysis of such data still remains a difficult and manual task. Specifically in the area of road mapping, traditional methods require time-consuming and tedious manual work which does not meet the increasing demands and requirements of current applications. Although considerable attention has been given on the development of automatic road extraction techniques it still remains a challenging problem due to the wide variations of roads (urban, rural, etc) and the complexities of their environments (occlusions due to cars, trees, buildings, etc).

In this work we focus on the automatic and reliable detection and extraction of transportation networks from remote sensor data including aerial photographs, satellite images, and LiDAR. We present an integrated solution that merges the strengths of perceptual grouping theory (Gabor filters, tensor voting) and segmentation (global optimization by graph-cuts), under a unified framework to address the challenging problem of automated feature detection, classification and extraction. The proposed approach leverages the multi-scale, multi-orientation capabilities of Gabor filters for the inference of geospatial features, the effective and robust handling of noisy, incomplete data of tensor voting for the feature classification and the fast and efficient optimization of graph-cuts for the segmentation and labeling of road features.

2. Related work

A plethora of work has been proposed for solving the complex problem of extracting road networks from remote sensor data. Almost all of the existing work shares similar processing pipeline and relies on the combination of pixel-based, region-based and knowledge-based techniques. However, several distinctions exist between the different processing components. Below is an overview of the state-of-the-art in this area. Mayer et al. (2006) offers a comprehensive survey on the state-of-the-art road extraction techniques from a variety of different datasets.

In Baumgartner et al. (1999) lines are extracted in an image with reduced resolution as well as road-side edges in the original high resolution image. Using both resolution levels and explicit knowledge about roads, hypotheses for road segments are generated and

[*] Corresponding author. Tel.: ++1 3102109787.
E-mail addresses: charalambos@poullis.org (C. Poullis), suyay@graphics.usc.edu (S. You).

0924-2716/$ – see front matter © 2009 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2009.10.004

are grouped iteratively into larger segments. Although the results seem promising, the proposed method is focused on extracting road networks for rural areas.

Lisini et al. (2004) presents a system which relies on adaptive filtering to determine predominant orientations of the roads. The response of the filtering is then used to extract linear segments which are then connected based on tolerances determined by the spatial resolutions. This approach relies on various hard thresholds and data-dependent parameters thus requires considerable user interaction to tune the parameters prior to processing the data.

A different approach which tries to reduce the number of tunable parameters is presented in Laptev et al. (2000). The authors propose the integration of the well-established techniques of multi-scale image processing and active contour models to resolve the complex problem of road extraction. They use a multi-scale ridge detector for the detection of lines at a coarser scale, and then use a local edge detector at a finer scale for the extraction of parallel edges which are optimized using a variation of the active contour models technique (snakes). The results indicate that the approach performs very well especially for rural areas.

Similarly, Wessel (2004) employs Steger's differential geometry approach (Mayer and Steger, 1998) for the extraction of linear segments. Context information about road networks is then used to connect the linear segments into roads. Steger's differential geometry approach is also employed in Bacher and Mayer (2005) for the extraction of linear segments from multi-spectral images. The extracted lines are then used for training through an automatic supervised classification to produce a road class image which can be used to verify road hypotheses. The approach has been shown to perform well on rural areas only.

The authors in Barsi and Heipke (2003) present an approach for extracting road junctions. To achieve this they train a feed-forward artificial neural network to learn a junction model which supports junctions of up to four arms. The training is performed interactively and the junctions are extracted using a Deriche operator for the edge detection with an added hysteresis threshold, followed by an edge smoothing using the Ramer algorithm. Although the result is not a complete road network the approach seems to perform very well for rural areas.

The system in Zhang et al. (2001) integrates knowledge processing of color image data and information from digital geographic databases, extracts and fuses multiple object cues, thus takes into account context information, employs existing knowledge, rules and models, and treats each road subclass accordingly. Clode et al. (2005) uses a rule-based algorithm for the detection of buildings at a first stage and then at a second stage the reflectance properties of the road. Similarly, Zhang and Couloigner (2006) uses reflectance as a measure for the image segmentation and clustering. Explicit knowledge about geometric and radiometric properties of roads is used in Wessel (2004) to construct road segments from the hypotheses of road-sides. In Barsi and Heipke (2003) the developed system can detect a variety of road junctions using a feed-forward neural network, which requires collected data for the training of the network. Peteri et al. (2003) take high resolution images as input along with prior knowledge about the roads e.g. road models and road properties.

In Porikli (2003) the authors present an approach based on point-wise Gaussian models. A set of quadruple line filters is applied on the image to extract linear segments. Additionally, road points which are not perceptible by the line filters are enhanced using the likelihood of each image point as being part of a road. The results are impressive however, this approach only deals with images where the roads appear as thin linear features and have no width.

A method which relies on elevation data is presented in Clode et al. (2005). LiDAR data provides accurate elevation information which can be used to resolve problems occurring using optical imagery such as road overlaps due to bridges. A region growing algorithm is used to segment the road segments from other points in the data such as buildings, trees, etc. The road candidates are then vectorized using a phase-coded disk which allows the extraction of roads of different widths and different orientations.

The importance of scale-space processing is described in the work of Mayer and Steger (1998). Building on similar concepts, the authors in Heller and Pakzad (2005) present a concept to automatically adapt road models for high resolution images to models appropriate for images of lower resolution with similar spectral characteristics. Additionally, in Heuwold (2006) the author presents a framework for the verification of the automatic adaptation of object models consisting of parallel line-type objects parts to a lower image resolution. Similarly, in Hinz and Baumgartner (2003) the authors present an automatic road extraction technique by integrating detailed knowledge about roads and their context using explicitly formulated scale-dependent models. A slightly different approach which combines a scale-space processing framework with the introduction of Markov random fields is presented in Tupin et al. (2002).

On a different note, the authors in Mena and Malpica (2005) present an automatic method for road extraction which uses a new technique, named Texture Progressive Analysis and consists of a fusion of information streaming from three different sources for the image. The approach was successfully applied on rural as well as semi-urban areas with successful results.

Zhou et al. (2007) present a user-guided image interpretation system which integrates inputs from human experts with computational algorithms in order to learn road tracking. Although the results seem promising, the goal of completely eliminating the need for human intervention and interactions is still not achieved.

An approach which combines a line-based road extraction and area-based color segmentation techniques is presented in Ziems et al. (2007). They show that the incorporation of prior information into the line-based road extraction algorithm allows the robust estimation and automatic tune-up of parameters that control the contrast between road and background, the homogeneity within the road objects and the global threshold for masking out non-road areas.

The aforementioned work clearly indicates that the predominant approach for addressing the complex problem of road extraction involves the multi-scale processing of the input data. In addition to the scale-space processing, an imperative part of road extraction systems is the elimination of data-dependent parameters since this directly affects the applicability of the system. Although very impressive and promising results have already been reported as mentioned above, the majority of the existing work in the area focuses on particular types of datasets (i.e. LiDAR or satellite images) and/or particular types of scenes (i.e. rural, urban, forest, etc). The result is road extraction systems which perform well for one type but fail for another unless numerous parameters are fine-tuned.

Hence, the goal of our work is to design and develop a system which relies on well-established computer vision techniques, incorporates scale-space processing, requires no (or minimal and stable) parameter tuning and can simultaneously process various remote sensor data such as LiDAR, intensity response and satellite imagery. The solution to these problems is sought in the development of a novel system which combines the strengths of perceptual grouping (Gabor filters, Tensor Voting) and global optimization (Graph-Cuts) for the geospatial feature inference and classification. As a result, the proposed system has no data dependencies and requires minimal parameters which were found to be stable and remain fixed for all the examples presented (scale factor for the Tensor voting, number of labels and smoothness factor for the optimization). The results shown in Section 7 indicate the high success rate of our system on all types of datasets and scenes, and verify the validity of the approach.

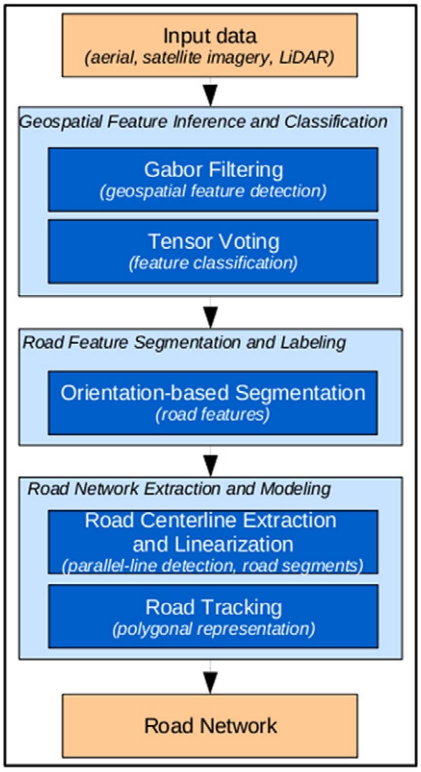

Input data
(aerial, satellite imagery, LiDAR)
Geospatial Feature Inference and Classification
Gabor Filtering
(geospatial feature detection)
Tensor Voting
(feature classification)
Road Feature Segmentation and Labeling
Orientation-based Segmentation
(road features)
Road Network Extraction and Modeling
Road Centerline Extraction and Linearization
(parallel-line detection, road segments)
Road Tracking
(polygonal representation)
Road Network

Fig. 1. System overview.

3. System overview

Although many different approaches have been proposed and developed for the automatic extraction of road networks, it still remains a challenging problem due to the wide variations of roads e.g. urban, rural, mountainous etc and the complexities of their environments e.g. occlusions due to cars, trees, buildings, shadows etc. For this reason, traditional techniques such as pixel- and region-based have several problems and often fail when dealing with complex road networks. Our proposed approach addresses these problems and provides solutions to the difficult problem of automatic road extraction. Fig. 1 summarizes our approach.

Firstly, geospatial feature inference and classification. Local orientation information is extracted using a bank of Gabor filters, which is encoded into a tensorial representation. This representation can simultaneously capture the geometric information of multiple feature types passing through a point (surface, curve, junction) and an associated measure of the likelihood of that point being part of each type. A tensor voting is then performed which globally communicates and refines the information carried at each point. An important advantage of combining Gabor filters and tensor voting for the classification is that it eliminates the need for hard thresholds. Instead, the refined likelihoods of each point give an accurate estimate of the dominant feature passing through that point, and are therefore used for the classification into curve and junction features. Furthermore, it removes the limitation of tensor voting to work only with binary images and extends its application to grayscale images.

Secondly, road feature segmentation and labeling. A novel orientation-based segmentation using graph-cuts is performed. An important aspect of this segmentation is that it incorporates the orientation information of the classified curve features and favors towards keeping those curves connected. The result is a binary segmentation into road and non-road candidates.

Finally, road network extraction and modeling. A pair of gaussian-based bi-modal and single mode kernels are developed for the automatic detection of road centerlines and the extraction of width and orientation information from the segmented road candidates. Linear segments resulting from the application of an iterative Hough transform on the road centerlines, are validated and refined (merge, split, approximate, smooth). Using the automatically extracted width and orientation information, a tracking algorithm converts the refined linear segments into their equivalent polygonal representations.

4. Geospatial feature inference and classification

4.1. Gabor filtering

A 2D Gabor function $g(x, y)$ in spatial frequency domain is given by,

$$g(x, y) = c(x, y) \times e(x, y) \tag{1}$$

where $c(x, y)$ is a complex sinusoidal, known as the carrier, and $e(x, y)$ is a 2D Gaussian function, known as the envelope.

The complex sinusoidal carrier is defined as,

$$c(x, y) = e^{j(2\pi(u_0x+v_0y)+\phi)} \tag{2}$$

where (u_0, v_0) is the spatial frequency and ϕ is the phase of the sinusoidal. The spatial frequency can also be expressed in polar coordinates as magnitude F_0 and direction ω_0. The 2D Gaussian envelope is defined as,

$$e(x, y) = Ae^{(-\pi(s_x^2(x-x_0)_\vartheta^2+s_y^2(y-y_0)_\vartheta^2))} \tag{3}$$

where A is a scale of the magnitude, (s_x, s_y) are scale factors for the axes, (x_0, y_0) is the peak coordinates and ϑ is the rotation angle.

An attractive characteristic of the Gabor filters is their ability to tune at different orientations and frequencies. Thus by fine-tuning the filters we can extract high-frequency oriented information such as discontinuities and ignore the low-frequency clutter.

We employ a bank of Gabor filters tuned at 8 different orientations θ linearly varying from $0 \leq \theta < \pi$, and at 5 different high frequencies (per orientation) to account for multi-scale analysis. The remaining parameters of the filters in Eq. (3) are computed as functions of the orientation and frequency parameters as in Manjunath and Ma (1996).

The application of the bank of Gabor filters results in a total of 40 response images (8 orientations ×5 frequencies) as shown in the Table 1. The response images corresponding to filters of the same orientation and different frequency are added together. The result is a single response image per orientation (total of 8) which is then encoded using a tensorial representation as explained in Section 4.2.

4.2. Tensor voting

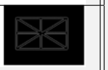
Tensor voting is a perceptual grouping and segmentation framework introduced by Medioni et al. (2000). A key data representation based on tensor calculus is used to encode the data. A point $x \in \mathbb{R}^3$ is encoded as a second-order symmetric tensor T and is defined as,

$$T = \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \begin{bmatrix} \vec{e}_1^T \\ \vec{e}_2^T \\ \vec{e}_3^T \end{bmatrix} \tag{4}$$

$$T = \lambda_1\vec{e}_1\vec{e}_1^T + \lambda_2\vec{e}_2\vec{e}_2^T + \lambda_3\vec{e}_3\vec{e}_3^T \tag{5}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ are eigenvalues, and $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ are the eigenvectors corresponding to $\lambda_1, \lambda_2, \lambda_3$ respectively. By applying the spectrum theorem, the tensor T in Eq. (5) can be expressed as a linear combination of three basis tensors (ball, plate and stick) as in Eq. (6).

Table 1
Gabor filters are applied at 8 different orientations and 5 different high frequencies. Output images of the same orientation (and varying frequency) are grouped together resulting in a total of 8 images (one for each orientation) as shown in the last column. Similarly, the 8 images can then be grouped together resulting in a single image depicting the detected edges.

Freq.($\phi \rightarrow$)-Orient.($\theta \downarrow$)	ϕ_0	ϕ_1	ϕ_2	ϕ_3	ϕ_4	$\sum_{i=0}^{4}(\phi_i)$
θ_0						
θ_1						
θ_2						
θ_3						
θ_4						
θ_5						
θ_6						
θ_7						
$\sum_{j=0}^{7}\theta_j$	×	×	×	×	×	

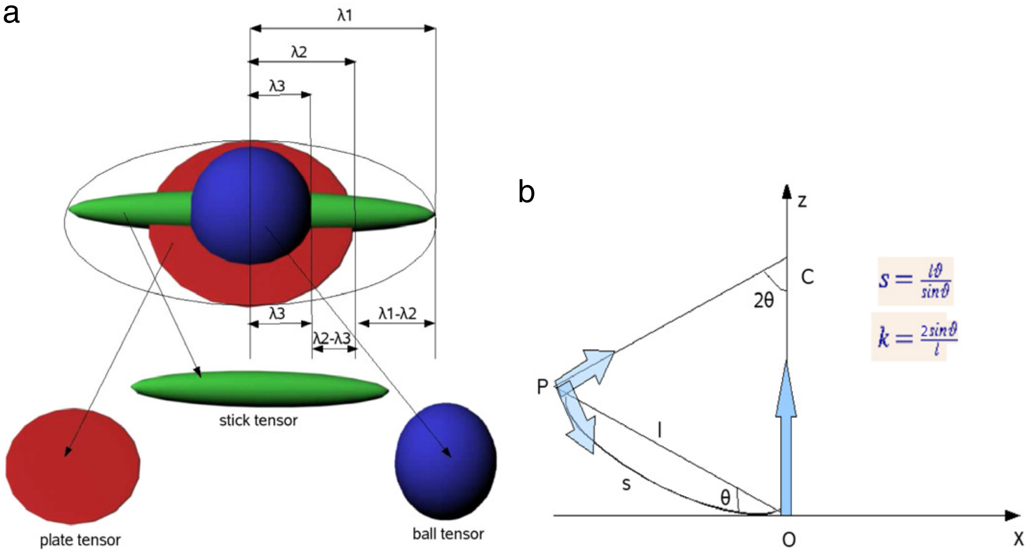


Fig. 2. (a) Tensor decomposition into the stick, plate and ball basis tensors in 3D. (b) Votes cast by a stick tensor located at the origin O. C is the center of the osculating circle passing through points P and O.

$$T = (\lambda_1 - \lambda_2)\vec{e}_1\vec{e}_1^T + (\lambda_2 - \lambda_3)(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T) + \lambda_3(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T + \vec{e}_3\vec{e}_3^T). \quad (6)$$

In Eq. (6), $(\vec{e}_1\vec{e}_1^T)$ describes a stick (surface) with associated saliency $(\lambda_1 - \lambda_2)$ and normal orientation $\vec{e}_1$, $(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T)$ describes a plate (curve) with associated saliency $(\lambda_2 - \lambda_3)$ and tangent orientation $\vec{e}_3$, and $(\vec{e}_1\vec{e}_1^T + \vec{e}_2\vec{e}_2^T + \vec{e}_3\vec{e}_3^T)$ describes a ball (junction) with associated saliency λ_3 and no orientation preference. The geometric interpretation of tensor decomposition is shown in Fig. 2(a).

An important advantage of using such a tensorial representation is its ability to capture the geometric information for multiple feature types (junction, curve, surface) and a saliency, or likelihood, associated with each feature type passing through a point.

Every point (x, y) in the Gabor filter response images computed previously is encoded using Eq. (4) into a unit plate tensor (representing a curve) with the orientation $\vec{e}_3$ aligned to each filter's G_i orientation and is scaled by the magnitude of the response of that point $(G_i \otimes I)_{x,y}$. The resulting eight tensors for each point are then added together which produces a single tensor $T_{(\mathbf{x},\mathbf{y})}$ per point

Fig. 3. (a) Successful handling of discontinuities. Before (left) and after (right) the tensor voting process. (b) Original image of Copper Mountain area in Colorado. (c) Saliency map indicating the refined likelihoods produced by the tensor voting. Green indicates curve-ness ($\lambda_2 - \lambda_3$), blue indicates junction-ness (λ_3). (d) Classified curve features derived from 3(c). Note that no thresholds were used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

capturing the local geometric information and is given by,

$$T_{gabor} = \sum_{i=0}^{8} ((G_i \otimes I)_{x,y} * T_{x,y,i}). \tag{7}$$

Using the tensor decomposition Eq. (6), all pixels for which $(\lambda_2 - \lambda_3) > \lambda_3$ are classified as part of curves with tangent orientation $\vec{e}_3$. Similarly all pixels for which $\lambda_3 > (\lambda_2 - \lambda_3)$ are classified as junction points with no orientation preference.

For example, if a point p_c lies along a curve in the original image its highest response will be at the Gabor filter with a similar orientation as the direction of the curve. Encoding the eight responses of pixel p_c as unit plate tensors, scaling them with the point's response magnitudes and adding them together results in a tensor where $(\lambda_2 - \lambda_3) > (\lambda_1 - \lambda_2)$, $(\lambda_2 - \lambda_3) > \lambda_3$ and the orientation $\vec{e}_3$ is aligned to the direction of the curve i.e. a plate tensor. Similarly a tensor representing a point p_j which is part of a junction will have $\lambda_3 > (\lambda_2 - \lambda_3)$, $\lambda_3 > (\lambda_2 - \lambda_3)$ i.e. a ball tensor.

The encoded points then cast a vote to their neighboring points which lie inside their voting fields, thus propagating and refining the information they carry. The strength of each vote decays with increasing distance and curvature as specified by each point's stick, plate and ball voting fields. The three voting fields can be derived directly from the saliency decay function (Guy and Medioni, 1997) given by

$$DF(s, \kappa, \sigma) = \mathrm{e}^{-\left(\frac{s^2 + c\kappa^2}{\sigma^2}\right)} \tag{8}$$

where s is the arc length of OP, κ is the curvature, c is a constant which controls the decay with high curvature (and is a function of σ), and σ is a scale factor which defines the neighborhood size as shown in Fig. 2(b). The blue arrows at point P indicate the two types of votes it receives from point O: (1) a second-order vote which is a second-order tensor that indicates the preferred orientation at the receiver according to the voter and (2) a first-order vote which is a first-order tensor (i.e. a vector) that points toward the voter along the smooth path connecting the voter and receiver. The scale factor σ is the only free variable in the framework.

After the tensor voting the refined information is analyzed and used to classify the points as curve or junction features. An example of a mountainous area with curvy roads is shown in Fig. 3(b). A saliency map indicating the likelihood of each point as being part of a curve (green) and a junction (blue) is shown in Fig. 3(c). The saliency map is used for the classification of the curve points which are shown in Fig. 3(d). A point with $(\lambda_2 - \lambda_3) > \lambda_3$ is classified as a curve point and a point with $\lambda_3 > (\lambda_2 - \lambda_3)$ is classified as a junction point. Intuitively, a greener point is a curve and a bluer point is a junction.

A key advantage of combining the Gabor filtering and tensor voting is that it eliminates the need for any thresholds therefore removing any data dependencies. The local precision of the Gabor filters is used to derive information which is *directly* encoded into tensors. The tensors are then used as an initial estimate for global context refinement using tensor voting and the points are classified based on the their *likelihoods* of being part of a feature type. This unique characteristic makes the process invariant to the type of images being processed. In addition, the global nature of tensor voting makes it an ideal choice when dealing with noisy, incomplete and complicated images and results in highly accurate estimates about the image features. This is demonstrated in Fig. 3(a) where the original image shows a polygon with many gaps of different sizes in white and the recovered, classified curve points are shown in yellow. As it can be seen most of the discontinuities were successfully and accurately recovered.