Rapid Extraction and Update of Road Network to Caltrans Database

Final Report

METRANS Project 09-17

January 2010

Suya You
Ulrich Neumann

Computer Science Department
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation, University Transportation Centers Program, and California Department of Transportation in the interest of information exchange. The U.S. Government and California Department of Transportation assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views or policies of the State of California or the Department of Transportation. This report does not constitute a standard, specification, or regulation.

ABSTRACT

The goal of this research is to develop new approach and technique to improve and extend the capabilities of creating, modeling and maintaining accurate and up-to-date road infrastructure databases for transportation managements and services. Our research efforts are to assess, define, and use the unique spatial and spectral characteristics of the new, advanced sensor techniques from aerial imagery and LiDAR for automated road extraction and road quality mapping.

A number of theoretical and experimental studies lead us to pursue an innovative approach that merges the power of perceptual grouping with sensor cues, geometric invariants, and machine learning under a unified framework to tackle these problems. This new approach has the potential for automating the extraction and mapping of complex road networks from remote sensing data. In addition, the same process also allows for a constrained optimal estimation of various terrain features and attributes, thereby producing hierarchical data representations under a consistent framework. Most important, we believe that the process of labeling the model elements as buildings, vegetation, roads, and terrains will be possible within this framework. We anticipate a significant step reduction in the human time and effort required to produce and update accurate road models to transportation infrastructure databases.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

DISCLOSURE

This project was funded in entirety under this contract to California Department of Transportation.

ACKNOWLEDGEMENTS

We acknowledge the United States Department of Transportation, California Department of Transportation and METRANS for their interest and generous support to this research. We would also like to thank Airborne I Corporation to provide us certain LiDAR data for this research.

1. INTRODUCTION

Creating and maintaining an accurate and up-to-date road infrastructure database is crucial to many transportation applications including transportation infrastructure management, traffic situational awareness, safety analysis, and mission planning and tactical decision-making for incident and emergency responses. Nowadays, Intelligent Transportation Systems (ITS) continuously gain ground in transportation management. A digital topographic database is an essential part of the ITS, which requires accurate, high-density spatial models of road infrastructures. In addition, accurate road maps and databases are high public demand for travel planning, route guidance, and real-time travel navigation, etc.

A complete road network model consists of road segments and intersections that join the road sections. In particularly, accurately identifying and labeling of the road intersections in road networks is important, especially for safety management. Intersection safety has become a serious problem in the United States. Intersection and intersection-related crashes consistently make up a high proportion of total fatal crashes. For example, in 2004 more than 2.7 million intersection-related crashes occurred, accounting for more than 45 percent of all crashes in the United States [FHWA]. FHWA has been initiating a variety of researches and strategies focus on improving intersection safety, including development of a comprehensive database of road intersections for major US roadways. Many states including California also started to conduct similar efforts and strategies: to develop road intersection databases, match traffic crashes to the intersections, calculate crash rates for various types of intersections, and identify and improve the intersection infrastructures with highest crash rates.

However, current road infrastructure databases and the methods to produce such databases are insufficient to meet these needs in terms of accuracy, confidence, completeness, and automation. Many existing digital maps and road databases are still generated from the old paper based topographical maps, which may contain significant spatial errors. Accurate road databases do not yet

exist for vast areas, particularly in areas with rapid expansion. Many existing infrastructure databases need to be updated to capture new road condition and expansion.

Recently, the US Department of Transportation (USDOT) has extended great efforts in use of GPS and advanced remote sensing technologies for road mapping and condition assessment to meet these information needs [USDOT]. These methodologies, though effective, poses new challenges in science and technology related to road mapping and feature extraction. GPS is useful for tracking and localization of road features such as road intersections, but it is time-consuming and costly, because every road within a DOT's jurisdiction must be field-visited to obtain accurate localization information. Remote sensing techniques such as measures from aerial imagery and LiDAR (Light Detection and Ranging) provides one means by which large areas can be rapidly mapped with high standard of accuracy, but technology that use these sensor data for detection and mapping of road and transportation networks is still in its infancy. Existing techniques and approaches are typically characterized by poor detection rates, low confidences, and a requirement for extensive operator interaction. The USDOT has therefore extended high demand for developing new technologies that can rapidly and effectively create accurate and up-to-date road infrastructure models and databases to facilitate advanced transportation management and public traveling requirements.

This report summarizes the research efforts and accomplishments for the project: Rapid Extraction and Updating Road network to Caltrans Database. The goal of this research is to pursue a step change in the approach and technique for improving and extending the capabilities of creating, modeling and maintaining accurate and up-to-date road infrastructure models for transportation managements and services. Our research efforts are to assess, define, and use the unique spatial and spectral characteristics of the new, advanced sensor techniques from aerial imagery and LiDAR for automated road feature extraction and road quality mapping.

2. LITERATURE REVIEW

Road feature extraction and mapping is a long-standing and difficult problem. Historically, the road maps are mainly produced during the road survey and construction stages, requiring time-intensive and tedious manual processing by skilled operators. Later on, the techniques from remote sensing have been widely employed for road extraction and mapping. Over the years, a wealth of research, employing a variety of sensing technologies, has been reported from geospatial and computer vision communities. Technically, the methods for road extraction from remote sensing data (aerial photographs, satellite images and LiDAR) can be classified as three categories: pixel-based approach, region-based approach and knowledge-based technique.

2.1 Pixel-based approach

In [Bau99], lines are extracted in an image with reduced resolution as well as roadside edges in the original high-resolution image. Similarly, [Lis04] uses a line detector to extract lines from multiple scales of the original data. [9] applies the edge detector on multi-resolution images and uses the result as input to the higher-level processing phase. [Wes04] applies Steger's differential geometry approach for ine extraction. [Bar03] uses a Deriche operator for edge detection with an added hysteresis threshold, followed by an edge smoothing using the Ramer algorithm.

[Lap00] uses a multi-scale ridge detector to detect lines at coarser scale, and then uses a local edge detector at a finer scale to extract parallel edges which are optimized using a variation of the active contour models technique (snakes). [Del05] presents a technique where a directional adaptive filter is used for the detection of pixels with particular orientation. Similarly, [Por03] achieves excellent results by using a Gaussian model based approach. In order to extract the road magnitude and orientation for each point, they use a quadruple orthogonal line filter set.

2.2 Region-based approach

[Zha06] uses predefined membership functions for road surfaces as a measure for image segmentation and clustering. Likewise, in [Clo05] they use the reflectance properties, from the ALS data and perform a region-growing algorithm to detect the roads.

Hierarchical network can be used to classify and segment objects. A slightly different approach is proposed in [Lis04] where a line detector and a classification algorithm are applied on multiple scales of the original data and the results are then merged.

2.3 Knowledge-based approach

In [Wes04], human input is used to guide a system in the extraction of context objects and regions with associated confidence measures. The system in [Zha01] 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. [4] uses a rule-based algorithm for detection of buildings at the first stage and then at the second stage the reflectance properties of the road. Similarly, [Zha06] uses reflectance as a measure for the image segmentation and clustering.

Explicit knowledge about geometric and radiometric properties of roads is used in [Wes04] to construct road segments from the hypotheses of roadsides. In [Bar03] 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. [Pet03] takes high resolution images as input along with prior knowledge about the roads e.g. road models and road properties.

While considerable attention has been recently given to the development of automated techniques for road network extraction and mapping, it still remains a challenge due to the wide variations of roads and the complexities of environments. Most of current methods applied to extract roads from open or rural areas were successful to some extent due to the relative simple scene and

road model. For the extraction of roads in dense urban areas, especially from high-resolution imagery, there are primary obstacles which lead to unreliable extraction results: complicated image scene and road model, furthermore, occlusions caused by buildings, trees, vegetations and their shadows. The lack of sufficient topographic information about the scene, especially three-dimensional information is the principle difficulty in obtaining the road information with high reliability and accuracy in urban areas.

2.4 LiDAR (Light Detection and Ranging)

Recently, airborne LiDAR has become a rather important data source for generating high quality 3D topographic models. A LiDAR sensor system permits an aircraft flyover to quickly collect a height field for a large urban environment with an accuracy of centimeters in height and sub-meter in ground position. Multiple passes of the aircraft are merged to ensure good coverage. Due to its advantages as an active sensing technique for reliable 3D determination, LiDAR offers a fast and effective way to acquire topographic models for a large urban site. The USDOT has suggested the LiDAR technique as a new resource with great potential in transportation infrastructure mapping. In fact, the USDOT and Caltrans have conducted many related projects to use the LiDAR technology for transportation. For example, the Caltrans' GeoResearch Group, in collaboration with the Department of Conservation, successfully used LiDAR technology to map landslides along two heavily forested highway corridors in Humboldt and Del Norte Counties [CADOT]. AHMCT researchers are developing a coordinated set of standards and specifications for the use of laser scanning in Caltrans projects [AHMCT]. METRANS has conducted projects of using LiDAR for generating terrain DEM models and for feature extraction to Caltrans databases [METR].

Over the years, we have conducted a range of researches related to the 3D modeling, building extraction and modeling, and geospatial feature extraction and mapping from remote sensing imagery and LiDAR [Pou07, Pou08, Wan07, Hu06, Hu03]. This developed work is built on these results by extending our work with new capabilities to rapidly detect and extract urban road networks from

various sensor resources including aerial photographs, satellite images, and LiDAR. We developed a novel vision-based system for automatic detection and extraction of complex road networks from remote sensing data. Uniquely, our approach merges the power of perceptual grouping theory and optimized segmentation techniques into a unified framework to address the challenging problems of road feature detection and classification. The extensive experimental results show the feasibility and advantages of our approach in terms accuracy, confidence, completeness, and automation.

3. METHODOLOGY

The main efforts of our research focus on developing new techniques to rapidly detect, extract, and map complex road networks from high-resolution aerial imagery and airborne LiDAR. A number of theoretical and experimental studies lead us to pursue an innovative approach that merges the power of perceptual grouping (tensor voting, Gabor wavelets) and classification (global optimization using graph-cuts) with sensor cues, geometric invariants, and machine learning under a unified framework to tackle these problems. The new approach has the potential for automating the process of extraction and mapping of complex road networks from data acquired from varied data sources. In addition, the same process also allows for a constrained optimal estimation of various terrain features and attributes including edges, curves, junctions, and surfaces and their relationships, thereby producing hierarchical data representations under a consistent framework. Most importantly, we found that the process of labeling the model elements as buildings, vegetation, roads, and intersections is possible within this framework.

3.1 System Overview

Figure 1 depicts the architecture of our approach and its major functions. Technically, the approach consists of three major components: geospatial feature inference and classification, road feature detection and labeling, and road network extraction and identification.

We developed a **Gabor Wavelets based technique** to extract the significant spatial and frequency information of sensor data, producing multi-channel feature maps (orientation and location) that clearly indicate the signatures of geospatial features at different spatial and frequency resolutions. The multi-resolution representation is important not only for the processing of multiple sensor data that have varied resolutions, but also for the problem of road detection itself. Roads typically vary in size from narrow local lanes to wide highways. There are methods that work well at one scale but fail at others. The proposed approach

enables us to apply the same algorithm at different scales to extract and label a wide range of road types.

We introduced a novel technique called **Tensor Voting** for road feature detection and intersections identification. Tensor voting is a new perceptual grouping theory, motivated from a perceptual organization formalism that addresses the problems of structure inference from sparse or noisy data [Med00]. Over the past several years, many researchers have successfully applied the theory to tackle many low-level vision problems in a unified computational framework that implements a smoothness constraint to generate object descriptions in terms of surfaces, regions, curves, and labeled junctions. The tensor voting approach can simultaneously capture the varying geometry elements (i.e. road surfaces, curves, and junctions) and their local structure information (i.e. locations, orientations, and confidences).

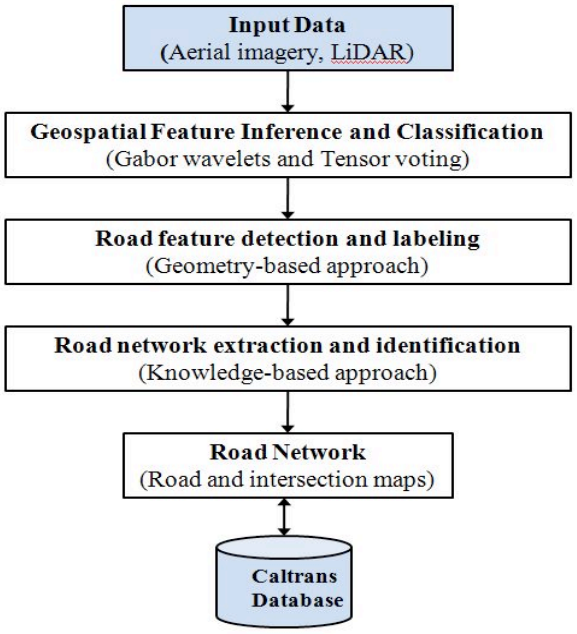


Figure 1: Architecture of the proposed road extraction system.

We employed **high-level scene knowledge** about roads to enhance the robustness and scalability of road network extraction and modeling. Many existing road extraction techniques rely on low-level information such as the magnitude and orientation of edges to link the individual edge points. Smoothness constraints on local orientation are useful for road network tracking because roads are often straight or smoothly curved. However, the assumption of intensity smoothing is not applicable to road tracking because moving vehicles, people, and shadows from trees or houses can produce rapid intensity changes. The high-level scene knowledge about roads, such as local surface characteristics, topology, and spectral properties can provide powerful

constraints to enhance the robustness of road network tracking by resolving local ambiguities that occur when a region has multiple similar or overlaid features. The following sections detail the research advances for each component.

3.2 Extraction and Classification of Geospatial Features

We integrate the mathematics of Gabor Wavelets and Tensor Voting to automatically extract the significant signatures of geospatial features at different spatial and frequency resolutions.

3.2.1 Gabor Filtering

Gabor filtering theory has received considerable attention in visual information processing, because the characteristics of certain cells in the human visual cortex of some mammals can be approximated by Gabor filters. In addition, the Gabor filters have been shown to posses optimal localization properties in both spatial and frequency domain and thus are well suited for the road feature extraction problems.

We employ a bank of Gabor filters tuned at multiple orientations and frequency resolutions (8-orientations and 5-frequencies) to produce multi-channel geospatial feature maps (orientation and location) from imagery and LiDAR data. A two-dimensional Gabor filtering function $g(x,y)$ in space domain is given by:

$$g(x,y) = e^{j(2\pi(u_0 x + v_0 y) + \phi)} \times \kappa e^{(-\pi(s_x^2(x-x_0)_\theta^2 + s_y^2(y-y_0)_\theta^2))}$$

where (u_0, v_0) is the spatial frequency, ϕ is the phase of the sinusoidal, κ 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. The remaining parameters in the equation are computed as functions of the orientation and frequency parameters.

The application of the bank of Gabor filters results in a total of 40 response images (8 orientations x 5 frequencies). The response images corresponding to

the filters having the same orientation and different frequency are added together. The result is a single response image per orientation (8), which is then encoded using a tensorial representation as explained in next section. Figure 2 shows a result of the Gabor filter responses to an urban road image.

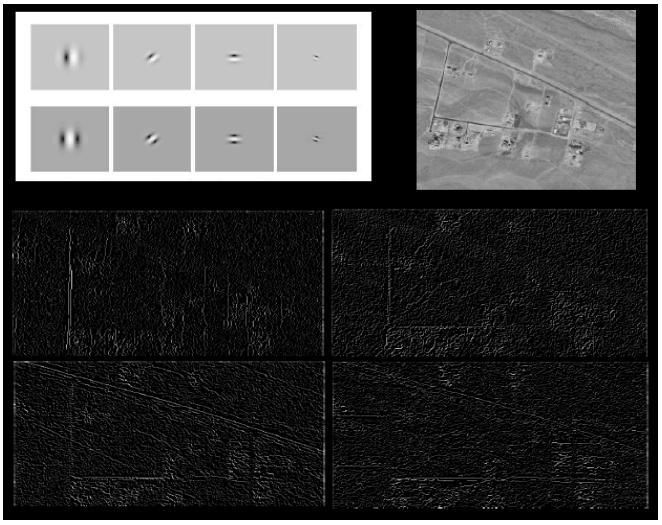
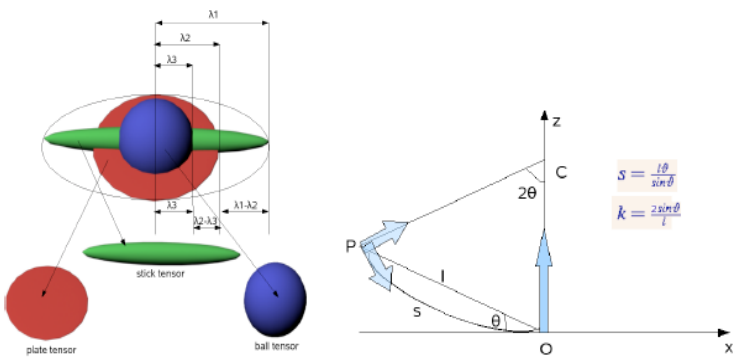


Figure 2: Use Gabor filter bank to extract geospatial features.

3.2.2 Tensor Voting

Tensor voting is a new perceptual grouping introduced by [Med00]. Mathematically, tensors are a set of functions that encode the desired information of data structure. For example, in 2D, a very salient curve element at a point can be represented by a thin ellipse whose major axis represents its tangent direction and whose length reflects its saliency of the point. The advantage of using a tensor representation over a vector representation is its powerful capacity to encode locally curved geometry as orientation and confidence or saliency. Consider a simple case of a symmetric tensor T with two perpendicular eigenvectors $\vec{e}_1$, $\vec{e}_2$ and two corresponding real eigenvalues $\lambda_1 > \lambda_2$. Depending on the eigenvalues, the points $\vec{x}$ that satisfy equation $\vec{x}^T T\vec{x} = k$ for a constant k give rise to three different interpretations: (i) if $\lambda_1 = \lambda_2$, the equation corresponds to a circle; (ii) $\lambda_1 = 0$ or $\lambda_2 = 0$, all points coincide with one of the eigenvectors; and (iii) if $\lambda_1 > 0$ and $\lambda_2 > 0$, the equation corresponds to an ellipse. Using the tensor voting, we can quickly infer the structure for each above case: locally un-oriented regions in the first case; a perfectly oriented region in the second case; and all the cases in between as the third case.

We employ the tensor calculus as a tensorial representation to encode data.



(a) (b)

Figure 3: Geometrical interpretation of tensor decomposition. (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*.

A point $x \in R^3$ encoded as a second order symmetric tensor T is defined as:

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

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

By applying the spectrum theorem, the tensor T in above equation can be expressed as a linear combination of three basis tensors (ball, plate and stick):

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

In this equation, $(\vec{e}_1\vec{e}_1^T)$ 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 geometrical interpretation of tensor decomposition is shown in Figure 3 (a).

3.2.3 Combining Gabor Wavelets and Tensor Voting

Every point in the Gabor filter response images computed previously is encoded using the tensor equation into a unit plate tensor (representing a curve) with the orientation $\vec{e}_3$ aligned to the filter orientation and is scaled by the magnitude of the response of that point. The resulting eight tensors for each point are then added together which produces a single tensor per point capturing the local geometrical information. To summarize, 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 the point p 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 that is part of a junction will have $\lambda_3 > (\lambda_2 - \lambda_3)$, i.e. a ball tensor.

The encoded points then cast a vote to their neighboring points 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. Figure 3 (b) depicts geometrical interpretation of the voting process. 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 indicating 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.

After the tensor voting, the refined information is analyzed and used to

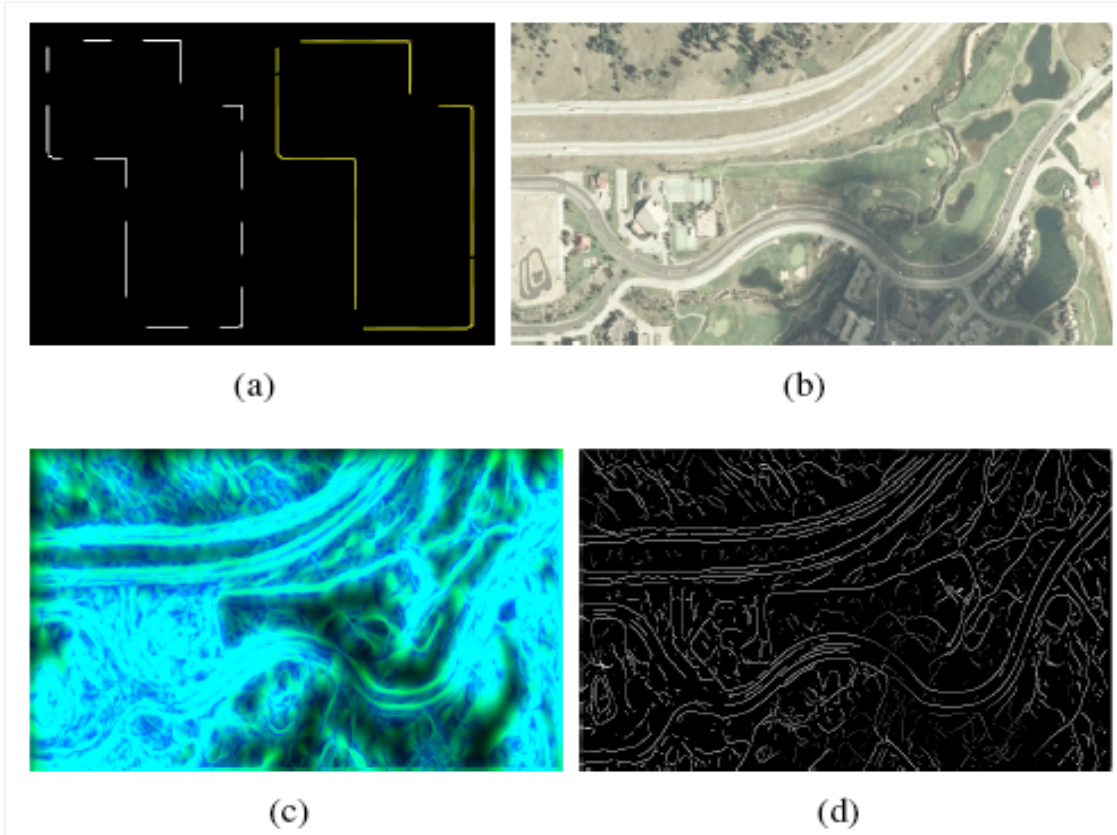


Figure 4: Using the proposed approach to automatically extract the significant signatures of geospatial features. (a) Simulated polygon demonstrates the approach to handle discontinuities: before (left) and after (right) applying the tensor voting process. (b) Original image of Copper Mountain area in Colorado. (c) Saliency map indicating refined likelihoods produced by tensor voting. (d) Classified curve features derived (c). Note that the processing is fully automatic, and no threshold parameter requires tuning.

classify the points as curve or junction features. Figure 4 shows an example of a mountainous area with curvy roads. The extracted saliency map indicating the likelihood of each point as being part of a curve (green) and a junction (blue) is shown in Figure 4 (c). The classification of the curve points based on the saliency map is in Figure 4 (d). A point with $(\lambda_2 - \lambda_3) > \lambda_3$ is classified as a curve point and a point with $\lambda_3 > (\lambda_2 - \lambda_3)$ has been classified as a junction point. Intuitively, a greener point is a curve and a bluer point is a junction.

The developed approach of combining the Gabor wavelets and tensor voting has been proved to be very efficient in dealing with noisy, incomplete, and complex datasets. It is able to not only describe unambiguously varying geometry elements (i.e. point, curve, junction, and surface) under a unified framework, but it also encodes considerably richer structure information for feature inference and classification. The Gabor filters produce multi-channel geospatial feature maps at different spatial and frequency resolutions. The tensor voting effectively captures the first-order differential geometry and its singularities (feature locations, orientations, and confidences). The feature inference is based on a voting communication process that is governed by a perceptual-field, encoding the constraints and rules of how a point receives/casts votes from/to its neighbors. The accumulation of votes at each point provides an accurate estimate of feature saliency at the point. By combining the voting results for all channels, the extracted geospatial features are further classified, and finally converted to road vector maps (road tracks, junctions, and surfaces).

3.3 Detection and Labeling Road Features

The above step of geospatial feature inference can extract the significant signatures of geospatial features and their attributes from sensor data. However, these extracted geospatial features may also contain non-road features such as buildings, vegetations and their shadows. A further filtering step is therefore needed to classify and label the road features of interest from the extracted geospatial feature sets.

Feature space clustering is currently a popular technique for feature extraction and classification, in which a feature vector of local properties (such as intensity or texture) is computed at each pixel. The feature space is then clustered, and each pixel is labeled with the cluster that contains its feature vector. A major limitation of this technique to separate the feature space clustering and pixel labeling as two independent phases is that the feature space clusters generally lack sufficient spatial coherence for effective pixel labeling. A good cluster in feature space will often not be coherent in image space.

Similarly, the correct segmentation in image space may not correspond to a highly distinctive group of feature vectors. The space-clustering algorithm could miss potential clusters which are less distinctive in feature space, but which give rise to spatially coherent segmentations. Furthermore, most of current approaches that use only position information tend to result in poor feature segmentation and classification, especially when the desired features have similar geometry structures as their neighbors or backgrounds.

In this research, we developed an effective approach for road feature detection and labeling that operates simultaneously in feature space and in image space. The approach is an improved graph-cut technique that incorporates both the feature location and orientation information inferred by the tensor voting into the processes of road feature detection and labeling. We define an energy function over both a set of feature clusters and a labeling of pixels with clusters. In our approach, a pixel is labeled with a single cluster, rather than, for example, a distribution over clusters. The energy function penalizes clusters that are a poor fit to the data in feature space, and also penalizes clusters whose pixels lack spatial coherence in image space. The energy function is efficiently minimized using graph-cuts technique to extract and label the road structures from geospatial feature sets.

3.3.1 Graph-cut to Segment Road Features

Technically, we can treat the image feature segmentation as a graph partition problem. Given an input image I, an undirected graph $G =< V, E >$ is created, where each vertex $v_i \in V$ corresponds to a pixel $p_i \in I$ and each undirected edge $e_{i,j} \in E$ represents a link between neighboring pixels $p_i, p_j \in I$. In addition, two distinguished vertices called terminals V_s, V_t are added to the graph G. An additional edge is also created connecting every pixel $p_i \in I$ and the two terminal vertices, e_i, V_s and e_i, V_t. For weighted graphs, every edge $e \in E$ has an associated weight w_e. A cut $C \subset E$ is a partition of the vertices V of the graph into two disjoint sets S, T, where $V_s \in S$ and $V_t \in T$. The cost of each cut C is the

sum of the weighted edges $e \in C$. The minimum cut problem can then be defined as finding the cut with the minimum cost which can be achieved in near polynomial-time.

The graph-cut process is performed to segment only the road features from the classified curve features. The geometric structure of the curve features combined with color information extracted from image is used to guide an orientation-based segmentation using optimization by graph-cuts, which produces a labeling of road and non-road candidates.

3.3.2 Labeling

The above binary case can easily be extended to a case of multiple terminal vertices. We create two terminal vertices for foreground O and background B pixels for each orientation θ for which $0 \le \theta \le \pi$. In our experiments, we have found that choosing the number of orientation labels in the range $N_\theta = [2,16]$ generates the best results. Thus the set of labels L is defined to be $L = \{O_{\theta_1}, B_{\theta_1}, O_{\theta_2}, B_{\theta_2}, ...O_{\theta_N}, B_{\theta_N}\}$ with size $|L| = 2 * N_\theta$.

3.3.3 Energy Minimization Function

Finding the minimum cut of a graph is equivalent to finding an optimal labeling $f : I_p -> L$, which assigns a label $l \in L$ to each pixel $p \in I$, where f is piecewise smooth and consistent with the original data. Thus, our energy function for the graph-cut minimization is given by:

$$E(f) = E_{data}(f) + \lambda * E_{smooth}(f)$$

where λ is the weight of the smoothness term.

Energy data term: the data term provides a per-pixel measure of how appropriate a label $l \in L$ is, for a pixel $p \in I$ in the *observed* data and is given by:

$$E_{data}(f) = \sum_{p \in I} D_p(f(p))$$

The initial seed points are used twice. Firstly, we use the seeds to compute an intensity distribution (in our case of color distribution, using the Gaussian mixture models) for the background and foreground pixels. A measure of how appropriate a labeling is, is then given by computing the negative log-likelihood i.e. $-\ln(P(I_p \mid f(p)))$. Secondly, we used the seeds to encode the hard constraints for the segmentation. Foreground and background pixels are assigned the lowest and highest value of the function $D_p(f(p))$, respectively. For all other pixels, D_p is computed as:

$$D_p(f(p)) = \frac{1 - ln(P(I_p|f(p)))}{2 - ||\theta_p - \theta_{f(p)}||^2}$$

Thus, the energy data term then becomes:

$$E_{data}(f) = \sum_{p \in I} (\frac{1 - ln(P(I_p|f(p)))}{2 - ||\theta_p - \theta_{f(p)}||^2})$$

Energy smoothness term: the smoothness term provides a measure of the difference between two neighboring pixels $p,q \in I$ with labels $l_p, l_q \in L$ respectively. Let I_p and I_q be the intensity values in the *observed* data of the pixels respectively. Similarly, let θ_p and θ_q be the initial orientations for the two pixels recovered as explained in Section 3.2.2. We define a measure of the *observed* smoothness between pixels $p,q \in I$ as:

$$\Delta_{p,q} = \frac{1 + (I_p - I_q)^2}{2 - ||\theta_p - \theta_q||)^2}$$

In addition, we define a measure of smoothness for the global minimization.

Let $I_{f(p)}$ and $I_{f(q)}$ be the intensity values under a labeling f. Similarly, let $\theta_{f(p)}$ and $\theta_{f(q)}$ be the orientations under the same labeling. We define a measure of the smoothness between neighboring pixels under a labeling f as:

$$\widehat{\Delta_{p,q}} = \frac{1 + (I_{f(p)} - I_{f(q)})^2}{2 - ||\theta_{f(p)} - \theta_{f(q)}||^2}$$

Using the smoothness measure defined for the observed data and the smoothness measure defined for any given labeling, we can finally define the energy smoothness term as follows:

$$E_{smooth}(f) = \sum_{\{p,q\} \in N} V_{\{p,q\}}(f(p), f(q))$$

$$E_{smooth}(f) = \sum_{\{p,q\} \in N} K_{p,q} * \widehat{\Delta_{p,q}}$$

where N is the set of neighboring pixels, $K_{p,q} = [e^{-\frac{\Delta_{p,q}^2}{2*\sigma^2}}]$, and σ controls the smoothness uncertainty. Intuitively, if two neighboring pixels have similar intensity and similar orientation in the observed data, then $\Delta_{p,q}$ will be small and thus there is a high probability of $\hat{\Delta}_{p,q}$ being small. To summarize, the function $E(f)$ penalizes heavily for severed edges between neighboring pixels with similar intensity and orientation, and vice versa.

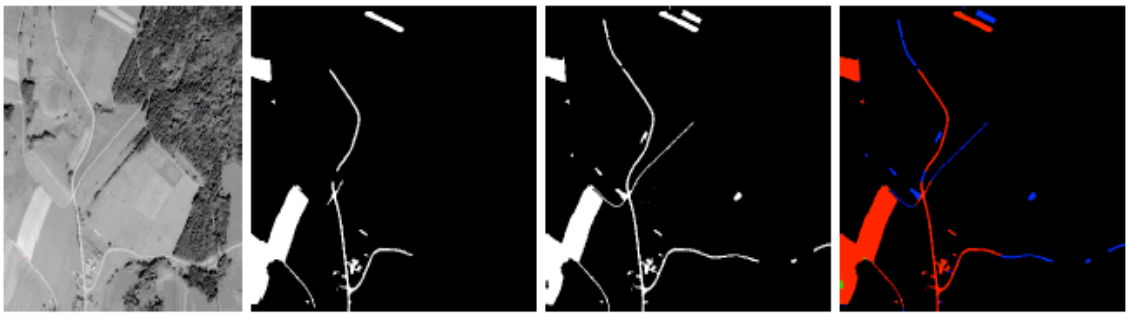


Figure 5: Comparison between the traditional intensity-based segmentation approach and the developed orientation-based technique. (a) Original image. (b) Intensity-based segmentation. (c) Orientation-based segmentation. (d) Color-coded segmentation difference (Red: common points, Green: only in intensity segmentation, Blue: only in orientation segmentation).

An advantage of the developed orientation-based segmentation is that by incorporating orientation information in the optimization process, it ensures that linear segments are not severed, even in the case where the color difference between neighboring pixels is relatively significant. By using the classified curve features to guide the segmentation process, we combine the fast computational times of graph-cuts and the high-accuracy of the information derived using the perceptual grouping to produce results with better defined boundaries compared to traditional segmentation techniques, as shown in Figure 5.

3.4. Extraction of Road Network and Identification of Intersections

Once detected road features, we extract complete road network and their attributes (road widths, centerlines, and road intersections), and convert them to road vector representations.

We developed a knowledge-based technique to enhance the robustness and scalability of road network extraction and modeling. A bi-modal Gaussian model is used to model the road segments and to extract road widths and centerlines.

The road widths are estimated by detecting two road-side points with the bi-modal Gaussian filter (i.e. road-side filter). In addition, a single mode Gaussian road-area filter is also applied to the labeled road feature sets to ensure that the feature points having the maximum responses to the road-side filter are indeed a part of road segments.

3.4.1 Road Centerline Extraction and Linearization

The extraction of the road centerlines is performed using a set of Gaussian-based filters. A bi-modal filter is employed to detect parallel-lines and is defined as a mixture of Gaussian kernels given by:

$$G_b = \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}}[e^{-[\frac{(x-\frac{w}{2})_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}]} + e^{-[\frac{(x+\frac{w}{2})_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}]}]$$

where there $(...)_r$ subscript stands for a rotation operation such that

$$(x-\frac{w}{2})_r = (x-\frac{w}{2})cos(\phi) + ysin(\phi)$$

$$y_r = -(x-\frac{w}{2})sin(\phi) + ycos(\phi)$$

where ϕ is the orientation of the filter and w is the distance between the peaks. The bi-modal filter is shown in Figure 6 (a).

The bi-modal filters of different orientations ϕ and widths w are applied to the classified curve features computed previously as explained in Section 3.3. In order to overcome the problems arising from the coindicidental presence of two

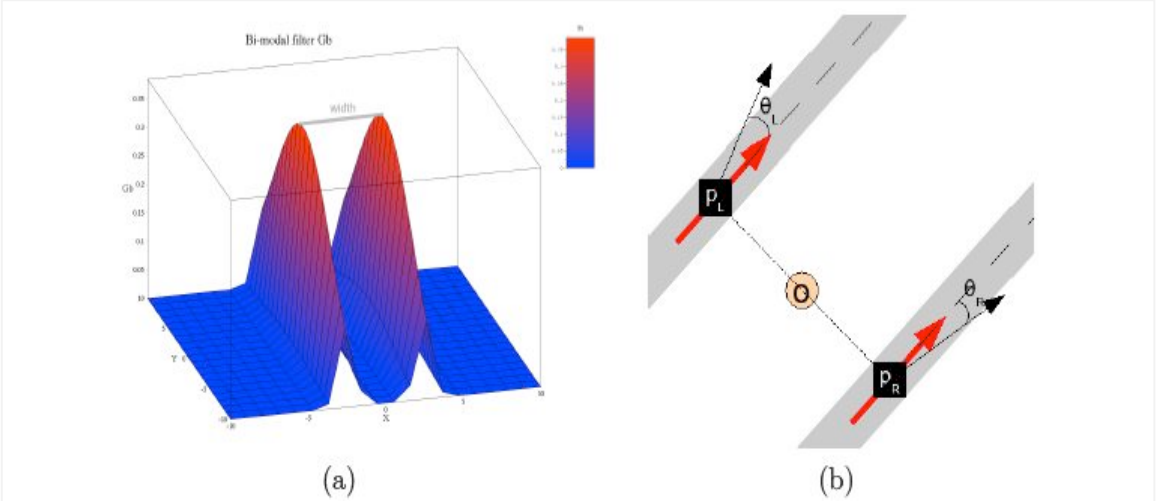


Figure 6: Bi-modal Gaussian model used for extracting road widths and centerlines. (a) The bi-modal filter G_b is applied to the classified curve features. (b) Red arrows: filter orientation (at peaks). Black arrows: actual pixel orientation.

curve pixels along the filters' peaks, the orientation information is used to weigh the response. This ensures that the maximum response only occurs when both pixels have the same orientation and are aligned to the filter's orientation. Figure 6 (b) demonstrates the application of a bi-modal filter to a point O. The orientations θ_L and θ_R of the left and right road side points p_L and p_R respectively are used to scale the response. Thus, the above Gaussian kernel

$$G_b = \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}}[cos(\theta_L)e^{-[\frac{(x-\frac{w}{2})_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}]} + cos(\theta_R)e^{-[\frac{(x+\frac{w}{2})_r^2}{\sigma_x^2}+\frac{y_r^2}{\sigma_y^2}]}]$$

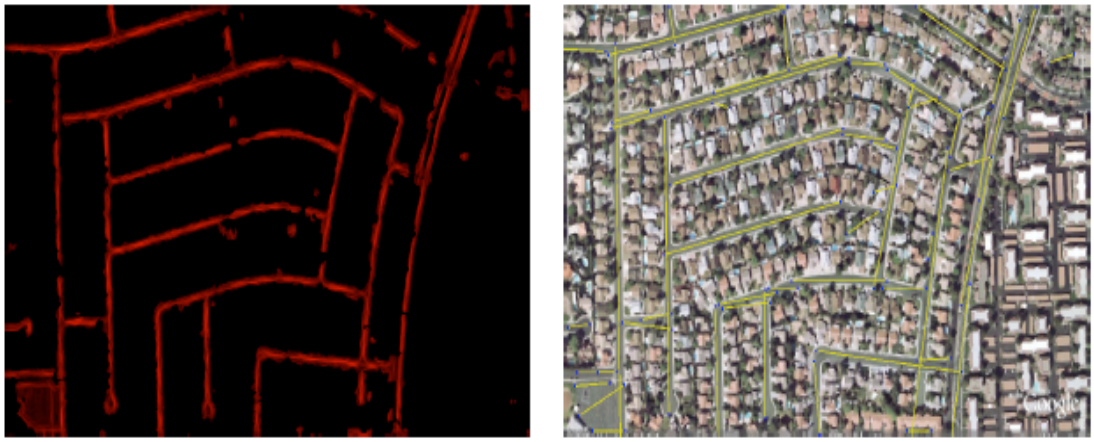
equation becomes

In addition to the bi-modal filters, single mode Gaussian filters are applied to the segmented binary image containing the road candidates. This ensures that the area between any parallel lines is indeed a part of the road and therefore should appear in the result of the segmentation.

Single mode and bi-modal filters of different widths and orientations are

combined as $G_t = G_b * G_s$ and are used for the extraction of centerline information. A point along the centerline of a road of orientation θ_R and width w_R will have a maximum response to a filter with the same or similar orientation and width. Thus, for each pixel we record the filter parameters (orientation, width) for which it returns a maximum response.

Finally, the centerline response magnitudes are used as votes in an iterative Hough transform. This has the significant advantage that no input parameters are required for the Hough transform, such as number of peaks, minimum vote thresholds, etc. therefore making the linearization process entirely automatic. The result is a set of lines representing the segments of the road network as shown in the examples of Figure 7 and Figure 8. The majority of the centerlines are correctly extracted automatically. However, some false positives still exist, which needs following road tracking process.



(a) (b)

Figure 7: Automatic extraction of road network and attributes including road widths (green), centerlines (yellow), and road intersections (blue) for the Oregon dataset. (a) The response magnitude map computed by the Gaussian filters is used for the voting of Hough transform. (b) The majority of road network and attributes are automatically extracted.

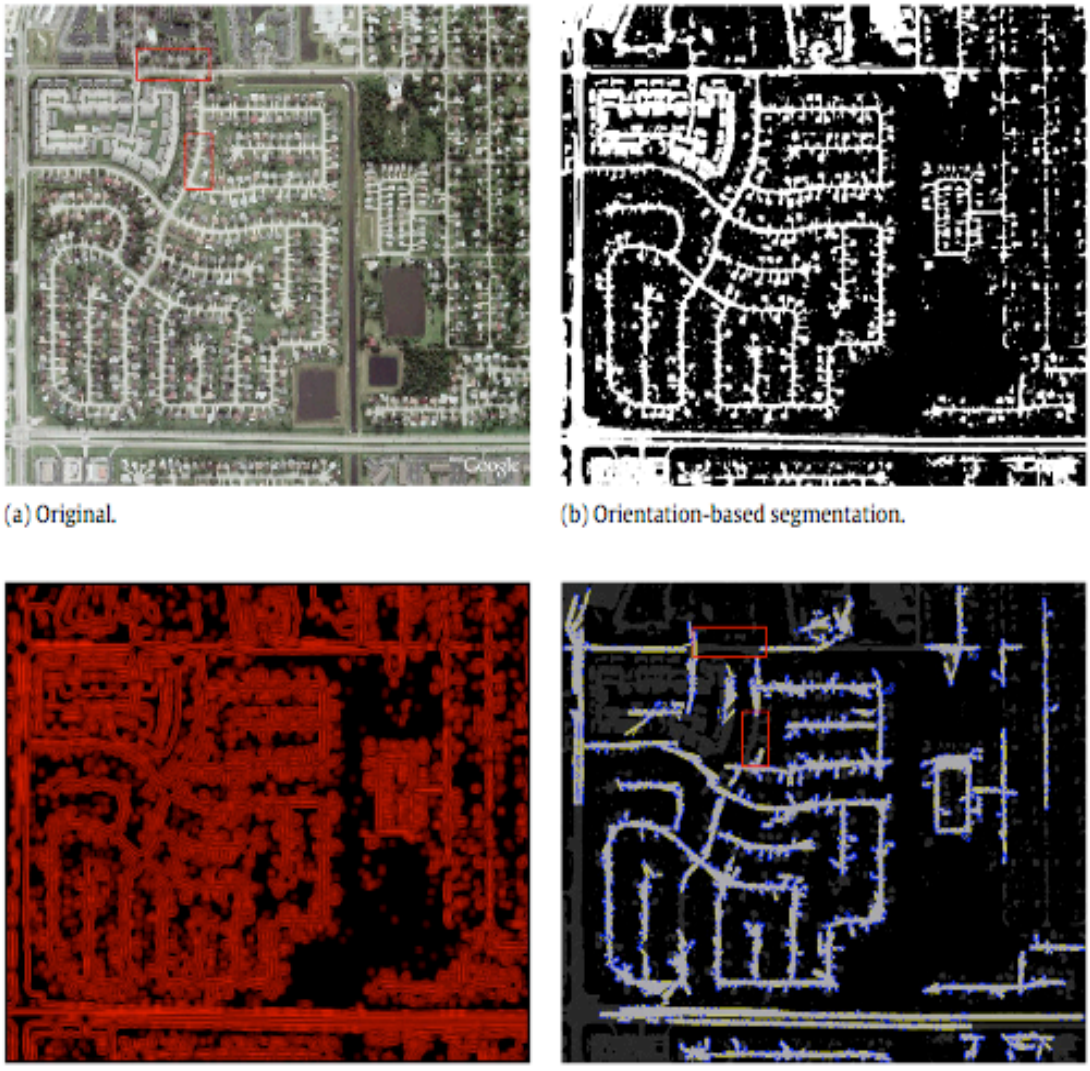


Figure 8: Automatic extraction of road network from high-resolution satellite image of an urban site (Kentucky dataset). The marked areas in (d) show the miss-detected roads due to occlusions and shadows.

3.4.2 Detection Road Intersections

Road intersections are extracted by using a model-based approach applied to the geospatial feature sets detected by tensor voting. As stated above, the tensor voting is able to not only estimate feature saliency, but it can also identify unambiguously the types of varying geometry elements such as point, curve, junction, and surface. The detected geospatial features represented as tensorial representation encode richer structure information for possible intersection features together with associated confidence/uncertainty measures. This leads us a natural way to merge the power of the tensorial representation with model-based approach to solve the problem of road intersection detection.

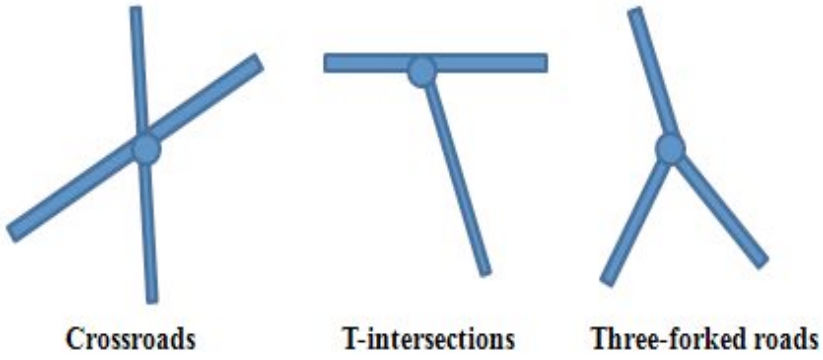


Figure 9: Basic road intersection models used for detecting road intersections. Any complex intersection can be modeled as a combination of the three basic models.

We establish road intersection models based on the road geometric characteristics. Each road segment is represented as an elongate rectangle that has constant width and length. The center position of junction is defined as a road intersection where two or more roads segments either meet or cross at grade. Under this definition, we classify the road intersections as three basic types: crossroads: representing the intersection of two road portions, T-intersections: consisting of one straight road and connected branch, and three-forked roads: having three road segments where each branch has different direction (Figure 9). Any complex intersection can be modeled as a combination of the three basic models.

Once defined, the road intersection models are matched to those possible junction points detected by tensor voting. The matching measure is based on the saliency measure encoded in the feature's tensorial representation, including orientation, confidence, and junction type. A point having a maximum match score to the road intersection models is identified as a road intersection. Figures 7, 8 show the results of extracted road network and its attributes including road widths, centerlines, and intersections.

3.4.3 Road Tracking and Network Completion

Using the automatically extracted width and orientation information computed by the filters, a tracking algorithm converts the linear segments into their equivalent polygonal representations i.e. road segments. In some cases where the road network is particularly complex, the automatically extracted linear segments may contain false positives and false negatives. For such cases, we developed an interactive approach for the further refinement, which can have the form of several actions outlined below.

1. Adding a seed point: Once a seed point is added the filters are applied to derive the width and orientation information. The system then recursively performs a local neighborhood search to find a candidate pixel that minimizes the function:

$$f(x,y) = argmin(w_d * D_{(x,y)} + w_\theta * O_{(x,y)} + w_w * (W_{(x,y)}))$$

where $D_{(x,y)}$ is the Euclidian distance between the candidate and the seed point, $O_{(x,y)}$ is the orientation difference, $W_{(x,y)}$ is the width difference, and w_d, w_θ, w_w are weights corresponding to each term, respectively. The weights are experimentally determined. In all of our experiments, their values are defined as constants: $w_d = 0.4, w_d = 0.3, w_w = 0.3$. This process is recursively repeated and each candidate point that minimizes $f(x,y)$ is added to the current line until no more neighboring points are found.

2. Adding or editing a centerline: Once a centerline is added the filters are applied at a fixed orientation aligned to the specified centerline's slope.

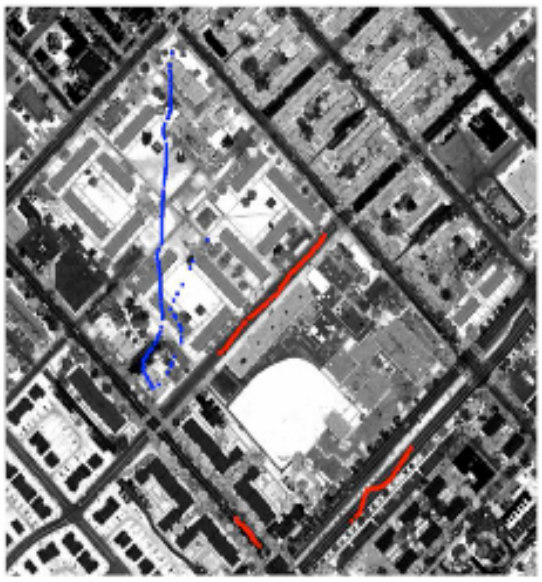
3. Merging of two centerlines. Given two centerlines a Hermite spline is fit between the most appropriate endpoints resulting in a single merged centerline.

4. Deleting a centerline: delete the road centerline and all the associated points.

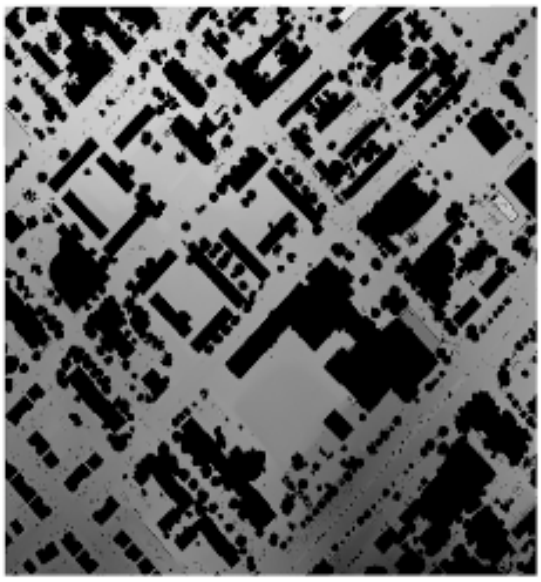
5. Smoothing. The centerline vector is converted to dense points. A snake is then used to refine the spatial position of those points using the centerline magnitude map as an external force.

6. Approximation/Point reduction: A centerline consisting of dense points is approximated using Iterative End-Point Fit, thus reducing the number of points.

Finally, a set of polygonal Boolean operations is applied to the road segments, resulting in a polygonal representation of the entire road network. Figures 10, 11 show the results of applying the approach to process an airborne LiDAR data (a section of Baltimore downtown area) and aerial imagery (a section of Las Vegas downtown area).

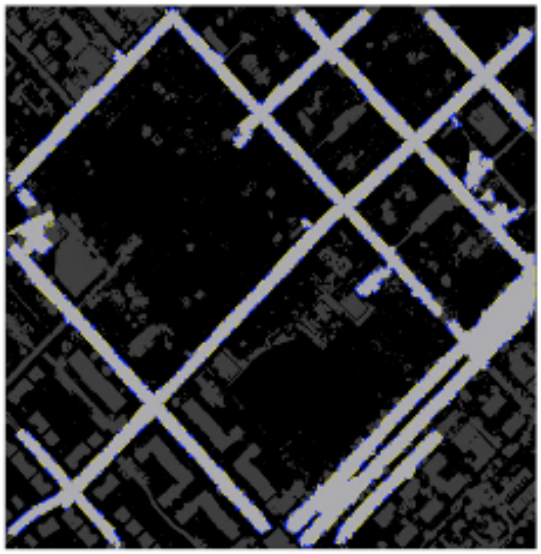


(a) LiDAR intensity data



(b) LiDAR depth data

(c) Detected road areas



(d) Extracted road network

Figure 10: Apply the approach to process an airborne LiDAR data (a section of Baltimore downtown area). The blue and red points in (a) are the seed points used to indicate background and foreground objects.

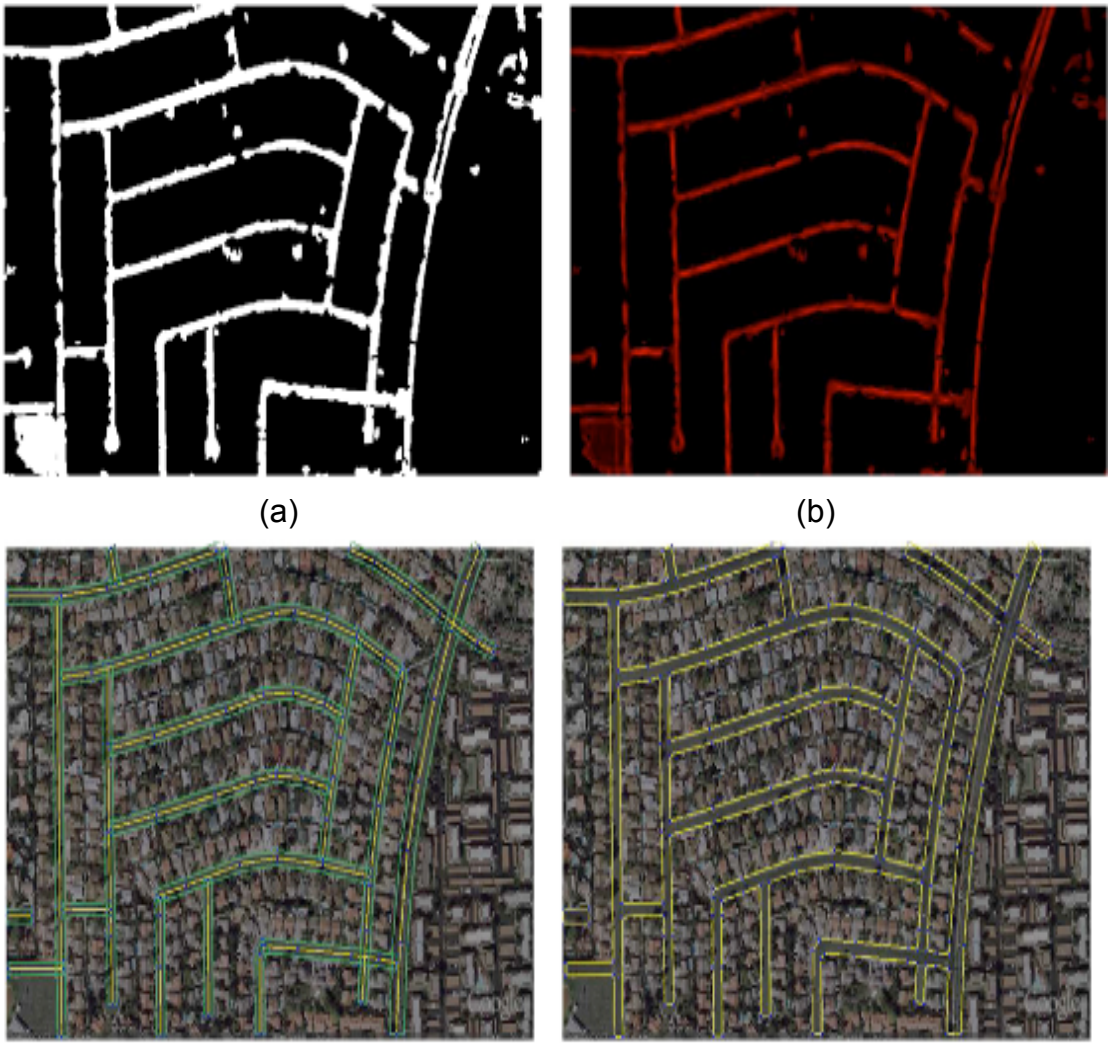


Figure 11: Apply the approach to process an aerial imagery (a section of Las Vegas downtown area). (a) Detected road areas, (b) extracted road centerline areas (only the magnitude is shown), (c) extracted road network and attributes, and (d) polygonal representation of the extracted road network.

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

We implemented the road extraction system and extensively evaluated its performance with various datasets acquired by different remote sensors, including aerial photographs, satellite images and airborne LiDAR. The datasets cover various urban, suburban and rural scenes with various road and imaging conditions. Our research efforts are to assess, define, and use the unique spatial and spectral characteristics of the advanced sensor techniques for rapid road feature extraction and road quality mapping.

4.1. Algorithm and System Implementation

We implemented two versions of the algorithms and systems on Microsoft Windows and Linux platforms. In the first step, we developed the core algorithms and constructed an experimental system on Linux platform because the Linux has the capability to directly handle large size data on system memory. Upon success, we then transferred the developed algorithms to Windows platform. We employed several advanced memory management and database techniques such as page memory, tree data structure to optimize the system's capability to process large datasets on COTS hardware platform.

The developed system comprises following main modules:

Data acquisition: accessing the data input to the system. The system supports multiple data sources in many standard formats, including optical color and grey images, LiDAR point clouds, and LiDAR intensity maps.

Data pre-processing: performing necessary data conversion and data processing tasks required by the core algorithm processing.

Road extraction: performing main tasks of geospatial feature detection, road extraction and road network completion.

Post-processing: providing useful tools for verification and refinement of the processed results.

4.2. Experimental Data Collections

We collected a number of testing datasets of designated sites. The datasets covers various urban suburban, and rural terrain scenes and conditions acquired by aerial imagery and LiDAR sensors, including the areas of USC campus and surround areas, Los Angeles downtown, San Diego, San Francisco, Denver, and Baltimore downtown. We used those datasets to evaluate our developed road extraction technique. The results are shown in next section.

4.3. Performance Evaluation and Results

We extensively evaluated the developed road extraction techniques with various collected datasets. We first tested the performance of every individual algorithm step-by-step on urban and rural terrains using both high-quality and lower-quality sparse/noisy imagery and LiDAR datasets. We then evaluated the performance of overall system that integrates all the individual algorithms into an overall architecture.

4.3.1 Individual Algorithm Evaluations

The three main components are tested individually with the collected imagery and LiDAR datasets: geospatial feature inference and classification, road feature detection and labeling, and road network extraction and identification.

The input to the system is the original data that could be optical images, LiDAR 3D point clouds, LiDAR intensity maps or their combinations. Once the data is loaded, the system automatically identifies the data type, preprocesses the data, executes core processes of road extraction, and output the results. The system also provides a toolbox that allows users to post-process the results or verify and refine existing roads to provide update road network maps.

Depending on the input data type, the system needs to perform necessary preprocessing step such as data conversion, noise filtering, contrast and illumination enhancements, and segment operations (select, copy, cut, past, delete). Our system treats all the data (whether they are LiDAR 3D point clouds or imagery) as 2D image data to process. The raw LiDAR point cloud data are converted to 2D depth images by hole-filling, noise filtering, and grid re-sampling.

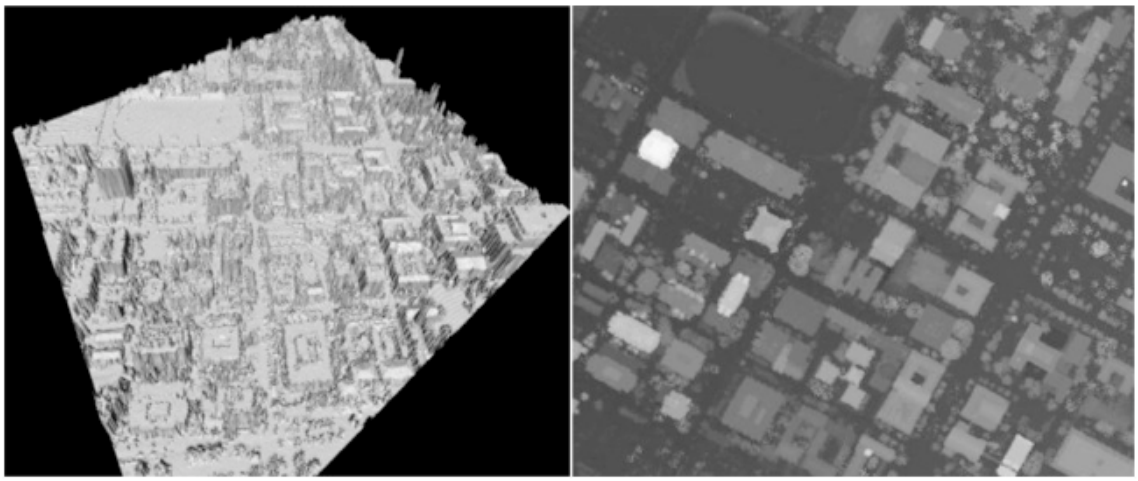


Figure 12: LiDAR data acquired for USC campus: (left) raw 3D LiDAR point cloud data, and (right) produced 2D depth image after hole-filling, noise filtering, and grid re-sampling.

Figure 12 shows a raw LiDAR point cloud data of USC campus area and its 2D depth image by the data preprocessing.

Once the input data is pre-processed, the system executes core processes that include detecting and extracting geospatial features, segmenting and labeling road features, and extracting complete road network and their attributes. Two types of results are generated: road segments map that shows all the extracted roads with segments and their attributes (road widths, centerlines, and road intersections) (Figure 13b), and road connection map that shows the connected road networks and their attributes (Figure 13c). The system allows user to select the output format and individual road attribute. All the output results are in vector representation, which we believe is important for further integration of the road extraction system and results into a general geospatial system and infrastructure database that include various geospatial elements such as 3D terrain models, buildings, vegetations, and transportation networks.

As stated above, our approach uniquely combines the Gabor wavelets and tensor voting to extract the significant signatures of geospatial features. The experiments have proved that the approach is very efficient in dealing with noisy, incomplete, and complex urban scenes. It is able to not only find unambiguously

varying geometry elements (i.e. point, curve, junction, and surface) under a unified framework, but it also encodes considerably richer structure information for feature inference and classification. The Gabor filters produce multi-channel geospatial feature maps at different spatial and frequency resolutions. The tensor voting effectively captures the first-order differential geometry and its singularities (feature locations, orientations, and confidences). The feature inference is based on a tensor voting communication process that is governed by a perceptual-field, encoding the constraints and rules of how a features receives/casts votes from/to its neighbors. By combining the voting results for all channels, the extracted geospatial features could be effectively classified and labelled (road tracks, junctions, and surfaces). Further the approach has no data dependencies, eliminating the need of threshold tuning and adjustment. Figure 16 shows a result of applying the approach to a rural terrain scene acquired by satellite imaging sensor, where the geospatial features have varying geometric (buildings, junctions, linear and curve roads) and radiometric (noise, material reflection, and contrast) properties. The results we obtained with the unique integration of perceptual grouping (tensor voting, Gabor wavelets) and geospatial data processing are clearly more complete road extractions than with prior methods and the approach provides a natural framework for fusing multiple data sources and features.

The above step of geospatial feature inference can extract the significant signatures of geospatial features and their attributes from sensor data. However, these extracted geospatial features may also contain non-road features such as buildings, vegetations and their shadows. A further filtering step is therefore needed to classify and label the interested road features from the extracted geospatial feature sets.

We tested the effectiveness of our graph-cut based technique to segment and label the road features. The approach is an improved graph-cut technique that incorporates both the feature location and orientation information inferred by the tensor voting into the processes of road feature detection and labeling. It overcomes the problems of traditional pixel-based approach that can effectively

segment the road features and label them in feature space and image space simultaneously. Figure 17 shows a result of applying the approach to an aerial image to extract the rural roads. Figure 18 is the result of applying the approach to a complicated urban scene (Baltimore downtown) acquired by LiDAR sensor, where the scene contains varying geospatial features and objects such as buildings, vegetations, parking lots, local streets, and highway roads. The low contrasts, varying shadows, and homogenous surface reflections contained in this scene make it extremely difficult to apply the pixel-based approach.

4.3.2 Overall System Evaluations

We evaluated the overall system performance that integrates all the individual algorithms and modules into an overall architecture. The system architecture was conceived and used as a framework for integration of the algorithms and modules. Initial usability and interface enhancements were designed, implemented, and evaluated.

Tests were performed on a number of datasets for designated sites including, Los Angeles downtown, San Diego, San Francisco, Denver, Oregon, Kentucky, Las Vegas, and Baltimore downtown areas. The sample results are shown in Figure 13 – 26. Note that all the output results are in vector representation. We believe that using vectors to represent the road networks is important for further integration of the road extraction system and results to a general geospatial system and database that include various geospatial elements of 3D terrain models, buildings, vegetations, and transportation networks. Figure 26 shows the example of merging the extracted road network with 3D building models under a consistent framework.

Quantitative evaluations were conducted using the evaluation metrics introduced in [Wie99], in terms of completeness, correctness and quality.

- **Completeness**. The completeness is defined as the ratio of true positives from the sum of the true positives and false negatives given by:

$$E_{Completeness} = \frac{TruePositives}{TruePositives + FalseNegatives}.$$

- **Correctness**. The correctness is defined as the ratio of true positives from the sum of the true and false positives given by:

$$E_{Correctness} = \frac{TruePositives}{TruePositives + FalsePositives}$$

- **Quality**. The quality is a measure of the ``goodness'' of the final result, and is given by:

$$E_{Quality} = \frac{TruePositives}{TruePositives + FalsePositives + FalseNegatives}$$

Table 1 lists evaluation results using above metrics for the datasets of Oregon, Kentucky and Las Vegas. It is shown that the system performs very well in terms of completeness, correctness and quality for the rural and urban areas.

It is worth to mention that the overall system performance primarily depends on the performance of the low-level grouping and mid-level segmentation processes. The grouping and segmentation are the two essential components that can drastically affect the outcome. The use of tensor voting framework significantly improves the grouping results, since it eliminates any ‘‘guessing’’ (i.e. using thresholds) when deciding the neighboring features.

Table 1: Quantitative performance evaluations

Dataset	Completeness (%)	Correctness (%)	Quality (%)
Oregon (Figure 7)	82	80	68.6
Kentucky (Figure 8)	86	75.5	67.3
Baltimore (Figure 10)	83	66	58.8
Las Vegas (Figure 11)	71.4	80	60.6

5. CONCLUSIONS AND RECOMMENDATIONS

The research efforts of this project are to develop new approach and technique to assess, define, and use the unique spatial and spectral characteristics of advanced remote sensor techniques from aerial imagery and LiDAR for automated road extraction and road quality mapping. The methods developed and tested in this research have shown the ability to extract road features from a wide range of terrain. Urban and rural areas with wide and narrow roads and highways are successfully extracted, as are dirt or gravel roads. Road attributes including road widths, centerlines, intersections and overpasses are estimated.

We focused on the development and testing of new algorithms based on new mathematical frameworks for road feature segmentation, detection, road tracking and verification/assessment. Specifically, we explored the integrated use of perceptual grouping (Gabor wavelets and Tensor Voting) with sensor cues, geometric invariants, and machine learning under a unified framework to tackle the problems, which has not been used for this application before. This proved to be a very important component of our techniques and overall technical approach. The results we obtained with the unique technique are clearly more complete road extractions than with prior methods. Our technique provides a natural framework for fusing multiple data sources and features.

This new technique has demonstrated the potential for automating the extraction and mapping of complex road networks from advanced remote sensing data. In addition, the same process also allows for a constrained optimal estimation of various terrain features and attributes, thereby producing hierarchical data representations under a consistent framework. Most importantly, we believe that the process of classifying and labeling the important geospatial elements such as buildings, vegetation, terrains, and transportation networks is possible within this framework.

The technique has been implemented and extensively tested on urban, suburban, and rural terrains using both high-quality and lower-quality

sparse/noisy imagery and LiDAR data. The algorithms and modules were integrated into an overall architecture that has several important features:

- the ability to fuse multiple data sources in many standard formats, including optical images, LiDAR 3D point clouds, LiDAR intensity maps and their combinations.
- the ability to operate independently on subregions of the data and to merge all results.
- the ability to apply multiple operator options suited for various data and terrain characteristics.
- the ability to store and load interim processing results. Processing errors, manual errors, or system failures do not cause loss of prior step results.
- the ability to process data automatically. If needed, the results can been further refined or corrected with the developed editing tools.
- the ability to process large data sets using standard platforms (MS Windows and Linux) and benefiting from their CPU, memory, and graphics performance improvement curves.

Technical literatures are collected and analyzed to determine the research direction and relevant technologies essential for this research.

A large number of imagery and LiDAR datasets have been collected for algorithm development and evaluation. Tests were performed on the datasets for designated sites including, Los Angeles downtown, San Diego, San Francisco, Denver, Oregon, Kentucky, Las Vegas, and Baltimore downtown areas. The datasets and tests are the most extensive in this research area, to the best of our knowledge.

The new findings resulting from the research have been documented and published in peer-referred journal and international conference including:

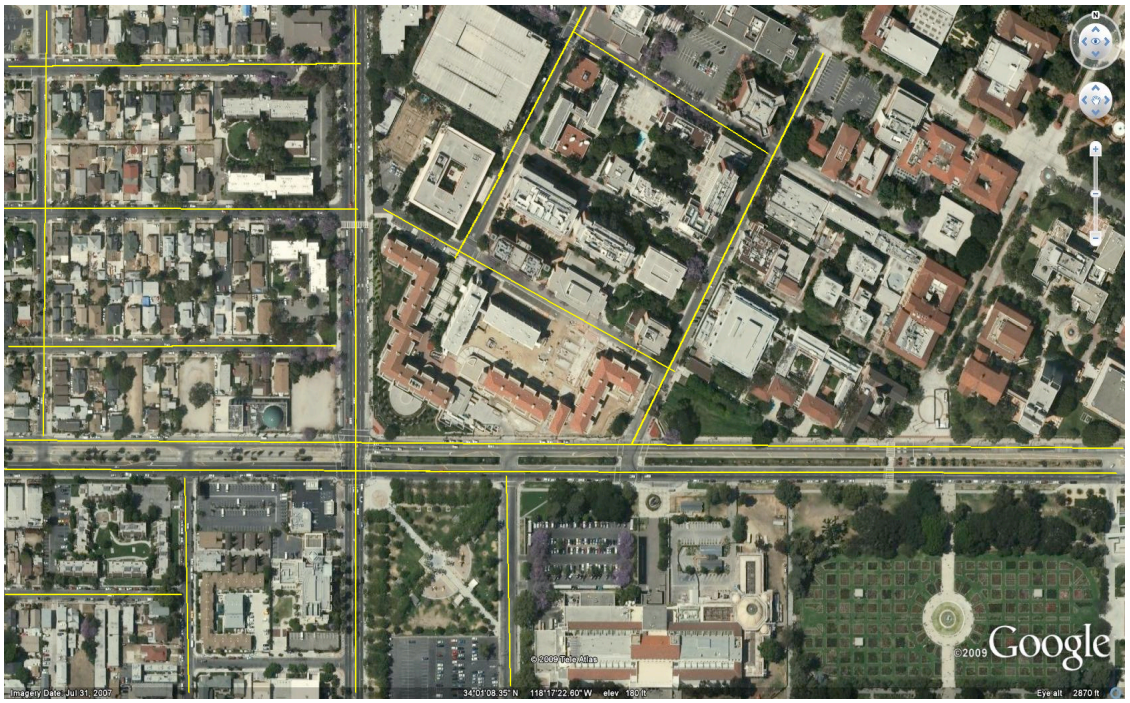
- "Delineation and Geometric Modeling of Road Networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, Elsevier, 2009 (in press)
- "A Vision-based System for Automatic Detection and Extraction of Road Networks," *IEEE Workshop on Applications of Computer Vision (WACV),*

2008, Copper Mountain, Colorado

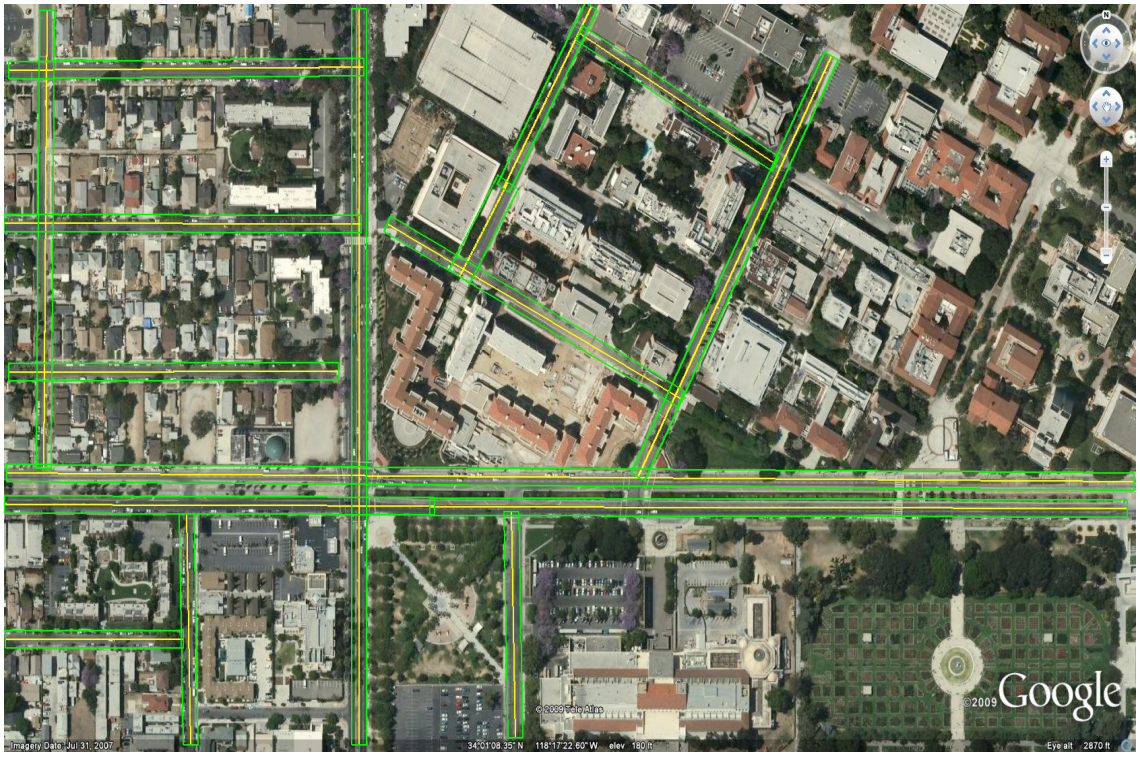
Two students (one Ph.D. student and one Master student) worked on this project and graduated. This project offers a significant opportunity to engage the students in transportation research activities. The METRANS needs and applications are excellent driving problems that focus the research efforts on producing tangible results, helping the students understand the need for balance and coordination in pursuing intellectual ideas and producing tangible results.

In summary, this project has demonstrated, based on our research efforts and results, that using advanced remote sensors combining with advanced computer vision technology for rapid extraction of complete road networks is efficient. Remote sensing techniques such as measures from aerial imagery and LiDAR provides one effective means by which large-area topographic models can be rapidly mapped with a high standard of accuracy. Advanced computer vision theory and approach offer powerful tools to process such data, having advantages in terms of accuracy, confidence, completeness, and automation.

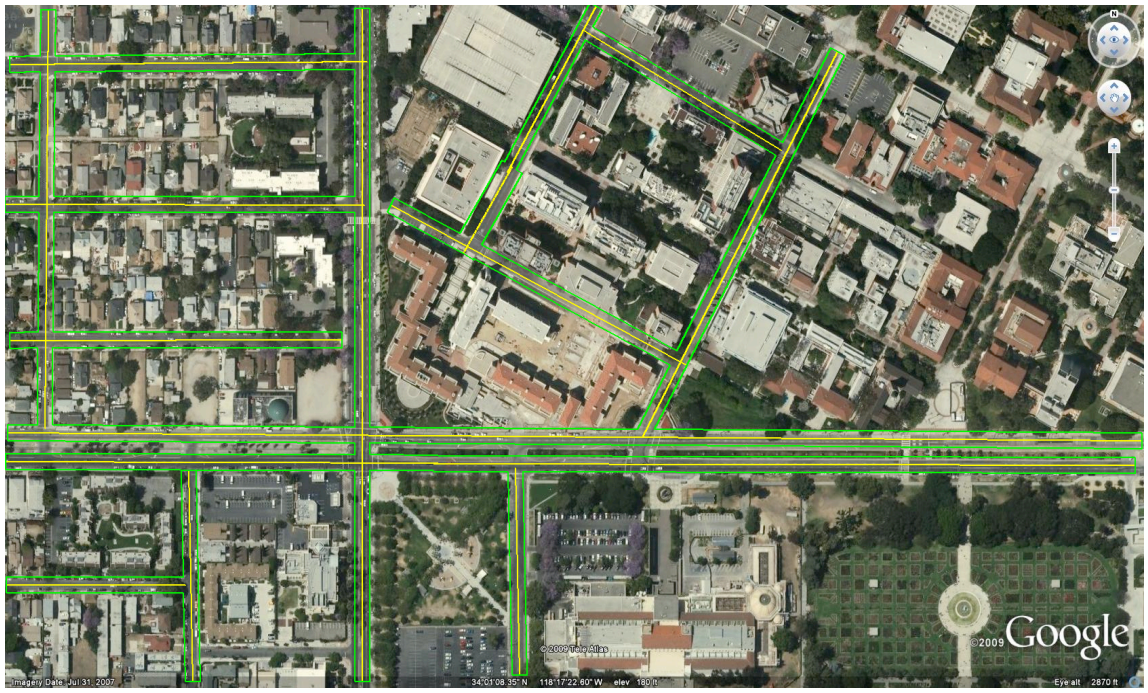
The research, development, and testing indicate that the our approach is sound, the risk of moving forward with development of a complete end-to-end prototype application is low, and the opportunity for impacting the current state of commercial products and markets is there. In addition, extension of the developed approach to extract other related transportation networks, such as rivers, waterways, bridges, and rail lines could be pursued.



(a)



(b)



(c)

Figure 13: Experimental results of southern portion of USC dataset: (a) extracted road centerlines, (b) extracted roads segments and their attributes (road widths, centerlines, and road intersections), and (c) complete road connection network and attributes.

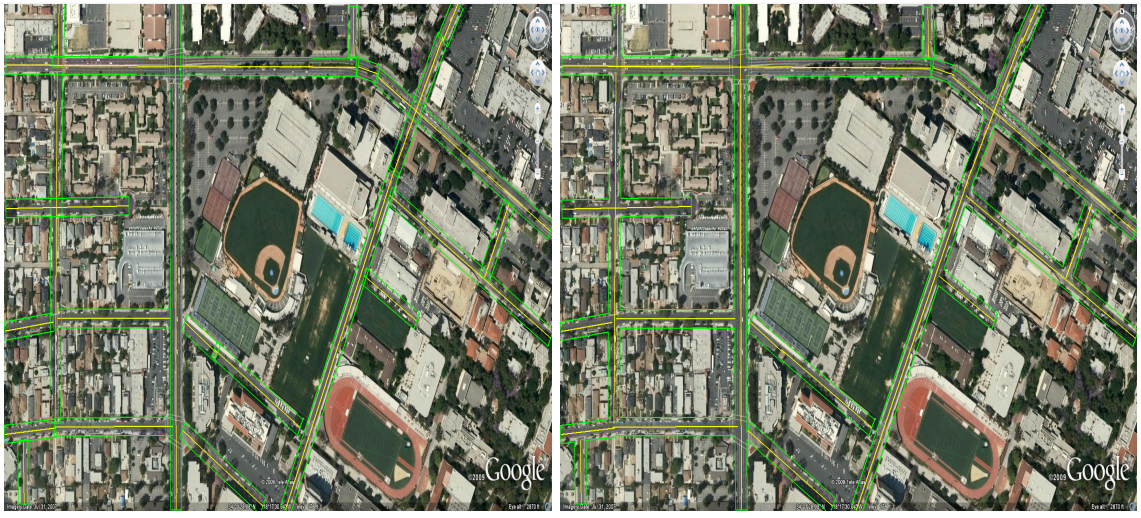
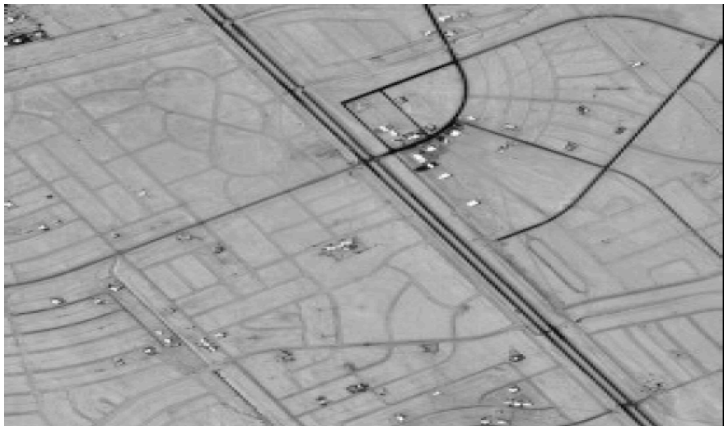
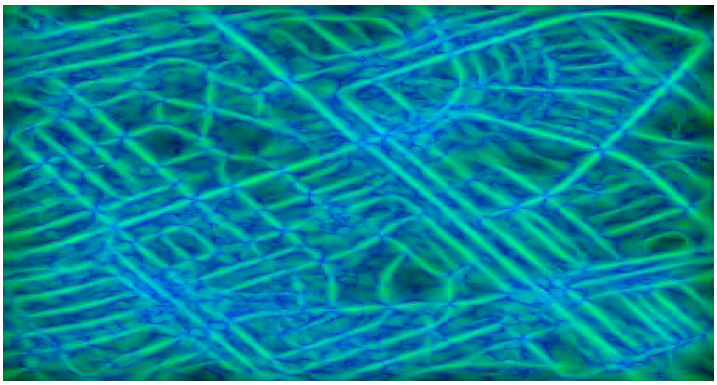


Figure 14: Experimental results of northern portion of USC dataset: (left) extracted road segments map, and (b) complete road connection map.

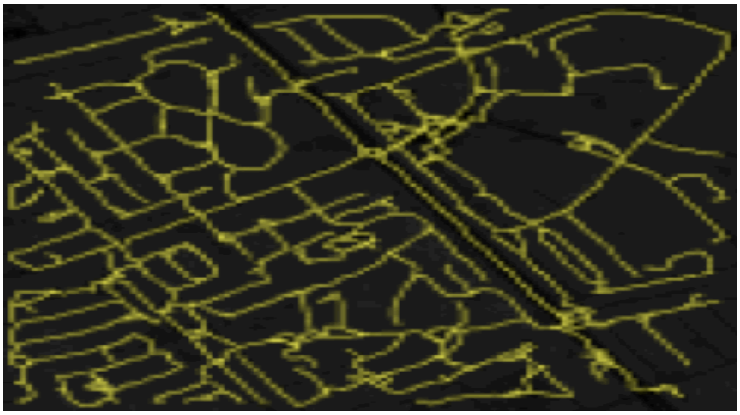
Figure 15: Extracted road network of entire USC campus and surround areas.



(a)



(b)



(c)

Figure 16: Robustness test: using the hybrid Gabor-Tensor Voting approach to extract the significant signatures of road features having varying geometric and radiometric properties. (a) Original data acquired by satellite imaging sensor, (b) saliency map indicating the feature likelihoods captured by Gabor-Tensor Voting, and (c) inferred road features with confidences. Note that the processing is fully automatic, and there is no threshold parameter that requires tuning.

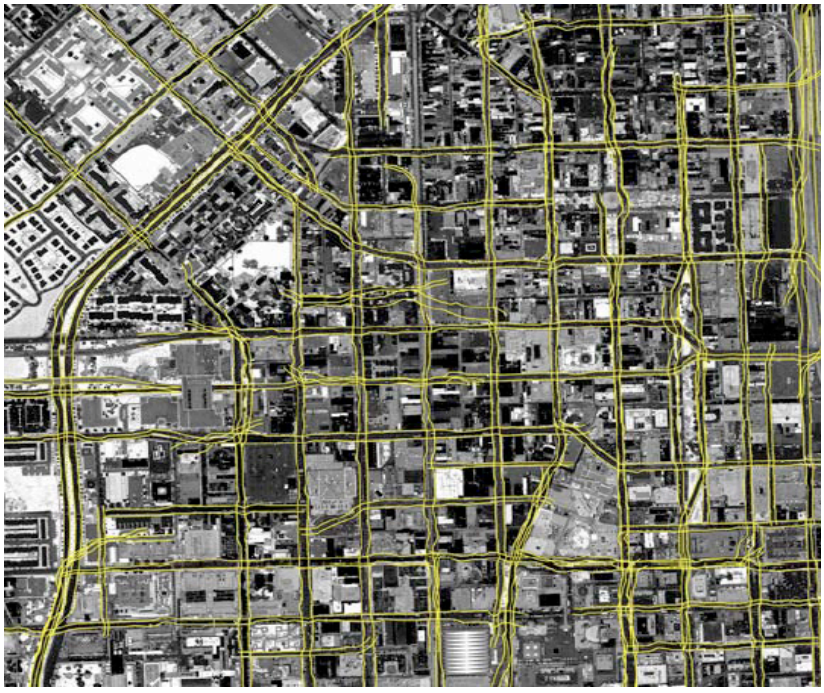
(a)

(b)

(c)

Figure 17: Robustness test: using an improved graph-cut technique to segment and label the dirt and gravel roads in rural scene. (a) Original aerial image, (b) segmented and labeled road features, and (c) extracted roads overlaid on the original image.

(a)



(b)

Figure 18: Robustness test: using an improved graph-cut technique to segment and label a urban scene acquired by LiDAR sensor. (a) Segmented and labeled road features, and (b) extracted roads overlaid on the original image.

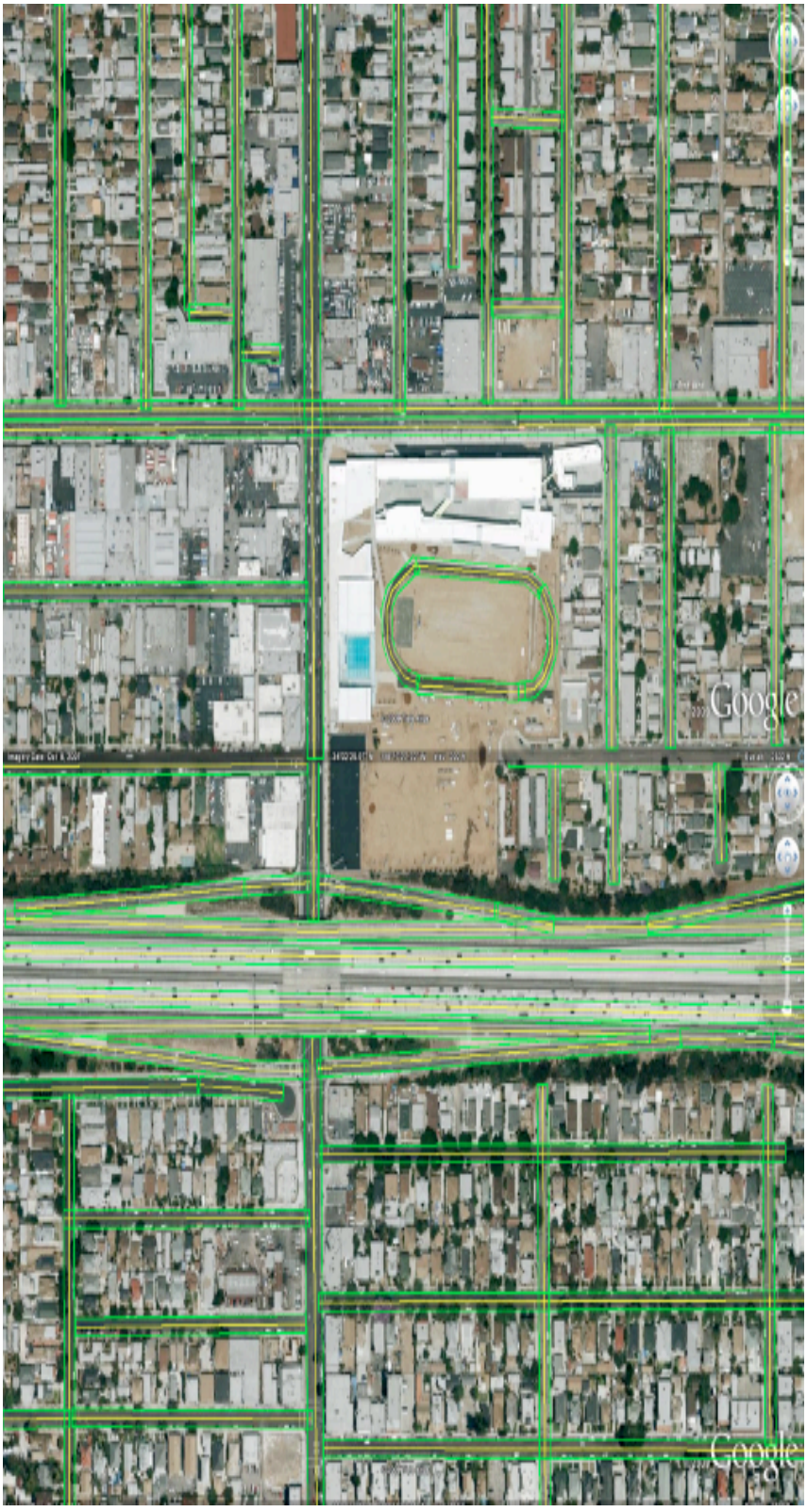


Figure 19: Extracted road network of Los Angeles downtown areas using aerial images.

Figure 20: Extracted road of portion of San Francisco areas using aerial images.

Figure 21: Extracted road network of San Francisco areas using aerial images.

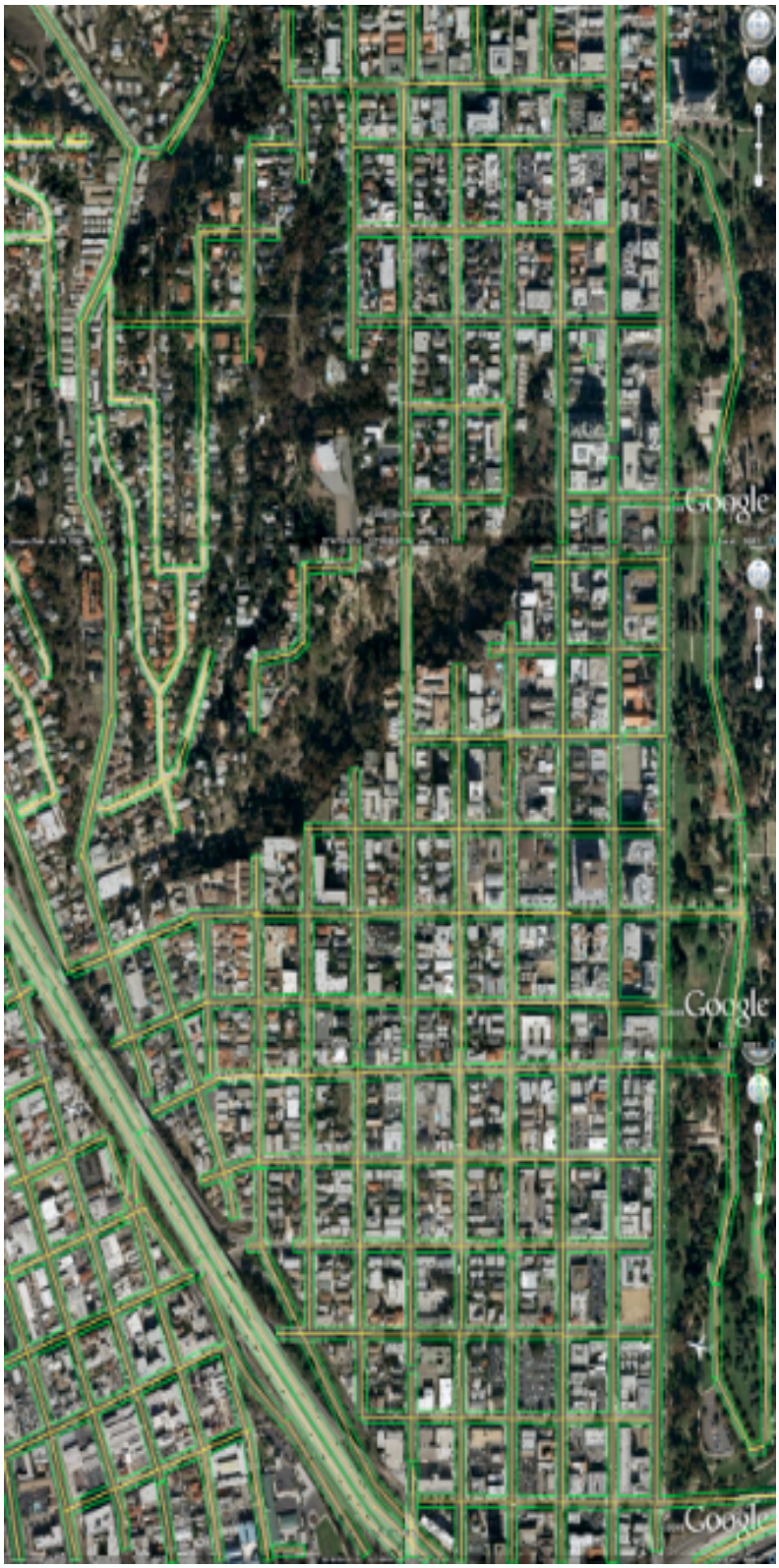


Figure 22: Extracted roads of portion of San Diego areas using aerial images.

Figure 23: Extracted road network of portion of Denver areas using aerial images.

(a)

Figure 24: (a) Extracted roads of portion of Baltimore downtown from LiDAR, and (b) road attributes (widths, centerlines, and intersections).

Figure 25: Polygonal representation of the extracted road network (Baltimore downtown) is useful for integration of the road extraction system and results to a general geospatial system and database.

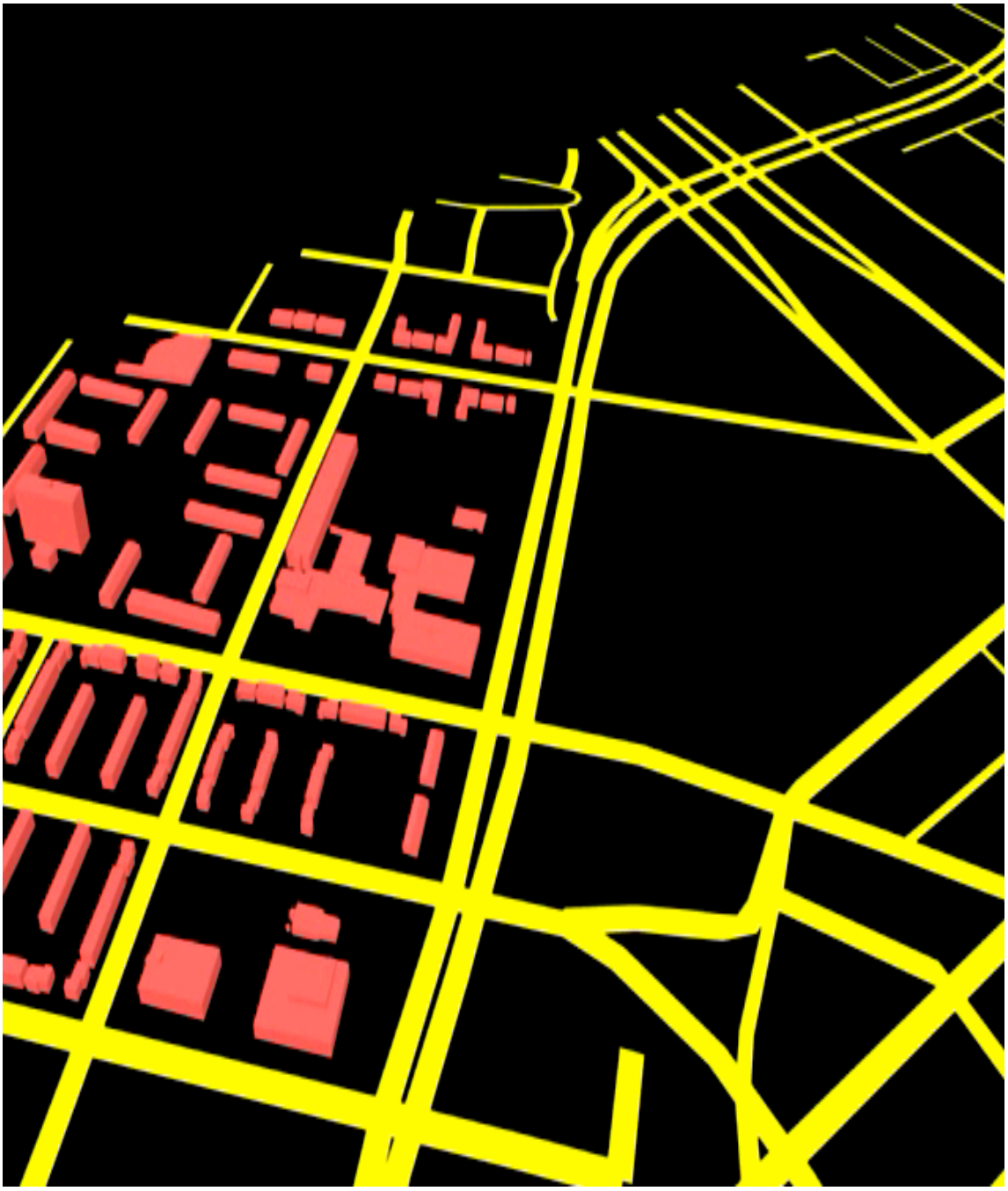


Figure 26: Integaring the extracted road network of Figure 22 to a general geospatial system and database that include various geospatial elements of 3D buildings models and road transportation models.

6. REFERENCES

[AHMCT] http://www.caltrans.ca.gov/newtech/researchreports/twopage_summaries/ahmct_3d_laser_scanning_5-2007.pdf

[Bar03] A. Barsi and C. Heipke. Artificial neural networks for the detection of road junctions in aerial images, In ISPRS Archives, Vol. XXXIV, Part 3/W8, Munich, 17-19, Sept., 2003

[Bau99] A. Baumgartner, C. T. Steger, H. Mayer, W. Eckstein, and H. Ebner. Automatic road extraction in rural areas, In ISPRS Congress, pages 107–112, 1999

[CADOT] http://www.dot.ca.gov/newtech/researchreports/two-page_summaries/resnotes_lidar.pdf

[Clo05] S. Clode, F. Rottensteiner, and P. Kootsookos. Improving city model determination by using road detection from lidar data. In Inter. Archives of the PRSASI Sciences Vol. XXXVI - 3/W24, pp. 159-164, Vienna, Austria, 2005

[Del05] F. Dell'Acqua, P. Gamba, and G. Lisini. Road extraction aided by adaptive directional filtering and template matching, In ISPRS Archives, Vol. XXXVI, Part 8/W27, Tempe, AZ, 14 – 16, March, 2005

[FHWA] http://safety.fhwa.dot.gov/intersections/intersectionsap.htm

[Hof] P. Hofmann. Detecting buildings and roads from ikonos data using additional elevation information, In Dipl.-Geogr.

[Lap00] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner. Automatic extraction of roads from aerial images based on scale space and snakes, Mach, Vis. Appl, 12(1): 23–31, 2000

[Lap00] Laptev I, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes," Machine Vision and Applications, 12(1), pp.23-31, 2000

[Lis04] G. Lisini, C. Tison, D. Cherifi, F. Tupin, and P. Gamba. Improving road network extraction in high-resolution sar images by data fusion. In CEOS SAR Workshop 2004

[Med00] G. Medioni, M. Lee, and C. Tang, “A Computational Framework for Segmentation and Grouping,” Elsevier Science B.V., March 2000

[METR] http://www.metrans.org/research/index.php?year=&status=ALL

[Hu03] J. Hu, S. You, and U. Neumann, “Approaches to Large-Scale Urban Modeling,” IEEE Computer Graphics & Applications, 23(6), November/December 2003

[Hu04] Hu X., C.V. Tao, “Automatic Highway Extraction from High Resolution Imagery by an Energy Minimizing Based Perceptual Grouping Method”, Geomatica, 58 (1), 2004

[Hu06] J. Hu, S. You and U. Neumann, “Integrating LiDAR, Aerial Image and Ground Images for Complete Urban Building Modeling,” Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT), June 2006

[Pou07] C. Poullis and S. You, “Linear Feature Extraction Using Perceptual Grouping and Graph-Cuts,” ACM GIS, Seattle, Washington, 2007

[Pou08] C. Poullis, S. You and U. Neumann, “Rapid Creation of Large-scale Photorealistic Virtual Environments,” IEEE Virtual Reality (VR), 2008

[Pet03] R. Peteri, J. Celle, and T. Ranchin. Detection and extraction of road networks from high resolution satellite mages. In ICIP03, pages I: 301–304, 2003

[Por03] Porikli and F. M. Road extraction by point-wise gaussian models. Technical report, MERL, July 2003. SPIE Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery IX, Vol. 5093, pp. 758-764

[USDOT] http://www.tfhrc.gov/about/02101/index.htm

[Wan07] L. Wang, S. You, and U. Neumann, “Semiautomatic registration between ground-level panoramas and an orthorectified aerial image

for building modeling," VRML07 workshop in conjunction with ICCV, October 2007

[Wes04] B. Wessel, "Road network extraction from sar imagery supported by context information," in Proceedings of ISPRS, 2004

[Wie99] Wiedemann, C., Hinz, S., "Automatic extraction and evaluation of road networks from satellite imagery," International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 32 (Part 3/2W5), 95–100, 1999

[Zha01] C. Zhang, E. Baltsavias, and A. Gruen. Knowledge-based image analysis for 3d road construction. In Asian Journal of Geoinformatic 1(4), 2001

[Zha06] Q. Zhang and I. Couloigner, "Automated road network extraction from high resolution multispectral imagery," in Proceedings of ASPRS, 2006