

Ortho-Image Analysis for Producing Lane-Level Highway Maps

Young-Woo Seo
The Robotics Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
ywseo@ri.cmu.edu

Chris Urmson
Google
1600 Amphitheatre Parkway
Mountain View
CA 94043
curmson@google.com

David Wettergreen
The Robotics Institute
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
dsw@ri.cmu.edu

ABSTRACT

This paper presents new aerial image analysis algorithms that, from highway ortho-images, produce lane-level detailed maps. We analyze screenshots of road vectors to obtain the relevant spatial and photometric cues of road image-regions. We then refine the obtained patterns to generate hypotheses about the true road-lanes. A road-lane hypothesis, since it explains only a part of the true road-lane, is then linked to other hypotheses to completely delineate boundaries of the true road-lanes. Finally, some of the refined image cues about the underlying road network are used to guide a linking process of road-lane hypotheses.

We tested the accuracy and robustness of our algorithms with high-resolution, inter-city highway ortho-images. Experimental results show promise in producing lane-level detailed highway maps from ortho-image analysis – 89% of the true road-lane boundary pixels were successfully detected and 337 out of 417 true road-lanes were correctly recovered.

Categories and Subject Descriptors

I.4 [Image Processing and Computer Vision]: Segmentation—*Partitioning*; I.4.8 [Scene Analysis]: [Object Recognition]

General Terms

Algorithms

Keywords

Ortho-Image Analysis, Lane-Level Map Generation

1. INTRODUCTION

This paper proposes new aerial image analysis algorithms that produce a map of road-lanes appearing on a given highway ortho-image. The output of this procedure is cartographic information about road-lanes in a set of pixel coordinates of road-lanes' centerlines and lateral road-widths.

Because our target images are depicted in high-resolution (i.e., 15-centimeter ground resolution), such image objects as lane-markings and road image-regions contain significant

variations in their appearances, such that an object appears differently based on the condition of an image acquisition process and road surface materials.

To effectively tackle these challenges, we develop a hierarchical approach to three tasks: gathering road boundary image cues, generating road-lane hypotheses, and linking the hypotheses. To this end, we first scrutinize input images, to harvest two types of image cues about the underlying roads: road image-regions and the geometry of underlying road-lanes. These collected image cues about road surface and geometry will provide strong evidences of the true road-lanes. In particular, these cues facilitate a road-lane hypothesis generation and guide a linking of these hypotheses to build an accurate map of road-lanes. For the problem of linking road-lane hypotheses, we formulate it as a min-cover problem [3, 15] where we look for a set of hypotheses about the unknown true road-lanes to maximally cover the estimated road image-regions with a minimum sum of costs.

2. RELATED WORK

In the GIS community, aerial image analysis has played a crucial role in maintaining existing cartographic databases [1, 2, 4, 6, 9]. This is primarily because the topological relations among spatial objects (e.g., intersection between road segments) appearing on aerial images are invariant over a long period of time, even after natural disasters [14]. An aerial image analysis also provides an alternative way of maintaining existing cartographic databases without virtually going out to the regions of interest. The clear difference between these work and ours is the ground resolution. Most of these work analyzed low-resolution aerial images in which the ground resolution was greater than one meter [1, 2, 4, 5, 6, 9]. Variations of object appearances in a low-resolution aerial imagery are not as significant as those of high-resolution imagery.

We detect overpasses to identify potentially complex road geometry. To recover such 3-dimensional road structures, researchers have directly accessed a road vector or utilized 3D data such as air-borne point clouds [11, 12]. By contrast, our method, without using any of these specialized data, detects overpasses by using screenshots of road-vectors.

3. HARVESTING ROAD-BOUNDARY IMAGE CUES VIA BOOTSTRAPPING

This section explains how to analyze input images to extract low-level image features and how to refine such features to produce task-specific image features that can be directly used to execute other sub-tasks. Our algorithms take two images as the input: highway ortho-image and the input image's road-vector screenshot as shown in Figure 1.

We parse two input images, to extract photometric im-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM SIGSPATIAL GIS'12, November 6-9, 2012, Redondo Beach, CA, USA, Copyright (c) 2012 ACM ISBN 978-1-4503-1691-0/12/11...\$15.00



Figure 1: A screenshot of the road-vector of the input image.

age features such as lines [7] and superpixels [8]. To extract the useful geometric information of the underlying roads, we analyze a road-vector screenshot to extract image regions of road-vector sketches, (i.e., yellow drawings in Figure 1), and then further analyze each of the road-vector sketch fragments, to obtain their geometric properties, such as extremity and bifurcation points.

Next we refine these low-level features to produce more relevant and useful features, such as a segmentation of a road image-region, an estimation of some legitimate driving directions of roads appearing on the input image, a lane-marking detection, and locations of interesting road-structures, such as intersections and overpasses.

For identifying road image-regions, we first convert an input ortho-image into a superpixel image and then represent each of the superpixels by a combination of color histogram and textons [10]. Next we perform, by using a combination of Gaussian Mixture Model (GMM) and Markov Random Fields (MRF), a binary classification that takes superpixels as input and assigns each superpixel with one of two class labels: road or non-road. Figure 2 shows a result of image road-region segmentation.

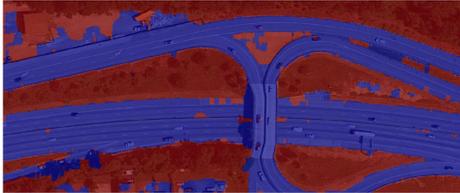


Figure 2: Results of road image region segmentation. The blue regions represent identified road image-regions and the red regions represent non-road image-regions. Although some non-road image-regions are labeled as road, for the most part segmentation results correctly depicted road image-regions.

To detect the driving direction from a given image, we use line extraction results that each of the extracted lines partially explain as the contour of roads in a given image. We first partition the input image into a number of grid cells. For each grid cell, we identify extracted lines which pass by it and use them to approximate the driving direction of the grid cell by using the vector sum method. Figure 3 shows a result of driving-direction estimation.

For detecting lane-markings, we perform a binary classification problem of discriminating non lane-marking pixels from true lane-marking pixels. To this end, we tried six different classification setups with ortho-images separated from the images for generating lane-level highway map and found that AdaBoost outperformed all others. Figure 4 shows a result of lane-marking detection.

For detecting overpasses, we analyze the road-vector screenshot image to convert it into a set of road-vector fragments. For each of the road-vector fragments, we extend each of the



Figure 3: Results of driving direction estimation. The blue lines indicate the estimated driving direction of a grid cell and the non-drivable regions are depicted by red circles.

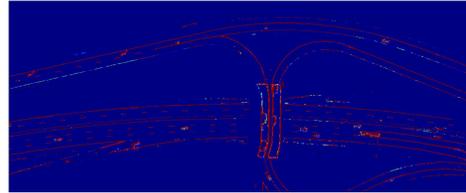


Figure 4: Lane-marking detection results. Because the outputs of our lane-marking classification are probabilistic, the results are shown in a heat-image where the color closest to red represents the highest probability.

extremity points in the direction of the fragment. To localize a potential overpass, we then identify any intersection with other fragments if their intersection angle is greater than a given threshold (e.g., $\pi/3$). Finally, we search for any of the closest extracted lines to identify boundaries of the detected overpass. Figure 5 shows the final result of overpass detection.

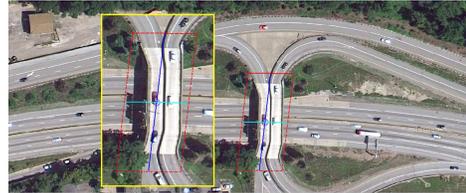


Figure 5: Results of overpass detection. A red parallelogram represents the boundary of the detected overpass, and two (blue and cyan) lines inside the polygon depict two principal axes.

In the next section we detail how these four refined features are used to generate road-segment hypotheses and how to link them to build lane-level detailed highway maps.

4. ROAD-LANE HYPOTHESIS GENERATION AND EVALUATION FOR DELINEATING TRUE ROAD-LANE BOUNDARIES

The previous steps of extracting image cues about the true road-lanes provide us a better understanding of road-lanes appearing in the input image. In particular, we know which image sub-regions are most probably road-regions, which pixels within the road-regions are likely parts of lane-markings, how the roads are laid out, and where overpass structures occur. Based on this understanding, we are generating road-lane hypotheses and linking them, in order to delineate road-lane boundaries.

A road-lane is modeled by a piecewise linear curve that consists of multiple control points along the centerlines of

the road-lane and their properties, such as lateral width and orientation. We use lane-marking pixels to approximate locations of control points along the true road-lane.

A true lane-marking pixel has many neighboring lane-marking pixels regularly-spaced longitudinally and laterally (or orthogonal to the longitudinal direction). We are looking for lane-marking pixels that have strong supportive (or neighboring) patterns in longitudinal and lateral directions of the roads. For each superpixel, we investigate whether each of the lane-marking pixels has a sufficient number of neighboring lane-marking pixels in longitudinal and lateral directions on the roads. Any lane-markings with more than the predefined threshold remain in the candidate list for generating road-lane hypotheses. For each road-width cue (or road-width hypothesis), we draw two lines, longitudinally, from the center of the two lane-marking locations and group together any road-width cues within extending line segments. This forms a road-lane hypothesis. The longitudinal direction corresponds to the driving direction estimated earlier from extracted lines. This search results in grouping the neighboring road-width cues around the input road-width hypothesis.

A road-lane hypothesis, since it explains only a part of the true road-lane, is then linked to other hypotheses to completely delineate boundaries of the true road-lanes. We formulate the problem of linking hypotheses as a min-cover problem in which we search for a new set of road-lane hypotheses of linking the generated road-lane hypotheses based on the previously obtained local evidences of the unknown true road-lanes with a minimum sum of linking costs. Our formulation is motivated by two previous studies [3, 15]. For our case, we generate a set of hypotheses about unknown true road-lanes to cover approximated true road image regions. The previous studies generated hypotheses to delineate object contours [3] and to cover road regions in a LIDAR intensity image [15].

To find approximate solutions to these cost functions, we devise two linking functions. The first linking function considers a potential connection between any two hypotheses purely following geometric constraints. And the second function investigates any photometric constraints of a potential link. The optimal link between two road-lane hypotheses would be one that locally minimizes these two constraint functions. Unlike previous work of the min-cover algorithm applications [3, 15], where their solutions were explicitly searching for a sequence of hypotheses, we look for a set of hypothesis pairs such that their potential, geometrically plausible, links are sequentially traced by photometric image cues to cover road image-regions.

5. EXPERIMENTS

This section details experiments conducted to investigate the robustness of our approach to extracting a lane-level highway map and the accuracy of the resulting maps.

From Google’s map service¹, we collected 50 ortho-images that are sampled from the route between the Squirrel Hill Tunnel to the Pittsburgh International Airport. We also saved road-vector screenshots of the ortho-images and manually drew boundary lines of individual road-lanes in each of the collected images for the ground truth.²

To the best of our knowledge, no prior work or image data was available on extracting road-lane boundaries that we could have used for comparison. Hence, we had to come up with reasonable ways of evaluating our results. We evaluate resulting road-lane boundary delineation in two-ways: accuracy of matching between output and ground truth pixels

¹<http://maps.google.com>

²The complete experimental results of these 50 test images are available from [13].

and counting the number of correctly recovered road-lanes in the final outputs. Matching pixel to pixel aims at investigating the performance of our approach at the micro-level; counting the number of road-lanes aims at revealing the accuracy of the resulting geometries.

To evaluate our results at a pixel-to-pixel level, we utilized the method from evaluating performance of object boundary detection [10]. Similar to [10], we regard the extraction of road-lane boundaries as a classification problem of identifying boundary pixels and of applying the precision-recall metrics using manually labeled road-lane boundaries as ground truth. For resolving the correspondence between output pixels and ground truth pixels, we utilized the Berkeley Segmentation Engine’s³ performance evaluation scripts. We also used, as a baseline method, BSE’s probabilistic boundary detection outputs. BSE was developed to detect generic object boundaries, not road-lane boundaries. In addition, since training BSE with our image data is not possible, it may fall short of being a fair comparison. But since anyone can think of such probabilistic boundary outputs as a starting point of delineating road-lane boundary lines, we compared it with our output. Table 1 presents an averaged performance difference between the two outputs over fifty test images.

	F-measure	Precision	Recall
Ours	0.82	0.77	0.89
BSE’s	0.44	0.38	0.54

Table 1: An averaged precision-recall measure of micro-level performance between the two outputs.

In achieving our goal, the performance evaluation by a pixel-to-pixel matching for road-lane boundary extraction outputs might be insufficient. The pixel-to-pixel measure counted a match when an output boundary pixel was located to a true boundary pixel within a predefined distance threshold (e.g., 10 pixels). Therefore a collection of boundary pixels would not necessarily correspond to a road-lane boundary. To be useful, these detected boundary pixels must be interpreted as parts of a road-lane. In other words, the desirable output for our purpose, is one that treats a road-lane as a polygon, bounded by a closed path and image boundaries, where we can estimate lateral road widths, curvature, and other interesting geometric properties along the centerline of a road-lane polygon. To measure such macro-level performance, we first visually inspected our outputs and the input image to resolve the correspondence between the resulting road-lanes and true road-lanes appearing on the input image. We then counted the number of correct and incorrect output road-lanes and missed true road-lanes. If the area of overlap between a road-lane output and a true road-lane was roughly greater than 80%, then we counted it a correct match. This counting resulted in a two-contingency table for the performance of each test image. Table 2 shows a macro-level performance that is obtained by merging individual contingency tables over fifty test images. An averaged performance was then computed by using this table, precision = $0.792 = \frac{337}{337+88}$, and recall = $0.771 = \frac{337}{337+100}$, meaning that 79% of the resulting road-lanes were correct and 77% of true road-lanes appearing on the test images were correctly recovered.

Examples of resulting maps are shown in Figure 6. Figure 6 shows some of the most accurate results with all of the true road-lanes appearing on test images recovered correctly. While processing these images, our approach successfully tracked high-curvature ramps, correctly connected road-lane

³The BSE and related information are available at <http://www.cs.berkeley.edu/~fowlkes/BSE/>

		Ground Truth	
		Road-lane	Not road-lane
Output	Road-lane	337	88
	Not road-lane	100	×

Table 2: A contingency table is used to measure the macro-level performance of our highway map generation methods.

boundaries around overpasses, effectively handled variations in road-surface materials, and partial image distortions.

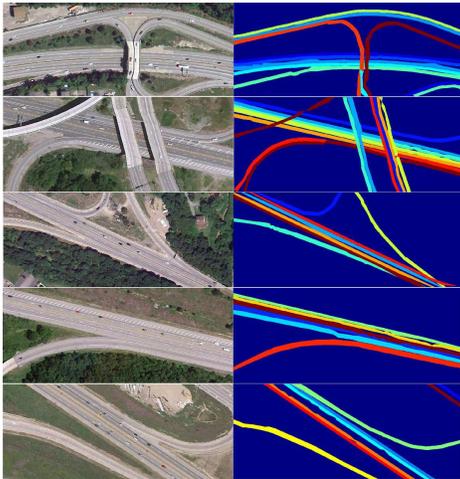


Figure 6: Some of resulting maps. There are two sub-figures in each row. The figure on the left is a test image and the figure on the right is our output, where each road-lane output is depicted in a different color and the background is depicted in blue. These figures are best viewed in color.

6. CONCLUSION

This paper presented a new approach to extracting lane-level detailed highway maps from a given ortho-image. We chose high-resolution, inter-city highway ortho-images as target images because pixels along road-lane boundaries must be visually and computationally accessible. To effectively address photometric and geometric challenges appearing on our target images, we developed a hierarchical approach to three tasks: to collecting road boundary image cues via bootstrapping, to generating hypotheses about the unknown true road-lanes, and to linking hypotheses with respect to the photometric and geometric constraints imposed by the collected image cues and prior information.

We tested our algorithms with 50 challenging arterial highway images. The results were evaluated according to two aspects: pixel-to-pixel matching and counting correct and incorrect outputs. Our approach demonstrated promising results in that, overall, 79% of the resulting road-lanes were correct and 77% of true road-lanes appearing on the test images were correctly recovered.

Although we believe our test images pose sufficient challenges for the task of producing lane-level detailed highway maps, for future work, we would like to test our algorithms with more challenging aerial images.

7. ACKNOWLEDGMENTS

This work is funded by the General Motors-Carnegie Mellon University Autonomous Driving Collaborative Research Laboratory (AD-CRL).

8. REFERENCES

- [1] E. Baltsavias and C. Zhang. Automated updating of road databases from aerial imagery. *International Journal of Applied Earth Observation and Geoinformation*, 6:199–213, 2005.
- [2] C.-C. Chen, C. A. Knoblock, and C. Shahabi. Automatically conflating road vector data with orthoimagery. *GeoInformation*, 10:495–530, 2006.
- [3] P. Felzenszwalb and D. McAllester. A min-cover approach for finding salient curves. In *Proceedings of IEEE Workshop on Perceptual Organization in Computer Vision*, 2006.
- [4] M. flavie Auclair Fortier, D. Ziou, C. Armenakis, and S. Wang. Automated updating of road information from aerial images. In *Proceedings of American Society Photogrammetry and Remote Sensing*, pages 16–23, 2000.
- [5] G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. In *Proceedings of European Conference on Computer Vision*, 2008.
- [6] J. Hu, A. Razdan, J. C. Femiani, M. Cui, and P. Wonka. Road network extraction and intersection detection from aerial images by tracking road footprints. *IEEE Transactions on Geoscience and Remote Sensing*, 45(12):4144–4157, 2007.
- [7] P. Kahn, L. Kitchen, and E. Riseman. A fast line finder for vision-guided robot navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11):1098–1102, 1990.
- [8] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan. Detecting ground shadows in outdoor consumer photographs. In *Proceedings of European Conference on Computer Vision*, pages 322–335, 2010.
- [9] F. Leberl, H. Bischof, H. Grabner, and S. Kluckner. Recognizing cars in aerial imagery to improve orthophotos. In *Proceedings of International Symposiums on Advances in Geographic Information Systems*, 2007.
- [10] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1), 2004.
- [11] C. Qian, B. Gale, and J. Bach. Earth documentation: Overpass detection using mobile lidar. In *Proceedings of IEEE International Conference on Image Processing*, pages 3901–3904, 2010.
- [12] J. Schpok. Geometric overpass extraction from vector road data and dsms. In *Proceedings of ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 3–8, 2011.
- [13] Y.-W. Seo. *Augmenting Cartographic Resources and Assessing Roadway State for Vehicle Navigation*. PhD thesis, The Robotics Institute, Carnegie Mellon University, April 2012.
- [14] B. Soleimani, M.-H. Z. Ashtiani, B. H. Soleimani, and H. Moradi. A disaster invariant feature for localization. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1096–1101, 2010.
- [15] Q. Zhu and P. Mordohai. A minimum cover approach for extracting the road network from airborne lidar data. In *Proceedings of ICCV Workshop on 3D Digital Imaging and Modeling*, pages 1582–1589, 2009.