Image Driven GPS Trace Analysis for Road Map Inference *

Jiangye Yuan
Computational Sciences & Engineering Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
yuanj@ornl.gov

Anil M. Cheriyadat
Computational Sciences & Engineering Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831
cheriyadatam@ornl.gov

ABSTRACT

The trace data generated from GPS enabled vehicles is highly valuable for applications such as map inference and traffic analysis. However, the data tends to be noisy due to signal interference. In this paper, we introduce aerial images in GPS trace analysis. Computer vision techniques are developed that effectively integrate image information with GPS data to generate road networks. An image is first segmented by an efficient factorization-based algorithm. A structure tensor approach is proposed to measure the orientation difference between a trace segment and the corresponding image patch. The segmentation result and orientation measures lead to significantly reducing the traces not aligning with roads. The traces are further processed to produce high-quality road networks. We show that our method produces promising results for very noisy GPS data with a low sampling rate and also outperforms the leading method of map inference from GPS traces.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; D.2.8 [**Software Engineering**]: Metrics—*complexity measures, performance measures*

General Terms

Algorithms, Experimentation

*This work was supported in part by U.S. Department of Energy/National Nuclear Security Administration under Grant DOE-NNSA/NA-22. This manuscript has been authored by employees of UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the U.S. Department of Energy. Accordingly, the United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

SIGSPATIAL'13 Nov 05-08 2013, Orlando, Florida USA
ACM 978-1-4503-2521-9/13/11. http://dx.doi.org/10.1145/2525314.2525465.

Keywords

GPS, Aerial image, Road map, Segmentation

1. INTRODUCTION

GPS receivers are widely deployed in everyday vehicles, which generate large volumes of GPS trace data. This data, which consists of sequences of location and time information, attracts a great deal of research, including traffic condition analysis, hotspot analysis, and map inference [3, 8]. Despite prior work, a major problem remains in exploring GPS data, which is the noise in the data caused by GPS errors. The massive GPS data is often generated by low-cost devices. As a result, position records are of limited accuracy. The positioning errors can reach over 100 meters in areas with severe signal interference. Moreover, given the number of devices generating the data, it is infeasible to analyze the noise statistics [8]. Another problem of large-scale GPS trace data is the low sampling rate. For a large GPS data management system, data has to be collected at a relatively low frequency (e.g., once per minute), due to energy consumption and communication cost. Many current GPS analysis methods typically deal with data recorded every 10-30 seconds and cannot perform well when the sample rate decreases.

In this paper, the main goal of GPS trace analysis is to construct road maps. A number of methods have been developed to achieve this goal. These methods mainly adopt three strategies including k-means clustering, trace merging, and kernel density estimation [1]. The assumption behind using these strategies is that the traces aggregate more densely on actual roads than in other areas. Unfortunately, this assumption can be substantially violated in real-world data. Due to measurement errors and low sampling rates, there are cases when only few traces correspond to roads while a significant amount deviate from roads. Such trace data is very difficult, if not impossible, to deal with for the methods solely relying on GPS data.

We introduce a different source of data, aerial images, into GPS trace analysis. With the advancements in remote sensing technologies, aerial images are highly available in terms of coverage and cost. Aerial images complement the trace data with spectral and spatial information that characterizes ground objects. Worth mentioning is that, although aerial images alone are widely used to extract roads and other objects, it has been recognized that automatically extracting high level information from image data is challenging. For example, road extraction from remote sensing images has been extensively studies for decades, but reliable performance is still difficult to achieve, especially for the images of urban areas [6].

Given GPS traces and aerial images, we leverage computer vision techniques to identify traces that lie in potential road regions in images. In particular, we use a factorization-based segmentation method and propose a new method based on structure tensor to examine whether a trace segment is aligned with the corresponding image patch. The two techniques make effective use of image information to filter out the traces deviating from roads. We design a simple approach to obtain high-quality road networks from the filtered traces. Our strategy greatly facilitates the task of inferring road maps from GPS traces, even when raw data are very noisy.

The rest of the paper is organized as follows. The proposed method are discussed in detail in Sections 2 and 3. In Section 4 we conduct experiments on large datasets and provide quantitative evaluation. We conclude in Section 5.

2. IMAGE BASED TRACE FILTERING

2.1 Image segmentation

The first technique we use is image segmentation, which partitions an image into homogeneous regions. Since road regions tend to be grouped in a reasonable segmentation, a GPS trace on roads should not cross many image segments. We define an indicator, which equals to the length of a trace divided by the number of segments it spans. A small value suggests the trace traversing a large number of different regions, which is unlikely to be on roads.

Despite the large number of existing algorithms, segmenting aerial images remains a challenging task, especially for a complex scene containing various ground objects. We employ a factorization-based segmentation algorithm [10]. It has been shown to be effective to segment images with great efficiency. For completeness sake, we give a brief description of the algorithm. An image is convolved with a bank of filters, and a feature at each pixel location is computed, which consists of histograms of different filter responses within a local window. Such a feature can capture the appearance of local window, and a homogeneous region has a representative feature that is similar to other features in the region. Each feature in an image can be approximated by a linear combination of representative features, and combination weights indicate the region ownership of the corresponding pixel. A factorization-based image model can be expressed in the following equation,

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{1}$$

$\mathbf{Y}$ is a feature matrix consisting of columns representing features at all pixel locations. $\mathbf{Z}$ contains columns corresponding to representative features. Each column of $\boldsymbol{\beta}$ is the combination weights at each pixel location. The largest weight in each column indicates the segment the corresponding pixel belongs to. $\boldsymbol{\varepsilon}$ represents the noise.

Based on this image model, the segmentation algorithm aims at factoring $\mathbf{Y}$ into two matrices. By applying singular value decomposition to $\mathbf{Y}$, the number of segments can be estimated, and a subspace can be revealed where all features reside. Initial representative features are estimated by analyzing the feature distribution in the subspace. A nonnegative matrix factorization algorithm is then applied to obtain the factored matrices that give segmentation.

Figure 1(a) shows raw GPS traces containing significant noise. Darker pixels represent more traces overlaid. Figure 1(b) displays the result of applying this filtering step to the data. The number of noisy trace segments is clearly decreased. Here we discard the traces with indicator values less than 20 meters per segment. This value is fixed for other experiments in this paper.

2.2 Orientation estimation

As can be seen from Figure 1(b), there still exist a noticeable amount of noisy traces, which happen to lie in large image segments resulting from either large regions or undersegmentation. Different types of information is required to remove those traces.

Inspired by Harris corner detector [5], we propose a structure tensor approach to examine whether the alignment of a trace segment is consistent with the image content in its surroundings. From an aerial view, we can observe that most objects on or near a road, including vegetation, pavement markings, vehicles, and buildings, spread along roads. Therefore, if we shift the image patch containing a road and compute the pixel difference, the largest difference occurs when the shift is perpendicular to the road orientation, and the smallest difference occurs when it is parallel to the road orientation. We use structure tensors to find the orientation.

Given an image I and an image window W, we slightly shift the window by $(\Delta x, \Delta y)$, and the sum of square differences between W and the shifted window is written as

$$S = \sum_{W} \left[I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)\right]^2, \tag{2}$$

where (x_i, y_i) is the pixel location in the window. The second term in the equation, denoting the shifted window, can be approximated by the first order term of Taylor expansion. Then, Equation 2 can be rewritten as

$$S = [\Delta x, \Delta y]\, A \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \tag{3}$$

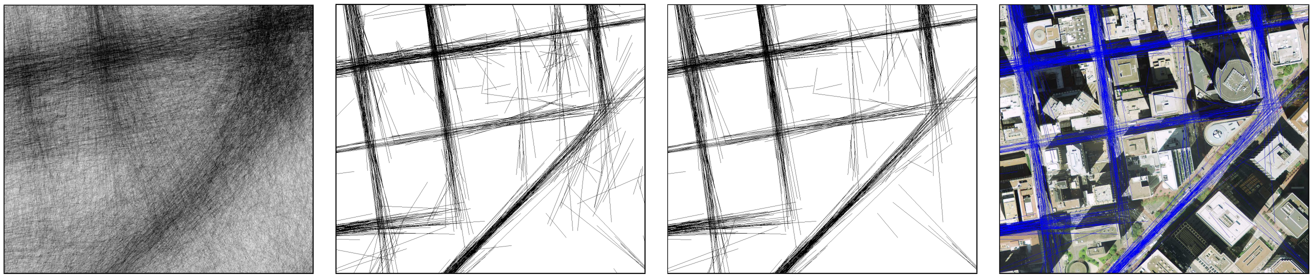
where A is a matrix called structure tensor and takes the following form

$$A = \begin{bmatrix} \sum_W I_x^2 & \sum_W I_x I_y \\ \sum_W I_x I_y & \sum_W I_y^2 \end{bmatrix}. \tag{4}$$

Here I_x and I_y denote the partial derivatives in x and y, which are gradients in the image sense. For corner detection, the relationship between two eigenvalues of A indicate whether the image window contains a corner.

Let U denote the shift vector $[\Delta x, \Delta y]^T$. According to the Rayleigh quotient [4], $\frac{U^T A U}{U^T U}$ reaches the minimum value, which equals to the smallest eigenvalue of A, when U is the corresponding eigenvector. Likewise, the maximum value equals to the largest eigenvalue. It is reasonable to assume a fixed norm for the shift vector. Thus, the minimum value of S is given by the smaller one of two eigenvalues of A, and it is achieved when the shift vector is the corresponding eigenvector. Based on the observation of road appearances, this direction should agree with the road orientation.

Now we can easily examine whether the orientation of a trace segment is consistent with the orientation estimated from the image. For each trace segment, the image patch is defined as a rectangular area around the trace that has the same length as the trace segment and a width of 30 meters. The gradients within the patch are used to construct the



(a) (d) (c) (d)

Figure 1: Image-based filtering. (a) Raw trace data. (b) Traces filtered based on the segmentation result. (c) Traces filtered by examining the alignment with respect to the image content. (d) Filtered traces (blue lines) overlaid on the aerial image.

structure tensor, and the eigenvector corresponding to the smaller eigenvalue gives the orientation of the patch. In the case when the image patch contains no roads, the orientation can still be estimated, but the trace segment is unlikely to have a similar orientation. We only retain the trace that has less than 15 degree difference from the orientation of image patch. Figure 1(c) shows the filtering result after this step. Compared with the original data in Figure 1(a), the noise is significantly reduced, and the road patterns are now clearly visible. To provide a better illustration, we overlay the traces on the aerial images, shown in Figure 1(d). As we can see, the filtered traces are mostly located on the roads.

3. ROAD NETWORK GENERATION

By treating traces as foreground pixels, we obtain a binary image. With the filtered traces well describing roads, we apply morphological operations to the binary image to generate a road network. A closing operation is first performed to fill small gaps among traces. Closing includes a dilation operation where each background pixel next to an object pixel is turned into an object pixel, and an erosion operation where each object pixel next to a background pixel is turned into a background pixel. Then, a thinning operation is used to extract the medial axes, or skeletons, which continuously removes boundaries pixels but preserves the extent and connectivity of foreground objects. A medial axis point is the center of a circle that touches the object boundaries at two or more points. The medial axes extracted from the binary image represent the road network.

Based on the extracted medial axes, we find all the intersection points and end points. We prune those points by merging intersection points close to one another and removing small branches. The Douglas-Peucker algorithm is then used to reduce the path between points into line segments. A graph can be constructed, which provides a typical representation of road network.

4. EXPERIMENTS

In our experiments, we use the GPS trace data of taxi cabs in San Francisco, CA [7]. It contains the GPS coordinates of over 500 taxis in one month. In the dataset, the traces in the downtown areas are highly prone to measurement errors due to tall buildings. The time intervals between two sample points are varying, most of which are 60 seconds. A vehicle can pass several different roads within such long intervals. We use the geo-referenced color images covering the same areas, each of which is a 5000×5000 tile. The spatial resolution is 0.3 m. We select trace data corresponding to two image tiles to test our methods. Figures 2(a) and (b) show the traces, which respectively correspond to a residential area, where the traces are relatively clean, and the downtown area, where the traces are very noisy. Given the space limit, only cropped areas are shown here.

We feed the datasets into the proposed method with fixed parameter values. The results are illustrated in Figures 2(c) and (d), where the road networks are overlaid on the aerial images. For the dataset in Figure 2(a), although the density of traces on different roads varies largely, our method produces a quite complete road network. The roads in the dataset in Figure 2(b) are highly difficult to extract from either the GPS data or the image. Many traces corresponding to roads are completely buried in noise. Road regions in the image are occluded by vehicles and shadows. It can be seen that our method generates very promising results by exploiting the information from both data sources.

Biagioni and Eriksson recently proposed a method that combines several existing techniques to generate road network from GPS data and reported start-of-the-art performance [2]. For comparison, we apply their method, which will be referred to as the BE method, to our GPS dataset. We use the code distributed by the authors.

For quantitative evaluation, we use two indices, completeness and correctness [9], which are commonly used to assess road extraction results. Completeness is the percentage of the extracted roads among the roads in ground truth, and correctness the percentage of the correct extractions in the extracted roads. A buffer width is defined around road vectors for matching two sets of road vectors. We use the road vector data acquired from OpenStreetMap[1] as ground truth.

We apply the BE method to the dataset in Figure 2(a). With the buffer width of 5 pixels. the completeness and correctness are 0.35 and 0.45 for the BE method, while the two indices are 0.58 and 0.91 for our method. By increasing the buffer width to 15 pixels, the BE methods gives 0.53 and 0.62, while our method gives 0.71 and 0.99. Clearly, our method achieves higher scores. The BE method constructs an initial road network based on the trace density and applies map matching to the network to remove the edges with very few matched traces. When applied to this dataset, it confuses many crowded noisy traces as roads, which cause the low correctness rate. Also, because the initial network includes a large number of incorrect edges, the traces are matched to those edges and miss the edges corresponding

[1] http://www.openstreetmap.org/

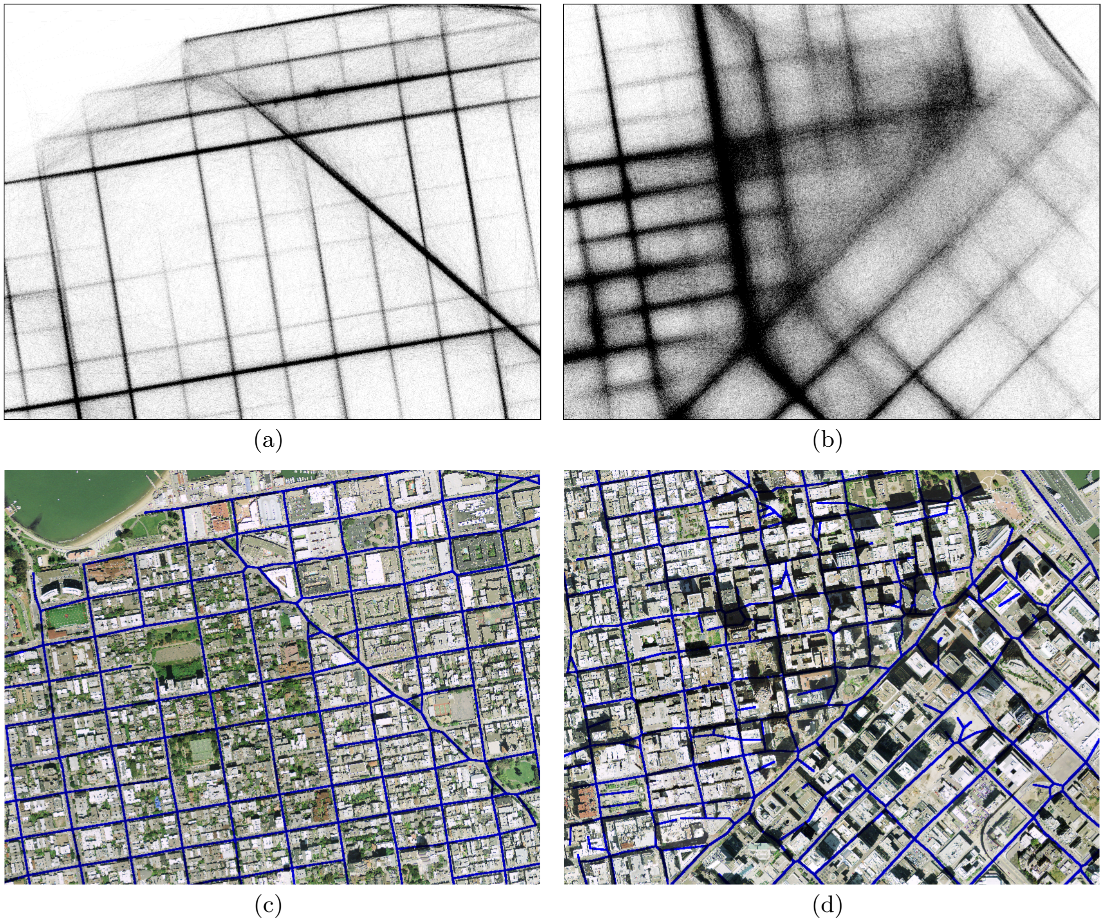


Figure 2: GPS data and generated road networks. (a) The data covering a residential area. (b) The data covering the downtown area. (c) and (d) The corresponding results of road networks.

to actual roads, which decrease the completeness rate. In contrast, our method benefits from the use of image information and identifies much more roads with very few false detections. The correctness rate of our result is particularly high. The reason is that the detected roads are actually verified by both GPS and image information. We have also applied the BE method to the dataset in Figure 2(b), but found that it fails to produce a reasonable result.

5. CONCLUSIONS

We have presented a new method that integrates aerial images with GPS trace data for map inference. Applying computer vision techniques to aerial images produces valuable information for removing noisy traces while preserving the useful ones. Our experiments demonstrate that our method can tolerate significant noise in GPS data and produce accurate results. The comparison shows that integrating images significantly extends the reach of map inference methods using GPS traces.

6. REFERENCES

[1] J. Biagioni and J. Eriksson. Inferring road maps from GPS traces: survey and comparative evaluation. In *Transportation Research Board, 91st Annual*, 2011.

[2] J. Biagioni and J. Eriksson. Map inference in the face of noise and disparity. In *ACM SIGSPATIAL GIS*, 2012.

[3] L. Cao and J. Krumm. From GPS traces to a routable road map. In *ACM SIGSPATIAL GIS*, 2009.

[4] G. Golub and C. V. Loan. *Matrix Computation*. The Johns Hopkens University Press, 1996.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conferencel*, 1988.

[6] H. Mayer. Object extraction in photogrammetric computer vision. *ISPRS Journal of Photogrammetry and Remote Sensing*, 63:213–222, 2008.

[7] M. Piorkowski, N. Sarafijanovoc-Djukic, and M. Grossglauser. A parsimonious model of mobile partitioned networks with clustering. In *COMSNETS*, 2009.

[8] Y. Wang, Y. Zhu, Z. He, Y. Yue, and Q. Li. Challenges and opportunities in exploiting large-scale GPS probe data. *Technical Report HPL-2011-109*, 2011.

[9] C. Wiedemann, C. Heipke, H. Mayer, and O. Jamet. Empirical evaluation of automatically extracted road axes. In *Empirical Evaluation Techniques in Computer Vision*, 1998.

[10] J. Yuan and D. L. Wang. Factorization-based texture segmentation. *Technical Report OSU-CISRC-1/13 -TR01*, 2013.