Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, October 6-9, 2013

Generating Lane Level Road Data from Vehicle Trajectories Using Kernel Density Estimation

E.R.I.A.C.M. Uduwaragoda, *Member, IEEE*, A.S. Perera, S.A.D. Dias, *Member, IEEE*

Abstract— **An informative digital map is a prerequisite for future Intelligent Transportation System (ITS) applications such as lane level navigation systems. Preparing and updating such detailed digital maps by using existing methods such as surveying and image digitization is not practical due to the time and cost involved. We propose a method that statistically mines Global Positioning System (GPS) trajectory data obtained from vehicles to generate a lane level digital map of the road. The proposed method is capable of generating a digital map of the road which contains lane centerlines. The proposed method is independent of the lane width, lane parallelism and can handle lane splits and merges. In addition, a method that can be used to generate a map of lane boundaries is also presented. The proposed methods have been tested using the GPS data collected using vehicles equipped with GPS enabled mobile phones. Results show that the proposed method for lane centerline generation is successful in different road geometries.**

I. Introduction

Next generation transportation applications widely referred to as Intelligent Transportation Systems (ITS) include applications such as lane departure warning, collision avoidance, lane level navigation, autonomous driving, dynamic road closure notification, real time traffic notification, road user charging and automatic speed control [2], [3]. Research oriented towards making these dream applications a reality has been the focus of many vehicle manufacturers and transport related organizations during the last two decades. Among the many advanced requirements, detailed digital map data is one of the core components required for making these ITS applications a reality [3]. A digital map fitting these requirements, unlike those available today, is expected to provide low level details about the road such as positions of lanes, turning restrictions, width of the road and geometry of junctions [2],[3].

Consequently, generation of informative digital maps commonly referred to as Enhanced Maps (EMaps) has become one of the main research challenges faced by the ITS research community. Currently, digital maps of road networks are generated in a number of ways such as digitizing printed maps, extraction of the road network using satellite images and aerial photographs and surveying. These methods are time consuming, very costly and do not allow real time map updating. Furthermore, generating detailed digital maps that contain road structure up to lane level details using these methods is even harder and more costly.

E. R. I. A. C. Manel(email: champikam@uom.lk) , A. S. Perera (e-mail: shehan@uom.lk) and S. A. D. Dias (e-mail: dileeka@ent.mrt.ac.lk), are with the Faculty of Engineering, University of Moratuwa, 10400 Sri Lanka. (phone: 94 112650301; fax: 94112650622)

Preparing and updating digital maps using Global Positioning System (GPS) traces obtained from the vehicles travelling along the road network has been an interesting research area over the last ten years. The removal of Selective Availability in 2000 and the availability of sophisticated GPS receivers, their improved accuracy and their increasing use in vehicles have enabled the collection of more accurate GPS readings, accelerating research in this area [4],[5]. In addition, advances in wireless technology and the availability of wireless infrastructure have enabled the collection of vehicle trajectory data both online and offline. The availability of GPS receivers as a common feature in almost every Smartphone today has further established the viability of collecting trajectory data on a day-to-day basis. Thus, accumulating and processing large volumes of trajectory data for generation and enhancement of maps seems viable both economically and technically. There have been many attempts to reach the goal of incorporating extra information to digital maps by analyzing GPS trajectory data, but the results are well below what is required for practical deployment. In spite of the many problems, current research approaches provide the basis for a range of future research topics.

In this paper, a novel approach that can be used to improve an existing road map using information derived from GPS trajectories of vehicles is presented. The proposed method begins with an existing road map which contains only the road centerline. It then infers lane level details from a collection of GPS trajectories obtained from vehicles travelling along the lane and generates an improved map which contains the lane centerlines. In addition, the positions of lane boundaries can also be computed using the proposed method. The lane level information is extracted by statistically analyzing the probability density distribution of trajectories across the road using non parametric Kernel Density Estimation (KDE) [16].

The rest of the paper is organized as follows. In Section II, we discuss various attempts for the generation and improvement of digital maps using GPS trajectory data. In Section III, the proposed approach is discussed in detail. The results of experimental evaluation of our method are presented in Section IV. Finally, Section V presents the conclusions.

II. Prior Research

Use of vehicle trajectory data for the generation and enhancement of digital road maps has received a great deal of attention over the last two decades and has been researched using many different approaches. In 1998, a research team at Daimler-Benz Research and Technology

978-1-4799-2914-613/$31.00 ©2013 IEEE

Center, Palo Alto, CA, predicted that the positioning data coming from vehicles and other sources could be used for map refinement [1]. They have identified many intelligent vehicular applications that are enabled by high accuracy detailed digital maps obtained in this way.

A more comprehensive research was carried out at the Daimler-Chrysler Research and Technology Center in 1999 to find a way to improve road maps using Differential GPS (DGPS) traces [6]. The main focus was to develop an improved road model by improving the positional accuracy of the road centerline and by computing the number and the positions of the lanes as offsets with respect to the road center line. They have used a weighted averaging method to refine the road centerline and a hierarchical agglomerative clustering procedure to calculate the number and the positions of lanes as offsets from the calculated road centerline. One of the drawbacks in this approach is that they have assumed that the lanes are mutually parallel and last for the length of the entire road segment [6]. However in reality, lane geometry is different in different parts of the road. Thus, the lane width is not always constant and there can be lane splits and merges at different locations of the road. Furthermore, one major assumption of this approach is that a position is more reliable if the position error estimate computed by the DGPS receiver is low. However, there can be cases where the error estimate is low but the position is far from the true position. Further research at Daimler-Chrystler has revealed more powerful algorithms for map matching and lane clustering using DGPS data obtained from vehicles [7]. The new findings include new algorithms for inferring the road centerline from scratch and new algorithms for lane clustering that are expected to handle lane splits and merges. A new lane clustering algorithm called COPKMEANS has been introduced. However, there is no evidence of experimental evaluation of the proposed lane clustering algorithm in roads that have lane splits and merges. Furthermore, the use of DGPS data rather than GPS data limits the applicability of this approach in general.

Cao and Krumm [8] present a method based on the physical attraction between trace points to identify road directions, two adjacent lanes that are going in the same direction and road splits (when a single road going in the same direction subsequently split into two separate roads). However this work does not go into detailed lane model extraction. Brüntrup et al. [9] have proposed an approach that uses a clustering algorithm to map GPS points in the input traces to an existing map, if available. If a map is not available, the algorithm generates it from the GPS traces making mapping possible for unknown terrains. Their approach is capable of inferring the road network with a considerable level of accuracy. However, they have not focused on lane level geometry inference. Davies et al. [10] have proposed an image processing based approach to generate and update digital maps in near real time by dividing the 2D horizontal plane into cells and allocating GPS points obtained from vehicles to appropriate cells in order to create an initial 2D histogram. This histogram is improved by going through different steps to finally generate a road map. They also have focused only on generation of the road centerline. Guo et al. [11] have suggested a novel approach by leveraging the fact that the real value of a location can be represented by the mean value of a large number of Gaussian distributed location data. This approach also is limited to road centerline inference. Image processing based approaches for generating the road centerline from GPS have been presented by Chen and Cheng in 2008 [12] and Shi et al. [13] in 2009. Recent work of Chen et al. [14] has revealed new algorithms for reconstructing a road network from GPS data by modeling the road network as a geometric graph. Edges of the graph have been modeled as polygonal curves and they are used to represent road fragments. Given this graphical representation of the road network, they search for parts for which reconstruction is feasible. After identification of such sections, the reconstruction algorithm reconstructs these selected sections and hence improves them. This work also is limited to road centerline generation.

Knoop et. al [20] have developed a technique called Precise Point Position GPS (PPP-GPS) to collect highly accurate position data. They have also developed a method that finds the locations of lanes by minimizing the difference between the probability density function that describes the lateral positions of the passing vehicles and the observed empirical distribution of the lateral passages. The probability distribution of lateral positions is assumed to be Normal and the lane width, which is a parameter of the assumed Normal distribution, is taken from the motorway handbook.

In analyzing the existing literature, it can be seen that the initial work on using GPS data for map generation and improvement has basically focused on the generation and improvement of the road centerline. Even the few attempts that tried to generate a lane level map use DGPS or improved GPS traces rather than ordinary GPS traces. Our contribution is a novel approach that uses ordinary GPS data, that does not use prior assumptions on the lane width and lane parallelism, and that can accommodate varying road width and different lane geometries such as lane splits and merges. Moreover, our approach estimates the probability density function of the trajectories across the road using a completely non-parametric approach.

III. THE PROPOSED APPROACH

The proposed approach, based on the probability density distribution of the vehicle trajectories, provides methods for estimating the locations of lane centers and boundaries for a given road segment.

A. Estimating the Number and Locations of Lane Centerlines

Usually, vehicles tend to travel at or near the lane centerline. Consequently, the trajectory density at the lane centers is expected to be considerably higher than that at lane boundaries. Visual inspection of trajectory density across a given road segment illustrated in Fig. 1 proves this assumption. Thus, the probability that a trajectory can be found at and near lane centers is higher than the probability that a trajectory can be found at lane boundaries. Therefore, the problem of finding the locations of lanes reduces to determining the locations of lanes across the road where the probability of finding a trajectory is high.

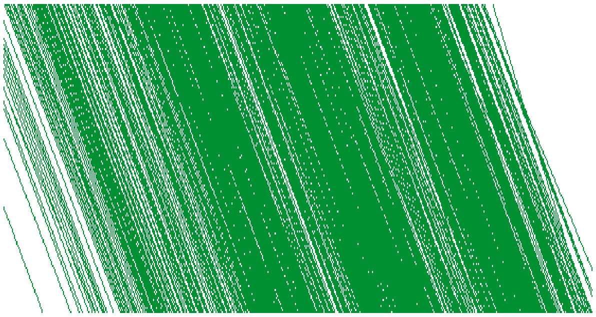

Figure 1. Trajectory density across the road

If a line perpendicular to a chosen segment of the road centerline is drawn through a selected point, it intersects vehicle trajectories corresponding to the selected segment of the road as illustrated in Fig. 2. For convenience and generality, the midpoint of the road centerline segment has been selected as the 'selected point' throughout the study. The intersection point density along the perpendicular line represents the trajectory density across the road segment.

Let P be an intersecting point and Xc and Yc denote its Cartesian coordinates. Xc and Yc are two independent random variables whose probability density functions are not known. However, the probability density function of Xc and Yc across the road represent the trajectory density. Therefore, the probability density functions of these variables are likely to generate peaks at locations on the road where trajectory density is high. Accordingly, the probability density function of Xc and Yc are expected to be multimodal distributions generating peaks at lane centers.

As the probability density distribution of Xc is not known in advance, a non parametric approach was used to estimate it. For this purpose, the well known nonparametric density estimation technique, Kernel Density Estimation (KDE) [16] was used.

Kernel Density Estimation is used to estimate the probability density function of a random variable whose probability distribution is not known in advance.

Let $x_1, x_2, \ldots, x_n$ be a sample of n data points on a d-dimensional space $\Re^d$, drawn from some distribution with

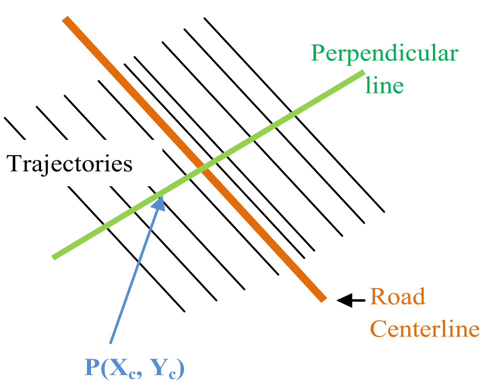

Figure 2. Trajectories intersecting with the perpendicular to the road centerline

an unknown density f. Then, the multivariate Kernel Density Estimator of f at X=x, denoted by $\hat{f}_h(x)$ is given by (1).

$$\hat{f}_h(x) = \frac{1}{nh^d}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

Where $K(u)$ is called the Kernel function and h is the bandwidth [16]. The Kernel function satisfies the following conditions.

$$\int K(u)d(u) = 1 \quad (2)$$

$$\int uK(u)d(u) = 0 \quad (3)$$

$$0 < \mu(K) < \int u^2 K(u)d(u) < \infty \quad (4)$$

Where $\mu(K)$ is a constant.

The kernel function determines the shape of the distribution [19]. Various kernel functions exist: Uniform, Triangle, Gaussian, Epanechnikov, Triweight are examples. The commonly used kernel function is the Gaussian function defined by (5). However, the choice of kernel function has little impact on the efficiency of the density calculation [19].

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right) \quad (5)$$

Choice of bandwidth (h) is very critical for the accuracy of the KDE. In order to generate the kernel density estimate for a given set of data, the bandwidth, which is a positive number, has to be provided as a parameter. Depending on the bandwidth, the number and positions of peaks differ. Inappropriate selection of the parameter would generate incorrect density estimates showing spurious bumps (under smoothing) or hiding important bumps (over smoothing). Various methods that can be used to select the optimal value for the bandwidth based on the dataset are available in literature. Most popular methods include but are not limited to rules of thumb, cross validation methods, direct plug-in methods and smoothed bootstrap [15]. However, these methods also have their pros and cons [15]. The proposed approach uses the new plug-in bandwidth calculation method proposed by Botev et. al. [15] because it is free from the arbitrary normal reference rules used in existing methods.

Our problem domain contains two dimensional (2D) or bivariate data points. The KDE defined by (1) can be simplified to 2D scenario as given in (6).

$$\hat{f}(x, y) = \frac{1}{nh^2}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}, \frac{y - y_i}{h}\right) \quad (6)$$

However, in this research, we consider the two dimensions separately. i.e. the KDE for Xc and the KDE for Yc are generated separately. We are currently investigating the 2D approach. The KDE defined in (6) can be further simplified for one dimensional data as in (7).

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{7}$$

The kernel density estimate (kde) for the random variable Xc can be generated by applying the Gaussian kernel function defined by (5) to equation (7).

The kernel density estimate of Xc generated using this approach for a given road segment is illustrated in Fig. 3. The values of Xc that corresponds to the peaks are the X coordinates of lane centers pertaining to the road segment under consideration. Accordingly, the number of lanes is equal to the number of peaks. The heights of the peaks are different due to the unequal number of trajectories found in each lane. The coordinates of the lane cluster centers can be determined by referring to the locations of peaks (local maxima/modes) of the estimated probability density function. The locations of peaks can be determined by referring to the gradient function of the kde. The gradient of the Gaussian KDE denoted by $\nabla\hat{f}(x)$ can be obtained by applying (5) in (1) and differentiating. This is given in (8).

$$\nabla\hat{f}(x) = \frac{1}{n(2\pi)^{d/2} h^{d+2}} \sum_{i=1}^{n} (x_i - x) \exp\left(-\frac{\|x - x_i\|^2}{2h^2}\right) \tag{8}$$

At local maxima (peaks) and local minima (valleys), $\nabla\hat{f}(x) = 0$. We have used this phenomenon to find the lane cluster centers and places of valleys by referring to the gradient function defined in (8).

The density distribution of Yc can also be generated in the same way. Accordingly, the Y coordinates of lane cluster centers of the road segment can also be determined. Finally, we have the (X,Y) coordinate pairs of the lane centers for a given road segment. The proposed method of finding lane centers for a given centerline segment is summarized in Fig. 4.

First, the road centerline is partitioned into segments with the aim of identifying lane cluster centers corresponding to each segment. Then, a centerline segment is chosen at a time to find the lane cluster centers of that segment. A perpendicular line (L) to the centerline is drawn at a selected point M on this segment. For convenience and generality, the midpoint of the centerline segment can be selected as M. The trajectory database is then searched to find the trajectory

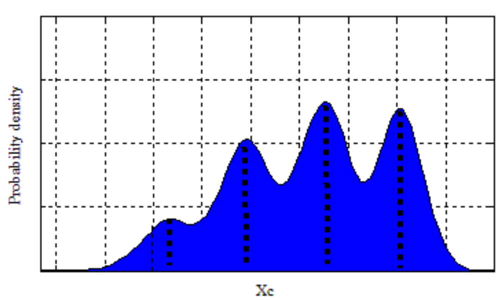

Figure 3. The kde of Xc for a selected road segment

segments that intersect with L. Once this trajectory segment subset is identified, the coordinates of the set of intersection points of each trajectory segment and L, (Xc, Yc), can be obtained. After that, the kernel density estimates for Xc and Yc are calculated. Finally, locations of the peaks of the kernel density estimate are calculated and taken as locations of lane cluster centers. This process is repeated for all the centerline segments to find their corresponding lane locations. The straight line joining adjacent lane cluster centers forms the lane centerline.

B. Estimating the Lane Boundaries

Kernel density estimate generated for finding the locations of lane centers can also be used to find lane boundaries and the lane width. In the same way as peaks of the kernel density estimate represent lane centers, we assume that the valleys represent lane boundaries. Therefore, from the kernel density estimates of Xc, and Yc it is possible to find the coordinates of the lane boundary positions.

IV. EXPERIMENTAL EVALUATION

A. Experimental Design

To experimentally evaluate the proposed method, we have used the 'Mobile Century' dataset [17], [21] from the

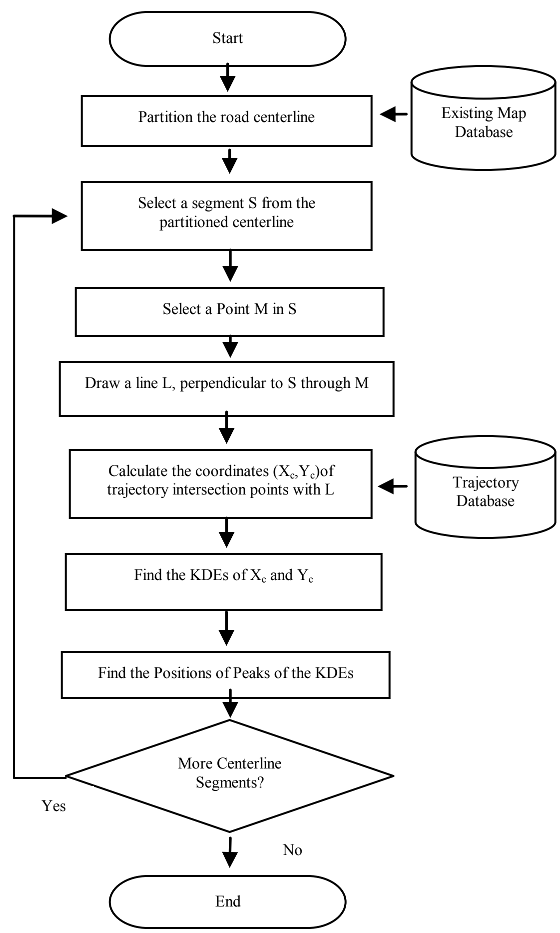

Figure 4. Process flow of finding lane cluster center

Mobile Millennium Project of the University of California, Berkley. The data has been recorded along the Nimitz Freeway (freeway I-880) near Union City in the San Francisco Bay Area. It contains trajectory logs recorded by vehicles equipped with GPS-enabled phones. Position data is available as geographical coordinates (latitude and longitude pairs) and are in degrees. The geographical coordinates were converted to planar Cartesian coordinates before further processing. To experimentally verify the proposed method, GPS data corresponding to three parts of the Nimitz Freeway were selected. The shape and the nature of the selected three parts are different from each other. Fig. 5 illustrates the satellite images of the selected road parts obtained from Google Earth© .

- Part 1 : A straight section

  The first part was a straight portion of the road in the northbound direction and contained eight (8) road centerline segments. Four constant width lanes can be seen in this part of the road in each direction

- Part 2 : A section with a bend

  The second part of the road contained a curved portion in the northbound direction and consisted of sixty (60) road centerline segments. There are five lanes of constant width in this part of the road in each direction

- Part 3 : A section containing a lane split

  The third part included a freeway exit in the southbound direction and contained thirty (30) centerline segments.

The outliers of the GPS data were removed using the two phase NECTOD algorithm [18] and the trajectories as well as the road centerlines were segmented into 20 m segments prior to using them for lane detection. Our method for lane extraction was then applied to the three selected sections of the road separately.

(a) Part 1

(b) Part 2

(c) Part 3

Figure 5. Road parts selected for the experiment

B. Determination of Lane Centerlines

Fig. 6 illustrates the generated lane centerline maps for the chosen direction (northbound direction for Part 1 and Part 2, Southbound direction for Part 3) of the selected road parts.

Fig. 6(a) shows the lane centerlines generated for road Part 1. Four lane centerlines representing the true lanes have been generated by the proposed method. Similarly, Fig. 6(b) shows five lane centerlines representing the five lanes of road Part 2. This proves that the proposed algorithm is capable of capturing curved parts of the road as well. In Fig. 6(c), five lane centerlines are shown representing the five true lanes. The leftmost lane which becomes an exit is clearly identified in the generated map. This verifies that the proposed method for lane centerline calculation is independent of the width of the lane. Even though the width of the lane may change along the road, the proposed approach looks at the trajectory density only and the dense regions are identified as lane centerlines. Furthermore, Fig. 6(c), also proves that lane parallelism is not expected by

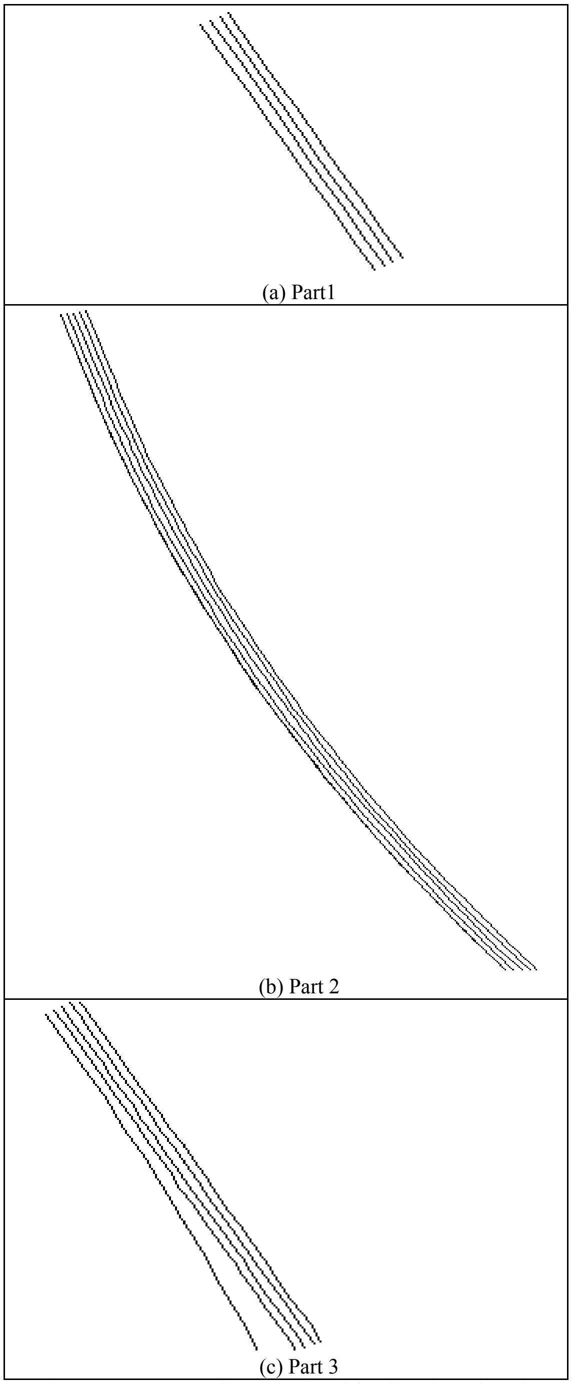

Figure 6. Lane centerlines generated for the selected road parts

the proposed method.

In order to further illustrate that lane centerline generation does not depend on the width of the lanes or on lane parallelism, changes in the kernel density estimate at different segments of the centerline of road Part 3 are illustrated in Fig. 7. Fig. 7 clearly shows how the left most lane gradually separates from the rest of the lanes and becomes an exit as the segment ID advances.

Fig. 8 analyses how the lane error varies with the number of trajectory traces used. The error is calculated by

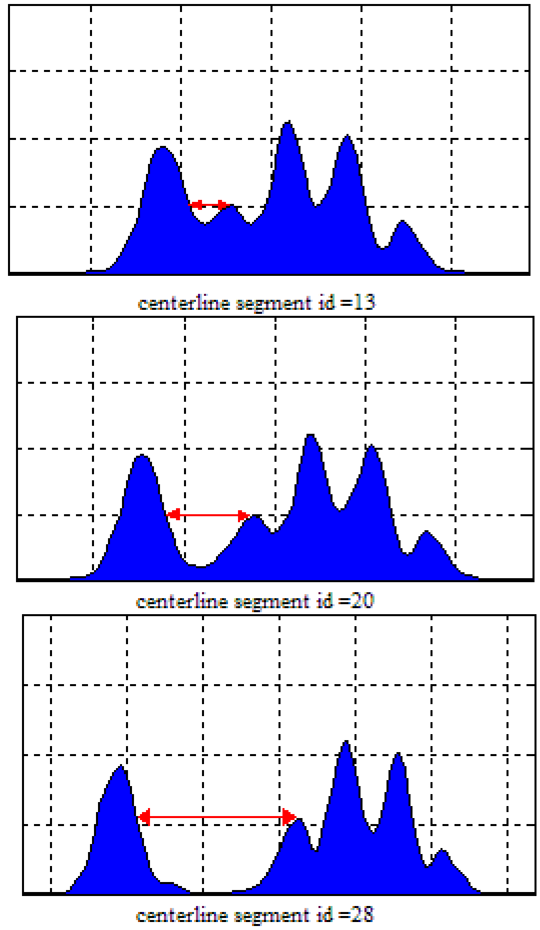

Figure. 7. KDE graphs for some selected centerline segments of road part 3.

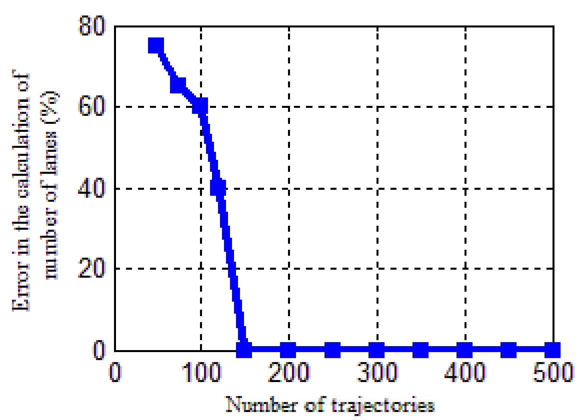

Fig. 8. Error in calculation of number of lanes vs. the number of traces.

comparison with the satellite image of the corresponding road segment. It is observed that a minimum of 150 traces is needed to obtain accurate results in the presence of 4-5 lanes.

A proper quantitative and qualitative evaluation for accuracy can only be done using a lane level map specifically prepared using high accuracy GPS receivers (eg. DGPS) as the ground truth. We were unable to obtain a map of the chosen road so prepared, and hence were unable to perform a proper quantitative evaluation of the positional accuracy. Although Google Earth is not a good reference to verify positional accuracy [22], we overlaid a portion of the generated lane centerlines on Google Earth as a means of visual comparison. The result is shown in Fig 9.

C. Determination of Lane Boundaries

The generated lane boundary maps for the selected road parts in the chosen directions (northbound direction for Part I and Part II, Southbound direction for Part III) are illustrated in Fig. 10.

Fig. 10 shows that although the generated boundary maps for road Part1 and road Part 2 tally with the corresponding satellite image in Fig. 5, it gives incorrect lane boundaries in road Part 3. If the lanes are numbered from left as L1, L2, L3, L4 and L5 for ease of reference, the gap that exist between the exit lane (L1) and its neighbor lane (L2), is not shown in the generated map. The reason for this incorrect calculation of boundaries can be explained using a sample of the kernel density estimate graph of the corresponding part of the road. The kernel density estimate graph in Fig. 11 shows that there are six valleys corresponding to six lane boundaries. The valley between L1 and L2 represents the gap between the two lanes, not the correct lane boundaries. This in turn inaccurately calculates the coordinates of the boundaries of L1 and L2. This also results in incorrect calculation of the lane widths. Therefore, the proposed method for calculating lane boundaries and widths does not perform satisfactorily when the lanes are not contiguous as in the case of a highway entrance or exit lane.

V. DISCUSSION AND FUTURE WORK

A novel method for extracting lane level road information such as lane centers, boundaries and width by estimating the probability density function of trajectories

Fig. 9. Portion of the generated centerline map overlaid on Google Earth.

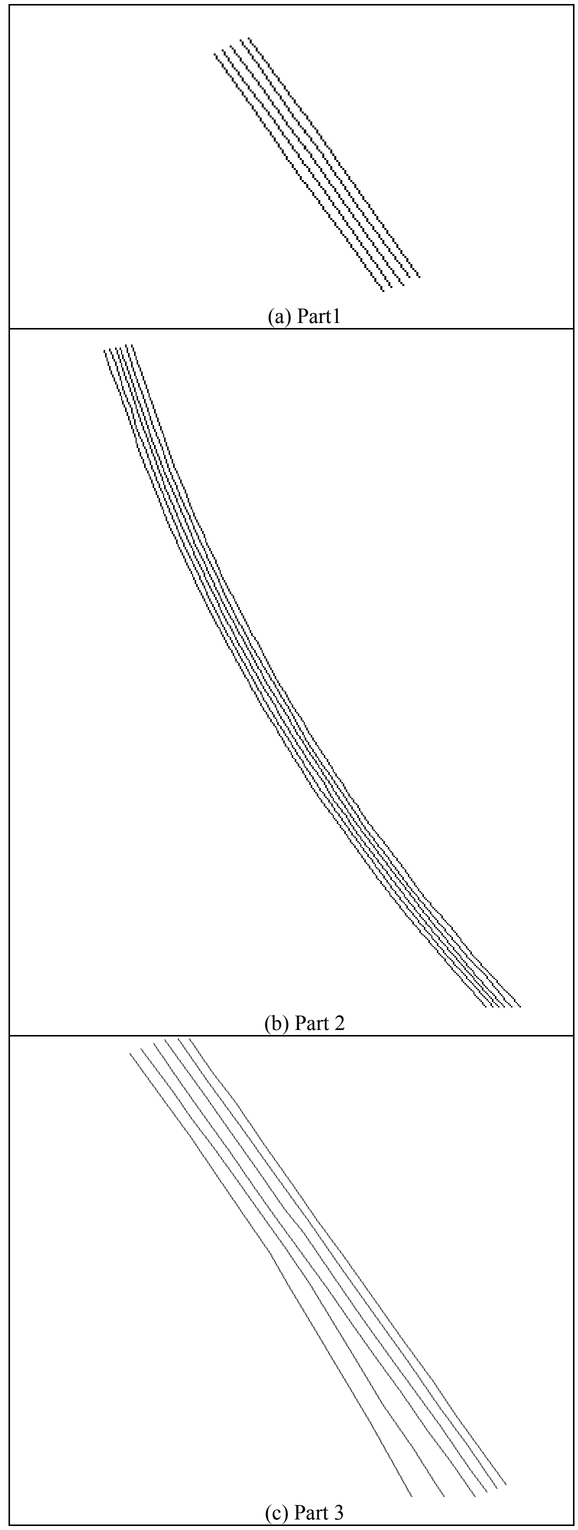

Figure 10. Lane boundaries generated for the selected road parts

across the road using a fully non-parametric approach was presented.

Unlike the existing methods that use highly accurate GPS data such as DGPS or PPP-GPS, the proposed method uses ordinary GPS data obtained from vehicles.

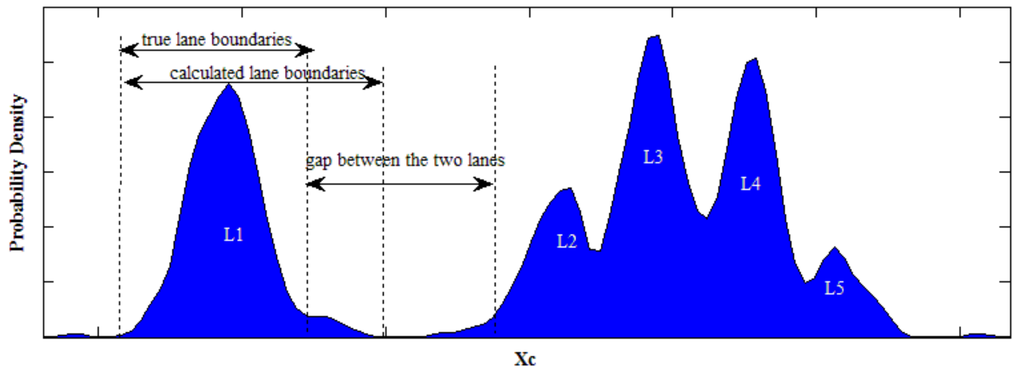

Fig. 11. The difference between actual and calculated lane boundaries.

Furthermore, it does not make stringent assumptions on lane parallelism or constant lane width. It was shown that the lane centerline could be accurately calculated if a minimum of 150 trajectories were present irrespective of the different road configurations such as straight sections, bended sections and sections with lane splits/merges. However, this method needs improvement in order to generate correct lane boundaries in the presence of gaps between lanes as in the case of entry and exit lanes. Our future work involves research on the use of the bivariate kernel density estimation for lane finding and the popular Mean Shift algorithm to find the lane clusters. Using principal curves for lane cluster identification is another future research objective.

ACKNOWLEDGMENT

The authors express their gratitude to the California Center for Innovative Transportation and the University of California Berkeley for providing the trajectory data set of the Mobile Century Field Experiment [17], [21].

REFERENCES

[1] C .K. H. Wilson, S. Rogers, and S. Weisenburger, "The potential of precision maps in intelligent vehicles," in Proc. *IEEE Int'l Conference on Intelligent Vehicles*, Stuttgart, Germany, Oct. 1998, pp. 419-422.

[2] EDMap team (2004, November 19). Enhanced Digital Mapping Project Final Report. [Online]. Available: http://www.nhtsa.gov/DOT/NHTSA/NRD/Multimedia/PDFs/Crash%20Avoidance/2004/FinalRept_111904.pdf.

[3] J.C. Pandazis. (2002). NextMap: Investigating the Future of Digital Map Databases. [Online]. Available: http://www.ertico.com/assets/download/nextmap/2_164v21.zip.

[4] D. Guo, S. Liu, and H. Jin, "A Graph-based Approach to Vehicle Trajectory Analysis," *Journal of Location Based Services*, vol. 4(3), pp. 183-199, Sep. 2010.

[5] G. Agamennoni, J. I. Nieto, and E. M. Nebot, " Robust Inference of Principal Road Paths for Intelligent Transportation Systems," *IEEE Trans.on Intelligent Transportation Systems*, vol. 12(1) , pp. 298-308, March 2011.

[6] S. Rogers, and S. Schroedl, "Creating and Evaluating Highly Accurate Maps with Probes Vehicles," in *Proc. IEEE Intelligent Transportation Systems*, Oct 2000, Dearborn (MI), U.S, pp.125-130.

[7] S. Schroedl, K. Wagstaff, S. Rogers, P. Langley, and C. Wilson, "Mining GPS Traces for Map Refinement," *Data Mining and Knowledge Discovery*, pp. 59–87, vol. 9(1), July 2004.

[8] L. Cao, and J. Krumm, " From GPS Traces to a Routable Road Map," in *Proc. 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Seattle, Washington, November 04-06, 2009, pp. 3-12.

[9] R. Brüntrup, S. Edelkamp, S. Jabbar, and B. Scholz, "Incremental Map Generation with GPS Traces," in *Proc. 8th International IEEE Conference on Intelligent Transportation Systems*, Vienna, Austria, September 13-16, 2005, pp. 413-418.

[10] J. J. Davies, A. R. Beresford, A. Hopper, "Scalable, Distributed, Real-Time Map Generation," *IEEE Transactions on Pervasive Computing*, vol. 5(4) , pp. 47-54, October 2006.

[11] T. Guo, K. Iwamura, and M. Koga, "Towards High Accuracy Road Maps Generation from Massive GPS Traces Data," in *Proc. IEEE Int'lGeoscience and Remote Sensing Symposium*, 2007, pp. 667-670.

[12] C. Chen, and Y. Cheng, "Roads Digital Map Generation with Multi-track GPS Data", in *Proc. 2008 Int'l Workshop on Education Technology and Training & 2008 Int'l Workshop on Geoscience and Remote Sensing* , December 2008, volume 01, pp. 508-511.

[13] W. Shi, S. Shen, Y. Liu, (2009a), "Automatic Generation of Road Network Map from Massive GPS Vehicle Trajectories," in *Proc. 12th Int'l IEEE Conference on Intelligent Transportation Systems*, St. Louis, MO, USA, October 3-7, 2009, pp. 48-53.

[14] D. Chen, L. Guibas, J. Hershberger, and J. Sun, "Road Network Reconstruction for Organizing Paths," in *Proc. 21st ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010, Hyatt Regency Austin, Austin, Texas, January 17-19, 2010, pp. 1309-1320.

[15] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, vol. 38 (5), pp. 2916-2957, 2010.

[16] B.W. Silverman, *Density Estimation for Statistics and Data Analysis*, London:Chapman & Hall, 1998, pp. 34-93.

[17] J. C. Herrera, D. B. Work, R. Herring, X. Ban, and A. M. Bayen. (2009). "Evaluation of Traffic data Obtained via GPS-enabled Mobile Phones: The Mobile Century field experiment,". [online]. Available: http://homepages.rpi.edu/~banx/publications/Herrera_GPS_TR18C_2010.pdf..

[18] C. Manel, A.S. Perera, and S.A.D. Dias, "Outlier Removal in Network Constrained Trajectories," in *Proc. ENTC Research Seminar*, vol 3 (9), Department of Electronic & Telecommunication Engineering, University of Moratuwa, October 2011, pp. 2-7.

[19] W. Zucchini. (2003, October). Applied Smoothing Techniques Part 1: Kernel Density Estimation. [Online]. Available: http://isc.temple.edu/economics/Econ616/Kernel/ast_part1.pdf

[20] V.L. Knoop, P.J. Buist, C.C.J.M. Tiberius, and B. van Arem, "Automated lane identification using precise point positioning an affordable and accurate GPS technique," in Proc. *15th Int'l IEEE Conference on Intelligent Transportation Systems (ITSC),* Hilton, Anchorage, AK, USA, 16 - 19 Sep *2012, p*p.939,944.

[21] Mobile Century. [Online]. Available: http://traffic.berkeley.edu/project/mobilecentury.

[22] K. Beck, and K. Ibrahim, " On the Positional Accuracy of the GoogleEarth Imagery", in *Proc. FIG Working Week 2011, Bridging the Gap between Cultures*, Marrakech, Morocco, 18-22 May 2011.