As mobile devices came into wide use, it became practical to collect travel data in personal logs. Many studies have been conducted to extract meaningful information from this trend. In this study, we present a system for monitoring road-geometry and crash-risk estimation, based on trajectories created using a smartphone-aided sensor network. The proposed system consists of a number of node vehicles with smartphone applications for GPS data collection and a map server which aggregates the collected GPS trajectories and estimates road conditions. In order to estimate road geometry and crash risk information, the trajectories were segmented and categorized into groups according to their headings. Based on the processed trajectories, the geometry of the road section was estimated using the principal curve method. The crash risk of the road section was estimated from the constructed road geometry and the density map of the trajectories. Our system was evaluated using bicycle trajectories collected from segregated bicycle tracks in Seoul, Korea. Constructed geometry and crash-risk information of the track was compared with real track geometry and crash data. As a result, the estimated road geometry showed over 74% similarity and the calculated crash risk (61%) matched the real crash data.

1. Introduction

As the technology of handheld devices and satellite based tracking systems has improved, creation of daily travel logs has become practical and popular. A travel log of a vehicle or person could become a source of meaningful information that might include kinematics (i.e., speed, acceleration, and heading) and semantic information (i.e., region of interest, purpose of trip, user behavior, and traffic environment).

A number of studies have been performed to extract and analyze information based on the trajectories of vehicles or persons [1–8]. One of the most popular issues with trajectory-based analysis is estimation of road geometry based on a large number of vehicle trajectories [1–4]. The trajectory of vehicles has also been used to analyze vehicle accidents and the behavior of drivers. Based on GPS-sensor data collected in real-time, crash-risk factors including roadway characteristics, environmental factors, and driver or bicyclist behavior were revealed [9–13].

In this study, we propose a system which estimates road geometry and crash risk using vehicles that perform the role of sensor nodes and which collects data about their movements using smartphones. The road geometry was estimated by applying the principal curve method to the vehicle trajectories. We expected that the factors affecting the crash risk would be estimated by analyzing aggregated, real-time trajectory data. We calculated the crash risk of a road section based on the distance traveled and shape of the road geometry and with a density map constructed with the trajectories. Compared to previous approaches, the proposed system can more easily cope with environmental or geometric changes of the roadway without additional information from maps or field investigation. This system can also support low-cost, periodic crash-risk analysis.

The components and procedures of the proposed system will be shown in Section 2 and its experimental results in Section 3. We present our conclusions in Section 4.
2. System Structure

The proposed system consists of a node vehicle and a map server. The node vehicle contains a smartphone application (SPA) designed to collect travel path using a GPS sensor and to communicate with the map server. This SPA also provides road-geometry and crash-risk information to a user.

The map server constructs and updates the road-geometry and crash-risk information using received sensor data and then distributes the updated road information to the node vehicles. Figure 1 shows how the proposed system operates. The SPA collects the position data of the node vehicle and sends the data to the map server via the Internet. Based on the data sent from the SPAs, the map server estimates the road-geometry and crash-risk information about a specific road section.

2.1. Node Vehicle. There were two types of node vehicles: motor vehicles and nonmotorized vehicles. Because the behavioral characteristics (i.e., changes in heading, acceleration, deceleration, and stop rate) of nonmotorized vehicles are much different from those of motor vehicles, the node-vehicle type is an important factor for the estimation of node-vehicle state and crash-risk information [8].

Using the GPS receiver in the smartphones, the node vehicle collects data on the travel path and determines its own moving state.

Communication between the SPA and the map server was established using TCP/IP. The SPA sent collected position data to the map server every 60 sec or when an abnormal state was detected. The driving state was determined based on the sequential positions of the node vehicle.

If sudden direction or speed changes occur, the SPA checks its current position and compares the changes with the road geometry. The accelerometers in the smartphones were also utilized to detect abnormal states. Figure 2 shows an example of an abnormal state. The adjusted headings \((hc_{1-4})\) were determined by the differences in the positions of the node vehicle \((p_{1-6})\) and the headings \((h_{1-6})\), defined as the angle of the vector from \(p_i\) to \(p_{i+1}\), with respect to the north. The headings in the figure are showing sudden changes compared to the road geometry. If accumulated change exceeds a certain threshold (i.e., 90 degrees) or the node positions are not within the road width, then the SPA defines the current state as an abnormal one. The threshold value varies for the type of node vehicle. In order to minimize the data sent from the node vehicle, the sensor data were compressed using a lossless compression algorithm.

The graphic user interface (GUI) of the SPA shows the current position and travel trajectories on map images.
Existing road section? 

Yes

Road segmentation

No

Update road geometry and density map

Update crash-risk information

END

FIGURE 3: Updating procedure of map server.

provided by Google Maps. The crash-risk information of the current bicycle position was shown on the top of the GUI with color and text. The application beeps when the crash-risk information is changed.

2.2. Map Server. The map server estimates the geometry and crash-risk information of the road section based on the GPS-sensor data collected from the node vehicles. Figure 3 shows the updating procedure of the map server.

Once the map server receives sensor data from the node vehicles, the server checks whether the sensor data belongs to the current road segment. If not, new road segments are constructed for the received sensor data. Based on the GPS-sensor data, the bidirectional road geometries and density map are constructed or updated, and, finally, the geometry and crash-risk information is recalculated for the road segment. The map server sends updated information to the node vehicles when they move onto different road sections or when its corresponding road segment is changed.

2.2.1. Road Geometry. As shown in Figure 2, each piece of GPS-sensor data collected by a node vehicle has its own direction at every point. Collected GPS data about a certain road section are categorized into two groups according to headings and then utilized to construct the road geometry.

The road geometry was constructed utilizing the latitude and longitude from the GPS data, with the principal curve method. The principal curve is a “self-consistent” smooth curve which shows a nonlinear summary of a set of given data points [14]. It has been applied to various fields of research including road geometry estimation [2, 3]. “Self-consistent” means that any point on the principal curve coincides with the average of the data points which were projected to that curve, and, therefore, that the curve passes through the middle of the data cloud. Let \( X = (X_1, X_2, \ldots, X_d) \) denote a \( d \)-dimensional random vector with finite second moments, and let \( f(t) = (f_1(t), f_2(t), \ldots, f_d(t)) \) denote a smooth curve parameterized by \( t \). For any point \( x \), the projection index \( t_f(x) \) is the largest parameter (value \( t \)) whose distance between \( x \) and \( f(t) \) is minimized and defined by

\[
  t_f(x) = \sup_t \left\{ t : \|x - f(t)\| = \inf \|x - f(r)\| \right\}.
\]  

A smooth curve \( f(t) \) is a principal curve if it is self-consistent and does not have any self-intersections or subsets of infinite length. The principal curve was estimated using the Hastie and Stuetzle algorithm [14].

In this way, two road geometries (in opposite directions) were constructed. Some GPS units fail to provide elevation data, so only latitude and longitude information was used to construct the road geometry. To enhance the present 2-D trajectories into 3-D ones, we needed the elevation of each geometric point. The elevation information for each position was attached by utilizing the DEM (digital elevation model) database.

2.2.2. Road Segmentation. The road geometry constructed with the collected GPS-sensor data was divided into a number of segments based on the length and geometric homogeneity of the segment. Namely, the road geometry was sequentially analyzed and divided into segments at geometric points where a sudden change in elevation or curvature occurred. The length of each segment was limited to 3 km, and if the length was shorter than 0.02 km, the segment was merged with a neighboring one with the most similar properties.

The GPS sensor data transferred from the node vehicles were assigned to a road segment according to position and heading and then utilized to update actions involving that segment. Heading was used to separate a crossroad into two different road segments.

2.2.3. Density Map. Vehicle density is closely related to the crash risk of a road section. Generally, a road section with high vehicle density is expected to have traffic congestion and higher crash risk than that with low density. For this reason, density maps based on collected sample data have been utilized to estimate the real density [15, 16]. Density maps have also been used to analyze and visualize the density of motor vehicles, airplanes, and vessels [17, 18]. The proposed system constructs two density maps for each road segment based on the GPS data groups, which were divided according to their headings.
The width and height of the density map were decided by examining the maximum and minimum coordinates among the trajectory points:

$$\text{map}_{(g,t,w)} = \max(p_{(g,i,\text{lng}}) - \min(p_{(g,i,\text{lng}}),$$

$$\text{map}_{(g,t,h)} = \max(p_{(g,i,\text{lat}}) - \min(p_{(g,i,\text{lat}}).$$

Here, $g$, $t$, and $i$ are the indices of the GPS data group, road segment, and trajectory point, respectively; $w$ and $h$ are the width and height of the density map; lng and lat denote the longitude and latitude of the point $p$. The trajectory points were transformed into density map coordinates and then the Gaussian kernel was applied to each point on the density map.

2.2.4. Crash-Risk Information. The crash risk of each road segment was determined using the estimated road geometry and density map.

First, we expected that node vehicles moving in one direction would influence those moving in the opposite direction. We assumed that this influence could be calculated by comparing two road geometries of a road segment constructed with grouped GPS-sensor data. In order to compare the road geometries, the differences between them, in distance and shape, were calculated using the Fréchet distance. The Fréchet distance was introduced by Alt and Godau [19] to measure the resemblance between curves and has been utilized in various fields to measure the similarity of objects [20]. The Fréchet distance is defined as

$$\delta_F(P,Q) = \min_{\alpha \in [0,1]} \max_{\beta \in [0,1]} \| P(\alpha(t)) - Q(\beta(t)) \|,$$

where $P$ and $Q$ are curves with length of $N$ and $M$, and positions on $P$ and $Q$ are expressed as a function of $\alpha(t)$ and $\beta(t)$; $t$ ranges from 0 to 1; $\alpha(t)$ and $\beta(t)$ are continuous and increasing functions with $\alpha(0) = 0, \alpha(1) = N, \beta(0) = 0, \text{and } \beta(1) = M$. The Fréchet distance $\delta_F(P,Q)$ is the minimum distance, among the maximum distances, between every possible $\alpha(t)$ and $\beta(t)$. The Fréchet distance between two polygonal lines is the minimum threshold $\epsilon$ satisfying $\delta_F(P,Q) \leq \epsilon$ and $\epsilon \geq 0$, and it is generally computed using the free space of the polygonal lines. The free space is the sublevel set of a distance function for the parameter space of two polygonal lines and is defined as

$$F_{\epsilon} := \{ (s,t) \in [0,p] \times [0,q] \mid d(P(s),Q(t)) \leq \epsilon \},$$

where $p$ and $q$ are the number of edges of $P$ and $Q$.

Second, the density map was utilized to estimate the crash-risk information. In order to find the influences mentioned above, an intersection map was calculated which shows the intersected density between two density maps of a road segment. Consider

$$\text{inter}_{(x,y)} = \begin{cases} d_{(1,x,y)} + d_{(2,x,y)} & \text{if } (d_{(1,x,y)} > 0, d_{(2,x,y)} > 0) \\ 0 & \text{otherwise.} \end{cases}$$

Here, $\text{inter}_{(x,y)}$ denotes the value of the intersection map at the position of $x$ and $y$, and $d_{(i,x,y)}$ is that of the density map with group index $i$.

Using the Fréchet distance between two road segments and the intersection map, the crash-risk information $C_t$ of a road segment was defined as

$$C_t = w_f (d_{\max} - d) + w_d m,$$

where $t$ is an index of the road segment index, and $w_f$ and $w_d$ are the weights for the Fréchet distance and density map, respectively. The normalized Fréchet distance of the segment is presented by $d$, and $m$ is the ratio of the intersection map points which exceeded the threshold value. Since we expected that the Fréchet distance would be negatively correlated with the crash-risk information of the road, it was subtracted from maximum Fréchet distance, $d_{\max}$.

3. Experiments

In order to evaluate the proposed system, we attempted to construct the geometry and crash-risk information of the road sections and compare them with real road geometry and crash data.

For the experiments, a nonmotorized node vehicle was selected. We expected that crash-risk estimation in this study (i.e., comparing road geometries and density maps between node vehicles moving in opposite directions) would be more suited to the behavioral characteristics of nonmotorized vehicles, particularly for bicycles. Therefore, for the first evaluation of our system, bicycles were used as the nodes.

Figure 4 shows an example of a node bicycle with the SPA. A smartphone is attached to the top of the handlebar to provide visual information to users during their rides.

3.1. Data Collection. For collection of sensor data, a 20.43 km bicycle path was selected in Seoul, Korea. It is a two-way bicycle route which traverses the center of Seoul City, from...
east to west, along the Han River. The path is also connected to roads leading to business centers in both northern and southern areas of the city. This bicycle path is the major one for bicycle commuters in Seoul because of its easy accessibility and separation from motor vehicles. The yearly number of bicyclists and pedestrians using the path has steadily increased from 7.2 million in 2009 to 10.2 million in 2010 [21]. Another advantage of this route is that there are no nearby long buildings that could interfere with the GPS signals; thus we expected less error in our GPS data.

From the bicycle path mentioned above, only a 20.43 km section was selected (for collecting sensor data) because it has relatively higher bicycle traffic than other parts. Our track section has six bicycle entrances, 15 junctions, and only one signalized intersection with motor vehicle traffic. Detours along the path were not included in the data collection. Korean government statistics show that the reported number of bicycle-related crashes on our track section was 89 from 2009 to 2012. In Figure 5, the bold line just south of the river shows the selected track section. The locations (i.e., latitude and longitude) of the east and west ends of the study section were 37.554785, 126.874924, 37.53396, and 127.030449, respectively.

The GPS-sensor data for the experiments were collected by Korean bicyclist groups. The sensor data was collected for six months (May to October, 2012), with the aforementioned smartphone application (SPA).

For each one-way trip made by an individual bicyclist, a set of GPS-sensor data was generated. Hence if a round trip was made, two sets of sensor data were generated. The collected GPS-sensor data were preprocessed before being sent to the map server, to exclude possibly erroneous data. During the preprocessing, GPS data showing speeds above 70 km/h were deleted. The remaining data showed a concentrative distribution along the track section.

3.2. Experimental Results. During the experiments, a total of 732 travel logs were recorded and 2% of them were discarded because of erroneous signals. The remaining 717 trajectories were divided into 190 road segments, with an average length of 0.10 km.

3.2.1. Road-Geometry Results. The latitudes and longitudes of estimated road-geometry points were compared with the points and lines between points of the real road geometry. If the shortest distance between an estimated and real point (or line) exceeded the threshold, the estimated point was considered dissimilar. The result was that 74% of the estimated points showed geometrical similarity with the real road geometry.

Figure 6 shows two examples of estimated road geometries. The blue line indicates the road geometry in the direction from down to up, and the red one represents the reverse. The road segments (a) and (b) share similar track geometry; however, their geometry similarities were different. This difference was mainly caused by environmental differences. Unlike (a), (b) had several nearby convenience facilities (circled areas in Figure 6(b): a convenience store, a public toilet, and benches with shades) which attract bicyclists and pedestrians. This example confirms that estimated road geometry may be different from the real one. However, the estimated geometry can reflect real traffic information and thus could be more useful for the crash-risk analysis.

3.2.2. Crash-Risk Results. Table 1 summarizes the average, minimum, maximum, and standard deviation of calculated Fréchet distance and of the intersection map with the number of crashes of the road segments.

|                      | Fréchet distance | Intersection map | Crash number |
|----------------------|------------------|------------------|--------------|
| Minimum              | 0.085            | 0.0              | 0            |
| Maximum              | 1.982            | 0.0041           | 3            |
| Average              | 0.436            | 0.0016           | 0.420        |
| Standard deviation   | 0.370            | 0.0007           | 0.737        |

The standard deviation of the number of crashes is relatively higher than the Fréchet distance and intersection map. This may be caused by excessive zero data, a condition frequently observed in crash data. The histograms of the Fréchet distance and intersection map are shown in Figure 7. Both histograms have similar, normal-distribution shapes.

The crash numbers of the road segment were used as the degree of the real crash risk and compared with the estimated crash risk of that segment. Both values were normalized
from zero to one. As a result, 61% of the crash risk of the road segments was matched with the corresponding crash numbers in the range of $-0.2$ to $+0.2$. Namely, the estimated crash risks have a 61% accuracy rate with a ±20% margin of errors.

Figure 8 shows the estimated crash risk with the crash numbers for each segment. We assumed that mismatched track segments were caused by (1) environmental crash-risk factors not reflected in the trajectory of the node bicycles, (2) relatively small numbers of crashes, and (3) excessive zero data for the number of crashes.

4. Conclusion

In this study, we proposed a system for road-geometry and crash-risk estimation based on GPS trajectories collected with smartphone-aided sensor networks. The proposed system consists of node vehicles which collect position data with GPS sensors and a map server which manages and updates the road geometry and crash risk. We tested our system with node bicycles on a segregated bicycle path and found the proposed system useful.

For future work, motor vehicles and nonmotorized vehicles will be utilized together for analysis of travel in a mixed-traffic environment.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

Professor Ilmin Kim was financially supported by Hansung University.

References

[1] Y. Zhu, X. Liu, and Y. Wang, “Pervasive urban sensing with large-scale mobile probe vehicles,” International Journal of Distributed Sensor Networks, vol. 2013, Article ID 762503, 7 pages, 2013.
[2] G. Agamennoni, J. I. Nieto, and E. M. Nebot, “Robust inference of principal road paths for intelligent transportation systems,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 1, pp. 298–308, 2011.
[3] C. Brunsdon, “Path estimation from GPS tracks,” in Proceedings of the 9th International Conference on GeoComputation, National Centre for Geocomputation, National University of Ireland, Maynooth, Ireland, 2007.
[4] A. K. Akasapu, P. S. Rao, L. K. Sharma, and S. K. Satpathy, “Density based k-nearest neighbors clustering algorithm for trajectory data,” International Journal on Advanced Science and Technology, vol. 31, pp. 47–57, 2011.
[5] D. Im, H. Yoon, and J. Lee, “Development of the walking trail applications on GPS-based smartphone utilizing the local narrative,” International Journal of Multimedia and Ubiquitous Engineering, vol. 8, no. 3, 2013.
[6] S. Hwang and D. Yu, “GPS localization improvement of smartphones using built-in sensors,” International Journal of Smart Home, vol. 6, no. 3, 2012.
[7] M. A. Qadeer, A. Chandra, and S. Jain, “Design and implementation of location awareness and sharing system using GPS and 3G/GPRS,” International Journal of Multimedia and Ubiquitous Engineering, vol. 7, no. 4, 2012.
[8] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W. Ma, “Understanding transportation modes based on GPS data for web applications,” *ACM Transactions on the Web*, vol. 4, no. 1, 2010.

[9] S. Cafiso, A. Di Graziano, G. Di Silvestro, G. La Cava, and B. Persaud, “Development of comprehensive accident models for two-lane rural highways using exposure, geometry, consistency and context variables,” *Accident Analysis and Prevention*, vol. 42, no. 4, pp. 1072–1079, 2010.

[10] J. Broach, J. Dill, and J. Gliebe, “Where do cyclists ride? A route choice model developed with revealed preference GPS data,” *Transportation Research Part A: Policy and Practice*, vol. 46, no. 10, pp. 1730–1740, 2012.

[11] N. Yiannakoulias, S. A. Bennet, and D. M. Scott, “Mapping commuter cycling risk in urban areas,” *Accident Analysis and Prevention*, vol. 45, pp. 164–172, 2012.

[12] S. P. Greaves and A. B. Ellison, “Personality, risk aversion and speeding: an empirical investigation,” *Accident Analysis and Prevention*, vol. 43, no. 5, pp. 1828–1836, 2011.

[13] J. Jun, R. Guensler, and J. Ogle, “Differences in observed speed patterns between crash-involved and crash-not-involved drivers: application of in-vehicle monitoring technology,” *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 569–578, 2011.

[14] T. Hastie and W. Stuetzle, “Principal curves,” *Journal of the American Statistical Association*, vol. 84, no. 406, pp. 502–516, 1989.

[15] T. K. Anderson, “Kernel density estimation and K-means clustering to profile road accident hotspots,” *Accident Analysis and Prevention*, vol. 41, no. 3, pp. 359–364, 2009.

[16] Z. Xie and J. Yan, “Kernel Density Estimation of traffic accidents in a network space,” *Computers, Environment and Urban Systems*, vol. 32, no. 5, pp. 396–406, 2008.

[17] R. Scheepens, N. Willems, H. Van De Wetering, G. Andrienko, N. Andrienko, and J. J. Van Wijk, “Composite density maps for multivariate trajectories,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2518–2527, 2011.

[18] O. D. Lampe and H. Hauser, “Interactive visualization of streaming data with Kernel Density Estimation,” in *Proceedings of the 4th IEEE Pacific Visualization Symposium 2011 (PacificVis '11)*, pp. 171–178, March 2011.

[19] H. Alt and M. Godau, “Computing the frederch distance between two polygonal curves,” *International Journal of Computational Geometry & Applications*, vol. 5, no. 1-2, pp. 75–91, 1995.

[20] D. Chen, A. Driemel, L. J. Guibas, and A. Nguyen, “Approximate map matching with respect to the Fréchet distance,” in *Proceedings of the 13th Workshop on Algorithm Engineering and Experiments*, San Francisco, Calif, USA, 2011.

[21] Seoul Metropolitan Government, Han River Project Headquarters Website, Korean only, http://hangang.seoul.go.kr.
