Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops

Long Tien Truong and Sekhar V. C. Somenahalli
University of South Australia

Abstract

This paper presents a GIS approach based on spatial autocorrelation analysis of pedestrian-vehicle crash data for identification and ranking of unsafe bus stops. Instead of crash counts, severity indices are used for analysis and ranking. Moran's I statistic is employed to examine spatial patterns of pedestrian-vehicle crash data. Getis-Ord Gi✳ statistic is used to identify the clustering of low and high index values and to generate a pedestrian-vehicle crash hot spots map. As recent studies have shown strong correlations between pedestrian-vehicle crashes and transit access, especially bus stops, bus stops in pedestrian-vehicle crash hot spots are then selected and ranked based on the severity of pedestrian-vehicle crashes in their vicinities. The proposed approach is evaluated using 13 years (1996–2008) of pedestrian-vehicle crash data for the Adelaide metropolitan area. Results show that the approach is efficient and reliable in identifying pedestrian-vehicle crash hot spots and ranking unsafe bus stops.

Introduction

Identifying pedestrian-vehicle crash hot spots, referred to as high pedestrian-vehicle crash locations, is important for understanding the causes of pedestrian-vehicle crashes and to determine effective countermeasures based on the analysis of the

causal factors. Recent studies (Clifton and Kreamer-Fults 2007; Hess et al. 2004) have shown strong correlations between pedestrian-vehicle crashes and pedestrian generators, more specifically, bus stops. Indeed, unsafe movements, encouraged by lack of necessary facilities close to bus stops such as pedestrian signals and crosswalks, result in pedestrian-vehicle crashes impacted by buses or other vehicles travelling on streets (Pulugurtha and Vanapalli 2008).

Common methods to profile crash hot spots use crash concentration maps that are based on Kernel Density Estimation (KDE) with absolute number of crashes. The KDE method calculates the density of crashes in a neighborhood around those crashes. However, these methods raise two potential issues. First, concentration maps are subject to different search bandwidths, also known as neighborhood sizes. Second, absolute counts of crashes may not truly indicate safety problems, while crash types and exposure measures, such as pedestrian and vehicular volumes, are neglected. Crash rates per unit exposure or crash rates based on severity would be more useful. This paper uses hot spot analysis to examine the spatial patterns of pedestrian-vehicle crashes and determines if they are statistically clustered, dispersed, or random. Criteria for analysis are severity indices that consider both the number and the severity of crashes. A hot spot is represented by a pedestrian-vehicle crash location with a statistically high severity index surrounded by other pedestrian-vehicle crash locations with high severity indices as well. Such spatial analysis can be performed using Geographical Information Systems (GIS) software programs such as ESRI's ArcGIS.

The main objective of this paper is to develop a GIS approach based on analysis of spatial autocorrelation of pedestrian-vehicle crash data to profile pedestrian-vehicle crash hot spots and to identify and rank unsafe bus stops in pedestrian-vehicle crash hot spots areas. The proposed approach is evaluated using 13 years (1996-2008) of pedestrian-vehicle crash data for the Adelaide metropolitan area.

Literature Review

Road crash hot spot analysis has been widely examined in the academic press, and various types of methods for identifying unsafe locations have been developed. Simple methods for identifying unsafe locations, where the number of crashes or the crash rate per unit exposure exceeds a given threshold, are routine and straightforward (Taylor et al. 2000). Austroads (1988) describes another method that uses critical crash rates to determine whether the crash record of each location is significantly

greater than the system wide average. Other statistical models, such as the empirical Bayes method, involve developing a statistical model based on the reference population and comparing the expected number of crashes with the observed number (Elvik 2008; Li and Zhang 2008). In addition to crash rates, unsafe locations are ranked according to their severity. Geurts et al. (2004) use the values of 1, 3, and 5 as the weights for a light, serious, or fatal casualty of a crash. Similarly, ranking methods also are made of a severity index, which is computed based on weights of 3.0 for fatal crashes, 1.8 for serious injury, 1.3 for other injury, and 1.0 for property damage only crashes (RTA 1994). In addition to these individual ranking methods, other composite methods that consider more than one factor at a time are also used. For instance, Vasudevan et al. (2007) use the average of the ranks based on frequency, weighted factor, pedestrian exposure, and traffic volume for ranking pedestrian hazardous locations. However, these methods, along with other traditional methods, focus on road segments or specific locations and thus produce results that are partially dependent on the length of road segment (Thomas 1996) and might not be able to capture area wide crash hot spots (Anderson 2009).

Numerous methods for studying spatial patterns of crash data as point events have recently been developed. One of the most widely used is KDE. Many recent studies use planar KDE for hot spot analysis, such as the study of high pedestrian crash zones (Pulugurtha et al. 2007), road crash hot spots (Anderson 2009), and highway crash hot spots (Erdogan et al. 2008). The goal of planar KDE is to develop a continuous surface of density estimates of discrete events such as road crashes by summing the number of events within a search bandwidth. However, planar KDE has been challenged in relation to the fact that road crashes usually happen on the roads and inside road networks that are portions of 2-D space. Road crashes are, therefore, needed to be considered in a network space, a simplification of the road network represented by 1-D lines. Several studies have extended the KDE to network spaces, which estimates the density over a distance unit instead of an area unit (Xie and Yan 2008; Yamada and Thill 2004). Neither planar or network KDE can be tested for statistical significance; this is a major weakness of these methods (Anderson 2009; Xie and Yan 2008).

Spatial patterns of crash data can also be analysed by spatial autocorrelation, statistics that take into account simultaneously discrete events' locations and their values. Moran's I Index and Geary's C Ratio are two popular indices for measuring spatial autocorrelation. Both of these methods combine two measures of attribute similarity and location proximity into a single index. The main difference between them is either

the similarity of attribute values of two points is computed from direct comparison in Geary's C Ratio or with reference to the mean value in Moran's I Index. Therefore, the two indices have different ranges and different statistical properties. Possible values of Moran's I range from -1 to 1. A positive value indicates clustering and a negative value indicates dispersion. Possible values of Geary's C range from 0 to 2. A Geary's C value near 0 indicates clustering, while a Geary's C value near 2 indicates dispersion. In contrast to the KDE, the statistical significance for both Moran's I and Geary's C can be calculated using z-score methods (Erdogan 2009; Wong and Lee 2005). In fact, Moran's I Index and Geary's C Ratio are global statistics since they are measures of the entire study area. To investigate the spatial variation and the spatial associations, it is necessary to rely on local measures such as the local Moran's I (Anselin 1995) and Getis-Ord Gi✻ statistics (Getis and Ord 1992). Particularly, Getis-Ord Gi✻ statistics are useful to identify cold/hot spots where their values are significantly low or high and be surrounded by other low or high values as well. Both Moran's I and Getis-Ord Gi✻ are available in commercial GIS software packages, such as ESRI's ArcGIS.

Study Area and Data

The required data for the study includes pedestrian-vehicle crash data, bus stop data, and street network data in GIS layers. Since bus ridership data are unavailable, bus stop data include location and bus route information only. This paper used 13 years (1996-2008) of pedestrian-vehicle crash data in Adelaide, the capital city of South Australia. However, due to the lack of information to determine pedestrian-vehicle crashes that involved bus users, this research used all pedestrian-vehicle crash data for identifying unsafe bus stops. This could be biased since some pedestrian-vehicle crashes might have happened as a result of other reasons. Another limitation of the data is the lack of exposure data such as annual average daily traffic (AADT) and alighting and boarding data.

Methodology

The methodology can be subdivided into the following steps:

1. Load pedestrian-vehicle crash data and aggregate the data at each location.

2. Compute a severity index at each location.

3. Examine the spatial patterns of the pedestrian-vehicle crash data.

4. Create a pedestrian-vehicle crash hot spots map.

5. Load the bus stop data and identify bus stops in hot spots areas.

6. Compute severity indices and rank unsafe bus stops.

ArcGIS software with the Spatial Statistic toolbox is used for mapping and spatial analyses.

1) Load pedestrian-vehicle crash data and aggregate the data at each location.

The street network theme and pedestrian-vehicle crash data are loaded in the ArcGIS software program (Figure 1). Each of the data points may represent a single pedestrian-vehicle crash or multiple crashes since a number of pedestrian-vehicle crashes may be reported at the same location such as an intersection. Therefore, the pedestrian-vehicle crash data are aggregated to produce counts for each type of crashes at each location.

2) Compute a severity index at each location.

Without weighted data, it is difficult to know whether high or low clustering exists. Counts of crashes are commonly used to evaluate safety problems at a location. There is a belief that the more severe crashes should have greater weights in identifying unsafe locations on the basis of crash costs. Additionally, results of the crash severity method are sensitive to various weighting systems (Geurts et al. 2004). Although there is no consensus on how the optimum weighting system should be developed, a compromise approach is becoming increasingly popular. The approach is to give weights to the more severe crashes, but not with the extreme high values computed in direct proportion to the crash costs. This research employs a crash severity weighting system in which the basic factor is tow-away crashes (RTA 1994). The severity index is computed by the following equation:

$$SI = 3.0 \times X_1 + 1.8 \times X_2 + 1.3 \times X_3 + X_4 \qquad (1)$$

Where:

$X1$= total number of fatal crashes

$X2$= total number of serious injury crashes

$X3$= total number of other injury crashes

$X4$= total number of property-damage-only crashes

This severity index is used as the criterion for spatial analysis in this research. Figure 2 shows the distribution of the severity indices. Visual observation of this figure indicates some high severity index zones. However, it needs a statistical significance test to examine the degree of clustering.
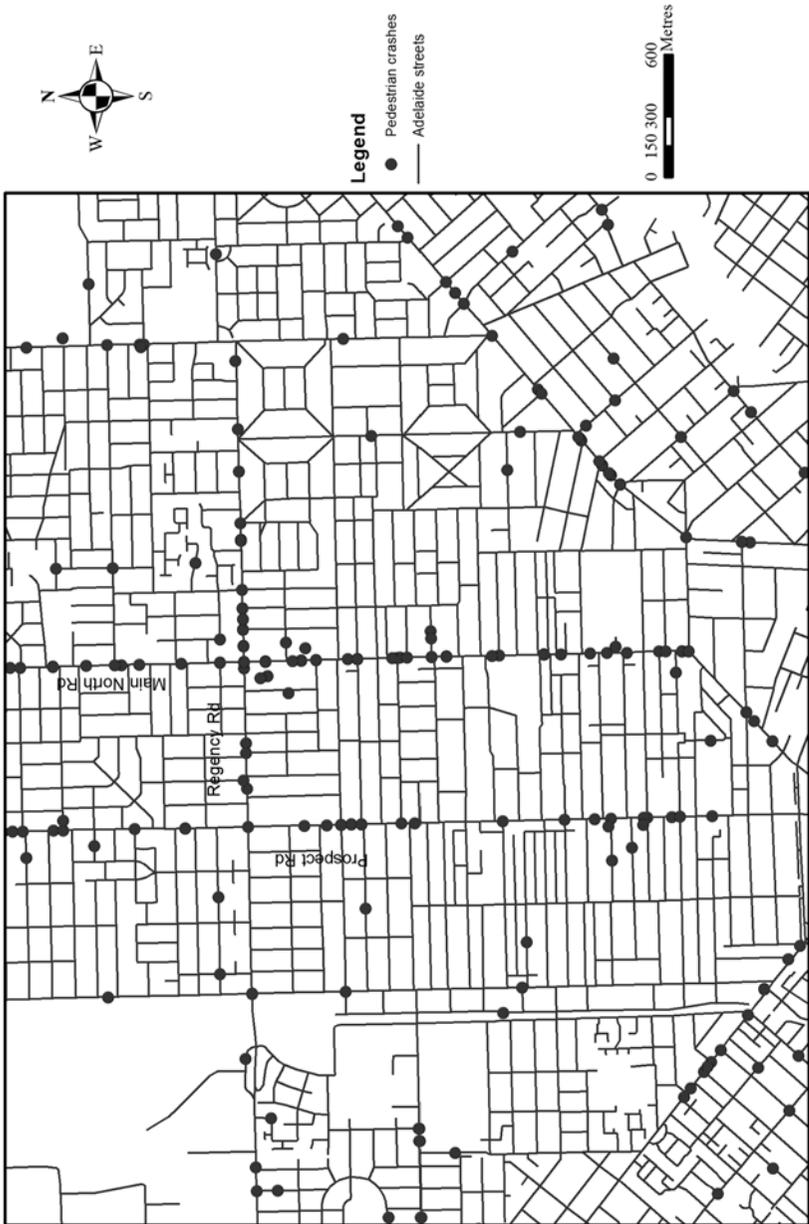
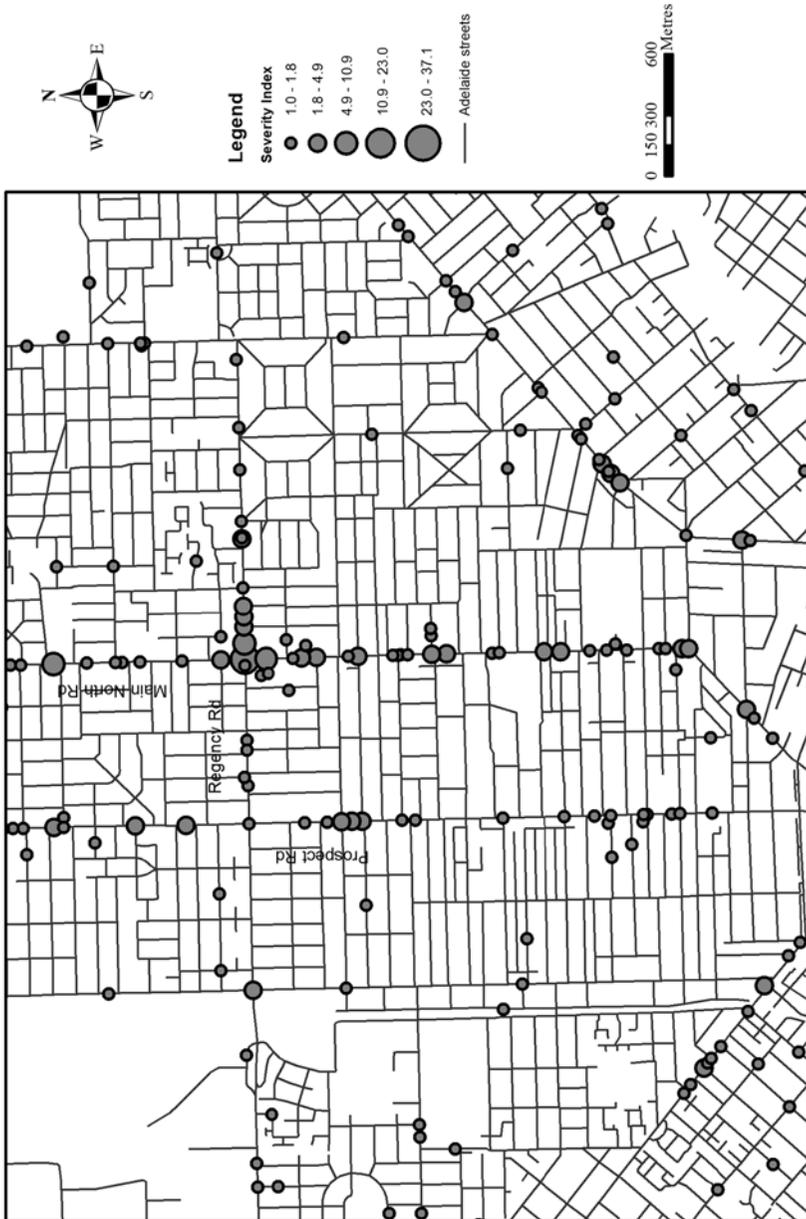**Figure 1. Locations of Pedestrian-Vehicle Crashes**

Figure 2. Distribution of Severity Indices

3) Examine the spatial patterns of the pedestrian-vehicle crash data.

To examine spatial patterns, this research makes use of the Moran's I Index to measure spatial autocorrelation. Moran's I combines the measure for attribute similarity and the measure of location proximity into an index. The location proximity weight between two points often is defined as the inverse of the distance between them. The attribute similarity of severity indices of two points is defined as the difference between each value and the global mean value (Wong and Lee 2005). Therefore, the index can be calculated using the following equation:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\left(\sum\left(x_i - \bar{x}\right)^2\right)} \tag{2}$$

Where:

w_{ij} = the proximity weight of location i and location j with $w_{ii}=0$

x_i = the severity index at location j

$\bar{x}$ = the global mean value

n = the total number of pedestrian-vehicle crash locations

The statistical significance for Moran's I can be calculated using z-score methods. Based on the expected values (E[I]) for a random pattern and the variances (VAR[I]), the standardized Z-score can be mathematically represented as follows:

$$Z = \frac{I - E(I)}{\sqrt{VAR(I)}} \tag{3}$$

In this research, the Spatial Autocorrelation tool was used to compute Moran's I statistics and z-scores. Since each data point is analysed in terms of its neighboring data points defined by a distance threshold, it is necessary to find an appropriate distance threshold where spatial autocorrelation is maximized. The Spatial Autocorrelation tool was run multiple times with different distance thresholds to find the distance with the maximum z-score. Table 1 shows that with a distance threshold of 1,000 metres, the z-score reaches the highest value of 136.2, which means the pedestrian-vehicle crash data is clustered until a distance threshold of 1,000 metres with a statistical significance level of 0.01.

Table 1. Spatial Autocorrelation by Distance Thresholds

Distance (m)	200	600	800	1,000	1,200	1,400	2,000
Z score	55.62	120.1	132.7	136.2	134.3	130.5	108.9
Sig.	0.01	0.01	0.01	0.01	0.01	0.01	0.01

4) Create a pedestrian-vehicle crash hot spots map.

In this step, Getis-Ord G_i^* statistic is used to identify pedestrian-vehicle crash hot spots. A high value of Getis-Ord G_i^* statistic represents a cluster of high index values (hot spots), while a low value represents a cluster of low index values (cold spots). Getis-Ord G_i^* statistic and its z-score are mathematically expressed by the following equations.

$$G_i^*(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j} \qquad (4)$$

$$Z(G_i^*) = \frac{G_i^* - E(G_i^*)}{\sqrt{VAR(G_i^*)}} \qquad (5)$$

Where:

w_{ij}= the weight for the target neighbour pair

d = distance threshold

x_j = the severity index at location j

The distance threshold of 1,000 metres associated with maximum z-score in the previous step is chosen for the Getis-Ord G_i^* analysis using the Hot Spot Analysis and Rendering tool. The result of the analysis is shown in Figure 3. It indicates that there are hot spots with z score greater than 1.96 at a statistical significance level of 0.05.

5) Load bus stop data and identify bus stops in hot spots areas.

This step makes use of the pedestrian-vehicle crash hot spots map to identify bus stops in the vicinities of the pedestrian-vehicle crash hot spot. In general, a person is willing to walk about 0.25 miles (400 metres) to reach a bus stop (O'Sullivan and Morrall 1996). Bus stop data are overlaid on the hot spots map to identify bus stops that are within 400 metres of each pedestrian-vehicle crash hot spot. This can be achieved using the Network Analyst tool to generate a 400-metre network buffer of the hot spots, which represents areas influenced by hot spots (Figure 4).

Figure 3. Pedestrian-Vehicle Hot Spots Map

Figure 4. Unsafe Bus Stops in Hot Spots Areas

6) Compute the severity indices and rank unsafe bus stops.

In this step, unsafe bus stops, identified in the previous step, are ranked based on the severity of pedestrian-vehicle crashes in their vicinities. The goal of this step is to compute an appropriate severity index for each potentially unsafe bus stop. It can be argued that pedestrian-vehicle crashes that occurred closer to a bus stop more likely could be related to that bus stop than others. Therefore, they should be given greater weights in computing severity index for that bus stop. Another issue is to select a buffer size so that only pedestrian-vehicle crashes related to unsafe bus stops are considered for calculations. A small buffer size of 100 feet (30 metres) is recommended in the study of hazardous bus stops (Pulugurtha and Vanapalli 2008). Arguably, a small buffer size might miss some crashes associated with bus users, while a greater buffer size might capture crashes that do not involve bus users. In the final choice, all pedestrian-vehicle crashes within a 100-metre network distance to an identified bus stop were used for computing the severity index for that bus stop. A weight of 1.5 was given to those within 50 metres as they are more likely to be related to the bus stop. The bus stop severity index is defined as follows:

$$SI_{bus\,stop} = 1.5 \times SI_{50} + SI_{50-100} \tag{6}$$

Where:

SI_{50} and SI_{50-100} are computed using Equation (1)

SI_{50} = severity index of all pedestrian-vehicle crashes within a 50-metre network buffer of the bus stop

SI_{50-100} = severity index of all pedestrian-vehicle crashes within a 100-metre network buffer, but outside a 50-metre network buffer of the bus stop

Finally, unsafe bus stops are ranked based on their severity indices.

Results and Discussion

The Moran's I analysis indicated a statistically significant clustering pattern of pedestrian-vehicle crash data for the Adelaide metropolitan area (Table 1). The Getis-Ord G_i^* analysis also detected pedestrian-vehicle crash hot spots with the 0.05 level of significance. The majority of hot spots were located near the intersections of Main North Road, Prospect Road, and Regency Road in Northern Adelaide (Figure 3). In this area, there were 3 pedestrian-vehicle crash hot spots at intersections and 10 hot spots at mid-block locations. This indicated pedestrian-vehicle crashes

at mid-block locations were more severe. To identify bus stops that are within 400 metres of a hot spot, influenced areas of the hot spots were generated using the Network Analyst tool (Figure 4). A total of 31 bus stops in the influenced areas were identified and ranked. Table 2 shows the top 10 unsafe bus stops with at least 3 pedestrian-vehicle crashes. It can be seen that some bus stops have extremely high severity indices, such as stop number 117 Regency Road and stop number 17 Main North Road. Site observations shows that pedestrian facilities such as pedestrian signals are available only at the intersection of Regency Road and Main North and the intersection of Regency Road and Prospect Road. In addition, many T-junctions with traffic from and to local streets in this area will generate considerable turning movements. In general, bus users need to cross streets at least once for a return trip. Therefore, it can be argued that the lack of pedestrian facilities, high turning movements, and pedestrian failure to yield are main reasons for pedestrian-vehicle crashes at mid-block locations in this area.

There are ways in which this study could be improved. It would be more reliable to apply a multi-criteria approach to rank unsafe bus stops depending on the

Table 2. Ranking Unsafe Bus Stops

Rank	Stop	Direction	Street	Within 50m				50-100m				S.I
				P.D	Fat	Seri	Oth	P.D	Fat	Seri	Oth	
1	117	Anti	Regency				4			4	7	24.10
2	17	Down	Main North				8				2	18.20
3	17	Up	Main North				8				2	18.20
4	116	Anti	Regency							4	6	15.00
5	14	Up	Main North			1	1			2	3	12.15
6	117	Clock	Regency				5	1			1	12.05
7	16	Up	Main North		1		1				4	11.65
8	15	Down	Prospect			1	1			1	2	9.05
9	15	Up	Prospect			1	1			1	2	9.05
10	14	Down	Main North			1	1				1	5.95

P.D. =Property damage only crashes; Fat = Fatal crashes; Seri = Serious injury crashes; Oth = Other injury crashes; S.I = Severity Index.

availability of data. For instance, bus stops are ranked based on an average rank of multiple criteria such as severity index and crash rates per million kilometres and per number of bus users boarding and alighting the bus. In addition, while an optimum search distance is statistically identified in hot spot analysis, sizes of buffers are issues in ranking unsafe bus stops. Different sizes may produce different severity indices and, therefore, different rankings. Sensitivity analysis should be conducted to find an appropriate buffer size.

Conclusions

This research presents a GIS approach based on the spatial autocorrelation of pedestrian-vehicle crash data for identification and ranking of unsafe bus stops. The merits of spatial autocorrelation statistics and GIS concepts and technology enable statistical evaluations of spatial patterns of pedestrian-vehicle crash data. Instead of absolute number of crashes, severity indices, considering both the number and the severity of crashes, are used for analysis and ranking.

The results from analyzing pedestrian-vehicle crash data in Adelaide indicate that the proposed approach can statistically detect spatial patterns of crash data and reasonably identify and rank unsafe bus stops in pedestrian-vehicle crash hot spot areas. The identification of pedestrian-vehicle crash hot spots is reliable and accurate because it is conducted using well-designed spatial statistics that consider both the locations of point events and their attributes. Spatial analysis based on raw data locations can provide more information to better capture safety indications. For instance, hot spot analysis shows that pedestrian-vehicle crashes at mid-block locations are more severe than at intersections. The issues of pedestrian-vehicle crashes at mid-block locations and bus stop safety are, in fact, raised by the lack of appropriate pedestrian facilities and high turning movements at the sites. The selection of buffer sizes used in ranking unsafe bus stops is an issue since it cannot be tested for statistical significance and the crash data do not indicate whether a crash is associated with bus users. Overall, this research offers a sound basis for identifying pedestrian-vehicle crash hot spots and prioritizing unsafe bus stops in order to study the causal factors, determine effective countermeasures, and plan a safer bus transit system.

Further research should include a multi-criteria ranking method and sensitivity analysis of buffer sizes. Such improvements could improve the accuracy and reliability of the prioritization of unsafe bus stops.

References

Anderson, T. K. 2009. Kernel density estimation and K-means clustering to profile road accident hot spots. *Accident Analysis & Prevention* 41(3): 359-364.

Anselin, L. 1995. Local indicators of spatial association-LISA. *Geographical Analysis* 27(2): 93-116.

Austroads. 1988. *Guide to Traffic Engineering Practice: Part 4, Road Crashes*. Sydney: Austroads.

Clifton, K. J., and K. Kreamer-Fults. 2007. An examination of the environmental attributes associated with pedestrian-vehicular crashes near public schools. *Accident Analysis & Prevention* 39(4): 708-715.

Elvik, R. 2008. A survey of operational definitions of hazardous road locations in some European countries. *Accident Analysis & Prevention* 40(6): 1830-1835.

Erdogan, S. 2009. Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of Safety Research* 40(5): 341-351.

Erdogan, S., I. Yilmaz, T. Baybura, and M. Gullu. 2008. Geographical information systems aided traffic accident analysis system case study: City of Afyonkarahisar. *Accident Analysis & Prevention* 40(1): 174-181.

Getis, A., and J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24(3): 189-207.

Geurts, K., G. Wets, T. Brijs, and K. Vanhoof. 2004. Identification and ranking of black spots: Sensitivity analysis. *Transportation Research Record* 1897: 34-42.

Hess, P. M., A. V. Moudon, and J. M. Matlick. 2004. Pedestrian safety and transit corridors. *Journal of Public Transportation* 7(2): 73-93.

Li, L., and Y. Zhang. 2008. Bayesian approach based on geographic information systems to identify hazardous roadway segments for traffic crashes. *Transportation Research Record* 2024: 63-72.

O'Sullivan, S., and J. Morrall. 1996. Walking distances to and from light-rail transit stations. *Transportation Research Record* 1538: 19-26.

Pulugurtha, S. S., V. K. Krishnakumar, and S. S. Nambisan. 2007. New methods to identify and rank high pedestrian crash zones: An illustration. *Accident Analysis & Prevention* 39(4): 800-811.

Pulugurtha, S. S., and V. K. Vanapalli. 2008. Hazardous bus stops identification: An illustration using GIS. *Journal of Public Transportation* 11(2): 65-83.

RTA. 1994. *Road TrafficAaccidents in NSW - 1993*. Sydney: Roads and Traffic Authority of NSW.

Taylor, M. A. P., P. W. Bonsall, and W. Young. 2000. *Understanding Traffic Systems: Data, Analysis and Presentation*. 2nd ed. Aldershot: Ashgate.

Thomas, I. 1996. Spatial data aggregation: Exploratory analysis of road accidents. *Accident Analysis & Prevention* 28 (2):251-264.

Vasudevan, V., S. Pulugurtha, and S. Nambisan. 2007. Methods to prioritize pedestrian high-crash locations and statistical analysis of their relationships. *Transportation Research Record* 2002: 39-54.

Wong, D. W. S., and J. Lee. 2005. *Statistical Analysis of Geographic Information with ArcView GIS and ArcGIS*. Hoboken: Wiley.

Xie, Z., and J. Yan. 2008. Kernel Density Estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems* 32(5): 396-406.

Yamada, I., and J. C. Thill. 2004. Comparison of planar and network K-functions in traffic accident analysis. *Journal of Transport Geography* 12(2): 149-158.

About the Authors

LONG TIEN TRUONG (*Long.truong@postgrads.unisa.edu.au*) is a graduate student in the Transport System Engineering in the School of Natural and Built Environments (NBE), University of South Australia, Adelaide.

DR. SEKHAR V.C. SOMENAHALLI (*Sekhar.somenahalli@unisa.edu.au*) currently teaches courses related to Geographic Information Systems (GIS) and their applications to Transport and Urban Planning disciplines in the School of Natural and Built Environments (NBE), University of South Australia, Adelaide. He has over 5 years of consulting experience and over 16 years of teaching and research experience and has worked in India, Malaysia, and Australia.