GIS for Road Accident Analysis in Hong Kong

Poh Chin Lai\(^1\) and Wing Yee Chan\(^2\)

\(^1\)Department of Geography, The University of Hong Kong, Pokfulam Road, Hong Kong, China
\(^2\)Ferry and Paratransit Division, Transport Department, Hong Kong Special Administrative Region Government, China

Abstract

Road accident analysis is not new. It has come to attention as traffic fatalities had an explosive increase in the 50's and 60's, especially in the industrialized world and during the intense period of economic growth. This paper examines the potential use of the Geographic Information Systems (GIS) technology in facilitating road accident analysis in Hong Kong. Various point-pattern techniques are illustrated. While GIS can extend the visualization and analytical capabilities, its implementation in Hong Kong would hinge upon more comprehensive planning in data acquisition and integration. A master digital database on road systems of Hong Kong with a standardized structure and coherent naming conventions in both Chinese and English is fundamental to further GIS development in the area of road accident analysis.

I. INTRODUCTION

It is startling to know that at least half a million people in the world die each year as a result of traffic accidents and about 15 million are injured (Hutchinson 1997). Road accidents do not happen without causes. Their causes can be traced and preventive measures can be devised. Trinca et al. (1988) proposed five different categories of strategies for reducing road accidents: exposure control, crash prevention, behavior modification, injury control, and post-injury management. Among these five strategies, road accident analysis is undertaken for a better understanding of the why's and how's (i.e. causes and casualties) of road-related accidents. Road accident analysis helps improve the safety of our roads and living environment with an aim to minimize the number of accidents and the human or financial losses.

Many countries utilize Geographic Information Systems (GIS) to manage existing transportation resources and enhance the effectiveness and efficiency of transportation operations (United States National Research Council 1992). Most researchers found that GIS brings efficiency in data handling and effectiveness in data analysis (Duranska 1998; Clark et al. 1998; Filian and Higelin 1995). The development of GIS applications on road accident analysis also has become a trend (Lupton and Bolsdon 1999). A GIS enables a new method in organizing geographical data through its database management system that facilitates the storage, organization and retrieval of digital records (Heywood et al. 1998).

Road accident analysis can benefit from the data management, cartographic and analytical functions offered by a GIS. The standardization of data allows accident data to be exchanged within a system or even among many different GISs. Indeed, the flexibility of data exchange has meant time saving and convenience (Lubkin et al. 1990). Road accident data are used not only as evidences for the court, but also as a tool for risk management. They are a key element in road accident analysis. The ability to manage these data is critical to the success of the analysis. However, such data have not been fully utilized in the past. The use of tabular data has been restrictive, and the production and use of graphical or mapped presentation have not been fully explored. This paper focuses on point pattern methods in presenting and visualizing accident data as a step forward in facilitating an understanding of accident patterns. More sophisticated analytical methods are available in GIS but their utility will depend on the successful integration of accident data into a visualization system.

II. THE SITUATION OF HONG KONG

Road accidents caused by vehicles have emerged as a vital transport issue of Hong Kong in recent decades. Road accidents have had a great increase since 1953 when there were 3,600 reported accidents causing injury (Road Safety and Standards Division 1983). This figure represented merely a quarter of about 15 thousand cases in 1997 (Road Safety and Standards Division 1998). The number of accidents in 1992 reflected nearly a six-fold increase from that in 1953. Even though accident occurrences have lessened in this decade, the figures have remained significantly high.

A great emphasis has been put on understanding road accidents not only because of the increasing numbers but also because of the severity of injuries. A ‘fatal’ accident is defined as that with one or more persons die within 30 days of the accident (Hong Kong Police Force 1996). About 7 percent (229 out of 14,776) of reported traffic accidents causing injury in 1997 were fatal cases, which computed to an equivalent of
37 fatalities per million persons in Hong Kong (Road Safety and Standards Division 1998). Between 1995 and 1997, Hong Kong ranked 14 out of 22 major cities in the world in terms of traffic fatality rate. Nonetheless, this statistic should not be taken lightly as it computed to at least one death from road accidents every one and a half days in 1997.

The occurrence of road accident has been shown to relate directly to the size of the population and the number of vehicles on the roads. The population of Hong Kong grew from 2.2 million in 1953 to 6.5 million in 1997 (Hong Kong Census and Statistics Department 1969; Road Safety and Standards Division 1998). Registered vehicles also increased from 18 thousand to 558 thousand during the same period (Traffic and Transport Survey Division 1998). Even though road lengths also increased from 948 kilometers in 1966 to 1,831 kilometers in 1997, the rate of growth was not as substantial as those of the population and the vehicles. The number of motor vehicles per kilometer of road in 1997 represented a 46.2 percent increase since 1978. It is not difficult to see why Hong Kong is getting more congested on the roads given the present trend.

Hong Kong relies heavily on her land transportation. Due to limited space, the government has already introduced other means of transport, such as the Mass Transit Railway (MTR) and the Kowloon-Canton Railway (KCR), to lessen the burden on roads. Citizens are encouraged to use public transports instead of private vehicles. Nonetheless, the number of private car licensed still doubled to over 303 thousand between 1978 and 1997 (Road Safety and Standards Division 1988 and 1998). The fact remains that most people in Hong Kong still rely on road transportation, particularly buses and minivans. This situation was reflected in the distribution of annual passenger trips in Figure 1. The average daily passenger trips between MTR and KCR (including East Rail and Light Rail) only occupied about 30 percent of total journeys in 1997 (Hong Kong Transport Department 1998). It is clear that the railway lines alone are not capable of providing sufficient transport services to Hong Kong. It also appeared that the risk of road accidents has persisted because passengers could not be effectively diverted from roads.

III. ROAD ACCIDENT DATA OF HONG KONG

There are standard procedures and established techniques for road accident analysis in Hong Kong (Hong Kong Transport Department, 1999). In general, the Hong Kong Police Force is responsible for collecting and recording road accident data while the Transport Department reviews accident cases and implements remedial measures for reducing accident cases in the future. Accident cases are considered in terms of their severity (how many people were injured and whether there were casualties) and their locations (at which major intersections or sections of roads did they occur). Road safety

![Figure 1. Annual passenger trips by various means of public transport, 1983-1997](image-url)
can be reconsidered after examining the causes and factors whereby remedial planning, design, and placement of road structures can then be formulated.

Road accidents should be reported to the police when it happened. The law in Hong Kong enforces accidents involving personal injury to be reported but there is no legal requirement to report damage only accidents. Road accident analysis therefore focuses only on cases involving personal injury. First-hand accident records are available within the jurisdiction of the Hong Kong Police Force based on minutes prepared by an Investigation Officer and his superior who would note down how the accident had happened. Accident data collected by the police can be classed into 3 types: (i) information concerning the general circumstances of the accident, (ii) details of the vehicle and the driver, and (iii) details of the casualties. They are recorded using the standardized accident report form and with data transferred from the accident file. However, only general information (including time, venue, weather conditions and schematic of the accident) can be obtained for this research in compliance with the data privacy ordinance. Personal data about the driver, the passengers, and the injured are withheld from public access.

The road accident data of Hong Kong have been computerized recently to facilitate further analysis. The raw data are sent by the Traffic Branch Headquarters of the Hong Kong Police Force to the Information Technology Services Department (ITSD) for processing. ITSD is responsible for inputting data, updating records, plotting accidents distribution map and compiling statistical tables. Statistical summaries are compiled according to the general norms in classifying accident types, happening sites and time, contributory factors and other relevant data. They are used in further analysis to identify common situations and general factors. Corrective measures are then introduced to improve road situations.

IV. CURRENT PRACTICE OF ROAD ACCIDENT ANALYSIS

The procedures for road accident analysis are similar amongst many countries in a broad sense. However, differences in the practice do exist due to dissimilarity in the availability of resources and the management structure. A computerized system of Traffic Accident Statistics in Hong Kong began in 1976 (Road Safety and Standards Division 1983). All data were converted from written accident reports into the computerized system. The system was used for data storage and to conduct accident analysis, such as accident investigation and prediction of the accident trend. The Transport Department (1999, section 2.5.1) summed up the general computer processing procedure to include “(a) data validation, checking for errors, updating of records, (b) plotting accident locations, and (c) tabulation of statistical tables and accident file lists.” The implementation of this computerized system is not without its problems.

Data consistency

The Hong Kong Transport Department (1999) indicated that data consistency has remained a key problem in the collection of road accident data. Not only is data consistency a problem in Hong Kong but it is also a major concern of many institutes in other countries. The problem roots in the judgment of the reporting officer at the accident scene and in the use of non-standardized reporting terms. Even though a standard accident reporting form is available in Hong Kong, much effort is needed to regulate input values for filling in the form. While a casual reporting scheme is convenient and can note down much detail about an accident, it also affects the utility value of the report.

Accident reports must be filed in a manner that facilitates cross-classification and retrieval. An objective judgment is also important in guaranteeing a reliable account of the incidence. Correct and justified contributory factors and an understanding of the environmental background of an accident are essential to the choice of appropriate counter measures. It is difficult in a computerized environment to recognize the same factors in varied qualitative terms or locations coded in different schemes. There was a suggestion to employ trained accident investigators for collecting data in much greater detail; however, the cost would be relatively high (Mak et al. 1998). Lubkin et al. (1990) also suggested the use of an Accident Report Entry System in standardizing accident data input. Some operator training was also needed if data entry was to conform to specific rules and standards.

Accuracy in placing accident data

It is important to have an accident plotted in its precise location to get a true picture of the distribution of various accident occurrences and to note the locations for corrective measures. The dependability of the computerized method in plotting an accident spot on a map is of concern. A coherent location coding system (or street naming system) will allow the computer to verify the street location accurately from the collected accident data, thus limiting errors during the matching process.

The proper matching of accident records has been a popular research topic in the United States. Levine and Kim (1998) suggested the use of GIS geocoding technique to encode locations of road accidents in Honolulu. They proposed a number of ways to improve geocoding, address matching, and spatial accuracy: (1) standardization should occur prior to data entry; (2) quality control procedures (including training of coders) need to be adopted to ensure consistency; and (3) data integration is essential to improve comprehensiveness and accuracy. Levine and Kim (1998) also suggested using the Global Positioning System (GPS) for address matching apart from the above recommendations.

The development of a more efficient geocoding system or
data entry program is recommended for Hong Kong (Hong Kong Transport Department, 1999). Limited by the high implementing cost for training, database management, and quality control and current weaknesses in the data construct, the above suggestions are not easily attainable in Hong Kong at this moment. Efforts are underway at various government departments to agree on a standardized list of street addresses with coherent naming conventions in both Chinese and English. The master list is essential in all reporting systems to facilitate address entry and enhance the accuracy of address matching.

V. UNDERSTANDING ACCIDENT PATTERNS

Lupton and Bolsdon (1999) suggested an object-based approach to reorganise existing database of accident sites in Britain. Lubkin et al. (1990) stated that “in addition to improving public safety, a good data analysis can facilitate effective engineering and law enforcement counter measures, and thus is a prime defense against tort liability. Local groups must learn how to use the information buried in accident data.” They also developed an Accident Report Entry System for tidying up input data to increase their utility and applicability.

Statistical charts or histograms and collision or stick diagrams have been shown to enhance the understanding of an accident situation (Hong Kong Transport Department 1999). Maps would be more revealing! Mapping is not simply a creative process of designing, compiling, and producing graphical displays (Dent 1990). It is expected that some interesting relationships or hidden associations between contributory factors and accident occurrences may emerge by overlaying two maps at a time. An examination of distributional patterns on the maps may trigger further analysis or systematic survey on selected sites that would have been neglected otherwise.

The cartographic functions of a GIS present new and exciting opportunities for users of spatial data. The presentation of graphical/mapped data has undergone great improvement with advances in the computing technology. Unlike conventional mapping, data stored in a GIS can be presented in many ways to offer new insights and to examine the relationships between different entities (Bernhardt 1992).

Cartographic analysis

The application of cartographic analysis and GIS was evident among local governments in the United States in the 80s and 90s (Clark et al. 1998; Kollin et al. 1998; Filián and Higelin 1995; Abbkowitz et al. 1990). Durcanska (1998) described various spatial analytical functions used in road safety improvements. He suggested a method to combine different maps layers for the identification of problematic areas. For example, high accident areas were examined along with the surrounding infrastructures, assessed in terms of the adequacy of sidewalks, streetlights, traffic signs, and markings. Simkowitz (1990) also demonstrated a similar example of the use of a GIS on pavement management. He proposed to use GIS as a framework for data integration because of its ability to relate data collected under various referencing systems, such as pavement condition surveys, skid resistance measurements, traffic counts, bridge inspections, and sign inventories.

Different mapping techniques must be considered in light of the data scaling and symbolization. The effectiveness of a map in uncovering hidden relationships or messages depends upon the integrative power of the designer.

Point pattern analysis

Road accident data are point source data of zero-dimension. Chou (1997) suggested “the distribution of point features can be described by frequency, density, geometric center, spatial dispersion, and spatial arrangement.” DeMers (1997) summed up the characteristics of spatial arrangement as “the placement, ordering, concentration, connectedness or dispersion of multiple objects within a confined geographic space.” Simply put, point patterns can be classified into regular, random or clustered patterns. It is by observing these distinctive distributional patterns that the hidden behavior of data is revealed. A clustered distribution is the concentration of points in one or a few segments of places. It helps to emphasize the problem or concentrated areas. A regular pattern shows an even distribution of points, which implies uniformity of the data. A random distribution means widespread problems without a pattern.

It is not difficult to find that a good description of accident data can be given by the spatial arrangement. Various statistics are available to measure divergence from randomness. The R-scale is based on the nearest neighbor analysis of points as compared with the Poisson probability function (Taylor 1977). Moran’s I and Geary’s C coefficients are measures of spatial autocorrelation which test for existence of systematic spatial variation in values across a map against the null hypothesis of zero spatial correlation (Sawada nd). The above measures can be used to evaluate if accidents are clustered, where they are clustered, to what scale they are clustered, and if they are closer to certain ground features (e.g. alcohol stores and night clubs) than expected.

Accident data can be of the nominal type in which the accident distribution map conveys the locations of different types of collisions - such as pedestrian and vehicle, vehicle and vehicle, or vehicle and built structure. Figure 2 employed different point symbols to represent road accidents by collision type. Here the map clearly revealed an exceptionally high occurrence of collision between vehicles and built structures within the locality. This finding would warrant a site investigation to identify specific structural causes or deficiencies leading to such collisions. The condition of the driver (whether sober or not) at the time (with or without daylight) of the accident could also be correlated against such occurrences.
Data of both interval and ratio scaling carry numerical values. The numerical values can be simplified into ordinal scaling for mapping. While the computer is capable of representing numerical values as they are, a map bearing unique symbols or symbols of varying sizes for each point location is overwhelmingly detailed and difficult to understand. Muehrcke and Muehrcke (1998) stated that simplification includes “eliminating unwanted features, smoothing the fine details from boundaries, abstracting the shape of features, aggregating features into clusters or regions, changing the dimensionality of features, reducing data measurement levels, and combining variables into ratios or indexes.” Simplification allows a map to be more comprehensible because of reduced clutter. Some may argue that simplification distorts the truth but the removal of “unwanted” details does seem to improve comprehensibility by highlighting the message intended for the readers.

Ordinal scaling helps classify objects into ranked orders. Objects are put into non-measurable categories of implied ranks. Some accident data, such as the level of severity of collisions, can be mapped to convey seriousness of the accidents and their distribution. Figure 3 used graduated circles to indicate collision severity at the scene whereby the larger circles meant more serious collisions. While the distribution of accident locations was scattered and distinctive, a few fatal accidents clearly stood out. This form of map representation is not only visually revealing but also easy to comprehend.

Proximity analysis

Proximity analysis is a feature manipulation process. Its operation concerns the inclusion and amalgamation of nearby features. The Buffer operation and Thiessen polygons are two common types of proximity analysis. A Thiessen polygon is useful in dividing a study area into smaller units to determine the spatial influence of point features. “Thiessen polygons are generated from a set of point features, and derived in such a way that each polygon represents the catchment area of a point, i.e., the area inside the polygon is closer to that point than to any other points” (Chou 1997). Figure 4 applies Thiessen polygons to examine the coverage of streetlights and how the facilities are spaced. A larger Thiessen polygon implies a wider spacing and possible deficiency in lighting. A more complete picture on the lighting situation will result had building structures included on the same map. Such a display is useful in assessing the need for better lighting where accidents have occurred.

Chou (1997) defined a buffer as “a polygon created through reclassification at a specified distance from a point, line, or area.” The buffering process is a method to identify surrounding area in all directions from a feature position. Buffers can be derived in four different ways. An arbitrary buffer is the creation of a buffer zone using an arbitrary distance in a series of trial and error exercises. A causative buffer requires apriori knowledge of a subject matter to arrive at a suitable distance; for example, a 500-metre buffer around a bus stop would indicate roughly a 15-minute walking distance from the bus stop. A measurable buffer employs specific distance measures for corresponding features; for instance, 20 meters from a tertiary road, 40 meters from a secondary road, and 60 meters from a primary road. Finally, a mandated buffer makes use of legalized distance as the measure; such as 1,000 feet for the designation of critical areas surrounding a wetland.

Road accidents are unlikely to occur at the exact same location. An accident black spot involves multiple occurrences and can be interpreted to mean accidents occurring within a stated proximity of a road junction. In this case, accident spots within the vicinity can be accumulated and aggregated.
in a GIS using the buffer operation. An accident black spot in Hong Kong is defined according to some criteria: (i) 6 or more pedestrian injury accidents per year; or (ii) 12 or more injury accidents of any description per year; or (iii) 9 or more injury accidents per year, of which over 50 percent, are fatal or serious (Hong Kong Transport Department, 1999, section 1.5.1). Figure 5 is a map that identifies accident black spots in accordance with the stated criteria. Symbol sizes are made to vary with the aggregated values that reflect the frequencies of accident in a locality, with a higher frequency of accidents clearly marked with larger symbols. This form of data aggregation is a useful technique to undertake generalization, especially when individualized data become overwhelmingly detail to interpret. Triangles are also superimposed on the same map to display locations identified by the Hong Kong Police as accident black spots, against which some disagreement or mismatch are apparent. The aggregation procedure thus can help target a selection of disputed locations to undergo remedial investigations subject to resource constraints.

Aggregation through the buffering operation is not to be confused with aggregation by areal units. A map of road accidents can be compiled for different zones on the basis of frequency; for example, the frequency of accidents for the four Tertiary Planning Units in our example can differentiate accident-prone areas and zonal variances (Figure 6). Figures 5 and 6 are both “frequency” maps but the former aggregates points based on proximity while the latter tally up occurrences within a specified area. Both can be used to guide the amount of planning and management of resources in specific areas.

**Figure 3.** Road accidents by severity of collision

**Figure 4.** Proximity analysis (Thiessen polygons) on the spacing of streetlights
subject to the preferred level of generalization. Frequency maps can also be made to observe variations (e.g. diurnal, seasonal, temporal, weather-bound, etc.) in accident patterns. Time series displays of frequency maps are useful in monitoring changes or more importantly revealing cases of improvement brought about by corrective planning and design measures.

Similar to the idea above, the density of accident cases in certain areas can be computed and mapped. Statistical calculations can also be performed based on the frequency and density of accidents. These maps based on the frequency and density of accidents can help substantiate findings from numerical and statistical analyses, which otherwise would not be effectively visualized.

**Associative analysis**

Associative analysis is an attempt to combine two map layers at a time in search of possible association or interaction. It is more commonly known as the overlay analysis whereby mapped elements are added or subtracted to derive a new
map. Tomlin (1990) explained the overlay analysis using map algebra and according to logical conditions specified in Boolean operations of AND (intersect), OR (union), XOR (exclusive or), and NOT (complement). The overlay operation can be applied on a combination of different features: point-on-point, point-on-line, point-on-polygon, line-on-line, line-on-polygon, and polygon-on-polygon.

The integrative ability of the overlay analysis has helped decision-making in many fields. Visual comparison of different data layers is one of the applications. The point-based accident data can be overlaid on a map of traffic lights to assess if there is possible association between accident occurrences and traffic lights in terms of their effectiveness from sequential or concentrated installation and the synchronization of signals. Figure 7 shows that while some positive association between accident occurrences and the locations of traffic lights are evident, about half of the accident sites are not accounted for whilst some traffic lights are free of traffic incidents. The accident data can also be overlaid on a map of major trunk roads to see if there is a positive correlation between the two, as shown in Figure 8. The coincidence and disagreement between the two features can be observed easily in these cases.

There are numerous possibilities of overlaying two map layers, limited perhaps by one's inspiration and creative vision. The overlay operation can be used to combine two map layers into a new layer. The operation is especially useful for undertaking data queries arising from or subject to spatial or locational connections. It can be used to uncover interlocking relationships by stripping off all but two variables for a closer examination. This method of query is powerful not only because of its visualization capability but also because of the simplicity of interpreting two maps at a time.

VI. RECOMMENDATIONS FOR HONG KONG

Road accident analysis in Hong Kong is handled jointly by the Hong Kong Police Force and the Transport Department. The Hong Kong Police is responsible for keeping computerized accident records since 1997 but is not charged with performing further analysis. The Department of Transport analyses road accident data, but none of their methodologies involves the use of GIS at the time. Weaknesses of the current computerized road accident analysis of Hong Kong were highlighted in this paper. Among the concerns were data consistency, accuracy in matching accident data, and understanding accident patterns; the last one was addressed from the cartographic standpoint in this study.

The computerized data are like a gold mine of information awaiting exploitation. Moreover, the accident database can be linked with related databases, such as traffic flows, for cross-referencing. This combination is considered fruitful as it furnishes a more complete picture toward a better understanding of the spatial variation of road accidents.

Traffic black spots deserve a closer look because of the potentially high frequency of accidents. Pedestrian injury and severity of accident are the two measures used by the Police and the Transport Department to assess a traffic black spot. The identification of traffic black spots using the 1997 accident cases of our study area was tested in a GIS environment according to criteria adopted by the Hong Kong Police and the Transport Department. Accidents involving public vehicles and pedestrians needed to be examined because of the potentially large number of human injuries involved. Data classification was employed to distinguish among accidents involving different types of vehicles, number of injuries and different levels of severity. The advantage of data classifica-

![Figure 7. An associative analysis of accident occurrences and the placement of traffic lights](image-url)
tion was to provide optional views concerning accident sites. Interesting patterns from a combination of different elements were observed and presented. The combination of different symbols helped to highlight different types of accidents occurrences.

The frequency of road accidents is another criterion used by the Police and the Transport Department to assess a traffic black spot. Accident occurrences at about the same locations were aggregated. The aggregation method was used to highlight places with a high frequency of accidents. The advantage of data aggregation is that high-risk places are highlighted by larger symbols. When we compared these places with the official black spots provided by the Police, we could find a few locations with a high frequency of accidents that have been omitted by the Police. There were also locations considered to be accident black spots but having significantly low traffic incidents in 1997. These discrepancies resulted because the classification of a traffic black spot involves several criteria, among which frequency of accidents is one. However, the classification of traffic black spots involving multiple criteria can be easily accommodated with the overlay functions in GIS.

The introduction of cartographic analysis in GIS may convey added information toward a better understanding of accident patterns. The feasibility of GIS to convert spatial realities onto maps to support cartographic analysis was demonstrated. Without a doubt, cartographic functions have maintained an important position in determining the success of GIS applications. A number of single-layer and multiple-layer operations for road accident analysis in Hong Kong have been suggested. Road accident distribution can be visualized from their spatial arrangement using single-layer operations. Point pattern analysis is useful in tracing the pattern of occurrences of road accidents. It shows whether there is clustering or dispersion of accident points and highlights the problem areas. The examples also illustrate the importance of map symbolization in highlighting unsafe intersections or districts.

A GIS has functions to present spatial realities in the form of maps and analytical abilities to track changes in the real world. The technology to implement the system described in the paper is available today. A comprehensive effort is required to harness the available technology and knowledge into an integrated and efficient information system which meets the needs of all stakeholders.

VII. CONCLUDING REMARKS

Road accident analysis is a relatively complex study that no formal academic training can cover the full spectrum (Harms 1993, p.1). The severity of road accidents in Hong Kong has stimulated our interests in the topic. The trend of persistent road accidents in Hong Kong has been attributed to the increasing number of population and vehicles. Inability to divert passengers to public transports from private vehicles may be another critical reason.

ACKNOWLEDGEMENT

Special thanks are extended to the Hong Kong Police Force (Information Office of the Traffic Kowloon West) and the Department of Transport for providing valuable data on road accidents and traffic in Hong Kong.

REFERENCES


