Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
SWUTC/05/167424-1		

4. Title and Subtitle	5. Report Date
Providing Personalized Traffic Safety Information to the Public: *Using Web-Based Geographical Information System Technologies*	February 2005
	6. Performing Organization Code
7. Author(s) Shaw-Pin Miaou, Raghav Tandon, and Joon Jin Song	8. Performing Organization Report No. Research Report 167424-1
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	10. Work Unit No. (TRAIS)
	11. Contract or Grant No. 10727
12. Sponsoring Agency Name and Address Southwest Region University Transportation Center Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	13. Type of Report and Period Covered Final Summary Report
	14. Sponsoring Agency Code

15. Supplementary Notes
Supported by general revenues from the State of Texas

16. Abstract

Due to their size and complexity, traffic crash records and related databases have not been particularly accessible to the general public. As more households, schools, and libraries are being equipped with high-speed desktop computers and as more subscribe to broadband communication services, providing Personalized Traffic Safety Information (PTSI) to individuals through the Internet will soon be a cost effective means of sensitizing and educating the public on traffic safety issues and on roadway conditions in areas and roads of personal interest to them. The main objective of this study was to explore and test the capability of existing web-based geographical information system (Web-GIS) technologies to personalize and disseminate traffic safety information to the public in a cost-effective manner. The Web-GIS technologies explored were those that could provide users with on-line access to safety databases and allowed users to visualize the distribution and statistics of traffic crashes geographically and learned about the nature of these crashes over the Internet. In particular, this study looked into the Internet Mapping (IM) technology that has been significantly advanced in recent years. To test these technologies, this project selected Brazos County, Texas, as a testbed.

At the time of this study, the locations of crash records and associated road inventory data were not georeferenced in Texas. To meet the research need, an important part of the study was an attempt to develop an efficient and accurate way of geocoding historical traffic crashes using Geographical Information Systems (GIS). In addition, under a joint effort between this and another research project, a prototype web-based traffic safety information system for Texas was developed by this research team. The system was named Web-Based Traffic Safety Information and Analysis System (W-TSIAS). It consists of a set of semi-automated GIS-based procedures for geocoding and is coupled with a suite of Internet mapping capabilities, which allow locations and attributes of geocoded crashes, road inventory, and related data to be securely accessed, viewed, and queried remotely through a typical web browser. One of the functionalites of W-TSIAS was to provide PTSI to the public.

17. Key Word	18. Distribution Statement
Personalized Traffic Safety Information, Traffic Safety Record, Geographical Information System, Geocoding, Internet Mapping	No Restrictions. This document is available to the public through NTIS: National Technial Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	43	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Providing Personalized Traffic Safety Information to the Public:
Using Web-Based Geographical Information System Technologies

By

Shaw-Pin Miaou, Ph.D.
Research Scientist
Texas Transportation Institute

Raghav Tandon*
Graduate Student
Computer Science Department
Texas A&M University

Joon Jin Song**
Graduate Student
Statistics Department
Texas A&M University

*Graduated from Texas A&M University in 2003
**Graduated from Texas A&M University in 2004

Report Number: SWUTC/05/167424-1

Sponsored by the
Southwest Region University Transportation Center
Texas Transportation Institute
Texas A&M University System
College Station, TX 77843-3135

February 2005

DISCLAIMER

The contents of this paper reflect the views of the authors, who are solely responsible for the accuracy of the information, opinions, and findings presented herein. The contents do not necessarily reflect the official views or policies of the Texas Transportation Institute (TTI), the Texas A&M University System, the University Transportation Centers Program, or the U.S. Department of Transportation. This document does not constitute a standard, specification, or regulation, nor is it intended for construction, bidding, or permit purposes. This document is disseminated under the sponsorship of the University Transportation Centers Program, U.S. Department of Transportation, in the interest of information exchange. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGEMENTS

This research was supported by a grant from the University Transportation Centers Program, U.S. Department of Transportation, to the Southwest Region University Transportation Center, which is funded 50% with general revenue funds from the State of Texas. The study was conducted mainly between September 2001 and August 2002. Due to some concerns that the information generated from this study, as contained in an earlier version of this report, could potentially be used in an unintended manner, the authors were advised to revise the report to avoid any potential misuse of the study. The report, as presented here, is the product from such revisions. Particularly, the revised report avoids the discussion of any individual traffic crashes and their locations. It should be noted that, at the time of this study, the locations of crash records and associated road inventory data were not georeferenced; but several developmental efforts to geocode them were known to be taking place at both Texas Department of Transportation and Texas Department of Public Safety. To meet the research need, an important part of the research effort was an attempt to develop an efficient and accurate way of geocoding historical traffic crashes using Geographical Information Systems (GIS).

ABSTRACT

Due to their size and complexity, traffic crash records and related databases have not been particularly accessible to the general public. As more households, schools, and libraries are being equipped with high-speed desktop computers and as more subscribe to broadband communication services, providing Personalized Traffic Safety Information (PTSI) to individuals through the Internet will soon be a cost effective means of sensitizing and educating the public on traffic safety issues and on roadway conditions in areas and roads of personal interest to them. The main objective of this study was to explore and test the capability of existing web-based geographical information system (Web-GIS) technologies to personalize and disseminate traffic safety information to the public in a cost-effective manner. The Web-GIS technologies explored were those that could provide users with on-line access to safety databases and allowed users to visualize the distribution and statistics of traffic crashes geographically and learned about the nature of these crashes over the Internet. In particular, this study looked into the Internet Mapping (IM) technology that has been significantly advanced in recent years. To test these technologies, this project selected Brazos County, Texas, as a testbed.

At the time of this study, the locations of crash records and associated road inventory data were not georeferenced in Texas. To meet the research need, an important part of the study was an attempt to develop an efficient and accurate way of geocoding historical traffic crashes using Geographical Information Systems (GIS). In addition, under a joint effort between this and another research project, a prototype web-based traffic safety information system for Texas was developed by this research team. The system was named Web-Based Traffic Safety Information and Analysis System (W-TSIAS). It consists of a set of semi-automated GIS-based procedures for geocoding and is coupled with a suite of Internet mapping capabilities, which allow locations and attributes of geocoded crashes, road inventory, and related data to be securely accessed, viewed, and queried remotely through a typical web browser. One of the functionalites of W-TSIAS was to provide PTSI to the public.

EXECUTIVE SUMMARY

Highway engineers and planners find crash histories and statistics produced from traffic crash records and related databases to be indispensable in making their operational, planning, and programming decisions to reduce the frequency and severity of crashes. Various population- and area-wide crash statistics have also been generated from these databases to sensitize and educate the public in safety campaigns regarding the magnitude and impact of important traffic safety problems. An example statistic used in campaigns against drunk driving may be: "Over 41,000 people were killed in traffic crashes in 1998, of which 38 percent were killed in crashes involving drunk drivers.... "

For individual drivers, however, one could contend that it would be more relevant, and thus more effective, to learn the traffic safety conditions of their neighborhoods and communities, and of the routes they take to work, schools, soccer games, and favorite parks, than to learn the population- and area-wide crash statistics. Under this contention, one could suggest that providing **P**ersonalized **T**raffic **S**afety **I**nformation (PTSI) to drivers and would-be drivers would be a more effective way of sensitizing them on traffic safety issues and, thus, have more impact on their driving behaviors than the traditional population- and area-wide statistics do. The PTSI could also be utilized to keep drivers and would-be drivers informed and educated of the roadway condition in their neighborhoods and communities and in the part of road networks of personal interest to them.

Due to their size and complexity, traffic crash records and related databases have not been particularly accessible to the general public. As more households, schools, and libraries are being equipped with high-speed desktop computers and as more subscribe to broadband communication services, providing PTSI to individuals through the Internet will soon be a cost effective means of sensitizing and educating the public on traffic safety issues and on roadway conditions in areas and roads of personal interest to them. The main objective of this study was to explore and test the capability of existing web-based geographical information system (Web-GIS) technologies to personalize and disseminate traffic safety information to the public in a cost-effective manner. The Web-GIS technologies explored were those that could provide users with on-line access to safety databases and allowed users to visualize the distribution and statistics of traffic crashes geographically and learned about the nature of these crashes over the Internet. In particular, this study looked into the Internet Mapping (IM) technology that has been significantly advanced in recent years. To test these technologies, this project selected Brazos County, Texas, as a testbed.

At the time of this study, the locations of crash records and associated road inventory data were not georeferenced in Texas. To meet the research need, an important part of the study was an attempt to develop an efficient and accurate way of geocoding historical traffic crashes using Geographical Information Systems (GIS). In addition, under a joint effort between this and another research project, a prototype web-based traffic safety information system for Texas was developed by the research team. The system was named Web-Based Traffic Safety Information and Analysis System (W-TSIAS). It consists of a set of semi-automated GIS-based procedures for geocoding and is coupled with a suite of Internet mapping capabilities, which allow locations and attributes of geocoded crashes, road inventory, and related data to be securely accessed, viewed, and queried remotely through a typical web browser. One of the functionalites of W-TSIAS was exactly to provide PTSI to the public

This report is organized as follows: Chapter 1 provides some background information about this project. In Chapter2, the developed procedures for geocoding Texas traffic crash records are introduced. In Chapter 3, the concept of the prototype system, i.e., W-TSIAS, which performs some visualization and analysis functions on-line, is presented and its applications illustrated. The presentation focuses on the system's Internet mapping capabilities and its limited capabilities in spatial analysis. Its use in providing PTSI is also included as part of the illustration. Chapter 4 discusses a possible extension of the developed system.

Here is an important note regarding this report. The study, as summarized in this report, was conducted mainly between September 2001 and August 2002. Due to some concerns that the information generated from this study, as contained in an earlier version of this report, could potentially be used in an unintended manner, the authors were advised to revise the report to avoid any potential misuse of the study. The summary report, as presented here, is the product from such revisions. Particularly, this report avoids any discussion of individual traffic crashes and their locations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Highway engineers and planners find crash histories and statistics produced from traffic crash records and related databases to be indispensable in making their operational, planning, and programming decisions to reduce the frequency and severity of crashes. Various population- and area-wide crash statistics have also been generated from these databases to sensitize and educate the public on the magnitude and impact of important traffic safety problems in safety campaigns. An example statistic used in campaigns against drunk driving may be: "Over 41,000 people were killed in traffic crashes in 1998, of which 38 percent were killed in crashes involving drunk drivers.... "

For individual drivers, however, one could contend that it would be more relevant, and thus more effective, to learn the traffic safety conditions of their neighborhoods and communities, and of the routes they take to work, schools, soccer games, and favorite parks, than to learn the population- or area-wide crash statistics. For example, one could arguably contend that knowing the answers to the following two safety questions would have more impact on individuals in their driving behaviors than knowing that over 41,000 people were killed in traffic crashes and many million more injured in this country every year:

- How many traffic crashes occurred on the route that I usually take to/from work in the last 3 years? When and where on the route did these crashes occur? How many crashes were related to speeding or the use of alcohols? How many pedestrians, children, and bicyclists were involved?
- How many traffic crashes occurred within the proximity (e.g., 3 miles) of where I live during the last 3 years? When and where did these crashes occur? How many crashes were related to speeding or the use of alcohols? How many pedestrians, children, and bicyclists were involved?

Under this contention, one could suggest that providing **P**ersonalized **T**raffic **S**afety **I**nformation (PTSI) to drivers and would-be drivers would be a more effective way of sensitizing them on traffic safety issues and, thus, have more impact on their driving behaviors than the traditional population- or area-wide statistics do. The PTSI could also be utilized to keep drivers and would-be drivers informed and educated of the roadway condition in their neighborhoods and communities and in the part of road networks of personal interest to them.

Due to their size and complexity, traffic crash records and related databases have not been particularly accessible to the general public, not even to local communities and educational institutions. As more households, schools, and libraries are being equipped with high-speed desktop computers and as more subscribe to broadband communication services, providing PTSI to individuals through the Internet will soon be a cost effective means of sensitizing and educating the public on traffic safety issues and on roadway conditions in areas and roads of personal interest to them.

The main objective of this study was to explore and test the capability of existing web-based geographical information system (Web-GIS) technologies to personalize and disseminate traffic safety information to the public in a cost-effective manner. The Web-GIS technologies explored were those that could provide users with on-line access to safety databases and allowed users to visualize the distribution and statistics of traffic crashes geographically and learned about the nature of these crashes over the Internet. In particular, this study looked into the Internet Mapping (IM) technology that has been significantly advanced in recent years. Specific technical capabilities evaluated include geocoding, spatial query, buffering and data streaming, data dissemination and downloading, dynamic mapping and visualization, scalability, and cartographic rendering.

To test these technologies, this project selected Brazos County, Texas, as a testbed. With many new students coming to the county to attend Texas A&M University and other colleges every year, this county is an interesting testbed for such a traffic safety oriented study. Some scenarios developed for testing in this study aimed at disseminating the PTSI to some hypothetical students in driving schools, colleges, and universities within the County, especially those live in City of Bryan and City of College Station.

Traffic crash records in Texas are administered and coded by Texas Department of Public Safety (TxDPS); while the road inventory database is maintained by Texas Department of Transportation (TxDOT). At the time of this study, the locations of crash records and associated road inventory data were not georeferenced; but several developmental efforts to geocode them were known to be taking place at both TxDOT and TxDPS. As the traffic crash records stood at the time of this study, three location referencing systems were used concurrently to indicate the location of a crash, depending on whether a crash occurred on or off state highways and which city the crash occurs if it occurs in a city on a non-state highway. A discussion of these location referencing systems will be provided in Chapter 2. To meet the research need, an important part of the study was an attempt to develop an efficient and accurate way of geocoding historical traffic crashes using Geographical Information Systems (GIS). The main sources of information employed in this study for geocoding Texas traffic crashes and some of the geocoding efforts made by this study are also reported in Chapter 2.

Under a joint effort between this and another research project entitled "A Highway Safety Performance Mapping Initiative: From Data to Visualization to Countermeasures," Project Number DTTS-00-G-G-B005-TX, funded by Bureau of Transportation Statistics (BTS), U.S. Department of Transportation," a prototype web-based traffic safety information system for Texas was developed by this research team and tested on a limited number of counties. The system was named Web-Based Traffic Safety Information and Analysis System (W-TSIAS). It consists of a set of semi-automated GIS-based procedures for geocoding and is coupled with a suite of Internet mapping capabilities, which allow locations and attributes of geocoded crashes, road inventory, and related data to be securely accessed, viewed, and queried remotely through a typical web browser. One of the functionalites of W-TSIAS was to provide PTSI to the public. Chapter 3 describes the concept of the system and illustrates its use in providing PTSI.

The study, as summarized in this report, was conducted mainly between September 2001 and August 2002. Due to some concerns that the information generated from this study, as contained in an earlier version of this report, could potentially be used in an unintended manner, the authors were advised to revise the report to avoid any potential misuse of the study. The report, as presented here, is the product from such revisions. Particularly, this report avoids any discussion of individual traffic crashes and their locations.

This report is organized as follows: In Chapter2, we introduce the developed procedures for geocoding Texas traffic crash records. In Chapter 3, we present the prototype system developed by the project team, i.e., W-TSIAS, which performs some visualization and analysis functions on-line. The presentation will focus on the system's Internet mapping capabilities and its limited capabilities in spatial analysis. Chapter 4 discusses a possible extension of the developed system.

Chapter 2. Geocoding Texas Traffic-Crash and Related Data

2.1 Introduction

Geocoding or georeferencing is the process by which we identify point location of an entity or event in geographical coordinates (i.e., longitude and latitude) from postal street addresses or addresses of other referencing systems for displaying the location of that entity or event on a map. It is sometimes described as the computer equivalent of pushing pins into a street map on our wall. For example, in studying traffic crashes, we are interested in (1) geocoding the location of a traffic crash on a local street by determining its longitude and latitude coordinates from the street address given in the police report, and (2) geocoding the location of traffic crashes on state (or, more precisely, state-maintained) highways from a location referencing system such as the control section-milepoint system often used by state highway departments (Smith et al., 2001). Once the point location of an entity or event is geocoded, its location and other attributes can be displayed on a (typically) planar map and its spatial relationship with other neighboring entities and events can be visualized, explored, and analyzed. For example, using statistical ranking and selection criteria, traffic engineers are interested in identifying sites, traffic zones, or a cluster of sites or zones on a road network, which exhibit unusually high crash rates or frequencies with given site characteristics or are overrepresented in certain types of crashes, for further engineering evaluation and safety improvements [Miaou and Song, 2004].

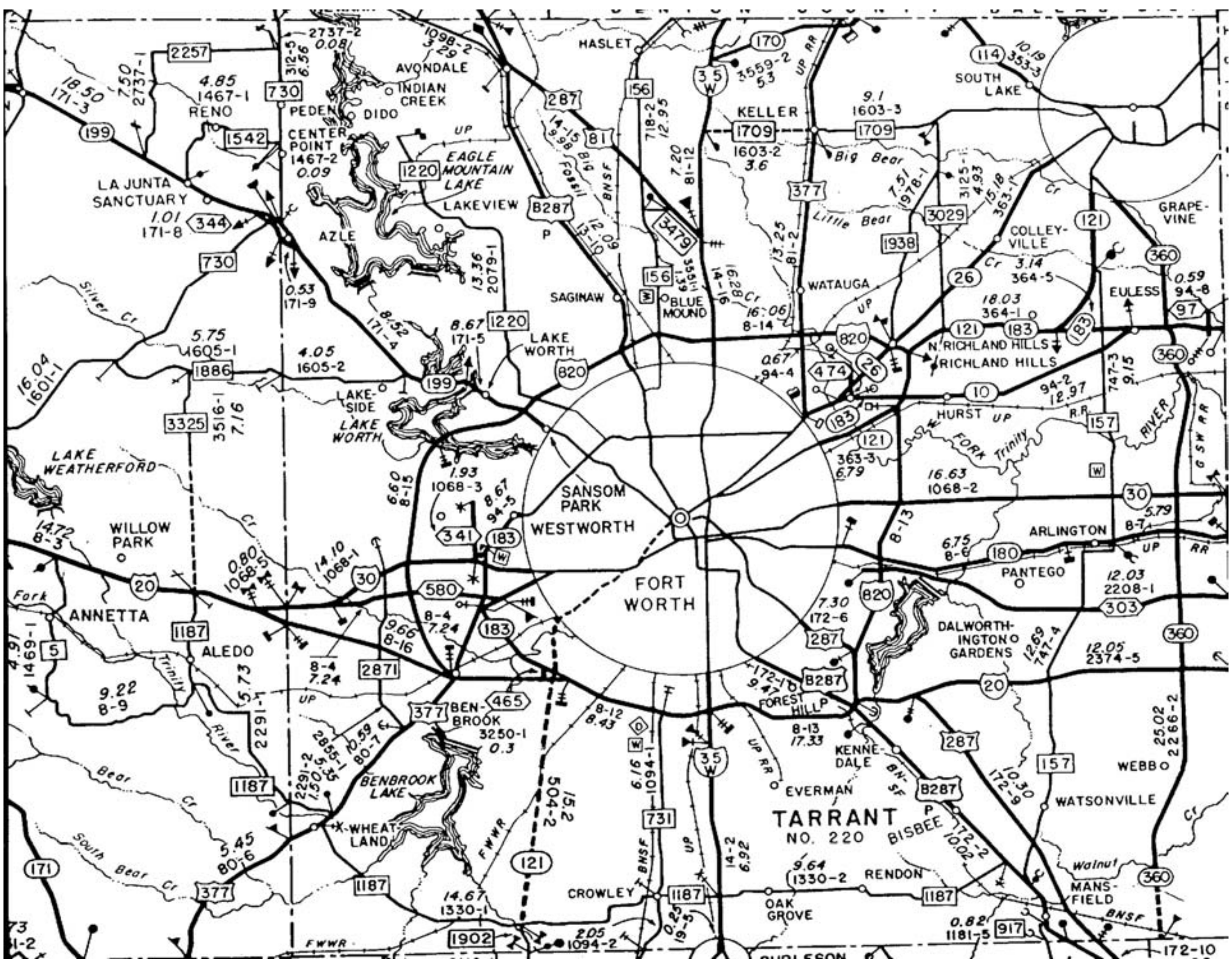
As indicated earlier, at the time of the study, the locations of traffic crash records were not geocoded. Instead, three location referencing systems were used concurrently to indicate locations of crashes, depending on whether a crash occurred on or off state highways and which city the crash occurred if it occurred in a city on a non-state highway:

- Traffic crashes occurred on state highways were coded using the Texas Deapartment of Transportation (TxDOT) control section-milepoint system across the state. Figure 2.1 shows an example control section map, in which highway numbers, control section numbers, and section lengths (or, more precisely, driving distances) in miles are marked along the highway. The control section-milepoint system is uniform across the state and control section numbers are unique in the state and, therefore, there is no confusion as to where the crashes actually occurred under the control section-milepoint referencing system. The main issue was that control sections were not geocoded and, thus, crashes could not be mapped with GIS in longitude and lattitude coordinates.
- For crashes occurred in a city on non-state highways, they were coded in either the first five letters of the street name or a unique five digit numerical code (or a mix of both), depending

on cities. Note that the five digit numerical code is unique only within that city. For crashes occur in cities on a mid-block, block numbers, in addition to the street code, are usually provided. The first-5-letter coding method can be problematic in that names of more than one street in a city may have the same first five letters, e.g., <u>Unive</u>rsity Drive and <u>Unive</u>rsal Street, or several streets in different parts of a large city may use the same street names, e.g., the popular "Main Street."

- For those occurred on county roads (outside of the city limit) that were not maintained by the State, the first five letters of the street name were typically given. However, for non-intersection crashes, the exact locations or block numbers of the streets might not be available. This is particularly problematic for geocoding crashes occur on a long stretch of roads with a single street name. Again, the first-5-letter coding method can be a problem when names of more than one county road have the same first five letters.

Figure 2.1 Example control section map for Tarrant County, Texas.



Without geocoded traffic crash records, traffic engineers, law enforcement officers, and other safety analysts have great difficuties mapping, visualizing, and analyzing the location and spatial characteristics of traffic crashes, which are essential in understanding the effect of roadway and roadside environment under which these crashes have occurred. An informal discussion with several researchers at Texas Transportation Institute (TTI), who have been involved in various highway safety improvement projects in Texas, suggested that typically about 60 to 85 percent of their project resources are expended on obtaining, preparing, and visualizing traffic crashes, road inventory, and related data. In addition, one common reason for the preparation of crash data to be so resource consuming is because of the lack of geocoding system in the crash record. As a result, crash data mapping, visualization, and linking with roadway and other related data for their projects are often painstakingly carried out manually. An exemplary story told by these researchers about this slow data preparation process is the need to use junior staff members and students to mark the location of each crash on an enlarged control-section or highway map by hand (often color-coded to indicate various levels of crash severities). Thus, having a geocoding and mapping system for the Texas crash record that can be deployed statewide can potentially save a lot of resources and enable the investigators to conduct their projects more efficiently and effectively.

The readers are referred to similar efforts on crash data geocoding and GIS-based mapping and analysis applications conducted at other states, e.g., New Mexico Traffic Safety Bureau, New Mexico (http://unm.edu/~dgrint/tsb.html), Kentucky State Police Records Branch, Kentucky (Carrico, 2002), and Honolulu, Hawaii (Levine et al., 1995). We also find Smith et al. (2001) to be a good reference for understanding how traffic crash records, road inventory, and other data can be integrated under a GIS platform using one or several linear location referencing systems.

2.2 Data Sources, Pre-Processing, and Control Points

The main sources of information employed in this study for geocoding Texas traffic crashes include: (1) traffic crash database from TxDPS, (2) state road inventory database from TxDOT, (3) control section maps for state highways from TxDOT (paper maps), and (4) TIGER/Line 2000 Files, which are available on-line from the Census Bureau, U.S. Department of Commerce. Note that "TIGER" stands for "Topologically Integrated Geographic Encoding and Referencing." Other sources used occasionally for verification purpose and for extracting additional location information include the National Transportation Atlas Databases (NTAD) 2001 from the Bureau of Transportation Statistics (BTS) and some free location information from on-line mapping service providers such as MapQuest.com™. Of these information sources, TIGER/Line data, NTAD, and maps from some on-line service providers are geocoded.

We choose the TIGER/Line 2000 data as the "basemap" in that it includes all classes of roads from primary highways, secondary roads, to local streets. Note that NTAD covers the national highway planning network, which includes only part of the state-maintained highways in Texas and does not have secondary roads and local streets. The use of TIGER/Line data as basemap also allows us to take advantage of the Census' data model, which facilitates topological integration with other Census' data layers including (1) Census units, such as blocks, block groups, tracts, and county boundaries, (2) other line features, such as railroads and hydro lines, and (3) various area-based demographic and socioeconomic indexes.

In developing the geocoding system for crashes occurred on state highways, the longitude and latitude coordinates of the two endpoints of each "TIGER line" (i.e., each TIGER road segment) in the basemap are assumed to be the "ground truth" coordinates. Basically, the system uses these coordinates from the basemap as the basis and extracts as much useful and accurate information as possible from various data sources mentioned above to establish as many "control points" as practicable on the basemap. These control points consist mainly of intersections and interchanges. Other point features of the road that serve as control points include boundary points between two political juridictions and bridges. These control points are used in "interpolating" the positions of control sections and crashes for which locations on the basemap are not as well informed. In general, the more control points that we can identify, the higher the location accuracy we can achieve in geocoding. However, it is desirable that the steps taken to establish control points be streamlined and automated to the extent possible to minimize the manual input required since this is by far the most time-consuming part of the geocoding process. At the time this report was written, the deveoped geocoding system was still not particularly "intelligent" as far as the automation goes, and many imporvements to speed up the process are possible.

Conventions for Numbering Control Sections

As mentioned earlier, traffic crashes occurred on state-maintained highways are coded using the TxDOT control section-milepoint system. A control section can be regarded as a "route" consisting of a number of consecutive road segments with varying lengths, which are numbered as ####-1, ####-2, etc. The control section numbber #### is unique within Texas and segment numbers 1, 2, … are identification numbers (id's) of segments within a control section. Each segment within a control section is given a beginning milepoint (bmp) and an ending milepoint (emp), the difference of which (= emp - bmp) indicates the centerline driving distance from the beginning to the end of the segment. Typically, the beginning milepoint of the first segment for a control section starts with 0.00, and the beginning milepoint of the second segment is the same as the ending milepoint of the first segment, and so on. However, from the data that we worked on, we do observe some control sections for which the first segment does not begin with milepoint 0.00 and/or exhibits “milepoint gaps” between consecutive segments. To be discussed below, for these control sections, special cares need to be exercised.

Connecting points between control sections are usually intersections, interchange points, or boundary points between two political juridictions. In addition, consecutive segments within a control section are usually split at these breakpoints as well. Other possible split points for segments within a control section inlcude entry/exit points on freeways, major bridges, over- or under-passes, and highway-rail crossings. These “breakpoints” between control sections and between segments can usually serve as control points. As will become clearer later, in the cases where milepoint gaps exist, more control points need to be determined and inserted into the basemap even when the location information is weak at these points.

Pre-Processing Traffic Crash Records

One lesson learned by the project team during the development of the geocoding system is the need to pre-process the crash records so that the information contained in the crash database can be best used to identify control points for state highways and be handled and geocoded more efficiently in general. In the pre-processing step, we usually screen about 3 to 5 years of crash records to identify crashes occurred on state-maintained highways (i.e., those crashes that have values in the control-section and related fields). This pre-processing step allows us to separate traffic crashes in the database into two major categories: on state highways and off state highways. We then group those crashes occurred on state highways into three subcategories and those off state highways into two subcategories:

(1) State-State Intersection Crashes:
This subcategory includes crashes occurred at intersections or related to intresections of two state highways, the location of which is identified by two set of control sections and milepoints. Symbolically, the intersection can be represented by the couple of two vectored attributes -- [(cs1, id1, mp1), (cs2, id2, mp2)], under the control section-milepoint system.

(2) State-Local Intersection Crashes:
This subcategory includes crashes occurred at intersections or related to intersections of one state highway and one “local” street (or, to be more precise, a “non-state highway”), the location of which is identified by the control section and milepoint of the state highway and a street code (either the 5-digit-numerics or first-5-letter code). Block

number is provided for the primary road of the two. Symbolically, the location can be represented by either [(cs, id, mp), (st. code, bn)] or [(cs, id, mp), (st. code)].

(3) State Non-Intersection Crashes:

This subcategory includes crashes occurred on a state highway that did not involve intersections. Their locations can be represented symbolically by a vectored attribute as (cs, id, mp).

(4) Local-Local Intersection Crashes:

This non-state highway subcategory includes crashes occurred on or related to the intersection of two non-state highways. Block number is provided for the primary street. Their locations can be symbolically represented as [(st. code 1, bn1), (st. code 2)].

(5) Local Non-Intersection Crashes:

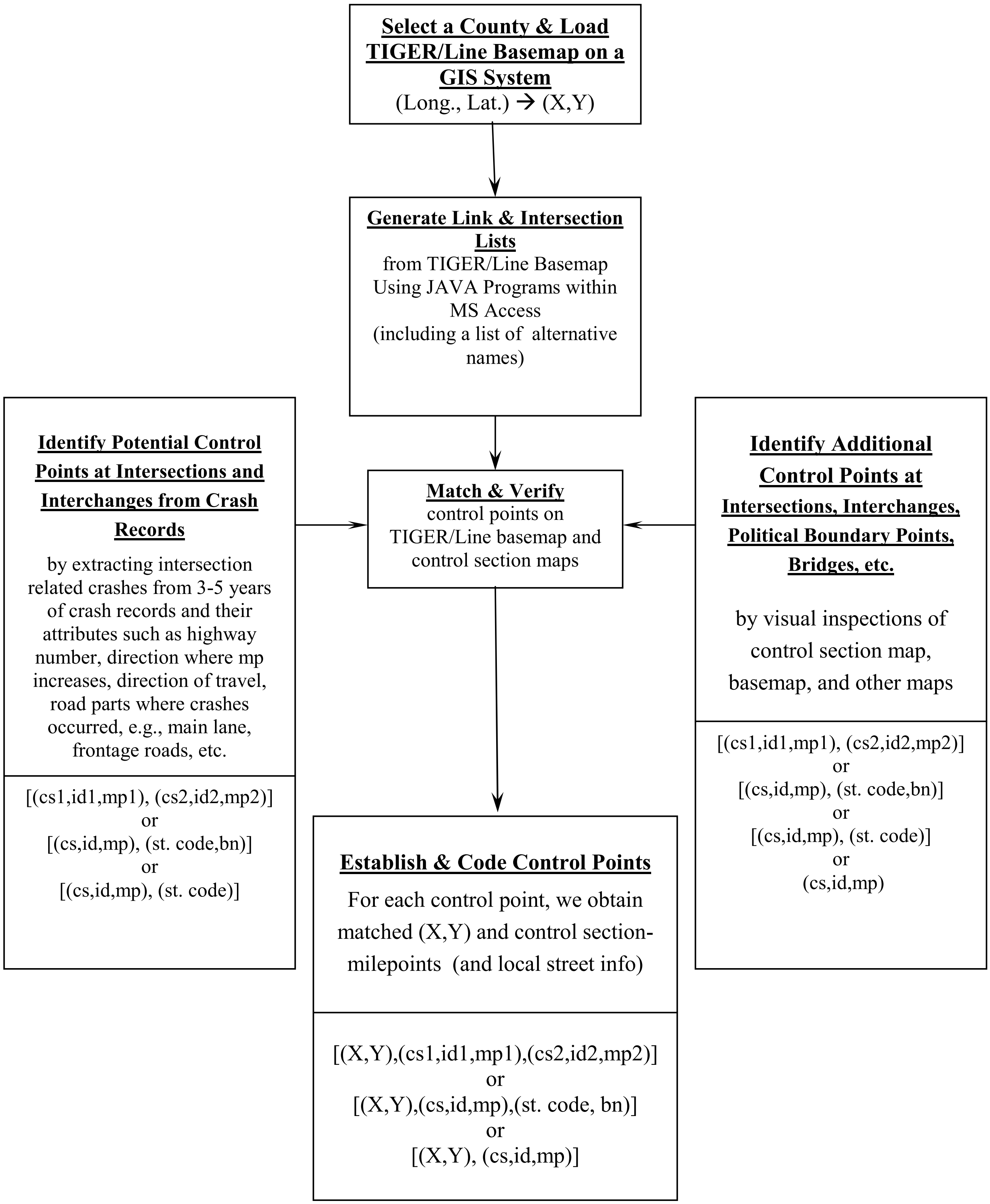
The subcategory includes crashes occurred on non-state highways and did not relate to intersections. They are represented symbolically as (st. code, bn).

Establishing Control Points on the Basemap

As discussed earlier, establishing as many accurate control points on the basemap as practicable is the most critical and time-consuming part of the geocoding process for crashes occurred on state highways. The goal is to find as many points on the basemap as possible for which we can match with good confidence their longitude and latitude, or (X,Y), with the corresponding control section-milepoint and/or street code-block number. Figure 2.2 shows a schematic diagram describing a typical process used by the project team to establish control points.

Our experience suggests that an efficient way of identifying many control points at once is to make use of the state-state intersection and state-local intersection crashes as screened in the pre-processing step. We found in all 4 counties that we have worked on a high percentage of intersections or interchange points had some crashes occurred during the 3 to 5 years period for which crash records were screened. These crashes are then sorted by the control section number, segment identification number, and milepoint of the first state highway indicated in the record. The locations of these crashes are further matched with the basemap loaded on a GIS system and visually verified with the control section map. Note that, other than the control section-milepoint inforamtion of these intersections, we also extracted some of their attributes, such as highway number, direction in which milepoint increases, direction of travel, road parts where crashes occurred, e.g., main lane, frontage roads, etc. These attributes are helpful in the matching and verification process mentioned above.

Figure 2.2 Establishing control points on basemap using location information from crash records and other data sources for state highways.



Not all connecting points between control sections and not all breakpoints between segments within a control section have a good number of crashes that are consistent enough in their location data to establsh them as control points. Extra effort, due diligence, and resourcefulness usually enable us to identify additional control points on the basemap. Boundary points between two political juridictions, for example, are usually either the beginning or the end of a control section. This part of pre-processing work is, however, quite ad hoc and time consuming, usaully involving a significant amount of visual inspection and confirmation.

2.3 Geocoding Approaches

Two automated approaches to geocoding are common in GIS: (1) dynamic segmentation and (2) address/intersection matching (or address/intersection geocoding) (Smith et al., 2001). The objective here is to geocode control sections from the state road inventory database and all traffic crashes as categorized in the last section. In geocoding control sections and crashes occurred on state highways, control points established in the last section are used as "anchors" for matching and linear interpolation. For local streets, street codes and block numbers are the key fields for matching and address interpolation. We found both approaches to be useful for geocoding control sections and crashes occurred on state highways, and the latter approach was very effective for geocoding crashes on local streets.

The way to use the automated address matching technique in GIS for geocoding crashes occurred on state highways is to trick the system to treat the "control section-segment ID" as the "street name" and the "milepoint" as the "block number." Note that, for this trick to work, "milepoints" for both control points and crashes need to have the same unit and be converted to integers (e.g., by multiplying the old milepoints with, say, 100). Readers are referred to Smith et al., (2001) for the concept and application of dynamic segmentation and address geocoding in traffic safety studies.

For local streets with the first-5-letter of the street name as the street code, the address matching capability in the ArcGIS software package does sort out possible matching street names for user's to inspect and choose the correct ones. Note that crash records with ambiguous addresses, new addresses, "vanity addresses", entry errors in address, and other types of errors sometimes do make the automated address matching difficult or impossible. For those non-intersection crashes occurred on state highways, some were given control section and segment identification numbers, but not the milepoints. For these crashes, the exact locations within the segment were not known and were artificially assigned to the middle of the segment. An "uncertain location flag" field was given a value of "1" to signify the uncertainty in location data (0 otherwise), and the "line range" within which the crash had occurred was recoded, i.e., the (X,Y)'s of the two endpoints of the segment. Note that, for those crashes with control sections reported without milepoint information, we found many of them did contain data on street names and block numbers, which could be used in geocoding with a special handling routine.

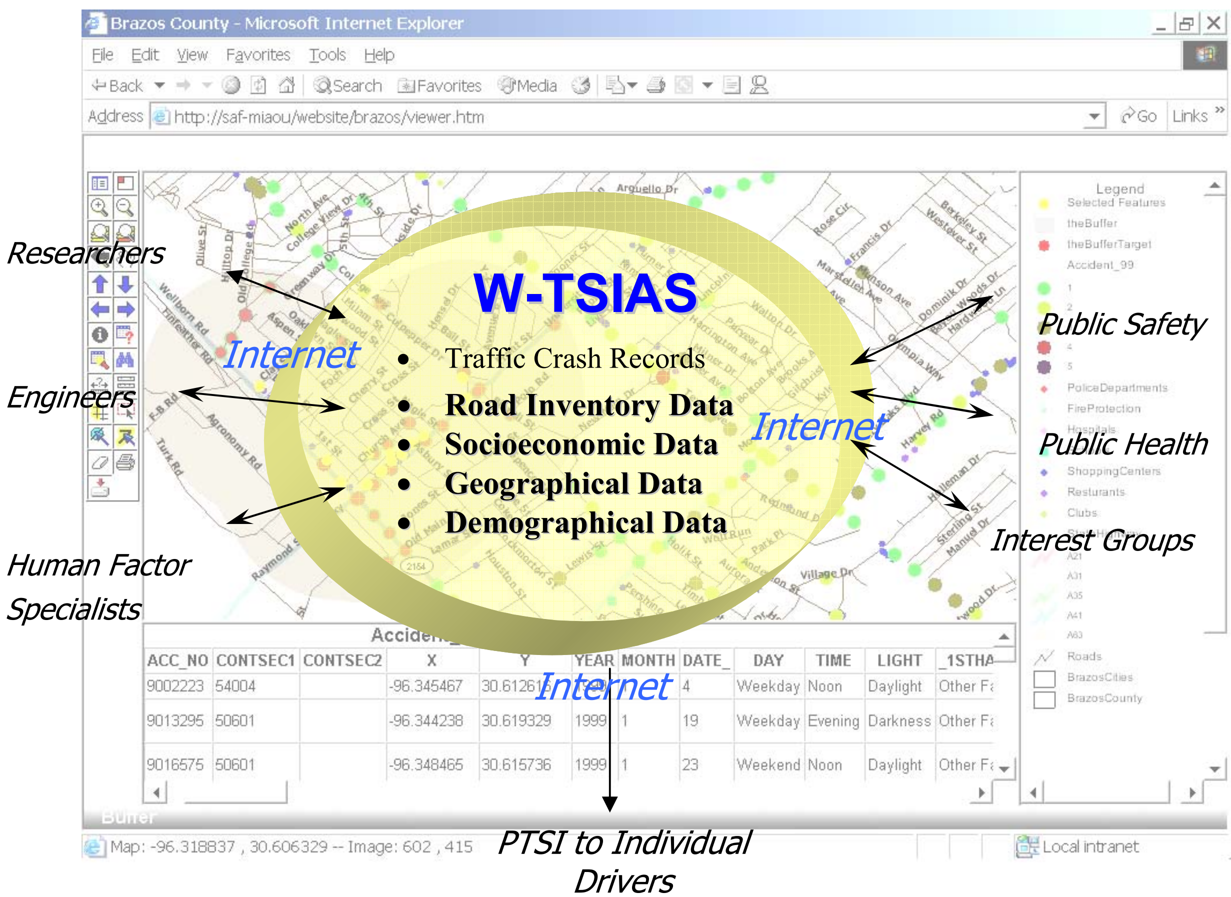
For statistical analysis purposes, it is important that the coder makes an effort to provide the best estimate and possible range of where the crash might have occurred when the exact location is unclear or unknown. The reason is that these crashes with missing or incomplete location data tend to be concentrated at certain control sections or streets or of a certain crash type.

In this chapter, we have discussed general principles and procedures used in geocoding Texas traffic crash records. As with all other real world traffic data, there are always unusual cases and some exceptions that need to be dealt with on a case-by-case basis. In an effort to geocode 3 years of crash records from 1997 to 1999 for Brazos County, the project team was able to achieve a success rate of about 85 to 90 percent with automated matching and interpolation methods discussed above. With additional manual efforts, we were able to correct many data reporting errors and deal with unusual cases, which push the success rate up to about 98 percent. Note that the first-5-letter street code is used in the Brazos County and the total number of reported crashes was 2,140, 2,211, and 2,251 crashes in 1997, 1998, and 1999, respectively.

CHAPTER 3. W-TSIAS AND EXPLORATORY SPATIAL ANALYSES

The Web-Based Traffic Safety Information and Analysis System (W-TSIAS) consists of a set of semi-automated GIS-based procedures for geocoding and is coupled with a suite of Internet mapping capabilities, which allow locations and attributes of geocoded crashes, road inventory, and related data to be securely accessed, viewed, and queried remotely through a typical web browser. W-TSIAS is intended to be a Web-based system allowing users, such as researchers, engineers, human factor specialists, public safety and health officials, and interest groups, to securely access, view, query, and analyze locations and attributes of crashes, road inventory, and related data, through an Internet browser for the purpose of devising countermeasures to imporve highway safety. With its current design, W-TSIAS can be easily extended to provide PTSI to the public. Figure 3.1 depicts the high-level concept of the system.

Figure 3.1 W-TSIAS concept: allowing crash data to be securely accessed, viewed, mapped, and analyzed through an Internet browser from anywhere by multiple users simultaneously.



Three Internet mapping tools were considered in developing the W-TSIAS prototype: ESRI's ArcIMS, MapInfo's MapXtreme® Java™ Edition, and Manifold 5.0 Internet Map Server. These tools were evaluated based on their system architecture and security management capabilities at the server side, supports of client-side browsers, and their out-of-the-box GIS tools and analysis features. A summary of the evaluation is presented in Table 3.1. Understanding system architecture and security management helps us judge whether some of the capabilities, such as data confidentiality and security management, video stream integration, on-line access of CAD-based highway contruction plans, and teleconferencing, that are desirable for the future version of W-TSIAS (but are not currently supported in the prototype) will be easy to implement. Knowing what client browers are supported by these tools helps to further determine how accessible can the system be for potential users. Examining out-of-the-box features supported by the Internet mapping tools helps us understand the functionalities that are useful and readily available to the user with minimal programaming efforts. (Note that, at the time of preparing this report, newer versions of the three GIS software packages evaluated in this study were released and we were unable to update our evaluation due to limited resource.)

Based on this evaluation, ESRI's ArcIMS was chosen to develop the W-TSIAS prototype. The ArcIMS Internet mapping system (http://www.esri.com/software/arcims/index.html):

- Provides the foundation for distributing high-end GIS and mapping services via the Internet
- Enables users to integrate local data sources with Internet data sources for display, query, and analysis in an easy-to-use Web browser
- Improve the way users can access and interact with Internet mapping and GIS data

In addition, key features of the system include:

- Integration with ESRI's ArcGIS Desktop products, including the ability to publish ArcMap and ArcPublisher documents on the Internet
- Ability to combine data from multiple sources
- Secure access to map services
- Wide range of GIS capabilities
- Highly scalable architecture
- Standards-based communication
- Support for a wide range of clients
- Useful metadata services for indexing and sharing geographic information

The out-of-the-box client-side functionalities supported by ArcIMS are listed in Table 3.2. A selected number of these functionalities are illustrated as follows.

Table3.1 Internet mapping capabilities: comparison chart

Company	ESRI	MapInfo	Manifold
Product	ArcIMS 4.0	MapXtreme® Java™ Edition	Manifold 5.0 (Integrated Internet Map Server)
Architecture	Scalable and uses ArcXML as a standard for communication	Highly Scalable Architecture with intelligent multi-threading, Support for enterprise XML	Simple ASP based with no middleware or plug ins required
Security	Offers support for Secure Socket Layers (SSL) and Secure Hypertext Transfer Protocol (HTTPS) for managing the security of your Web site.	Minimal, only for database connection	No information available
Client Language	Java Applets, JavaScript with HTML	Java Applets	HTML, ASP
Sever Side Language	Java Servlets, Cold Fusion, ActiveX	Java Servlets	ASP
Operating Systems Supported	Windows NT 4.0, Windows XP, Windows 2000, HP-UX 11.x, Sun Solaris (7.0, 8.0, 9.0), IBM AIX 4.3.3, RedHat Linux 7.1, IRIX 6.5.11	Windows NT 4.0, Windows XP, Windows 2000, HP-UX 11.x, Sun Solaris 2.7/2.8 (Sparc)	Windows XP, Windows 2000, Windows NT 4.0, Windows ME, Windows 98SE, Windows 98
Web Servers Supported	Apache (w/ Tomcat 3.3, Tomcat 4.0.1), IBM HTTP Server 1.3.12, iPlanet (4.0, 4.1), Sun ONE (iPlanet) 6.0, Oracle Application Server 9i, WebLogic (6.0, 6.1)	Apache (w/ Tomcat 3.3, Tomcat 4.0.1), IBM Web Sphere 4.0.2, Sun ONE (iPlanet) 6.0, Oracle Application Server 9i, BEA WebLogic 6.1	Microsoft IIS 5.0
Web Browsers Supported	Internet Explorer, Netscape Communicator[1]	Internet Explorer, Netscape Communicator[2]	Microsoft IE, Netscape, Mozilla and Opera[3]
Online Capabilities	Standard function (Zoom, Pan, etc.), image rendering, feature streaming, data extraction, geocoding, and the ability to	Standard function (Zoom, Pan, etc.), image rendering, animated images, attribute queries on	Standard function (Zoom, Pan, etc.)

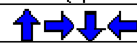
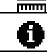
[1] Versions supported vary with operating systems

[2] Minimum version 5

[3] Version support varies with the type of template used. All templates support version 5 and above

	perform both spatial and attribute queries on underlying data.	underlying data	
J2SE JRE Version Supported	1.3.1_01 and 1.4.0[1]	1.22 and higher	Not required
Memory Requirements	256 MB RAM per CPU	64MB RAM	64MB RAM
Support for Handheld Devices	Yes	No information available	No information available
Multi-Language Support	No	Yes (American English, French, German, Italian, Japanese, Korean, Spanish, Swedish, as well as Simplified and Traditional Chinese)	No

Table 3.2 Out-of-the-box client-side functionalities available from ArcIMS.

No.	Property	Symbol	Description
1.	Zoom Out		View a larger area than current
2.	Zoom In		View a smaller area in greater detail
3.	Zoom Full Extent		View the entire extent of the map
4.	Zoom to Active Layer		View the area of the map that is covered by the active layer
5.	Pan		Free hand pan allows viewing area in any direction
6.	Unidirectional Pan		Allows to view area in particular direction
7.	Back to Last Extent		View the image seen before the setting (by zoom in, pan etc) were changed
8.	Overview Map Toggle		View/hide a small image that displays the overview map
9.	Measure		Measure distance in terms of scale units set.
10.	Set Units		Set scale units to feet, miles, degrees, kilometers etc.
11.	Identify		Identify the selected active feature
12.	Select by Line/Polygon		Select feature values in the given point, line or polygon
13.	Select by Rectangle		Select feature values within the specified rectangle
14.	Buffer Shape		Select feature values that like within some specified distance from the given point, line or polygon
15.	Buffer		Select feature values of the specified layer that lie within some specified distance from the selected active features.
16.	Find		Find a string that is specified in the data items of a given layer
17.	Query		Select feature values that satisfy a given query
18.	Stored Query		Select feature values that satisfy a predefined query that may have some specified values entered by the user.
19.	Clear Selected Features		Clear the feature values shown as selected on the image.
20.	Print		Generate a Print View of the current view of the map

Location, Frequency, and Density of Traffic Crashes

To show some on-line mapping capabilities and limitations, Figures 3.2-3.4 present locations of crashes of various levels of severities in different colors in 1999 on state highways for Dallas and Tarrant Counties–two large urban counties, and for state and local roads around the campus of Texas A&M University in College Station, Brazos County. Darker color circles indicate locations of crashes with greater severity levels. As we can see from these figures, displaying crash locations to get a good sense of how crash frequencies and densities vary on the network is not an easy task, given that they often overlap at intersections and at road segments under certain display resolutions. As will be seen later, one way to allow some visual feel of how crash frequencies vary by intersections is to let the diameter (and the color) of the circle be proportional to the number of crashes. Note that crashes at some control sections in Dallas and Tarrant Counties don't have milepoint information in the database and are not shown in these maps. Figure 3.4 also shows locations of police and fire departments, hospitals, schools, restaurants, shopping centers, and clubs serving alcohols, which are key public and commercial establishments of interest in studying traffic crashes on urban streets.

Cell-Based and Distribution Analyses

Cell-based analysis, also referred to as "grid-based" analysis, uses a grid or a cell to aggregate crashes and vehicle-miles, and other covariate values. Figure 3.5 gives an example of cell-based total crash counts in Brazos County for year 1997. Each grid is of the size 4.5 miles by 3 miles. Total number of crashes in each grid is first calculated and variation of crashes on the network is then represented by the relative size of circles, which are placed at the center of the grid. In the figure, the information displayed at the bottom of the screen reflects the lower-left red circle with 411 crashes occurred in that grid for the year 1997. Other information displayed is the total number of vehicles involved, the number of pedestrian related crashes, the number of rollover crashes, the number of crashes occurred under dry, wet, and snowy road surface conditions, the number of fatal crashes, the number of non injury crashes, the number of fatalities, and number of crashes in weekdays, weekends, nighttime, daytime, etc. These crashes could be adjusted by the corresponding vehicle miles took place in that grid. Since cell-based analysis aggregates data at a specified grid resolution, it would not be appropriate for site-specific spatial analysis.

Figure 3.2 Locations of crashes of various levels of severities in different colors in 1999 on state highways for Dallas Counties

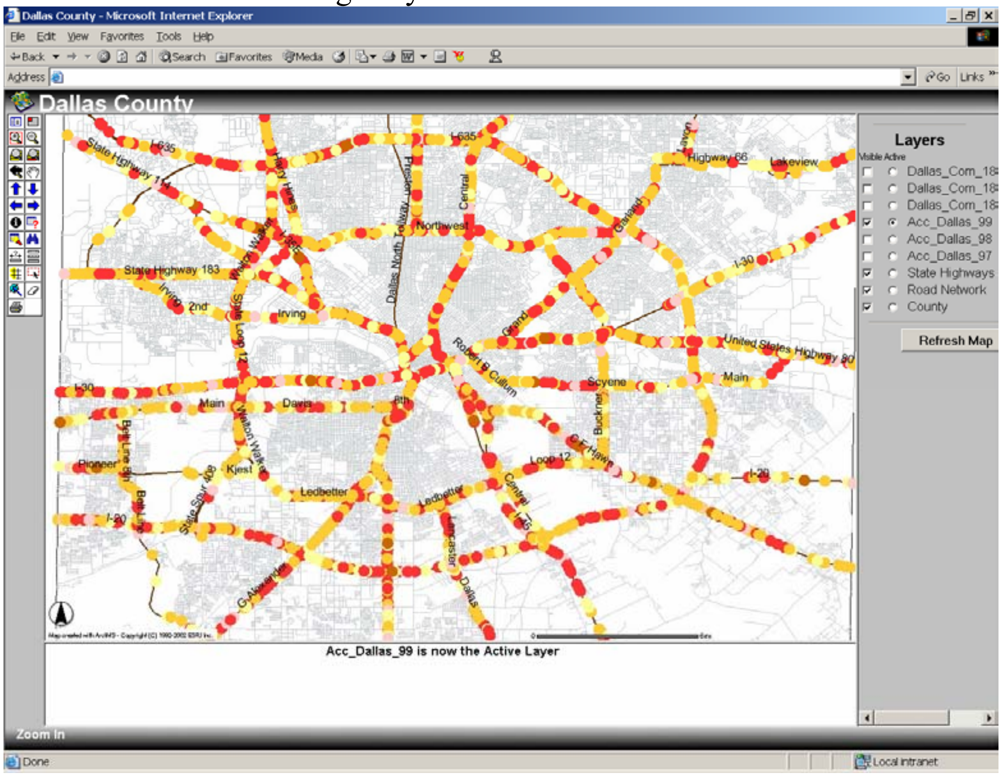


Figure 3.3 Locations of crashes of various levels of severities in different colors in 1999 on state highways for Tarrant Counties

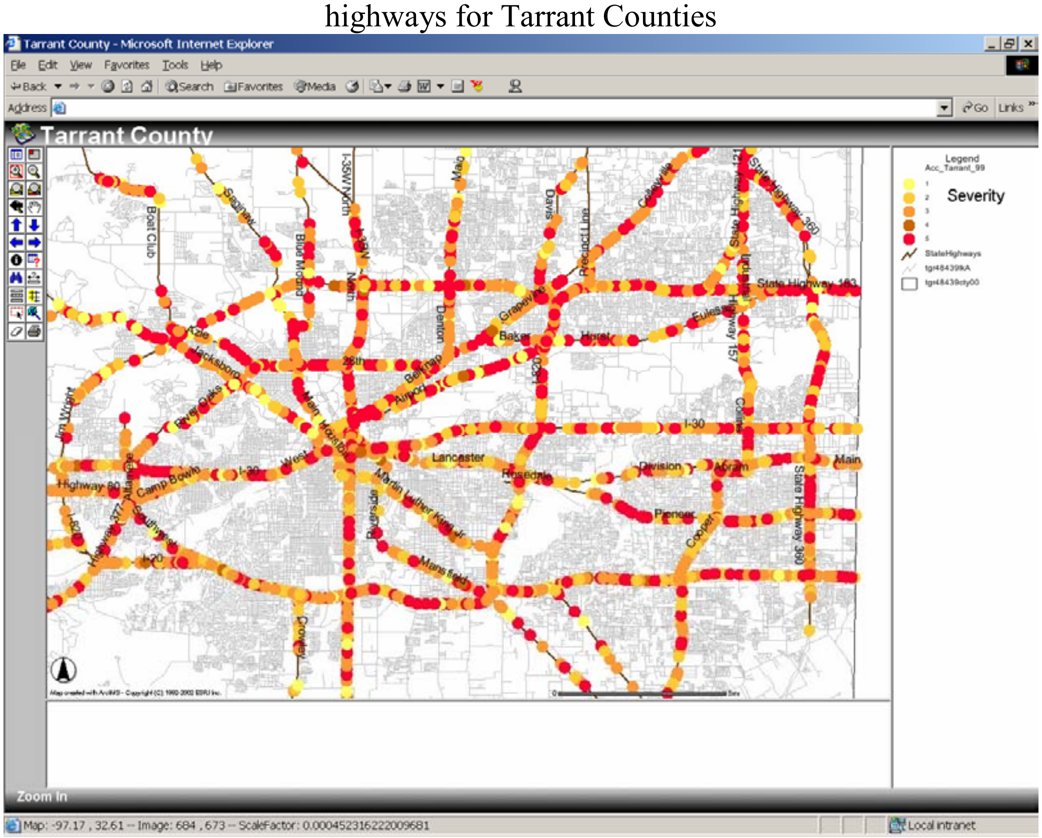


Figure 3.4 Locations of crashes on state and local roads in 1999 around the campus of Texas A&M University, College Station, Brazos County

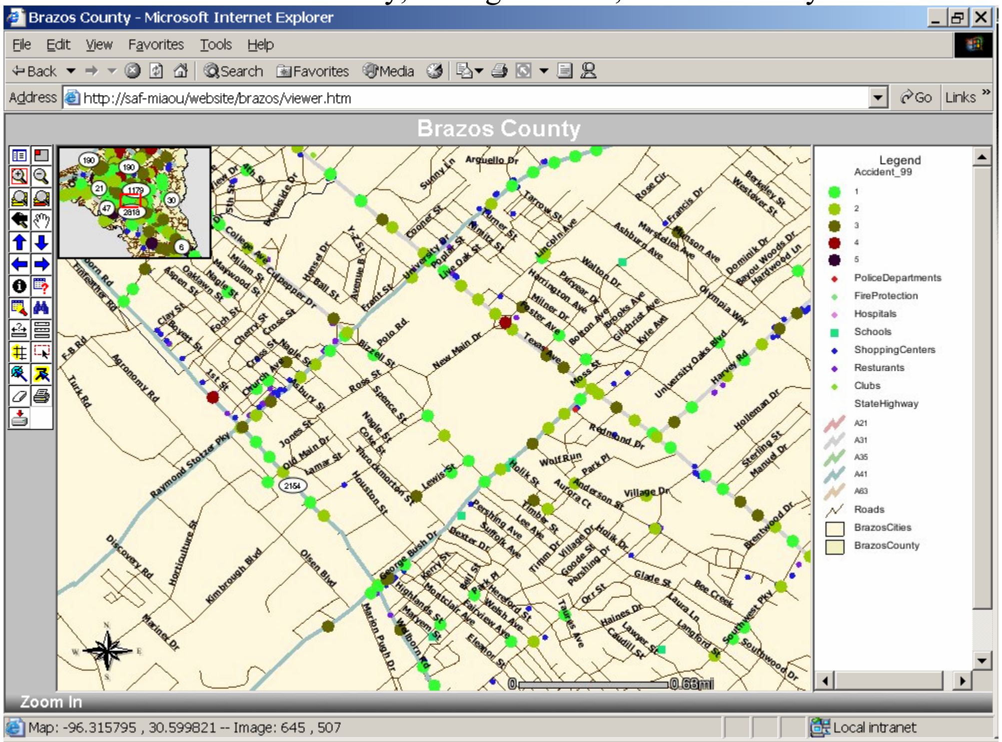
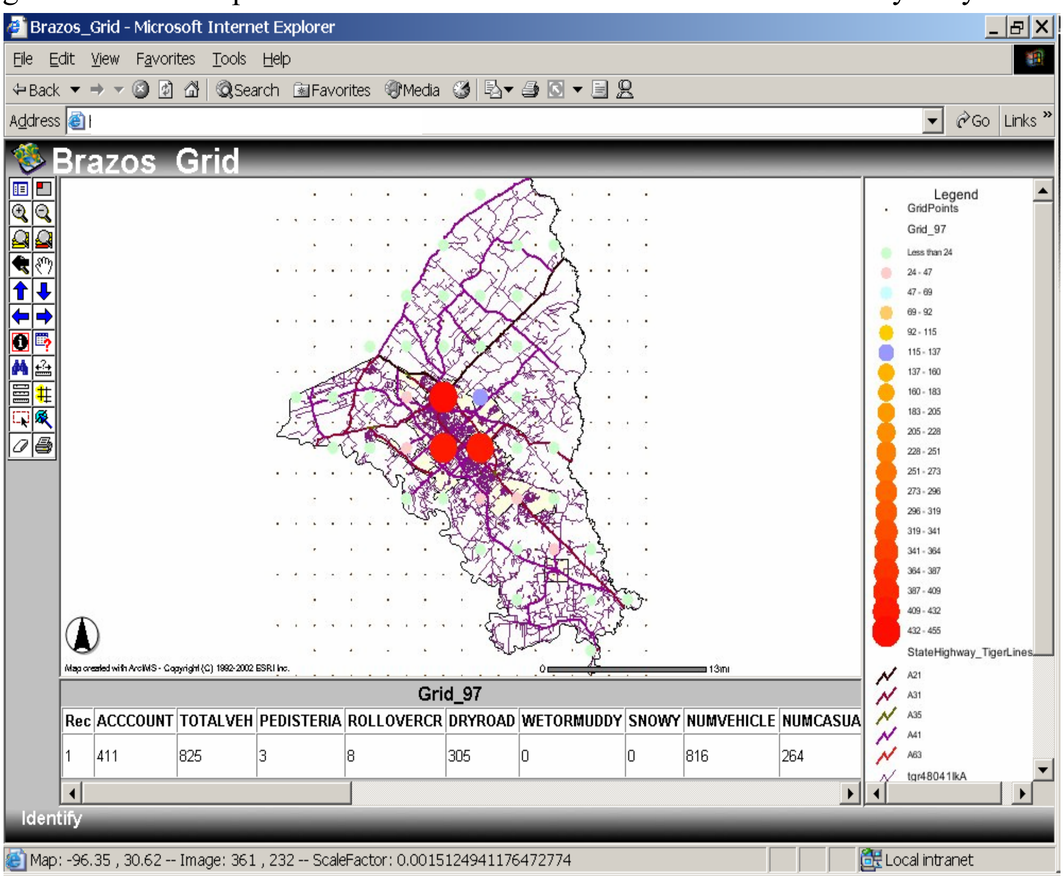
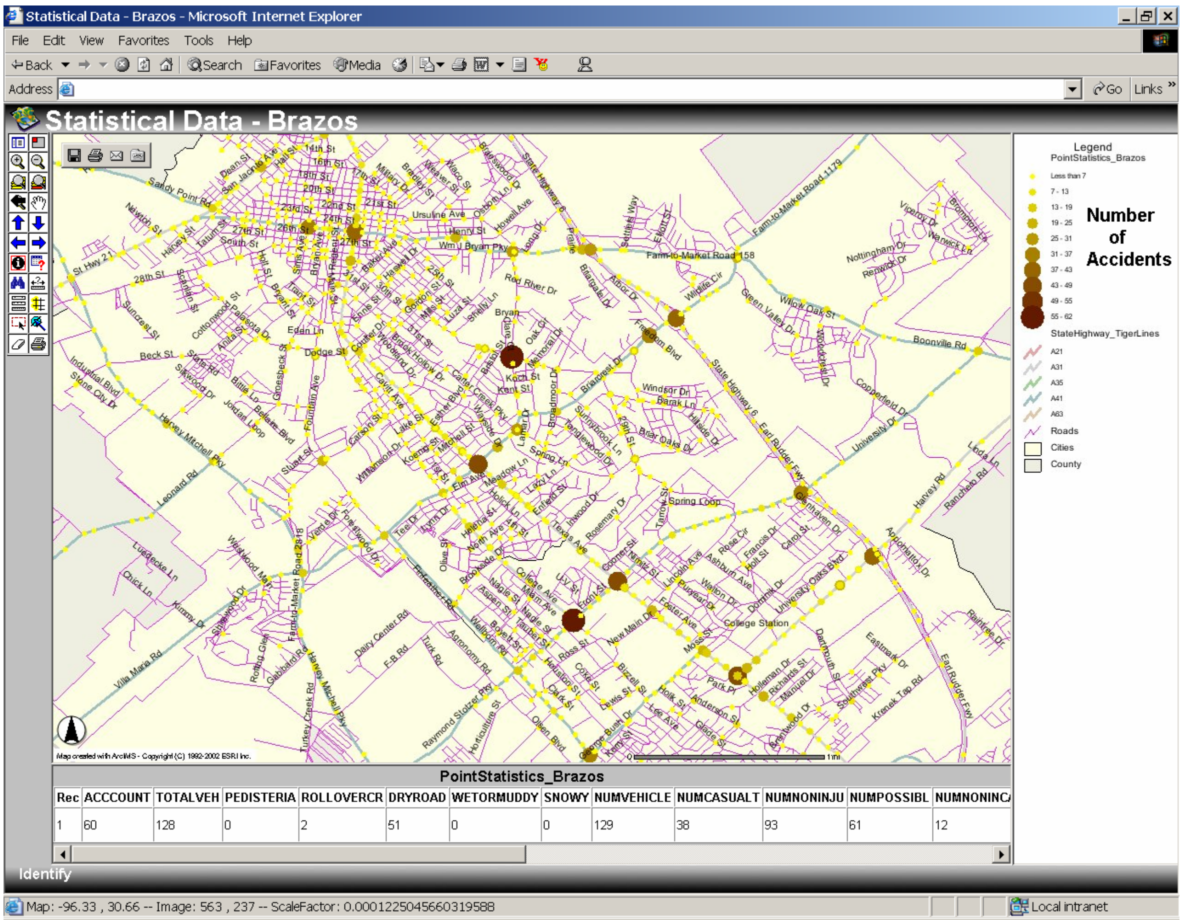


Figure 3.5 An example of cell-based total crash counts in Brazos County for year 1997



Unlike the cell-based analysis, distribution analysis generates descriptive information of certain features by location, e.g., number of crashes at intersections. Figure 2.8 shows the frequency of crashes at all sites. It helps to visualize crash-prone spots by using different symbols (in different colors and sizes) to represent the relative frequency of crashes.

Figure 3.6 Distribution analysis: number of crashes by sites



Spatial Query: Proximity and Network Analyses

The spatial query capability in ArcIMS gives users a means to obtain "quick answers" on some of the questions regarding the spatial distribution and characteristics of certain types of crashes. Typical analyses conducted under a spatial query include proximity analysis and network analysis. Proximity analysis searches features in all directions from a user-designated point (or spot), line, or area. Network analysis, on the other hand, is restricted to searching along a line or a path, such as a route, on the network.

Buffering is a means of performing spatial query to determine the proximity of neighboring features. Buffer selection allows features of specified layers that lie within some specified distance from pre-selected active features to be selected, highlighted, and listed. Figure 3.7 shows locations of several restaurants in a specified rectangular area. These restaurants serve as pre-selected active features. To select crashes occurred in 1999 within a buffer distance of 0.4 miles around these restaurants, Figure 3.8 shows the buffer layer selected as "Accident for year 1999," attributes of crashes in the buffer are listed at the bottom of the screen, and the buffer region is displayed.

Unlike proximity analysis that searches in all directions from a point, line, or area, network analysis is restricted to searching along a line or a path. This analysis allows a selected feature such as crashes that lie within a specified distance from a specified polygon shape. The polygon is specified by a set of line segments that may or may not form a polygon. On executing the request, the program first completes the polygon by joining the beginning point of the first and ending point of the last line segments. In Figure 3.9, the red line starts from the intersection of Wellborn Road and Raymond Stotzer Parkway and goes around Texas Avenue and ends at the intersection of George Bush Drive and Wellborn Road. On clicking "Complete Polygon & Select" the program completes the polygon by joining the beginning, intermediate, and ending intersections, and then selects the crashes that occurred within 1,000 feet of the polygon. Figure 3.10 shows the crashes in yellow, the attributes of which are listed at the bottom of the screen.

Other queries are possible using Query String. For example, Figure 3.11 shows a query to display all SUV-related crashes that occurred on wet surface conditions. On executing the query, Figure 3.12 displays all crashes meeting the condition in yellow and their attributes.

Figure 3.7 Proximity analysis: selecting restaurants (shown in yellow dots) in a specified rectangular area

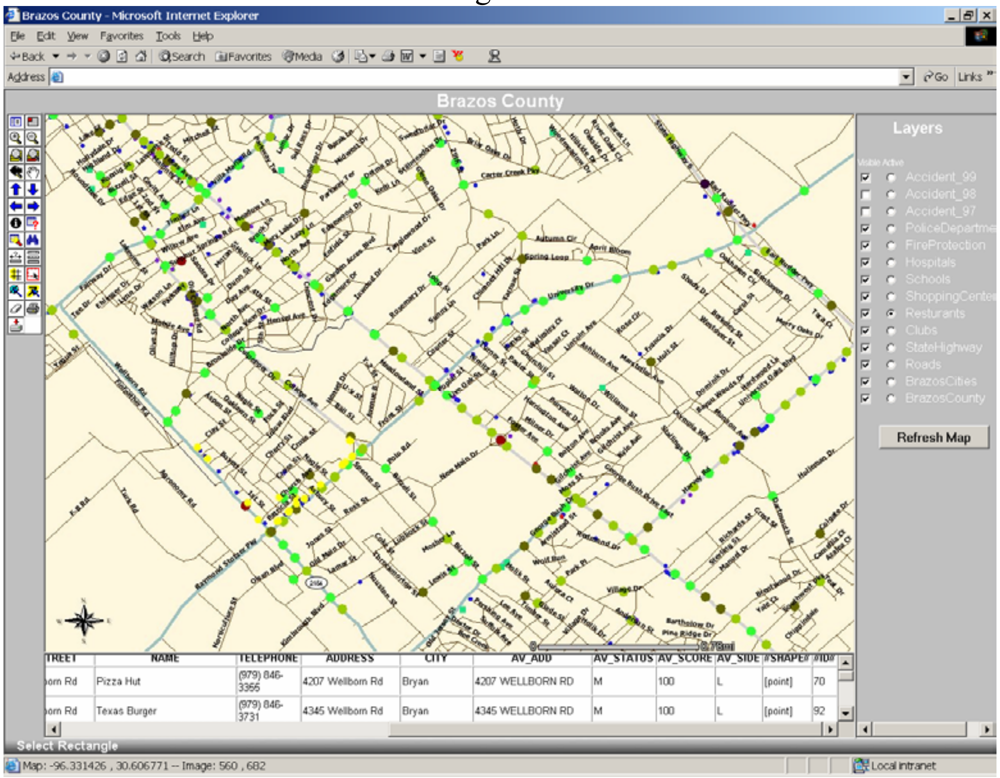


Figure 3.8 Proximity analysis: crashes within a buffer distance of 0.4 mi around pre-selected restaurants

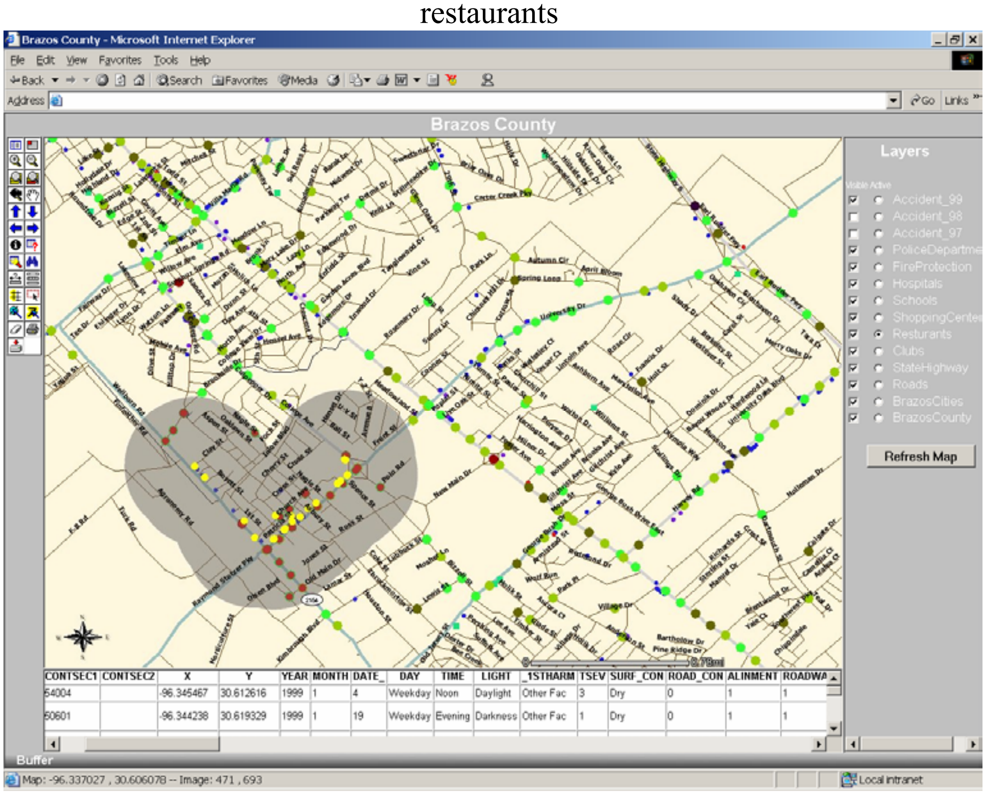


Figure 3.9 Network analysis: selecting a polygon

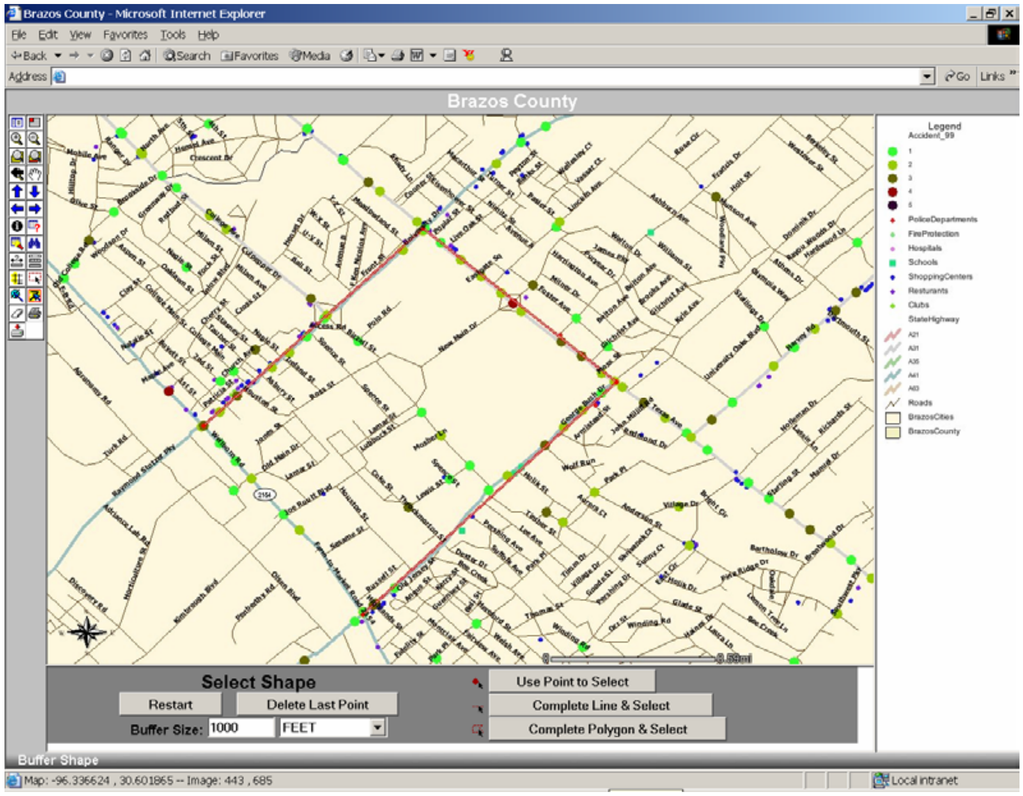


Figure 3.10 Network analysis: crashes occurred within 1,000 feet of the polygon—yellow dots

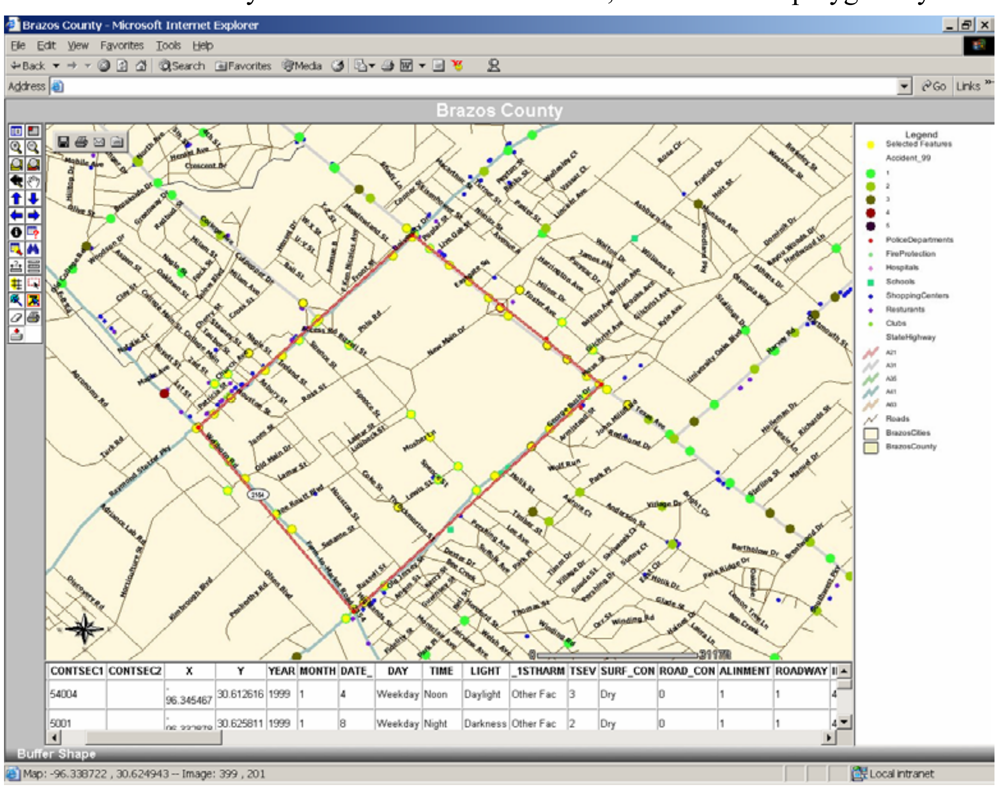


Figure 3.11 Specify a query using Query String

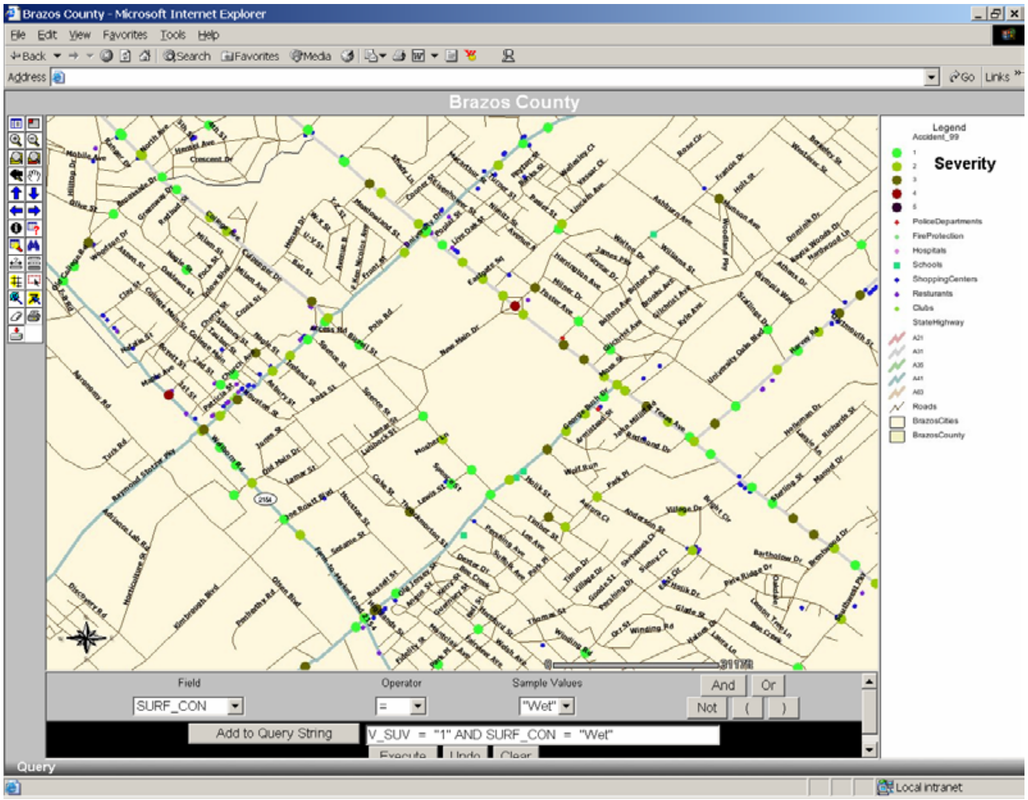
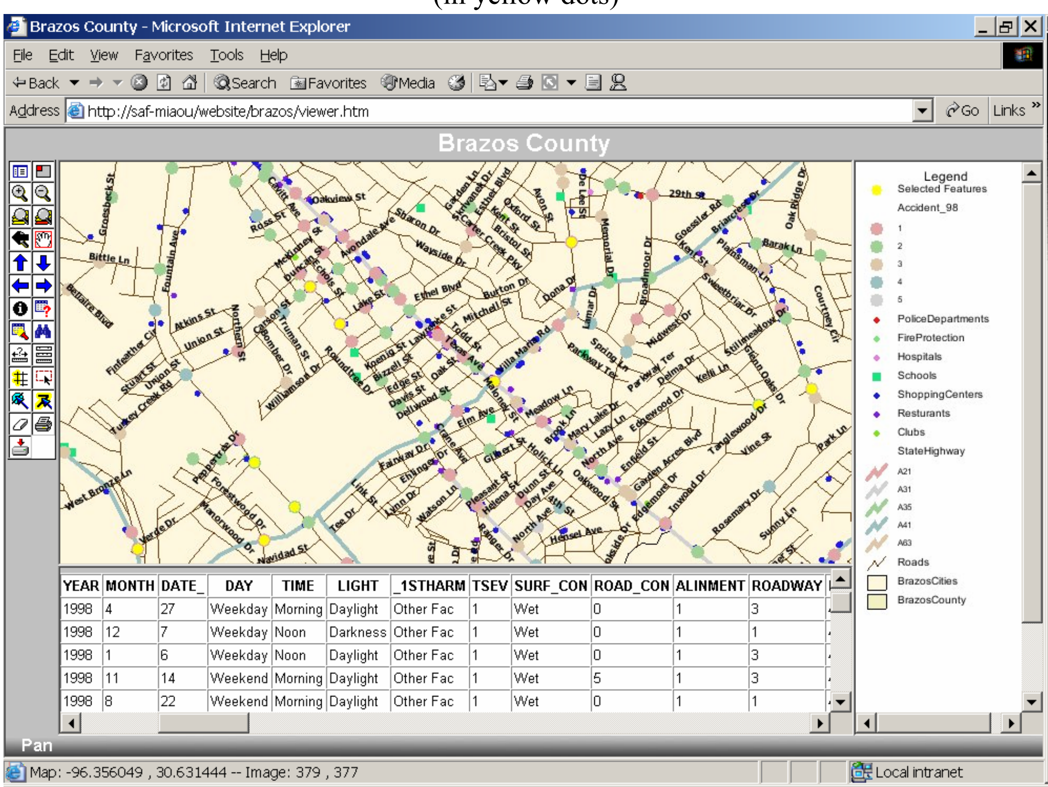


Figure 3.12 Example results of a query to show SUV-related crashes occurred on wet surface conditions (in yellow dots)



CHAPTER 4. FUTURE EXTENSIONS

Currently, the W-TSIAS prototype system has very limited statistical analysis capabilities. At the time of this study, more advanced spatial statistical modeling techniques and associated tools were the focus of a separate project supported by Bureau of Transportation Statistics (which was mentioned in Chapter 1). The development of these more advanced spatial statistical modeling techniques and their subsequent applications by the research team can be found in, e.g., Miaou et al., (2003), Miaou and Lord (2003), and Miaou and Song (2004).

As it stands, an off-line version of W-TSIAS is used in the pre-processing steps, such as data preparation, visualization, and model specification, and in the post-processing step of visualizing modeling results (e.g., showing model residuals for performance diagonstics). It is envisioned that enabling W-TSIAS to have a tighter coupling with these advanced spatial statistics tools for modeling and analysis will be a very fruitful extension of the system.

REFERENCES

1. Carrico, J.R. “Kentucky’s CRASH: An Enterprise Solution for Traffic Collision Records.” ***The Police Chief***, July 2002.
2. Levine, N., K.E. Kim, and L.H. Nitz. “Spatial Analysis of Honolulu Motor Vehicle Crashes: II. Zonal Generators.” ***Accident Analysis and Prevention***. 27(5): 675-685, 1995.
3. Miaou, S-P., and Lord, D. “Modeling Traffic Crash-Flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes Versus Empirical Bayes Methods.” Paper No. 03-3363, ***Transportation Research Record 1840***, Journal of Transportation Research Board, National Research Council, pp. 31-40, 2003.
4. Miaou, S-P., Song, J.J., and Mallick, B.K. “Roadway Traffic-Crash Mapping: A Space-Time Modeling Approach.” ***Journal of Transportation and Statistics***. Vol. 6, No. 1, pp. 33-57, Nov. 2003.
5. Miaou, S-P., and Song, J.J. “Bayesian Ranking of Sites for Engineering Safety Improvements: Decision Parameter, Treatability, Statistical Criterion, and Cost Function.” Paper No. 04-4963, Presented at ***Transportation Research Board’s 83rd Annual Meeting***, National Research Council, Jan. 2004.
6. Smith, R.C., D.L. Harkey, and B. Harris. ***Implementation of GIS-Based Highway Safety Analyses: Bridging the Gap***. FHWA-RD-01-039, Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2001.