Site-Based Video System Design and Development

DETAILS

96 pages | 8.5 x 11 | PAPERBACK
ISBN 978-0-309-12924-4 | DOI 10.17226/22836

AUTHORS

Gordon, Timothy; Bareket, Zevi; Kostyniuk, Lidia; Barnes, Michelle; Hagan, Michael; Kim, Zu; Cody, Delphine; Skabardonis, Alexander; an Vayda, Alan

The Second

STRATEGIC HIGHWAY RESEARCH PROGRAM

SHRP 2 REPORT S2-S09-RW-1

Site-Based Video System Design and Development

TIMOTHY GORDON
University of Michigan

ZEVI BAREKET, LIDIA KOSTYNIUK, MICHELLE BARNES, AND MICHAEL HAGAN
University of Michigan Transportation Research Institute

ZU KIM, DELPHINE CODY, AND ALEXANDER SKABARDONIS
California PATH

ALAN VAYDA
Soar Technology

TRANSPORTATION RESEARCH BOARD

WASHINGTON, D.C.
2012
www.TRB.org

Subscriber Categories

Design
Highways
Operations and Traffic Management
Safety and Human Factors

The Second Strategic Highway Research Program

America's highway system is critical to meeting the mobility and economic needs of local communities, regions, and the nation. Developments in research and technology—such as advanced materials, communications technology, new data collection technologies, and human factors science—offer a new opportunity to improve the safety and reliability of this important national resource. Breakthrough resolution of significant transportation problems, however, requires concentrated resources over a short time frame. Reflecting this need, the second Strategic Highway Research Program (SHRP 2) has an intense, large-scale focus, integrates multiple fields of research and technology, and is fundamentally different from the broad, mission-oriented, discipline-based research programs that have been the mainstay of the highway research industry for half a century.

The need for SHRP 2 was identified in *TRB Special Report 260: Strategic Highway Research: Saving Lives, Reducing Congestion, Improving Quality of Life,* published in 2001 and based on a study sponsored by Congress through the Transportation Equity Act for the 21st Century (TEA-21). SHRP 2, modeled after the first Strategic Highway Research Program, is a focused, time-constrained, management-driven program designed to complement existing highway research programs. SHRP 2 focuses on applied research in four areas: Safety, to prevent or reduce the severity of highway crashes by understanding driver behavior; Renewal, to address the aging infrastructure through rapid design and construction methods that cause minimal disruptions and produce lasting facilities; Reliability, to reduce congestion through incident reduction, management, response, and mitigation; and Capacity, to integrate mobility, economic, environmental, and community needs in the planning and designing of new transportation capacity.

SHRP 2 was authorized in August 2005 as part of the Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU). The program is managed by the Transportation Research Board (TRB) on behalf of the National Research Council (NRC). SHRP 2 is conducted under a memorandum of understanding among the American Association of State Highway and Transportation Officials (AASHTO), the Federal Highway Administration (FHWA), and the National Academy of Sciences, parent organization of TRB and NRC. The program provides for competitive, merit-based selection of research contractors; independent research project oversight; and dissemination of research results.

SHRP 2 Report S2-S09-RW-1

ISBN: 978-0-309-12924-4

Copyright Information

Note: SHRP 2 report numbers convey the program, focus area, project number, and publication format. Report numbers ending in "w" are published as web documents only.

Notice

The project that is the subject of this report was a part of the second Strategic Highway Research Program, conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for their special competencies and with regard for appropriate balance. The report was reviewed by the technical committee and accepted for publication according to procedures established and overseen by the Transportation Research Board and approved by the Governing Board of the National Research Council.

The opinions and conclusions expressed or implied in this report are those of the researchers who performed the research and are not necessarily those of the Transportation Research Board, the National Research Council, or the program sponsors.

The Transportation Research Board of the National Academies, the National Research Council, and the sponsors of the second Strategic Highway Research Program do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of the report.

SHRP 2 Reports

Available by subscription and through the TRB online bookstore:

www.TRB.org/bookstore

Contact the TRB Business Office:
202-334-3213

More information about SHRP 2:

www.TRB.org/SHRP2

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

SHRP 2 STAFF

ACKNOWLEDGMENTS

This work was sponsored by the Federal Highway Administration in cooperation with the American Association of State Highway and Transportation Officials. It was conducted in the second Strategic Highway Research Program (SHRP 2), which is administered by the Transportation Research Board of the National Academies. The project was managed by Walter Diewald, Senior Program Officer for SHRP 2 Safety.

The research reported on herein was performed by the University of Michigan Transportation Research Institute (UMTRI) in collaboration with California PATH and Soar Technology. Timothy Gordon was the project director and principal investigator. The authors acknowledge technical and logistical support from the Washtenaw County Road Commission, which owns and manages the intersection used in this study. The authors are grateful to the road commission for its help and support throughout this project.

FOREWORD

Walter Diewald, PhD, *SHRP 2 Senior Program Officer, Safety*

The goal of the SHRP 2 Safety program is to prevent or reduce the severity of highway crashes through more accurate knowledge of driver behavior and other factors. The Safety program's research is proceeding along two distinct but related tracks: (1) the in-vehicle, naturalistic driving study, which encompasses all types of driving, and (2) the site-based risk study, which focuses on vehicle trajectories at specific locations, such as intersections. This report describes the work that was done in the latter track to develop and test an on-site, video-based data collection system with the potential for widespread application by researchers and state and local authorities to examine intersection safety.

This report documents the development of a prototype system capable of capturing vehicle movements through intersections by using a site-based video imaging system. By tracking individual vehicles through an intersection, the Site Observer provides a basis not only for viewing crashes and near crashes but also for developing objective measures of intersection conflicts and collecting before-and-after data when design or operational changes are made at intersections. It also yields detailed and searchable data on the normal driving population so that exposure measures can be determined.

This research built on previous work on video-based systems to develop a system that is relatively inexpensive, portable, and flexible enough for installation at all types of intersections, as well as robust enough for use in locations with a wide range of environmental conditions. The system embraces modern machine vision cameras and draws from the large body of research on algorithms for extracting information from video streams, a key advantage where data must be collected continuously. It was tested at a location during fall and winter months and found to operate as designed. The Site Observer is a robust prototype system that is deployable as is but is also capable of further development and refinement for use in intersection safety assessment.

CONTENTS

Executive Summary

This report describes the design and development of the Site Observer, an automated site-based video system for capturing and analyzing vehicle trajectories for the purpose of highway safety research. Many highway safety problems currently are unsupported in terms of high-quality objective data that adequately describe real-world traffic conflicts. Although much effort has been devoted to instrumenting vehicles for naturalistic driving studies, the vehicle-based approach has several limitations, particularly for addressing site-specific safety questions. Some examples of limitations are (a) monitoring sites with specific geometries of interest, (b) evaluating the effects of site-specific countermeasures, and (c) addressing conflicts and crashes that result from the kinematics of several vehicles simultaneously (e.g., path crossing at intersections). In such cases, it is more efficient and effective to use fixed sensors and data acquisition strategically situated at a site of interest, provided they are capable of measuring the continuous positions, speeds, and accelerations of all relevant vehicles passing through the site. The purpose of the Site Observer system is to fill this technology gap and provide research data and actionable data to support future developments in areas such as geometric design, signal timing, road markings, and signage and systematically improve highway safety.

Although the site-based approach also may miss some variables, such as those associated with human factors (e.g., the status and individual actions of drivers and passengers), this limitation merely defines the site-based approach as being complementary to naturalistic driving studies and not a direct substitute. It also offers future opportunities for combining site-based vehicle trajectory measurement with vehicle-based naturalistic driving studies.

Much work has been done in the area of video-based tracking of moving objects, including the tracking of highway vehicles, but to date no system has been shown to be sufficiently accurate, flexible, and automated to be used as a routine tool for safety research. The Site Observer has been designed with this broad scope of safety research as its primary goal. This is in contrast to many research-level tracking systems, which typically make use of temporary single-camera installations and focus on algorithm development or in creating specific reference data sets. The progress in such research, reviewed in the main report, is important for advancing the science and technology of video-based vehicle tracking, but in the past such developments have not gone far enough to establish scalable architectures or robust hardware and real-time software. These kinds of development are crucial to enabling long-term development and routine implementation. Given the technical challenges associated with machine vision, the previous focus on algorithm development is understandable, and of course the supporting low-level video image processing algorithms are also important for the Site Observer. However, the philosophy of the current project has been to use state-of-the-art image processing as the starting point of the system design, albeit including some application-specific refinements; thus, the challenge has been to integrate the resulting information from multiple synchronized video streams and infer vehicle positions and velocities across time to provide the overall motion capture capability.

The Site Observer's capability is also in contrast to the many commercial video systems currently used for traffic management. One major difference is that the commercial systems are based almost exclusively on background subtraction (also a part of the Site Observer's operation) and make use of predefined image regions to determine occupancy—whether or not a vehicle is currently in those regions. This serves the important purpose of simplifying the vision/inference problem and effecting data reduction, but it removes too much critical information relevant to vehicle motion capture, and makes it very difficult to subsequently perform tracking. In addition, this typical approach requires well-defined camera locations relative to the highway, something that is feasible for traffic management but is not realistic for temporary site monitoring equipment that must be capable of dealing with low camera angles, oblique motion of vehicles, and relatively unstructured vehicle motions in and around intersections (where predictable lateral positioning within lanes cannot be assumed).

As mentioned, the Site Observer was designed to address key safety research questions posed by SHRP 2. A number of safety research topics are identified in this work, including those relating to path-crossing conflicts at intersections and the influence of highway factors on lane or road departures. But one crucial area that spans multiple crash types is in the relationship between conflicts and crashes and the potential to use conflict measures as surrogates for crash. Using validated conflict measures in place of actual crashes means that the influence of highway factors, countermeasures, and site improvements can be evaluated over relatively short monitoring periods, but the surrogates must faithfully represent crash type and crash risk. The Site Observer can be used for surrogate validation and for the use of surrogates in countermeasure evaluation, but validation should come first. To achieve this requires sampling the traffic flow in an unbiased way, so it is crucial that large volumes of vehicle motions can be captured with high fidelity so that conflict rates can be ascertained and then related to crash rates.

The focus of this report is on intersection conflicts and safety, mainly because intersections are sites of relatively high conflict rates and are also most challenging for the video tracking technology; however, the Site Observer can be applied to other sites of safety concern. Path-crossing conflicts are particularly challenging for the technology in terms of position, speed, and synchronization errors of different vehicle trajectories. Ideally the system should provide a high degree of positioning accuracy for the analysis of path-crossing conflicts; it has been shown that unbiased errors with root-mean-square (RMS) values on the order of 20 cm are needed at the conflict points (near the center of the intersection) and with negligible timing errors. The system performance was seen to have negligible synchronization errors (sub millisecond), but with positioning errors as far as 50 m from the intersection of approximately 40 cm RMS, which is in excess of the 20-cm target. In fact, it was not possible to fully and formally evaluate positioning errors because of the lack of a completely reliable, accurate, and independent benchmark measurement system. Manual review of sample trajectories showed some small sources of bias in lateral position because of shadows, but overall the RMS errors at the intersection center were of the same order of magnitude as in the stated requirement. Note that such levels of accuracy are completely impossible using radar technology. Worst-case errors occurred when only two cameras covered the vehicle movement within the intersection, whereas in most cases three cameras could simultaneously view at least part of the vehicle track, in which case positioning errors were low.

Among the broader range of design requirements, the most important is that the system should be fully automated. One of the most serious deficiencies in previous systems designed to track vehicles is that manual corrections are needed; operator intervention was used to correct for faults and guide the detection and tracking subsystems. Full automation may lead to less than 100% of all vehicles being tracked, but as long as errors are flagged, bad data excluded, and data loss shown to be unbiased, this has no impact on the usefulness of the system. It is possible that this consideration will be important when the Site Observer is used in locations with very dense traffic, but according to the pilot study in this project negligible data loss (less than 1%) of this type was experienced.

In addition to accuracy and automation, the third critical requirement for the system has been mentioned—the need for a flexible and expandable architecture, with relatively unrestricted camera positioning and the ability to add additional cameras for more complex or large sites of interest. Of course, many other practical considerations exist, and these are elaborated in the report. However, the system was found to be capable of running in a variety of weather and lighting conditions, including light snow and with low sun angles.

The design has been realized as a prototype system, which has been built and tested, and has captured representative data. It uses four machine vision cameras and is organized in a hierarchical way with early stage processing taking place on site, local to where each camera is installed. To ensure precise synchronization, camera shutters are triggered via pulses from local Global Positioning System (GPS) transceivers, and there is no need for the different cameras to communicate during operation, other than to send extracted image features to the site computer (ground station); this can happen at discrete time intervals, and although optical fiber links were used in this project, it is anticipated that wireless networks can be used in the future, reducing installation times.

In practice, camera positioning depends on the availability and access to building structures, lighting poles, and so forth, and although it may be necessary in some cases to install new mounting poles, it was considered a design requirement that only modest camera installation heights would be available. In this case, perspective changes in vehicle outline and the effects of road surface height changes were all considered as likely sources of error for vehicle tracking. Therefore, algorithms were implemented to include these effects in the 3-D mappings used when data from multiple cameras were combined. The processing method is organized into two distinct levels:

1. *Camera level.* Features are extracted in the 2-D camera frame from the image (pixel) data and grouped into clusters, in this case clusters of corner features. The grouping process is effective at creating long-lived tracks in the image frame, although a certain amount of dither is introduced as individual corner features are either lost or added. A second set of features is also captured, namely, the boundaries of the foreground regions (so-called "blobs") where candidate vehicles exist. These features are determined by background subtraction and fitted by convex polygons. Real-time data processing was developed for this task using an existing real-time software platform.
2. *Site level.* The recorded camera level features are projected into the 3-D space of the site. Although multiple cameras capture the same scene from different angles, there is no attempt to implement stereo vision techniques. Stereo vision requires cameras to have similar viewing angles, and the camera separation and calibration is crucial. The research team used a simpler technique (that requires fewer cameras) based on the fact that the foreground regions coincide with "same vehicle," whereas the cluster tracks do not. Fitting a 3-D candidate shape that is rectangular (in plan view) and aligned with the motion vector then allows a match to the intersecting blobs and provides a high level of vehicle localization, especially when there are three or more simultaneous camera views available. This requirement is for only one location on the site, and it was found to be true for most vehicle trajectories at the test intersection. Once localized in space, the absolute heights of the cluster tracks can be estimated and used as virtual markers to track the vehicles. There is no requirement for real-time processing for site-level data analysis.

This is the essence of the vehicle tracking system, and the result is a set of trajectories indexed by absolute GPS time. From the trajectory data, any kind of single-vehicle or multiple-vehicle conflict is found, such as path-crossing or turn-into-path conflicts. Intersection traffic signal states were captured simultaneously so that relevance to signal phase can be determined easily.

Hardware for the Site Observer comprised commercial off-the-shelf components based on a PC architecture. Custom circuit boards and assembly were part of the hardware system design, and dedicated cabinets were used together with temperature control for all-weather use. The system was installed at a suburban intersection in Ann Arbor, Michigan, capturing on the order of 17,000 vehicle trajectories over several sessions of daytime traffic monitoring. As previously mentioned, the vertical geometry of the site was important for carrying out the necessary geometric mappings from camera to site views. To assist this and the choice of camera locations, the project made use of a lidar survey. The current system uses four cameras and five computers (one per camera plus one to collect feature data and apply start/stop control). This choice was made in light of balancing hardware and installation costs against the need for full coverage of the site. The cameras were mounted on traffic signal mast arms at heights of approximately 6 m aboveground (i.e., a little less than 20 ft).

Limiting data volumes for storage and processing is also an important operational constraint. The data structures used for tracking exist at three stages, ranging from high volumes and low specific information content to low volumes and high information content. These are, from highest to lowest data volume:

1. *Camera input data.* Grayscale values are found at each pixel location, 1 byte per pixel (with 256 levels of gray). Although color cameras could have been used, they were not considered necessary and would increase the data volumes handled by the local camera computers.
2. *Camera output data.* Feature locations are stored in camera coordinates. These consist of 2-D positions and velocities for the clusters used for tracking, as well as the vertices of the polygons used for foreground blobs. Optionally compressed video can be exported for operator review.
3. *Site Observer output data.* This is the set of vehicle trajectories, augmented by estimates of vehicle length, height, and width.

The data volumes for the above were, for the system tested, of the order (1) 25 Mb per second, (2) 50 Kb per second (without compressed video), and (3) 1 Kb per second. This final value depends on traffic volumes, here assumed to be around 1,000 vehicles per hour passing the site (somewhat higher than the actual peak traffic flow rate at the trial intersection). Thus, the overall operation of the Site Observer can be viewed as squeezing meaningful trajectory information from vast volumes of raw video data via the two distinct processing levels. (Note that 25 Mb per second scales to roughly 2 Tb of image data during a 24-h period; by contrast the Site Observer output for 1 year of operation is a modest 0.3 Tb). As the number of cameras increases, the data volumes scale in proportion at stages 1 and 2, but so does the processing power because there is one image processor per camera station. For stage 3, there is no corresponding increase in data volumes, just an increase in positioning accuracy. Again the scalability of the system is clear.

In this project, a feature database was constructed, trajectories were extracted, and sample conflict analyses were undertaken for left-turn-across-path (opposite direction) and right-turn-into-path configurations.

The Site Observer is a robust prototype system that, while capable of further development and refinement, exists as a deployable system in its current state. The natural follow-on from this work is to deploy the system at a site of particular safety interest, where it is possible to test the extent to which configurations and relative frequencies of crashes match to the same patterns of conflicts. A second follow-on is to implement the system at a location where significant numbers of instrumented vehicles pass through, which would include the influence of driver factors on conflict measures obtained from the Site Observer.

CHAPTER 1

Introduction

The aim of this project was to develop and validate a new observation tool for vehicle safety research: an automated video tracking system. The system is to be capable of wide-scale use and capture detailed vehicle trajectories in traffic environments where safety performance of the highway is of particular interest. The captured trajectories can be used to find traffic conflicts and compute relevant metrics, such as gap times and distances between vehicles. The motivation for the study is the low fidelity and low frequency of historical crash data. Researchers have little objective information about vehicle speeds and positioning and timing just prior to a crash, and limited information about the contributing roles of driver, environment, highway design, and especially surrounding traffic when crashes occur. The hypothesis is that essential new information can be supplied by captured trajectories for conflicts and near crashes via objective measures of conflict, thus avoiding excessive delays waiting to gather sparse data on actual crashes. To assess risk factors, the tracking system is also to gather associated exposure data. Therefore, the challenge is to design and test an automated video tracking system that captures most or all vehicle trajectories at a particular site; this must be done with sufficient fidelity to compute conflict metrics (as well as related factor variables such as speed and traffic density). The current study includes a small field trial designed to establish the validity of the video tracking system as a viable research tool for safety research.

The concept design and system development are to be guided by the research agenda of the SHRP 2. Although the SHRP 2 Safety program is centered on a large-scale field study using instrumented vehicles, it includes the current thrust to develop a robust prototype system to capture vehicle motions from site-based video image processing. Vehicle and site-based approaches have their own strengths and weaknesses. In-vehicle studies provide high-quality data about the driver and the vehicle but only limited access to information about the traffic, highway, and environment. For incidents or events of interest, it is possible to supplement the vehicle-based data collection with static information, such as highway geometry; but for dynamic information, especially the kinematics of other vehicles, the approach is limited. By contrast, the site-based approach is directly focused on the dynamic traffic environment, and it is much easier to supplement this with detailed information about the local geometry, signage, and traffic signals; thus, it can capture interactions and conflict information that are unavailable to an in-vehicle study. In the current study, in which trajectories were captured at a signalized intersection, traffic signal states were recorded in parallel to the vehicle motions. Of course, the human factors information is at best limited for site-based systems.

The broad theme of the SHRP 2 Safety program is to identify risk factors, develop and validate surrogates for crash, and enable methods and data collection to answer a variety of research questions relating to highway safety. Particular emphasis is placed on intersection and road departure crashes. Given the strengths and weaknesses of the in-vehicle and site-based approaches, the current project is focused more on the intersection safety problem, where conflicts are spatially localized and interactions between vehicles dominate the safety problem. Intersection vehicle tracking is also the more technically challenging problem for video-based tracking, so the current project focuses directly on intersection safety.

Objective measures of intersection conflict, such as gap timing for path-crossing conflicts, are a particular focus of the video-based system. According to the SHRP 2 priorities, these metrics are to be used as surrogates for actual crashes, and statistical techniques will validate the approach by relating the patterns and influences of highway and other factors between conflicts and crashes. These aspects are to be the subject of future studies, in which larger data sets will be captured and analyzed. For this project, such aims help define the system requirements and evaluation criteria in terms of accuracy, level of automation, reliability, and system availability. This is the central aim of the system development, so that future deployment is enabled on a wider scale: crash frequencies and types can be associated with corresponding conflicts, and ultimately design improvements and other interventions can be implemented to systematically reduce both conflicts and crashes.

Accurate and searchable intersection trajectory data not only are required in the recording of events (for example, crashes, near crashes, and conflicts that require avoidance maneuvers), but also are important for recording accurate trajectories for nonevents to characterize the baseline flow of traffic, and this places strong demands on the video system. This is because risk analysis requires *detailed and searchable* data on the normal driving population to determine denominator or exposure measures. In the context of conflict metrics, it is far more powerful to have detailed trajectories for all vehicles than, for example, to use manual techniques to refine trajectories in the case of specific events.

Because of the complex nature of the data collection, which deals with multiple interacting vehicles, the video camera appears to be the most feasible sensor, and although this is a basic assumption of the research project, it is worth making a brief comparison with existing alternatives. On the positive side, video cameras are commonly installed on highways and are relatively inexpensive and portable, even high-precision machine vision cameras. The system development builds on earlier work, which includes research and development carried out previously by the S09 research team. There is also a relatively mature commercial technology for machine vision cameras and a large body of research on algorithms for extracting information from video streams. On the negative side, the video image provides only a projective 2-D rendering of the 3-D world; the video image has no sense of depth. Video images also suffer from large data volumes, image complexity, optical distortion, susceptibility to occlusions and stray light reflections, and basic lighting and weather variations. All of these factors provide challenges to the development of an automated video-based tracking system.

Alternative sensors capture depth information but with other limitations. The most common alternative sensor used for vehicle detection and tracking is 24 GHz or 77 GHz radar. These devices scan horizontally and lack vertical resolution, so the mounting position ideally is just above ground height, similar to where such sensors typically are mounted on vehicles (i.e., at bumper height). This increases occlusion problems, where a nearby vehicle blocks the reflection of a more distant vehicle. Radar also has limited angular resolution (at around 0.5° this is poor by video image standards), which is made worse by changes in reflection point location on a detected vehicle. Most important, radar offers a limited field of view, typically in the range 10–15°, which is wholly inadequate for covering a large intersection without using a very large number of radar systems.

Lidar systems use scanning laser light instead of microwaves to perform the ranging, and these provide a more serious option for vehicle tracking but have their limitations. Lidar systems are significantly more expensive than cameras, so it is less reasonable to consider such devices for long-term installation; it is hard to conceive of multiple laser scanners being left unattended at an intersection for periods of several weeks or months. As is the case with radar, lidar systems lack vertical resolution and are best mounted close to the ground, again increasing problems of occlusion. Angular resolution is better than that for radar, and the field of view generally is better.

Integrating video and radar is also a feasible option, but it adds to system complexity and in no way removes the need for video image processing. Likewise, other sensors such as loop detectors could provide data to support triggering and such in the video system, but in the case of loop detectors, there appears to be little advantage because using video image processing to provide virtual loops is already an established approach, and the additional issues of sensor latency and dependency on installed hardware and interfaces make it hard to justify. In fact, one of the strong points about the video image data is that triggering an electronic camera shutter can be accurately controlled to the point that sensor latency is of minimal concern.

For intersection safety, an additional sensor or data requirement is to acquire traffic signal states at the same time vehicle trajectories are recorded. This could be a digital input from suitably equipped traffic signal controllers or, as was used in the present project, using an analogue sensor to record activity on the signal load circuits. This was conveniently achieved using optical sensing of the LEDs on the load switches in the traffic signal cabinet, which has the advantage of excluding electrical connection between the monitoring system and the traffic signal circuits.

In this project, an operationally efficient option has been developed, namely, to *integrate video with video*. This means to use synchronized video streams from multiple digital cameras mounted at an intersection and combine the results from these video streams to reconstruct vehicle trajectories. The design concept includes the architecture of the system, whereby individual video streams are processed in parallel, based on feature extraction; data fusion takes place with extracted features, not the raw images, so there is no need to store or transmit large volumes of video data, and the design has been made with the intention of being completely scalable to systems with larger numbers of cameras and with all image processing taking place on site. The fact that image processing is automated is also essential to the scalability of the system; it is a basic assumption of the system design that no manual intervention is allowed in the image and feature analysis software, and manual review is used only to assess performance and quality of results.

In terms of tracking performance, the most critical step is to uniquely isolate vehicles from each other and from the background; this is the main purpose of the multiple camera approach, using consistency between features seen from the different camera perspectives. The multicamera view is well suited to this because it provides multiple cases of the same vehicle kinematics and offers the best opportunity to minimize the effects of occlusions. Also important is that stray features caused by glare and reflections cannot pass a basic validity check of feasible vehicle motions and may be excluded automatically. The system design concept, based on fusion of features from different camera perspectives, is inherently robust.

Camera locations are another important motivator for the multicamera approach. Because a stable and very high camera position normally is not possible, vehicle detection and tracking must be done while considering the full 3-D geometry of the intersection space; vehicle height and perspective effects must be accounted for, and this is one of the challenges for the postprocessing of extracted features.

The resulting site-based video system is referred to as the SHRP 2 Site Observer. In this report, it is applied to trajectory capture and conflict metric analysis at one particular intersection.

The pilot study collected data at a four-way signalized intersection in Ann Arbor, Michigan. Installation was carried out in fall 2009, and data collection took place between January and March 2010. During this time, weather conditions were variable and included snow cover and bright sunshine. Data collection was sufficiently extensive to enable a sample analysis of conflict metrics presented in this report. However, the main purpose was proof of concept and validation of performance and not to create a data set for extensive analysis of crashes and conflicts.

This report is structured as follows. Research questions relating to SHRP 2 Safety are reviewed in Chapter 2, and the applicability of the site-based video system to address such questions is considered. Chapter 3 presents a survey of previous systems developed for vehicle tracking using video camera technology, including commercial and research systems. As part of this survey the major challenges of video-based tracking are considered, in particular the technical challenges that stand in the way of a robust automated vehicle tracking system. Chapter 4 considers a broad range of system requirements for the Site Observer, while Chapter 5 provides more detail on the conflict metrics proposed for use in the analysis of surrogates for vehicle crash. The ideal accuracy requirements for the system are suggested by simulations, which are therefore appended to the general requirements. In Chapter 6 the full Site Observer design concept is proposed and in Chapter 7 the intersection site chosen for the study and the system installation, including camera calibration, are described. Chapter 8 sets out the main image processing techniques used in the project, including the features extracted and stored in the single-camera subsystem. Chapter 9 describes the critical steps for data fusion, whereby vehicle trajectories are extracted from the feature databases and vehicles are localized in the 3-D world. Trajectory refinement and estimation of vehicle velocities and accelerations are described in Chapter 10, and in Chapter 11, a number of validation tests are described and results presented. Chapter 12 offers a sample analysis of conflicts and Chapter 13 concludes the report with a summary and conclusions from the study.

CHAPTER 2

Safety Research Questions

The Site Observer design and development ultimately are tied to safety research questions. Unlike vehicle-based naturalistic driving studies, the Site Observer is particularly capable of capturing objective data from a variety of vehicle-to-vehicle conflicts, and because the most technically challenging scenarios relate to *intersection safety,* this application is the main focus of the study. Some consideration is also given to road departure safety because this is also a major theme of the SHRP 2 Safety program. However, the topic of road departure safety is a more natural one for vehicle-based data collection, so the need for the Site Observer to address this is less clear. In the future, other safety areas, such as lane change and merging conflicts, may prove amenable to analysis using automated video tracking, but these are not considered in the current study.

Intersection Safety

The ability to predict or model the occurrence of crashes at intersections has been a challenge to transportation engineers since the early days of motorized transportation, and there have been many empirical and theoretical efforts to model intersection crash occurrence. Most studies have used traffic volumes as exposure measures and either implicitly or explicitly incorporated risk into the models. The configuration of the intersection, the number of different types of turns that can be made, the type of traffic control, and driver compliance with the traffic control determine the encounters in which various types of crashes can occur.

Early studies of intersection crashes found that the intersection crash rate per volume of traffic was sensitive to changes in the proportion of traffic flow from the various legs of the intersection. An early and widely known study by Tanner (1953) using crash data from rural three-leg intersections found that the frequency of collisions between vehicles turning around either shoulder was approximately proportional to the square root of the product of the traffic volumes on the main road and around either shoulder. Other early studies of intersection crashes (McDonald 1953; Raff 1953; Webb 1955) indicated that an increase in traffic on the major facility has a small effect on the crash rate, whereas an increase in traffic volume or an increase in the percent of traffic from the minor facility results in a rapid increase in the crash rate.

Many studies have explored the relationship between crashes and descriptions of the features of intersections. For example, Hannah, Flynn, and Webb (1976) examined the relationship between crashes and characteristics of intersections in rural municipalities in Virginia. David and Norman (1975) analyzed the 3-year crash history of 558 intersections in northern California. Categorical analysis methods were used, and results indicated that sight distance obstruction, street signs, use of left-turn storage lanes, use of raised marker delineation, presence of bus loading zones, and multiphase signalization all affected crash rates.

Still other studies concentrated on developing statistical models of the relationship between traffic crashes and geometric features. Bauer and Harwood (1996, 2000) developed statistical models incorporating the effect of traffic control features and traffic volumes on intersection crashes, and statistical relationships between intersection crashes and characteristics (Bauer and Harwood 1996, 2000; Vogt 1999; Harwood et al. 2002; Washington et al. 2005) are the basis for the predictive model for intersection crashes in the Interactive Highway Safety Design Module (IHSDM) (FHWA 2006). IHSDM is a suite of software analysis tools developed by the FHWA to evaluate safety and operational effects of geometric design decisions during the design process.

Driver factors also influence crash risk. Information on drivers in state crash databases normally is limited; however, analyses have identified some patterns of driver-related factors in intersection crashes. Of particular note is the consistent finding that older drivers are overrepresented in intersection crashes (e.g., Insurance Institute for Highway Safety 2007). Driver distraction has also been found to be a factor in intersection crashes; for example, Eby and Kostyniuk (2004) report that drivers in approximately 30% of intersection crashes were considered to be distracted or inattentive to the driving task.

Surrogates for Intersection Crashes and Countermeasure Evaluation

Although statistical models of intersection crashes may be determined from crash data, they can hardly address detailed questions of whether particular conflicts carry more or less risk or how specific countermeasures, such as changes in signal timing, may affect driver behavior or crash risk. Relating crash risk to particular conflict types requires detailed crash and exposure data and detailed measurements of vehicle kinematics. In Chapter 5, intersection conflict measures are reviewed, but defined broadly, they are based on gap timing, range, and speed information that is not readily obtained using current technologies. The Site Observer opens up the possibility of capturing and evaluating conflict measures in great detail and thus relating their statistical patterns to those of actual crashes. This remains hypothetical, but once research can establish common patterns, metrics can be used as surrogates for crash. For example, to evaluate a particular countermeasure, before-and-after measures of conflict can be compared. What is crucial is that this can be expected to be achieved with sufficient statistical power in a few weeks or months, rather than after a wait of several years after crash numbers are reported. Arguably, such analysis may have immediate benefit; if the number and/or severity of traffic conflicts can be reduced using a countermeasure, it is reasonable to declare a safety benefit, even if the precise relationship to associated crash numbers is unknown.

Research Questions for Intersection Safety

Many safety research questions are relevant to the Site Observer, although all are centered on conflict metrics and other measures of risk and safety; there is no expectation that the observer will be installed in sufficient locations or for sufficient time to allow detailed analysis of large numbers of crashes. However, metrics derived from vehicle trajectory data will allow in-depth analysis of conflicts and provide an opportunity to address the basic questions about the relationship between conflicts and crashes.

Below is a broad indicator of the kinds of research questions that potentially can be addressed by the Site Observer. Other questions relate to unsignalized intersections, including roundabouts; Table 2.1 provides an additional summary, together with assessment of the likely applicability of the Site Observer.

One specific point worth mentioning is the possible use of the system to track pedestrians and bicyclists. The Site Observer was not conceived for this task, but it has the potential to do so. As can be seen in Chapter 8, the imaging system tracks pedestrians, even though these tracks are rejected as noise in the current vehicle tracking implementation.

Conflicts and Crashes at Intersections

- What are the differences in conflicts that result in crashes and those that do not?
- How do multiple drivers manage to resolve conflicts without crashing?
- What are the effects of volume, weather, and time of day on conflicts?
- Can distributions of conflict metrics at an intersection be directly related to safety and operation?
- What are the common factor dependencies of conflicts and crashes?

Signalized Intersections

Compliance with and Violations of Right-of-Way: Red Light Running

- Does the proportion of motorists violating red lights increase with the complexity of the intersection?
- What proportion of motorists violates the right-of-way rules by running a red light?
- What proportion of motorists who run red lights do so at the beginning of the red phase as compared to well into the red phase? This former indicates a deliberate action, whereas the latter probably is a nondeliberate action.
- How does the proportion of red light running affect crash risk?
- Does the proportion of motorists running red lights at an intersection vary with:
  - Approach speeds?
  - Approach traffic volumes?
  - Length of green phase of the signal?
  - Level of service of the approach?
  - Level of service of the intersection?
- Are there any changes in the signal phasing and timing that would reduce right-of-way violations at intersections with incidents of red light running?

Signalized Intersections: Right Turns

- What is the distribution of accepted gaps for right turns with permitted right on red? How does this vary by road type and traffic volume?

Table 2.1. Summary of Research Questions for Intersection Safety

Research Questions	Data to Address the Questions: Data from S09	Data to Address the Questions: Other Data	Methods to Devise Countermeasures	Evaluate Countermeasures
Compliance/violations of right-of-way (signalized intersections): Red light running (RLR) • What is the incidence of RLR? • What are the crash risks of RLR? • How can RLR be reduced?	Vehicle trajectories, kinematics of conflicting vehicles		Compare number and rates of incidents and crash risk across different road types, regions, traffic volumes, approach speed, levels of service	1. Change signal timing 2. Put in red light cameras (enforcement)
Compliance/violations of right-of-way (signalized intersections): Right turns on red (RTOR) • What is the incidence of RTOR when prohibited? • What is the incidence of not yielding to pedestrians? • What are the crash risks for RTOR compared with the benefits of increased throughput?	Signal phasing and timing		Compare number and rates of incidents and crash risk across different road types, regions, traffic volumes, approach speed, levels of service Calculate the decrease in delay from RTOR; compare monetary costs of crashes against costs of delay	1. Signage and signal for right lane 2. Enforcement cameras
Costs and benefits of left-turn lanes and phases • What are the crash risks of unprotected left-turn phases? • How much do dedicated left-turn lanes decrease crash risk?			Compare crash risks of unprotected left turns (and dedicated left-turn lanes) by road types, region, approach volumes, and levels of service	Change signal phasing to protected left-turn phase
Roundabouts • What is the incidence of right-of-way violations in roundabouts? • What are the crash risks of roundabouts? • What is the effect of roundabouts on pedestrian crashes? • What are the benefits of roundabouts in terms of safety and reduction of delay?	Vehicle trajectories, kinematics of conflicting vehicles and/or pedestrians	Calculation of delay, economic costs of delay	Compare incidence of row violation by region, approach volumes; calculate crash risk by region, approach volumes Compare to comparable crash risk and delay for intersections	
Gap acceptance (unsignalized intersection) • What is the distribution of gaps for vehicles on minor legs to unsignalized intersections? • What are the crash risks associated with those distributions? • Are there ways to decrease the crash risks for minor road vehicles (for example, increase sight distance standards)?			Calculate crash risk by approach volumes, region, and day-and-night conditions	Evaluate increases in sight distance
Access points near intersections • What is the crash risk of access points within 200 feet of intersections? • How far should the nearest access point be from an intersection so that it will not increase crash risk at the intersection?	Vehicle trajectories, kinematics of conflicting vehicles or pedestrians		Compare crash risk across different road types, regions, traffic volumes, and approach speeds	Evaluate closing off access points

(continued on next page)

Table 2.1. Summary of Research Questions for Intersection Safety (continued)

Research Questions	Data to Address the Questions		Methods to Devise Countermeasures	Evaluate Countermeasures
	Data from S09	Other Data		
Compliance/violations of right-of-way (unsignalized intersections): stop and yield controls • What proportion of vehicles on stop-controlled approaches do not stop? • Does this vary by region or weather? • What is the crash risk posed by this behavior? • What proportion of vehicles on yield-controlled approaches do not yield? • Does this vary by region or weather? • What is the crash risk posed by this behavior? • Can this behavior be reduced?			Compare incidence of violations and crash risk across different road types, regions, traffic volumes, and approach speeds	Evaluate use of enforcement cameras

- What proportion of drivers make right-on-red turns when such turns are prohibited? How does this vary by road type, sign location, and approach traffic volumes?
- How often is the right-of-way of a pedestrian violated by drivers making a right-on-red turn? What are the risks of pedestrian crashes from right turns at intersections? With permitted right on red? With prohibited right on red? By road type or traffic volume?

Signalized Intersections: Left Turns

- What is the distribution of gap acceptance on approaches with unprotected left turns? How does it vary with volume, presence of a left-turn bay, and complexity of intersection?
- What is the crash risk of unprotected left-turn phases?
- What are the safety benefits of dedicated left-turn lanes?
- What are the safety benefits of dedicated left-turn lanes with protected left-turn phases?
- What is the proportion of drivers turning without making a stop?

Access Points Near Intersections

- What is the effect on crash risk of commercial access points within 200 feet of intersections?
- How does crash risk vary with approach volumes and access point volumes?

Road Departure Safety

Because road departure safety is one of the key research areas for the SHRP 2 Safety program, it is worth considering the extent to which the system can contribute to this research. Single-vehicle road departure crashes include crashes with roadside objects, rollovers, and collisions with other vehicles if the vehicle first ran off the road then reentered and collided with another vehicle. Selecting a site with known road departure crash problems for a study may present technical, ethical, and legal problems. From the technical standpoint, the question is one of camera placement; crashes and conflicts can be distributed over considerable distances, so where is the best location? On the other hand, if a site were selected because of a serious safety record and a crash occurred during the study, the road commission or state DOT probably would face legal suits for knowing that this was a dangerous location and not fixing it. Experimental methods associated with the deployment of the Site Observer should take this into consideration, and selection of study sites most likely should be random from among typical examples of relevant road segments.

The research team considered the kinds of research questions that might be addressed via the Site Observer. The system may capture incidents of road departures, including events in which the vehicles almost ran off the road, ran off the road and stopped, ran off the road but recovered and either continued on the way or crashed into another vehicle, or ran off the road and crashed or overturned. Actual crash events will be rare, and even the events in which a vehicle runs off the road with no harm may not be sufficiently frequent to be captured in useful numbers. This again argues for the development and use of surrogates for running off the road events as an efficient and cost-effective way of studying the safety problem.

In a recent SHRP 2 study (Gordon et al. forthcoming), it was hypothesized that vehicle road departure crashes occur only under conditions of disturbed control, for which disturbed control is an interruption or delay in the process of

Table 2.2. Summary of Research Questions for Road Departure Safety

Research Questions	Data to Address the Questions: Data from S09	Data to Address the Questions: Other Data	Methods to Devise Countermeasures	Evaluate Countermeasures
• Can episodes of disturbed control be identified from vehicle trajectory data? • Can these episodes be used as surrogates for road departure events?	Vehicle trajectories, kinematics of conflicting vehicles	Vehicle-based data	Check if there are identifiers in vehicle trajectories that correspond to metrics of disturbed control from in-vehicle data	If disturbed control can be identified, it can be used in evaluations as a road departure surrogate
• What is the relationship between road departure events and road departure crashes?		Crash data for road segment	Compare rate of road departure events (or surrogates) and crash rate on segment	
Effects of changing conditions at one site • Effect of traffic on road departure events, crash risk? – Opposing lane, same lane? – Different mixes of vehicle types (trucks and passenger cars)? • Effect of weather on road departure events, crash risk? • Effect of light conditions on road departure crashes? • Effect of work zones on road departure crashes?	Vehicle trajectories, kinematics of conflicting vehicles	Monitor traffic volumes, weather, light conditions, and work zones	Compare incidence of road departure events and crash risk between the levels of conditions	
Effects of roadway features across sites • What is the effect of isolated horizontal curves on road departure events compared with that of a series of horizontal curves? • What is the effect of shoulder width on road departure events? • What is the effect of rumble strips on road departure events? • What is the effect of edge line markings on road departure events? • How does risk of road departure event vary with superelevation? • How does the presence of spiral transition affect risk of road departure event? • How does risk of road departure event vary with shoulder width, shoulder type?			Compare incidence of road departure events, and crash risk between categories of roadway features	
• How do changes in roadway features change incidence of road departure events and crash risk? • What is the effect of each of the following: – Changes in lane width? – Introduction of rumble strip? – Introduction of rumble strip to center line? – Change in shoulder width? – Change in shoulder type? – Change in pavement markings? – Change in advisory signs? – Change in speed limit?			Design experiments using before-and-after designs with controls or matched sites and conduct comparison analysis	

perception, recognition, judgment/decision, or action in the driving task. It was also hypothesized that crash surrogates for road departure crashes exist and are a combination of objective measures of disturbed control and highway geometric factors and off-highway environmental factors.

The question of development and validation of road departure surrogates was addressed in that study; a number of candidate surrogates were proposed, and some level of validation was carried out. One of the simpler candidate surrogates was time to edge crossing, the predicted time before a vehicle will leave the road (including passing across the shoulder) assuming it maintains its current path. In principle, the Site Observer can provide such surrogate information, although the accuracy in lateral position and velocity estimation is likely to be very demanding. By comparison, an in-vehicle study that makes use of a lane tracker has more direct access to vehicle kinematics and driver actions (especially steering wheel movement). One area of application that is more likely to favor the site-based observer is for before-and-after study of countermeasures, such as when rumble strips are installed or lane marker clarity is improved. Conversely, when driver factors and disturbed control are concerned, it seems unlikely that the site-based system can compete with the in-vehicle recording approach.

Table 2.2 summarizes a number of relevant research questions for road departure crashes. It seems clear that the site-based system most naturally complements what is possible with an in-vehicle approach in conditions in which it is hard to capture highway and environmental factors from the vehicle or from spatial databases that can be linked to the driving data. For example, the effect of rumble strips may be estimated by their influence on a validated surrogate, especially if it can be seen that surrogates associated with drift onto the shoulder are significantly reduced.

Overall, it seems plausible that the Site Observer can play a future role in analysis and countermeasure evaluation relevant to road departure crashes. However, for the current study, the focus is specifically on intersection safety problems.

CHAPTER 3

Existing Video-Based Vehicle Monitoring Systems

Many video-based systems currently exist for monitoring or tracking vehicles; however, most have limitations that the current system development seeks to overcome. Commercial systems typically are aimed at vehicle detection or with a basic level of tracking that the current project seeks to improve. Research-level systems are most typically not implementable systems; rather, they provide tools and algorithms to capture and convert video data to trajectory estimates with little regard for long-term implementation or automation. Some previous attempts at automated video capture are reviewed here.

Commercial Systems

The research team reviewed commercial and technical literature on products that use video to detect vehicle activity at highway intersections. The team focused on five such systems; other products were found, but none exceeded the current or likely future capabilities of the five selected systems. When possible, the team contacted the companies concerned and received additional technical background.

Autoscope

The Autoscope product line is made by Image Sensing Systems Inc. (Image Sensing 2008), located in St. Paul, Minnesota. Autoscope includes a wide product line sold and distributed by Econolite. This line includes two camera models and various hardware racks for data recording and processing. The company offers the Software Developer's Kit to interact with specific/custom applications.

Autoscope systems are widely implemented in the intelligent transportation systems field, with installations in more than 55 countries. These systems work with many signal controllers on the market, including SCATS, SCOOT, NEMA, Type 170/179 and others. There are a number of features of the system that Autoscope references in its documentation.

Camera and Communication

- Video detection using algorithms to simplify installation, setup, and ease of use;
- Streaming digital video via Ethernet;
- Dual-core processor for image processing;
- MPEG-4 digital streaming video output for review;
- Web browser communications for Internet access;
- Password protection for access control on shared networks;
- Camera integrated with machine vision processor in a single unit; and
- ClearVision technology (hydrophilic coating and faceplate heater) to ensure clean lens for high-quality video.

Data and Measurements Pertinent to Intersections

- General traffic management: Stopline and approach demand, turning movements, degree of saturation, queue measurement, speed and volume estimation.
- Occupancy: Autoscope detects whether a vehicle is present in predefined zones on the highway. The zones are polygons that project to rectangles in the ground plane but are sensitive to height variations in the target vehicles; provided the camera is sufficiently well aligned with the lane, this effect is not sufficient to invalidate the occupancy estimation.
- Incidents: Autoscope detects stopped vehicles in prohibited areas, red light runner detection.
- Nonintersection applications: Oversized vehicles exceeding safe speeds, vehicles driving the wrong way, work zone safety, railroad crossing safety, bus lane enforcement.

The system is flexible and capable within the limits of event detection but has limited capability for the extraction of motion variables (volumes and speed are estimated). No tracking capability is available sufficient to determine conflict metrics and detailed trajectories cannot be reconstructed.

EagleVision

The EagleVision video detection system is made by Siemens of Atlanta, Georgia (Siemens USA 2008) and has vehicle detection capabilities similar to those of Autoscope. The system has the advantage of simple setup and minimal calibration; installers mount the camera, aim the lens at the general target area, and connect the camera to the cabinet through a single cable. Fine-tuning of the detection area may be done remotely via a computer interface.

Other features of the system are support for as many as eight detector zones and outputs, IP communications via a single CAT-5 cable, color streaming video with GUI, and software operating via a Linux operating system. The system provides another good illustration of the state of the art in current site-based video technology for traffic applications; again, no tracking capability is included.

Vantage

The Vantage video detection system is made by Iteris in Santa Ana, California (Iteris 2008). Iteris has been using image processing to detect the presence of vehicles at intersections since 1993. The systems are mostly aimed at replacing inductive loop sensors. Worldwide, Vantage, and Autoscope are the market leaders in number of deployments of these “virtual loop” systems. The Iteris system shares the attributes just mentioned and has particular enhancements in terms of the detection of pedestrians and bicycles, as well as incident detection, ramp metering, and highway monitoring. Again, various aggregated traffic and congestion parameters can be collected and transmitted to a traffic management center.

Iteris offers an option to use a wireless IP camera with the Vantage system, which allows for simple cable-free installation and remote data retrieval. This option uses a license-free 2.4 GHz band to transmit live video from the CCTV camera to the controller cabinet. This wireless transmitter is integrated into the camera and has a 3-inch antenna.

The Iteris Vantage system can record counts, speeds, and three types of vehicle classifications. These data are developed by drawing zones in each lane of travel. A single zone drawn to 15 ft in length can store class, speed, count, and occupancy data by adjusting interval lengths. Each camera can be assigned as many as 24 zones with 8 count zones. The camera must be mounted to see oncoming traffic or outgoing traffic, and the data can be collected locally or downloaded via telephone line, wireless, or cable. Again, the detection and estimation are spatially referenced within the 2-D space of the camera image.

Traficon

Traficon, based in Wevelgem, Belgium, was founded in 1992 (Traficon USA 2008). According to its website, “more than 200 tunnels are already equipped with a Traficon Automatic Incident Detection (AID) system which equates to 40,000 detectors operating worldwide.”

Video data are fed into a detection unit, and at installation a number of detection zones are configured. When a vehicle crosses a predefined line or zone, vehicle detection is registered automatically (Figure 3.1).

Algorithms provide different types of traffic information, such as traffic data for statistical processing, incident-related data, or presence data. The communication board handles the compression of images and transmission of data, alarms, or images.

VideoTrak

VideoTrak by Quixote, based in Palmetto, Florida, is another video vehicle detection system for controlling intersections, monitoring freeways and tollways, and collecting traffic data (VideoTrak 2008). The VideoTrak system can detect vehicles, motorcycles, bicycles, and rail vehicles and uses a dedicated video camera design.

VideoTrak can be operated in two modes: (1) intersection operation and (2) highway management. During intersection operation, VideoTrak tracks targets (vehicles) down a tracking strip, normally associated with a roadway lane, and triggers a DC logic call to an intersection traffic controller, such as the Peek 3000E Controller, when a target’s lead pixel enters a designated detection zone. However, VideoTrak’s use of track strips is somewhat novel, with occupancy within the strip giving a semi-continuous 1-D track of the vehicle as the front of the vehicle moves along the strip (zone illustrations can be found at www.ustraffic.net/datasheets/videotrakdatasheet.pdf).

Each field of view can support 32 detection zones divided in any manner across the track strips. A VideoTrak 905 unit can

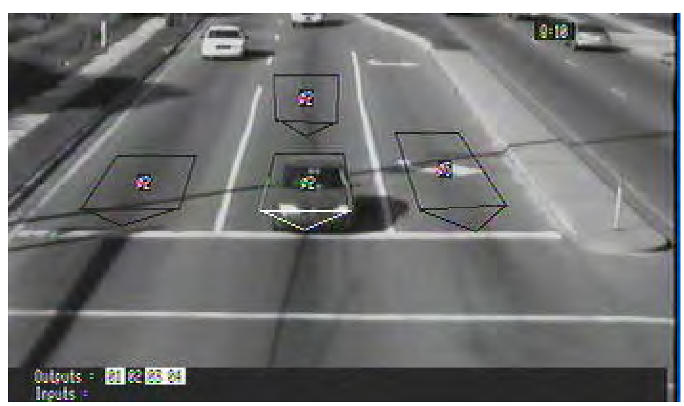


Image provided with permission from Traficon, courtesy of Control Technologies, Inc.

Figure 3.1. Traficon detection zones.

support four tracking/detecting cameras and one surveillance camera, whereas a VideoTrak 910 unit can support as many as eight tracking/detecting cameras and two surveillance cameras. According to VideoTrak, camera location is highly critical to proper detection, with the best location centered over the approaching lanes approximately 10 m above the roadway surface.

Summary

The preceding commercial systems share many similarities in their purpose, capabilities, and methods, although there are also substantial differences in terms of product definition and refinement. The common features of interest are:

- Information is extracted via occupancy detection within the 2-D space of the camera; such information largely can be derived using robust and simple background subtraction algorithms.
- Data output is suited to the traffic management application, rather than safety research: vehicle positioning is based on occupancy (and thus of limited resolution), and additional information is inferred by occupancy detection within zones to provide estimates of speed, volume, gaps, vehicle size, and so forth.
- Data sharing between camera subsystems is limited or nonexistent.
- Camera positioning is highly significant to detection performance, with preference given to alignment of the camera with the traffic lanes and maintaining a sufficiently high ratio between camera height and maximum detection range.

The last point is particularly important. It is only through careful choice of camera position and orientation (pose) that a 2-D sensor can be mapped into the 2-D world of ground vehicle motions on road surface. When cameras are poorly aligned or viewing angles are low, the vertical geometry of the vehicles has a substantial effect on the viewed image. Restricting information processing to the 2-D plane of the single-camera image is a constraint within all the above systems that fundamentally limits their capacity to be developed into vehicle tracking systems for capturing continuous data on vehicle trajectories. Recent work (Kanhere et al. 2010, 2011) shows that a limited degree of feature tracking, together with detection of edges and frontal area features, can enhance the capabilities of these types of commercial video systems and reduce sensitivity to camera mounting height. However, because none of these systems addresses the construction of detailed trajectories, the research team considered the design and performance of existing research-based systems specifically designed for the tracking application.

Research-Based Systems

A number of research groups have used video processing techniques for tracking applications. In almost all cases, the data collection and analysis system has been developed and used by the same research team that processed the resulting research data. Data collection effort normally is conducted on a limited scale, and there is a shortage of systems designed for long-term installation and larger scale deployment. The first two systems, SAVME (System for Assessment of the Vehicle Motion Environment) and NGSIM (Next Generation SIMulation), were developed by groups participating in this research project.

SAVME

The SAVME project (Ervin et al. 2000) was led by the University of Michigan Transportation Research Institute (UMTRI) and included two subcontractors, ERIM International and Nonlinear Dynamics. Data collections were conducted in 1996 and 1999, generating more than 30,000 individual vehicle trajectories. The hardware and software were delivered to the Department of Transportation and were enhanced by NHTSA starting in 2002. Additional data collections were conducted by NHTSA between 2002 and 2004. SAVME consisted of three subsystems:

1. *Data collection* consisting of cameras mounted on towers and computer equipment for collecting and storing video imagery.
2. *Trackfile production* to process video imagery and produce trackfiles containing trajectories for the vehicles in the imagery.
3. *Trackfile analysis* to process the trackfiles, store the track data in a relational database, and provide tools for access and analysis of the track data.

The initial installation covered 600 feet of a five-lane urban roadway, which included a small intersection at one end of the region. There were two cameras on 100-foot towers spaced 200 feet apart and 100 feet from the center of the road. The use of high towers aided the image processing but made the system intrusive and less flexible. Figure 3.2 shows a display of tracking results. Each vehicle has a red cross on it marking the tracking point, and the high camera mounting points allow for simple 2-D to 2-D transformations between image and ground plane, with vehicle height having little influence on the results.

Validation results showed that spatial accuracies typically were within 0.6 m, and velocity components typically were within 0.6 m/s of the true values. The collected database was used to explore a number of common driving scenarios,

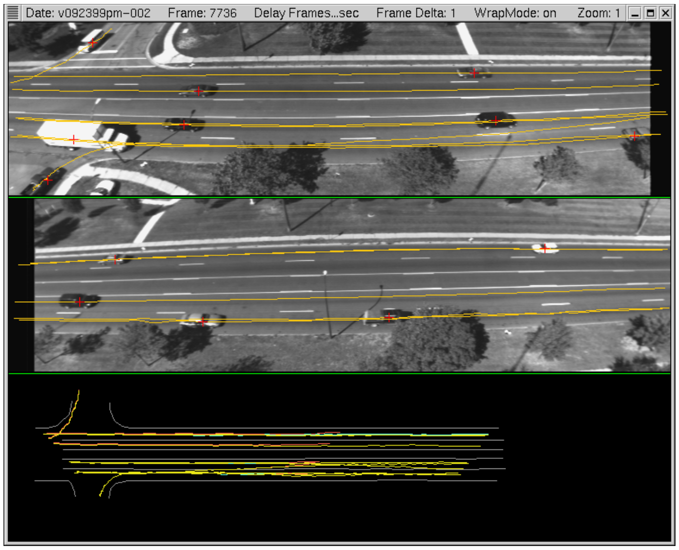


Figure 3.2. SAVME images and tracking display.

including flying passes, left turn across oncoming traffic, emerging from a signed intersection, queue formation and dispersal, and braking propagation along a vehicle string. Results include *X-Y* trajectories and motion time histories for individual vehicles and vehicle clusters.

The project was able to demonstrate the general feasibility of obtaining high-quality kinematic data from a site-based video system and the power of resulting data analyses. On the other hand, the project was limited in a number of key respects:

- Manual intervention was needed to define and correct candidate trajectories, requiring at least 10 h of operator time for each hour of video recorded.
- The methods were not scalable, in the sense that a video archive was required as an intermediate step, and for large-scale implementation there would be a massive expansion in the amount of video data recorded.
- To reduce perspective effects, and thus deterioration of location information, the cameras were mounted on high and intrusive towers, which greatly affects the feasibility and flexibility of installation.

NGSIM

The NGSIM program was instigated by FHWA and industry leaders who developed several commercial traffic simulation tools. The objective of the NGSIM program was to provide real-world data for the verification, calibration, and validation of traffic simulation models. A system was developed, taking video inputs from multiple cameras installed on a high building near the target road and generating long trajectories of individual vehicles with a relatively low level of user input. The source code of the software and the resulting trajectories are available to the public (NGSIM 2008).

Vehicle detection and tracking used a model-based algorithm (Kim and Malik 2003). The procedure is to fit wireframe models to horizontal and vertical line segments detected from the image (Figures 3.3 and 3.4). According to Kim et al. (2005), the detection rate is greater than 88%, with a false-positive rate of 1%. The detection algorithm was designed not just to maximize the detection rate but also to focus on finding the positions and dimensions of the detected vehicles as accurately as possible; the wireframe models perform the step of mapping the 2-D camera information back into the 3-D domain by correcting for perspective effects.

Tracking is based on template matching of the whole vehicle. An image patch for the vehicle is stored and matched in successive frames. Because perspective angles change gradually, the feature image also is modified gradually. The system operated offline and with user assistance; a human operator monitored the detection zone to correct any detection failures.

The first data set was collected and processed at the Berkeley Highway Laboratory. The overall video surveillance system consisted of eight digital video cameras with overlapping fields of view on the roof of a 30-story building overlooking a section of the I-80 freeway in the city of Emeryville, California (the San Francisco Bay Area). From a set of 30-min video clips, a prototype data set of 4,733 vehicle trajectories over a length of 2,950 ft (approximately 1 km) was collected. As reported by Kim et al. (2005), the accuracy of the data set was estimated at approximately 2 ft (60 cm) across the freeway

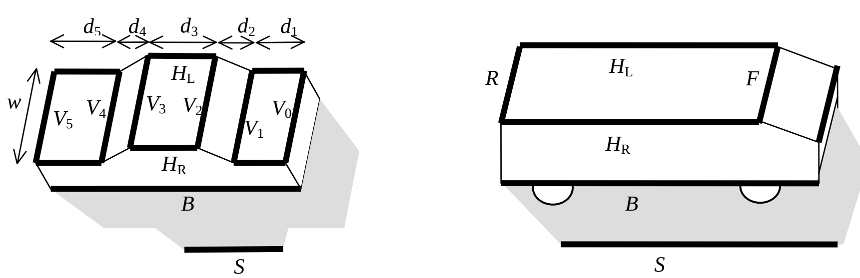


Figure 3.3. The wireframe models used by the NGSIM system to detect a passenger vehicle (left) and a container truck (right).

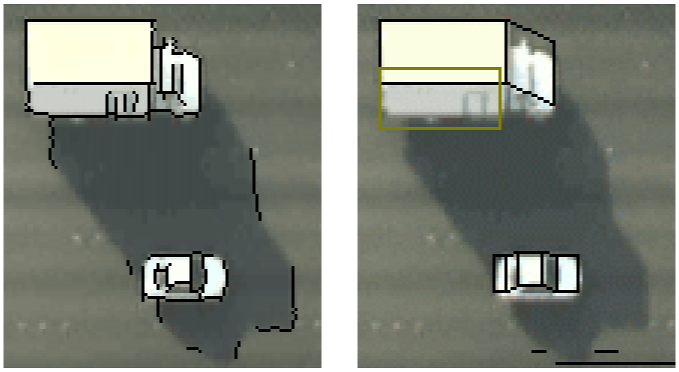


Figure 3.4. Detected line segments (left) and the detection result (right).

and 4 ft (120 cm) along the freeway, similar to the SAVME system data set.

Like SAVME, the NGSIM system's design is focused on extracting high-quality vehicle trajectories. Both are user assisted to ensure a 100% detection rate. NGSIM handles the effects of shadow through the 3-D model templates and is conceptually superior to methods based simply on background subtraction and centroid location. However, the template matching works only when the perspective angle of the vehicle changes gradually and when shapes are predictable. A fully automated system requires a method that is not so tied to predictable vehicle shapes, advantageous camera angles, or manual intervention.

IVSS Study

A recent study (Smith et al. 2009) conducted in Sweden provides a serious benchmark for the state of the art when using site-based video for vehicle tracking and safety evaluation. The study looked at intersection safety from a number of perspectives, including site-based tracking, in-vehicle data collection (small scale with a single instrumented vehicle), and multidriver simulators. The video image analysis art of this project is similar in scope and scale to SAVME, although using modern equipment and employing greater resources in terms of the amount of video analyzed and the automation of the algorithms. The study involved multiple partners: Autoliv, Chalmers Technical University, Linköping University (Computer Vision Laboratory), SAAB Automobile AB, Volvo Car Corporation AB, and the Swedish Road Administration (Vägverket).

In the IVSS study, the video tracking component is the most relevant feature; it was used at two intersections near Gothenburg, Sweden. The first application was at a three-way intersection with a speed limit of 50 km/h in an industrial area: 626 h of traffic were recorded on video. In the second application, 95 h of video were recorded at a 70-km/h, four-way intersection. A single camera was installed at the first intersection (using a tall building), whereas four cameras were installed at the second, with two cameras on each of two lighting poles, which gave high vantage points and unobstructed views of the intersection. In both cases video data were recorded and processed offline at a supercomputer facility. The similarities with SAVME and NGSIM are clear, but this project sought to address many of the issues raised in the current project, particularly in terms of developing a system that can perform accurate and automated tracking at intersections. The four-camera installation operated only during the hours of 09:00 and 15:00 on days with few shadows, from March 2007 to May 2008. Image frames capture was triggered using a GPS receiver so that trajectories could be stitched together after processing. Interestingly, the project also tested the use of laser radar, but the resulting data were not used. According to the authors, "analysis of the distance/angle data proved to be computationally intensive," from which the research team inferred that the large and complex data sets from laser radar were to be no easier to process than the video.

Video image processing was similar to the NGSIM approach, using background subtraction and box-type geometric model estimation, including shadow estimation. To reduce shadow effects, the data collection was limited to the middle of the day and mainly on overcast days. Because the primary aim of the study was data capture rather than system development, these restrictions seem entirely reasonable to the current research team. In terms of camera location, at the four-way intersection, preference was given to the use of very tall light poles, although this created problems of cameras swaying in the wind. The multicamera approach is similar to that used by SAVME, which had two cameras for the purpose of increased coverage, rather than sensor redundancy. Trajectories were estimated at the single-camera level and joined afterward, and it does not appear that overlapping regions were used to improve the underlying detection or vehicle position estimation.

In their report, the authors note substantial problems in making reliable and automated trajectory estimations, especially those arising from occlusion (and thus broken trajectories), tracking of fake objects (e.g., fleeting shadows), and transfer of track estimation from one object to another (e.g., when a small vehicle drives close to a larger vehicle). As with NGSIM and SAVME, identification of vehicles was based on background subtraction and box estimation. The algorithms developed included automated repair mechanisms to join broken tracks as well as additional algorithms to apply geometric corrections. The latter algorithms were found to be needed to remove an effect by which the estimated vehicle position is shifted systematically toward the camera, an effect that was site specific. The study amply demonstrates the problems of automated trajectory capture, even when the aim is not to develop a fully deployable system.

The three studies described indicate what has been achieved in terms of relatively large-scale research projects and how tracking system demands go beyond the standard capabilities of commercial video systems used in traffic management. Certainly, other researchers have conducted similar work on a smaller scale (e.g., Parkhurst 2006), but there remains two especially challenging components to the task undertaken in this project: to develop an installable and scalable hardware system, using an architecture and algorithms that make automated image analysis feasible and tolerant of extraneous effects such as reflections, shadows, and occlusions. To understand these effects in a more general way, we turn attention briefly to the wider area of object tracking.

Object Tracking Research

Tracking has been an area of interest in the computer vision community for many years, and technical advances and cost decreases in digital video cameras have boosted recent interest. Kim (2008) summarized the basic video image processing techniques commonly used to track vehicles, namely, background subtraction and corner feature tracking. It has been mentioned that background subtraction is commonly used in commercial systems. The underlying method is simple: some type of averaging over time is performed for each pixel so that an estimate of the background scene is obtained—fixed objects are retained and transitory objects are removed. Then, comparing any image to this background, regions of significant difference (where pixel brightness, so-called grayscale value, differs by more than some threshold) can be identified as containing candidate vehicles. Details of averaging method and selection of thresholds differ between implementations. As mentioned by Kim in 2008, tracking the resulting foreground regions is far from simple and prone to error, especially when there are occlusions. By contrast, corner features are highly localized and tend to persist over many frames, and thus are more suitable for tracking. A corner feature is a small region of an image where the spatial variation of grayscale value is large and where there is no single preferred direction for the gradient. If one direction predominates, an edge is obtained; edge features can also be useful for tracking, but they are not localized as well as corner features, and thus are more open to ambiguity when tracking.

Yilmaz et al. (2006) published a comprehensive review of object tracking research and included traffic monitoring in their list of six important object tracking tasks. They list eight issues that add complexity to the object tracking task; all eight apply to vehicle tracking to some extent. The research team augmented their list with its brief interpretation of the issues' relevance to vehicle tracking:

- Loss of information caused by projection of the 3-D world on a 2-D image: can be a significant problem depending on the camera angle and camera mounting height;
- Noise in images: less significant for vehicle tracking than for other applications;
- Complex object motion: not a major problem for vehicle tracking because vehicle motions are limited to a 2-D surface and orientation normally is aligned with the vehicle velocity vector;
- Nonrigid or articulated nature of objects: a limited problem, associated with vehicles towing trailers;
- Partial and full object occlusions: a significant problem for vehicles at intersections;
- Complex object shapes: a moderate problem because some vehicles (e.g., mobile cranes) have complex shapes;
- Scene illumination changes: a significant problem for background subtraction, both in the long time scale (night and day) and the short time scale (clouds and headlights); and
- Real-time processing requirements: real-time processing is desirable but not absolutely required.

Object representation is an important aspect, especially when the pixelated images are matched to object models. Two approaches are commonly used: primitive geometric shapes and object silhouette and contour. Both approaches are relevant to vehicle tracking. Four approaches to appearance representation are noted: probability densities of object appearance, templates, active appearance models, and multiview appearance models. All four approaches can be considered for vehicle tracking. However, vehicles can have complex shapes, and it seems best to avoid this problem of matching appearances if possible.

Another approach is to track image features rather than appearance. Four types of image feature are commonly used: color, edges, optical flow, and texture. Again, all are potentially relevant to vehicle tracking. Many tracking algorithms use combinations of features.

Object tracking is divided into three steps: object detection, frame-to-frame tracking, and behavior recognition. For object detection, four approaches are presented. They are point detectors, background subtraction, segmentation, and supervised learning. The detection of points of interest is a general approach that does not depend on lots of knowledge about what is being viewed, and there are algorithms that are viewpoint and intensity invariant. However, when points have been detected, the problem of grouping them into vehicles remains. The background subtraction approach is commonly used for vehicle tracking because the background typically is constant, whereas the vehicles typically are moving. Although not much knowledge about what is being viewed would seem to be required, in fact it is necessary to use such knowledge about the background and the vehicles to obtain the desired level of performance with this approach. The segmentation approach detects objects by their visual characteristics in individual images, rather than depending on characteristics related to multiple images. This approach works well with sparse traffic but becomes very challenging in dense traffic, where occlusions are common. The supervised learning approach is when a learning algorithm is presented with training data and learns to detect the objects automatically. This "black-box" approach again depends on appearance, though without any explicit model. In complex environments, it is unlikely to be successful.

For object tracking, three approaches are discussed. They are point tracking, kernel tracking, and silhouette tracking. Point tracking is not only relevant to point objects; there are algorithms that can cluster and segment sets of points associated with an object based on their motion and a rigid body assumption. Kernel tracking can use a variety of appearance representations, but a key factor is that the kernel is some feature that can be tracked consistently over multiple frames. The motion estimate is then based on the motion of the kernel, which leaves the problem of determining the motion of the object given the motion of the kernel. Silhouette tracking tracks the outline of the object, which can change over time, and is a common approach for vehicle tracking. The key difference between kernel tracking and silhouette tracking is that in silhouette tracking the complete region of the object in the image is tracked. This generally is desirable but can be a problem under occlusion. Again, once the object has been tracked, the problem of determining the motion of the object remains.

The survey paper (Yilmaz et al. 2006) identifies occlusion as one of the major challenges in object tracking and describes a variety of approaches that have been used to address the problem. It is clear that although much effort has been expended to develop principled approaches to detection and tracking, the approaches for dealing with occlusion are much more ad hoc and have been developed on a case-by-case basis by using the characteristics of the particular problem at hand.

The survey also identifies multicamera tracking as an important approach. As with occlusion, a variety of methods are described, but they are all basically ad hoc and use the specific characteristics of the problem at hand. Area coverage and depth estimation are the primary benefits mentioned for multicamera approaches, although tolerance of occlusions also is clearly of interest in the current application.

Addressing future directions in object tracking, the assumptions typically used to make the tracking problem tractable are violated in many realistic scenarios and thus limit the usefulness of trackers in many applications. Applying contextual information is a promising approach for addressing the tractability problem, something that certainly applies to the vehicle tracking problem.

CHAPTER 4

Performance Requirements

This chapter considers the broad range of performance and operational requirements expected of the "ideal" video tracking system. Given the unique importance of trajectory accuracy and resolution in both time and space, based on the need to compute accurate conflict metrics, Chapter 5 will consider these aspects in greater detail and discuss the likely applicability of these conflict metrics as surrogates for crash.

System requirements may cover a broad range of aspects (such as performance, operation, cost, maintenance, and user interface), and capturing requirements is an important step in the design of any system. For a prototype system, the emphasis is greater on the aspects related to performance and feasibility, but long-term viability for development into a deployable system remains important. In addition, there are many feasibility constraints that force the system design in a particular direction, so certain capabilities are derived rather than imposed: What tracking accuracy was achieved? What lighting conditions did it work under? Were there any speed or size limits, and so forth? Design requirements start with the overall system purpose, creating databases of vehicle trajectories that are sufficiently complete and accurate to support safety analysis. They also address the costs and effort associated with achieving that purpose, including hardware and software costs, manual effort in setting up the system, level of effort in running and maintaining the system, and the technical and financial risks. These considerations apply to both the initial development project and the future use of the system in extended field studies.

It is appropriate to think of the principal goal as being a proof of concept for a usable and widely deployable system, which may see expanded use in highway safety, rather than a time-limited study that generates a single fixed database. Location-specific safety issues, whether at intersections or other locations, clearly are too complex for any single data collection study to address the full range of problems. Ultimately, the requirement is to create a tool that can support a wide range of data collection, with many data sets made available to many safety analysts—in other words to create a new tool that can spark safety improvements in the field. The importance of the technology that can create the core information for these studies cannot be overstated.

The requirement for the tracking system is not solely to provide a new tool to address old problems, the very stubborn problem of traffic safety. As new technologies emerge, such as adaptive or intelligent traffic signal systems, the need for better evaluation techniques will only increase. Compared to technology developments in cars and trucks, where real-world testing and evaluation are becoming routine, site-based technologies lack the necessary safety monitoring and evaluation methods, and yet the need is no less great.

The preceding discussion captures the broad customer need for a system that:

- Produces accurate and reliable trajectory data;
- Is affordable and usable in the long term by the highway community, perhaps with modest levels of support from researchers or video systems suppliers;
- Is scalable for use in large and complex sites, with wide area coverage possible; and
- Creates data sets that can be integrated with diverse other information sources.

In the current project, while all major design decisions have been determined (and the prototype hardware built, installed, and run), it seems worthwhile to capture the motivations and considerations that went into the design. With this background and with emphasis on the robust prototype system developed, the research team developed a comprehensive list of operational and performance requirements.

The technical performance of the site-based video system is centered on the trajectory data that can be acquired and the conditions under which these data can be captured. At a high level, these should not be too specific; for example, it would be absurd to specify a requirement for “less than 5% of vehicle trajectories should display validation errors when the sun angle is less than 10° above the horizon.” In the following, the research team seeks to define a set of plausible benchmarks for testing performance. These can be used to assess performance and operation of the current system and provide a basis for future developments and comparison of rival systems.

Trajectory Information

Accuracy and resolution in spatial and time coordinates: This should be sufficient to extract conflict metrics in avoidance and near-crash events (and of course in actual crashes) as well as to characterize exposure and statistical patterns in conflict metrics. Numerical targets are developed in Chapter 5.

Error rates and sources of systematic bias: This relates to possible discrete errors in tracking—for example, when a trajectory is not resolved because of too few discernible features on the vehicle, sun glare, occlusion, intrusion of a flock of birds, or so forth. Clearly most trajectories need to be resolved under normal conditions of lighting, weather, and traffic density. A key point is that when an error occurs, it should not go unnoticed; validation checks are to be sufficiently robust to determine, for example, when some persistent moving feature (called a *cluster track* in the system design) cannot be associated with a vehicle, or to detect that a tracked vehicle enters the intersection but never leaves it. Under these circumstances, the error should be rectified using an automatic correction process or the error flagged so associated trajectories are not used for what might be an erroneous conflict analysis. In a complex intersection, we might expect to be collecting valid data at least 80% of the time; for the simple intersection used in the pilot study, a much higher figure would be expected. The tradeoff here is between a fully automated system that is prone to intermittent interruptions in sampling, compared with a manually supported system that is 100% available during the sampling times determined by project managers. The main concern is that sampling does not introduce bias, but otherwise it does not matter what process controls the sampling, provided the system is not prone to frequent interruptions. On the subject of potential bias, this will occur if the system cannot resolve certain specific situations relevant to crash frequencies (e.g., missing faster-moving traffic). To review missed vehicles or erroneous tracking, there is a requirement for compressed video to enable manual review of relevant flagged cases. Provided interruptions are random and unconnected to the safety problem (e.g., caused by random scattering of light), they will not affect the validity of the research data. If sources of bias exist, the algorithm or other aspects of the system must simply be improved.

Sensitivity to lighting and weather conditions: It must be expected that in poor lighting and weather, the availability (percentage time of error-free operation) will drop. In rare and extreme cases such as snow or dense fog, the vision-based system will certainly become inoperable. This does not present a source of bias provided researchers know and understand these effects. On the other hand, if availability is dropping significantly every time there is a change in ambient lighting, or when there is light rainfall, clearly this would be an unacceptable limitation. Nighttime use is another consideration, most easily addressed by suitable choice of camera type; in this project the focus was on daytime operation.

Vehicle Information

Vehicle size and type estimation: The basic requirement for trajectory capture is to extract the time history of the (x, y) coordinates of any suitable reference point on each vehicle and determine where this point is situated relative to the front, rear, and lateral boundaries of the vehicle. This means that vehicle length and width both need to be estimated, but nothing further need be considered an essential requirement. This provides the basis for at least an approximate classification into vehicle types, similar to the capability of some of the commercial systems discussed. When vehicle height or shape information is extracted, it is possible that more detailed and accurate information can be extracted relative to the existing commercial systems, but again this should be seen as a potential bonus, not as a fundamental requirement for automated processing. When some event is found to be of interest, however, recorded video should be available for a human reviewer to recognize vehicle types.

Additional vehicle characteristics (e.g., truck loading, numbers of passengers): Similar to the previous point, no such information should be regarded as a requirement of the automated data collection. Some information on truck loading might be inferred from acceleration performance, but the connection is only loose. It is preferred that stored video not be capable of resolving individual passengers to reduce concerns over possible invasion of privacy. Video is likely to be able to discern any very obvious external markings (e.g., conspicuous trade names on commercial vehicles) but with limited resolution; the inability to count passengers or recognize driver characteristic is an advantage in the long term.

Data Availability

Access to stored data: It is feasible that trajectory data can be made available to any team conducting relevant transportation research or development. A base server can host the trajectories in a relational database for easy access by researchers who are not necessarily part of the data collection team. Metadata and data dictionaries need to be formalized and should include information about periods with lost trajectories. Data volumes should not be large, and this seems a relatively simple matter for the long term.

Error Checking and Estimation

Error checks: The trajectory estimation involves a high degree of data fusion. As previously mentioned, it is important that the data fusion be open to validation checks. Comprehensive error checking should be feasible in the automated analysis, to provide a validation or confidence parameter to the trajectory analysis. In addition, a video viewing tool is required so that bounding boxes of computed vehicle trajectories can be overlaid on stored video images; thus, a reviewer can look at samples of data to confirm success or otherwise of trajectory estimation.

Derived information: Video processing can provide only so much information on the vehicle kinematics; the system should include additional estimation algorithms, including additional motion estimates, especially speed, longitudinal acceleration, and lateral acceleration.

Accuracy checks: Once a trajectory has been synthesized, the question of its accuracy remains. This checking is not required to form an integral component of the system operation because an independent reference is needed, which adds cost and complexity; however, such checks should be available, especially as part of the system development.

- *Instrumented vehicles*: Independent accuracy checks can be made using differential GPS (DGPS) and other on-vehicle instrumentation. The importance of onboard DGPS is clear, whereas onboard accelerometers, yaw rate sensors, and speed measurement (from wheel or transmission speed) can be used to test the accuracy of derived information.
- *Site-based radar*: This can be used to make an independent validity and accuracy test of the trajectory data. Radar may not provide an especially accurate alternative, but it does provide a method to benchmark performance.

Computational Aspects and Streaming Capability

Data buffering: Some video buffering is expected, to provide flexibility of processing, especially when traffic volumes are high. The volume of data storage is largely irrelevant to the overall system performance, although it may affect the time over which the system can run before interruption is needed. It is possible that in very complex installations, the system could go to the extreme operational mode of "capture by day, process by night," in which case-sufficient storage would be needed for approximately 12 h of uncompressed video for each camera station, something that is no concern for a hard drive but could be a challenge for solid state memory.

Scalability and parallel processing: Once the system is developed and proved for an intersection of modest complexity, the same hardware and software should be capable of scaling up to a larger and more complex installation; to avoid processing bottlenecks, the major portion of the video processing should be parallelized (e.g., processing each video stream in parallel).

Raw video: To avoid processing bottlenecks, it is a key requirement that raw video need not be exported from the data collection site.

Data streaming: A streaming capability is required so that processing of trajectories can take place at the same time new data are collected. Ideally, 24 h of trajectory information should be processed into trajectory data in a 24-h period. Achieving this 24/24 (100%) capability depends on a number of factors (especially how many image features must be processed per frame to enable the data fusion to achieve the required availability and accuracy), so for now the proposed requirement is for the prototype to achieve approximately 50% of this capability based on processing speed. This will provide a useful immediate capability of processing 12 h of trajectories in a 24-h period, plus a strong expectation that a more fully integrated and efficient second-generation system will fulfill the ideal 100% capability.

Modularity: Ideally, video image processing should take place adjacent to or within the camera itself, to remove the need for high-speed data networks at the intersection. For the prototype system, this level of integration is unnecessary, but the architecture is to be sufficiently modular to support such an integration step in the next generation of the system. If network bandwidth can be sufficiently reduced, this opens the opportunity of using purely wireless Ethernet connections linking cameras to one or more data storage hubs.

Archival storage: The principal requirement is that trajectory data be stored in a relational database to support conflict metric and other analysis on demand. When events of interest are discovered in the data, researchers are likely to be interested in other aspects not readily accessible from the trajectory data, so it is important that compressed video is also retained. The video should be stored in a form that is compatible with trajectory data so it can be displayed in synchronous fashion. Optionally, additional image-related primitives (e.g., cluster tracks) may be stored; this is not something the end user need be aware of, and the main reason will be to aid algorithm development.

User Interface

It is important that the system provides an end-to-end connection between vehicle motion and safety analysis. The system was conceived as a research tool, and the expectation is that researchers using the system typically will understand vehicle dynamics, traffic flow theory, and intersection design and have some experience in data analysis and relational databases. But they should not require detailed knowledge of video formats, image processing, or camera technologies. For this reason, some basic tools are required for users to interact with the data:

- *Data access*: Time history data should be stored in a standard database format, such as Microsoft SQL Server;
- *Analysis tools*: Basic tools are needed to query the data to calculate a subset of conflict metrics, including time to collision and gap times for path-crossing conflicts; and
- *Visualization tools*: A basic viewer should be provided so that events of interest, defined by event start and end time, can present video and time history data of that event.

All such tools should be part of the prototype, with the expectation that future generations will offer increasing user-friendly interfaces.

CHAPTER 5

Conflict Metrics and Crash Surrogates

As described in Chapter 2, the core use of the Site Observer is for addressing SHRP 2 Safety research questions via the use of crash surrogates (conflict measures that correlate with actual crashes) to provide a powerful tool to evaluate crash risk, factor dependencies, and the efficacy of any particular interventions. In this chapter the research team reviews the commonly used conflict metrics and uses a simulation study to evaluate the sensitivity of these metrics to track errors, thus providing benchmark requirements for tracking accuracy.

Intersection-Related Conflict Metrics

Defining and validating crash surrogates is by itself seen as an ongoing research issue, but in loose terms these are metrics that describe the normal vehicular interactions that are correlated with conflict severity and crash risk, although possibly in some context-dependent manner. As such, standard conflict metrics provide a suitable test case for the video tracking system; tracking errors should not be large enough to substantially affect distributions of conflict metrics such as the time to collision.

It is not the purpose here to seek to substantiate or validate the general method of traffic conflict theory. Rather, the current crop of metrics in use is summarized and it is noted that one important application of the SHRP 2 video tracking system will be a tool to enhance the future development and validity checking for the technique.

Table 5.1 gives a summary of the major conflict metrics in common use for intersection crashes. These are directed toward path-crossing conflicts, although time to collision (TTC) is equally applicable to conflicts relevant to rear-end collisions, which is the primary conflict type for the chosen pilot intersection (see Chapter 7).

A distinction should be made between two types of conflict metric: instantaneous and summary measures. The instantaneous measures are defined for each vehicle and for each instant of time; gap time and TTC are examples of these. Typically there is an assumption that the vehicles continue to move on a given trajectory and at a constant speed. On the other hand, summary measures provide surrogates on a per event basis (e.g., PET or postencroachment time) or on a per time basis (e.g., time-integrated TTC and time-exposed TTC).

The last four metrics in the table—signal encroachment time, signal transition deceleration time, signal transition acceleration time, and lateral encroachment time—are associated with crashes involving red light violations. The interest in these metrics is to support a better understanding of the relationship between the time when the red light running occurs and particular conflict types. Intersection conflict types were identified by Najm, Smith, and Smith (2001) and illustrated in Figure 5.1. The authors organized conflicts at intersections based on precrash movement and proposed the following five categories:

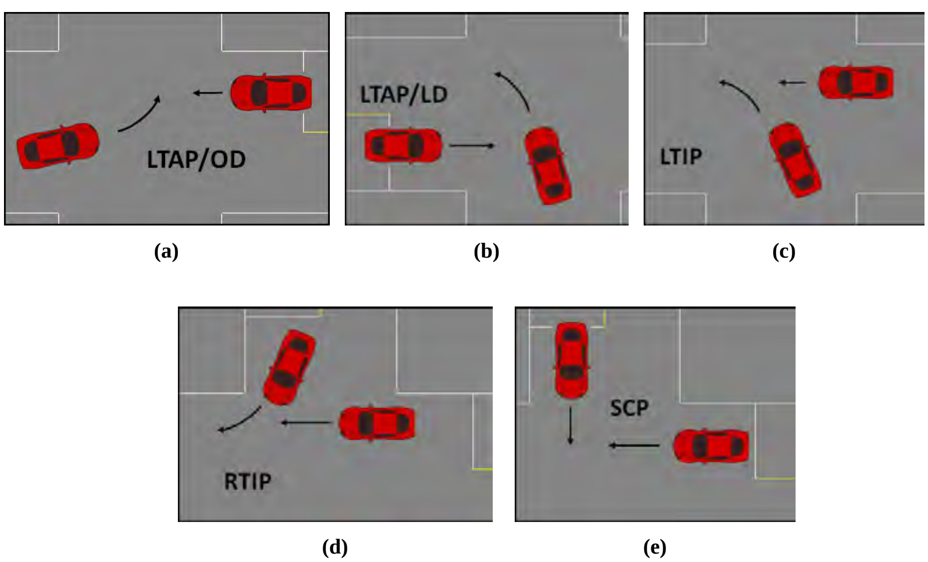
- Left turn across path/opposite direction (LTAP/OD);
- Left turn across path/lateral direction (LTAP/LD);
- Left turn into path—merge conflict (LTIP);
- Right turn into path—merge conflict (RTIP); and
- Straight crossing path (SCP).

Najm, Smith, and Smith also recommended considering the type of control device regulating the intersection and distinguished between signal, stop sign, no controls and others.

An example of how red light running and conflict type can be associated is found in Zimmerman and Bonneson (2005), who reviewed samples of crash data and distinguished between two conflicts involving red light running: LTAP/OD and SCP. The main difference between the two conflicts is found in the timing of the collision relative to the red light onset. In the case of LTAP/OD, most collisions occurred within the first 5 s after the transition, whereas most of the SCP collisions occurred after the first 5 s after the transition.

Table 5.1. Summary of Kinematic Conflict Metrics

Surrogate Conflict Measure	Description
Gap time (Gettman and Head 2003)	Time lapse between completion of the encroachment by turning vehicle and the arrival time of crossing vehicle if they continue with the same speed and path.
Encroachment time (Gettman and Head 2003)	Time duration during which the turning vehicle infringes upon the right-of-way of through vehicle.
Deceleration rate (Gettman and Head 2003)	Rate at which crossing vehicle must decelerate to avoid collision.
Proportion of stopping distance (Gettman and Head 2003)	Ratio of distance available to maneuver to the distance remaining to the projected location of collision.
Postencroachment time (Gettman and Head 2003)	Time lapse between end of encroachment of turning vehicle and the time that the through vehicle actually arrives at the potential point of collision.
Initially attempted postencroachment time (Gettman and Head 2003)	Time lapse between commencement of encroachment by turning vehicle plus the expected time for the through vehicle to reach the point of collision and the completion time of encroachment by turning vehicle.
Time to collision (TTC) (Gettman and Head 2003)	Expected time for two vehicles to collide if they remain at their current speed and on the same path.
Time-exposed TTC (Archer 2001)	The length of time that all vehicles involved in conflicts spent under a designated TTC minimum threshold during a specified time period.
Time-integrated TTC (Archer 2001)	Integral of TTC profile of drivers to express the level of safety over the specified time period.
Time to accident (Archer 2001)	Point at which the aversive action is taken. This measure, combined with the conflicting speed, allows determination of the level of severity of a conflict.
Signal encroachment time	Time lapse between the onset of red cycle and vehicle entering intersection.
Signal transition deceleration time	Time lapse between the transition of signal (green to amber or amber to red) and deceleration onset.
Signal transition acceleration time	Time lapse between the transition of signal and acceleration onset.
Lateral encroachment time	Time duration during which the violating vehicle infringes upon the right-of-way of through vehicle.



Source: Najm, Smith, and Smith 2001.

Figure 5.1. Intersection conflict types.

The implication is that, for LTAP/OD cases, the driver is waiting to make a turn and is under pressure to clear the intersection and might misinterpret the oncoming driver decision to stop at the red light.

From this perspective, the research team attempted to organize the potential crash surrogates relative to the conflict type and proposed causal factors in Table 5.2, where the causal factor could be observed via a set of surrogate measures departing from baseline driving.

It seems reasonable to conclude that any one of the metrics in Table 5.1 may be used for future analysis using the video tracking system results. Although other metrics might also be of interest (for example, time to lane crossing [TTLC]) when studying lane departures or road departures, the intersection conflicts provide a reasonable benchmark for the tracking system. The research team considers what level of time and space resolution and accuracy are needed to support their use in safety studies.

Table 5.2. Conflict Type and Surrogate Measures for Signalized Intersections

Conflict Type	Scenario	Causal Factors	Surrogate Set
LTAP/OD	A vehicle turning left at an intersection collides with a vehicle from the oncoming traffic.	Failure to yield right-of-way	
		Underestimation of oncoming traffic speed	• Gap time • Encroachment time • Deceleration rate • Proportion of stopping distance • Postencroachment time • Initially attempted post encroachment time
		Understanding of right-of-way	• Gap time • Encroachment time • Deceleration rate • Proportion of stopping distance • Postencroachment time • Initially attempted postencroachment time
		Run signal	
		Did not see light status	• Signal encroachment time • Signal transition deceleration time • Lateral encroachment time
		Vision obscured	• Distance to intersection • Signal transition deceleration time[a]
		Distraction	• Distance to intersection • Signal transition deceleration time[b]
		Tried to beat amber light	• Signal transition acceleration time • Signal encroachment time
		Deliberately ran red light	• Signal transition acceleration time • Signal encroachment time
LTAP/LD	A vehicle turning left at an intersection collides with a vehicle from the lateral direction.	Run signal	
		Did not see light status	• Signal encroachment time • Signal transition deceleration time • Lateral encroachment time
		Vision obscured	• Distance to intersection • Signal transition deceleration time
		Distraction	• Distance to intersection • Signal transition deceleration time
		Tried to beat amber light	• Signal transition acceleration time • Signal encroachment time
		Deliberately ran red light	• Signal transition acceleration time • Signal encroachment time

(continued on next page)

Table 5.2. Conflict Type and Surrogate Measures for Signalized Intersections (continued)

Conflict Type	Scenario	Causal Factors	Surrogate Set
LTIP	A vehicle turning left at an intersection collides with a vehicle in the flow in which it is inserting.	Run signal	
		Did not see light status	• Signal encroachment time • Signal transition deceleration time • Lateral encroachment time
		Vision obscured	• Distance to intersection • Signal transition deceleration time
		Distraction	• Distance to intersection • Signal transition deceleration time
		Tried to beat amber light	• Signal transition acceleration time • Signal encroachment time
		Deliberately ran red light	• Signal transition acceleration time • Signal encroachment time
RTIP	A vehicle turning right at an intersection collides with a vehicle in the flow in which it is inserting.	Failure to yield right-of-way	
		Underestimation of on-coming traffic speed	• Gap time • Encroachment time • Deceleration rate • Proportion of stopping distance • Postencroachment time • Initially attempted postencroachment time
		Understanding of right-of-way	• Gap time • Encroachment time • Deceleration rate • Proportion of stopping distance • Postencroachment time • Initially attempted postencroachment time
		Run signal	
		Did not see light status	• Signal encroachment time • Signal transition deceleration time • Lateral encroachment time
		Vision obscured	• Distance to intersection • Signal transition deceleration time
		Distraction	• Distance to intersection • Signal transition deceleration time
		Tried to beat amber light	• Signal transition acceleration time • Signal encroachment time
		Deliberately ran red light	• Signal transition acceleration time • Signal encroachment time
SCP	A vehicle going straight at an intersection collides with a vehicle from the lateral direction.	Run signal	
		Did not see light status	• Signal encroachment time • Signal transition deceleration time • Lateral encroachment time
		Vision obscured	• Distance to intersection • Signal transition deceleration time
		Distraction	• Distance to intersection • Signal transition deceleration time
		Tried to beat amber light	• Signal transition acceleration time • Signal encroachment time
		Deliberately ran red light	• Signal transition acceleration time • Signal encroachment time

[a] When the drivers' view of the signal is obscured by some static element (e.g., foliage, building), the data can be expected to show that drivers within specific distance from the intersection do not respond to signal transition.
[b] However, if there is no visual obstruction, the point of no response to signal change could be anywhere in the intersection.

Accuracy Requirements from Conflict Metric Analysis

Because the estimation of distributions of conflict metrics is a critical application for the system, these provide a significant test case for accuracy requirements: when errors occur in positions and velocities recorded in the trajectory record, these will induce errors in the relevant conflict metrics. To investigate this, one must have an unperturbed exact set of trajectories, which are here provided by simulation, together with an error model. From Monte Carlo simulation, a resulting distribution of conflict metrics can be obtained.

Based on the review of conflict metrics in Chapter 4, two particular intersection conflict types, and in each case, two conflict metrics are considered here. These are:

1. LTAP/OD (Figure 5.1a), with metrics of *gap time* and *postencroachment time;* and
2. LTIP (Figure 5.1c), with metrics of *time to collision* and *deceleration rate.*

For simplicity, no traffic signal timing is considered, and the arrival times of the vehicles have been set so that although no collision occurs, the postencroachment time (Case 1) and time to collision (Case 2) are short but realistic. Each scenario is considered here.

Left Turn Across Path

Figure 5.2 shows the simple intersection used for simulation. The subject vehicle (SV) has approached from the right, stopped briefly, and is now just starting to make a turn across the path of the principal other vehicle (POV) moving from the left. In this example, the POV initially is moving with constant speed and stays in its current lane but brakes for a short period to avoid hitting the turning vehicle. Figure 5.3 shows the speed and acceleration of the two vehicles during the maneuver.

Figure 5.3 shows the results of predicted arrival times at the intersection by the POV (where the front of the POV intersects the right boundary of the SV path) and the predicted exit time for the SV (when the rear of the SV clears the right boundary of the POV path). These predictions are made at each time instant and assume no change in path or speed, as is normally the case for the metrics chosen. The difference between these times is the gap time, and it can be seen that only as the conflict point is reached does the gap time rise above zero so the collision is avoided (Figure 5.4).

Although the (predicted) gap time has a time history over the entire interaction, the PET is a single value, namely, the actual time gap between the subject vehicle leaving the point of encroachment and the POV reaching it, and in the particular simulation, PET = 0.363 s. This is a suitable statistic to consider for sensitivity to path estimation error but clearly ignores the role of the POV driver in intervening to prevent the crash; thus, another pair of statistics is considered as well. These are the minimum and maximum gap time during the 2 s before the POV arrives at the zone of encroachment, which does include the effect of POV intervention; these values are $GT_{min} = -0.0060$ s, $GT_{max} = 0.363$ s. Not surprisingly, in this case GT_{max} equals PET, but this need not always be the case.

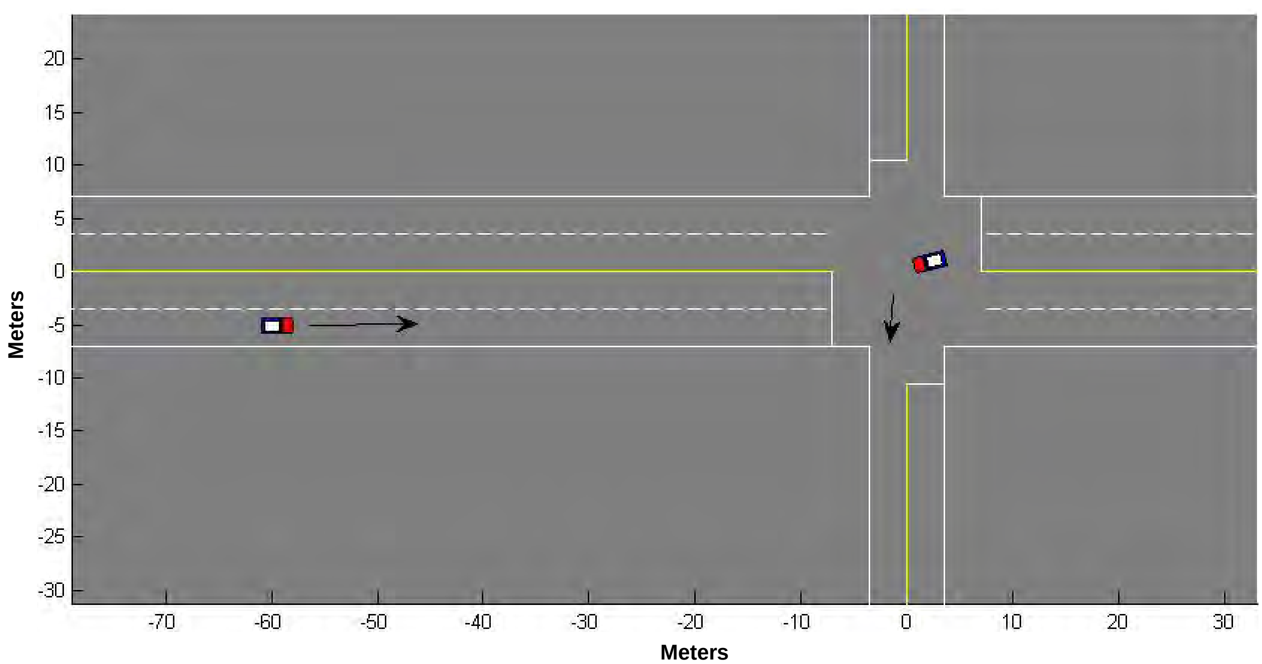


Figure 5.2. LTAP/OD scenario.

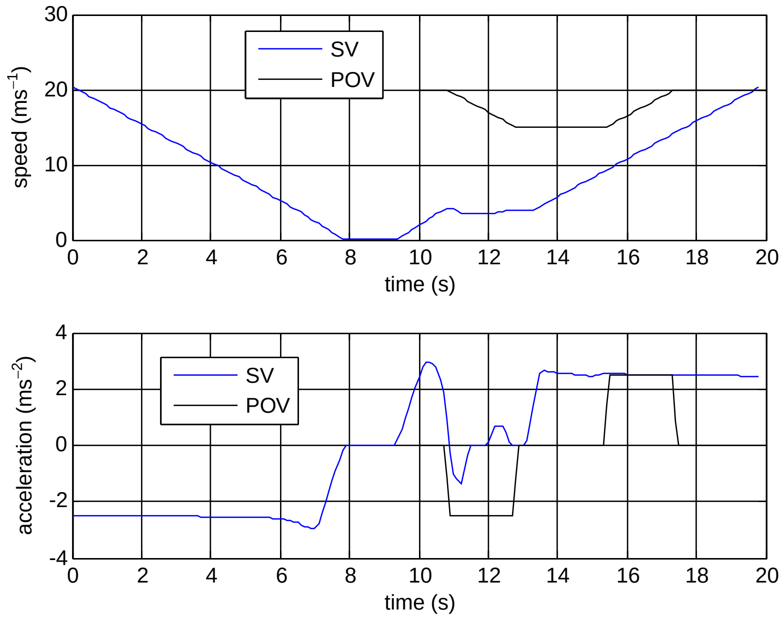


Figure 5.3. Simulated speeds and accelerations of the two vehicles.

Left Turn into Path

In this scenario, the same simple intersection geometry is used, but this time the SV approaches from the upper road, slows and does not stop, then turns into the same path as the approaching POV. In this simulation, the POV has a higher initial speed and again decelerates to avoid a collision. The conflict measures chosen in this case are more appropriate for the case in which paths are coincident (after a certain instant in time). Figure 5.5 shows the trajectories and Figure 5.6 presents speed and longitudinal acceleration data for the conflict.

Figure 5.7 shows the time variation of the selected metrics for the conflict. On the left is the time to collision, the time at which a collision would occur if the speeds and paths remained

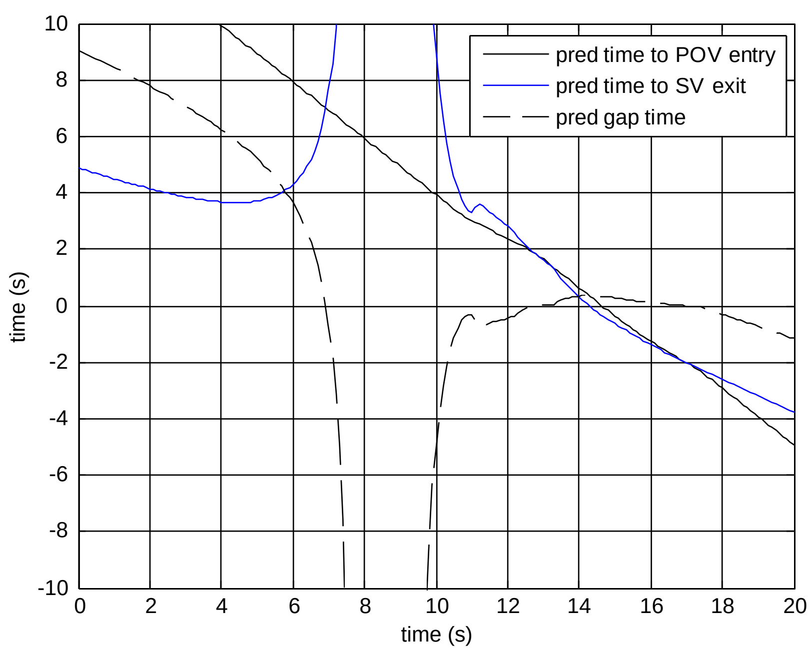


Figure 5.4. Predicted times of encroachment and gap time for LTAP/OD scenario.

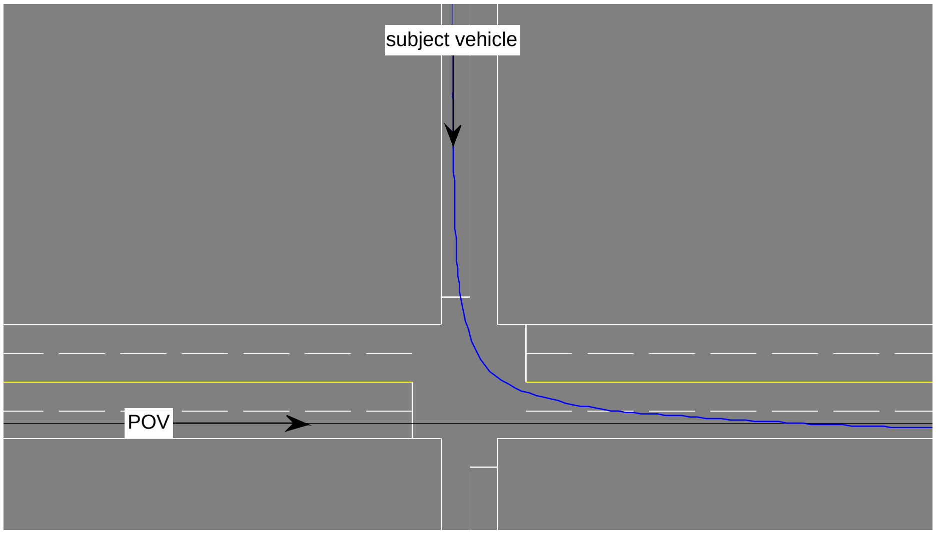


Figure 5.5. Vehicle's paths.

unchanged. The second was chosen as the acceleration rate of the turning (SV) required to avoid collision (assuming the POV maintains a constant speed at each instant). This is similar to the deceleration rate mentioned earlier, although in this case the SV clearly needs a positive acceleration to avoid a collision. The (predicted) time to collision falls to approximately 1 s, while at about the same time the acceleration required (AR) for the subject vehicle increases to almost 4 ms^{-2}. Here both metrics are continuous variables, so for the error analysis, TTC_{min} and AR_{max} are used as summary metrics of interest.

Sensitivity Analysis

The effect of tracking errors on the summary conflict measures is considered here. The basic idea is to corrupt the trajectories by adding some form of error to include in the "measured" locations or times of the trajectories. There are many ways in which this can be done, and here the aim is to keep the analysis relatively simple and not to deliberately amplify or overexaggerate the possible effects of the errors.

The simplest and perhaps most common form of error signal is a Gaussian variable, which is uncorrelated between samples;

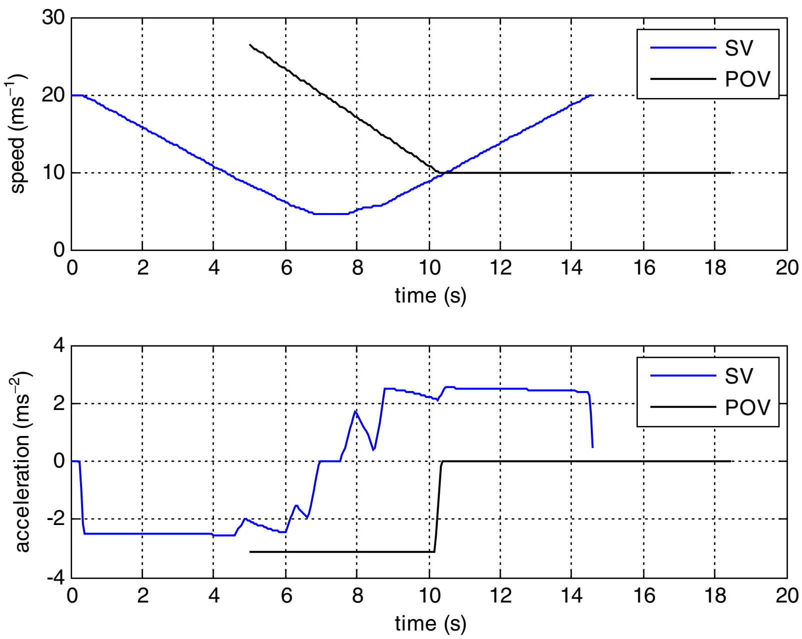


Figure 5.6. Vehicle's speeds and longitudinal accelerations.

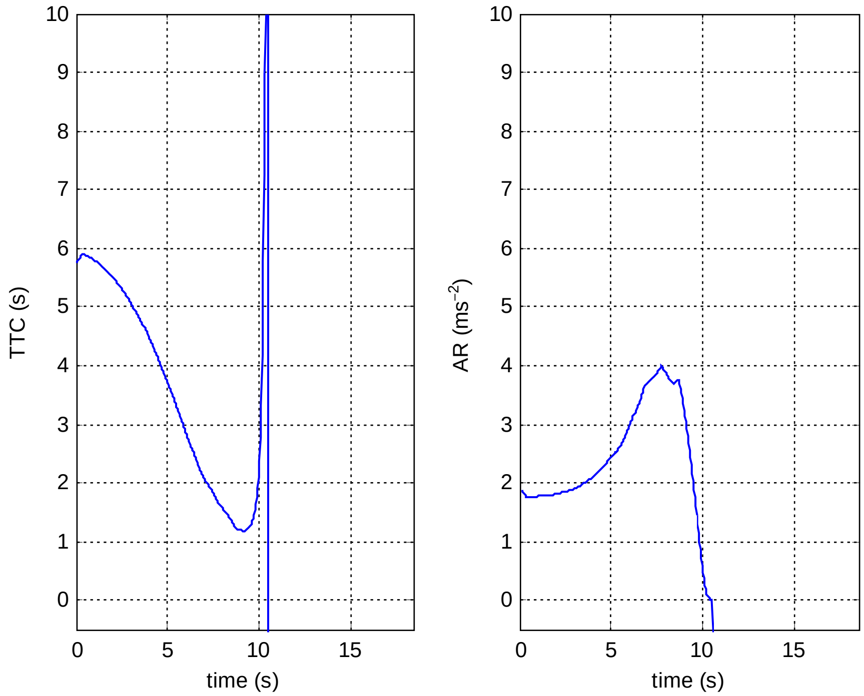


Figure 5.7. Conflict metrics for LTIP scenario.

that is, it is a discrete-time Gaussian white noise (GWN) error. In addition, when such errors arise in the data collection system, some form of filtering (e.g., Kalman filtering) undoubtedly will be applied (see Chapter 10), reducing the bandwidth and the RMS error induced. Put another way, GWN can show large errors that may be reduced because it is only the lower frequency components that can be confused with vehicle dynamic motions. Therefore, we will use a filtered form of the GWN, using a simple first order filter with a bandwidth of 1 Hz, which means that the errors become largely indistinguishable from the dynamics of a maneuvering vehicle. Two independent equal variance error signals are generated, applied to the x and y coordinates of the subject vehicle, while for simplicity the POV trajectory is left unchanged (a more realistic method might be to assign errors that are correlated as a function of vehicle speed and spatial location, but the detailed modeling of errors in the camera and video processing system rapidly become very complex indeed). The RMS errors of these signals are then scaled according to three assumptions about the measurement system: (1) low level, RMS = 0.2 m; (2) medium level, RMS = 0.5 m; and (3) high level, RMS = 1 m.

Figure 5.8 shows the effect on the estimated spatial trajectory when the medium level of error is used.

Figures 5.9 and 5.10 show the effects on the three summary statistics when low- and medium-level errors are used.

It may be noticed that the trajectory errors, although smoothed to some extent through the limited bandwidth of the filter, are still relatively coarse; it is certainly possible that an optimal filter can reduce the errors further. On the other hand, the errors shown are by no means worst case (recall that the POV trajectory is uncorrupted). If we were not looking at critical events, the spread in these distributions would not be unreasonable, but when the signal (gap-related timing) is close to zero, as in this case, and deviations from zero are highly significant (crash or no crash, for example), it is clear that even for medium error levels the noise-to-signal ratio is too high to be very useful. The conclusion is that when RMS errors in the vehicle location are on the order of 0.5 m or higher, the effects of errors on computed metrics are too high to be much use to the analysis of near-crash events. This conclusion depends on the simple assumptions made about the error process and would not apply, for example, to a camera calibration error that could be corrected in postprocessing. But for band-limited random errors in the vehicle path tracking, errors should not be any worse than around 20 cm in RMS. Presenting results for high-level errors has turned out to be unnecessary, but in this case the noise totally dominates the signal, and the resulting trajectories would be practically useless for the analysis of near-critical events.

The same analysis was applied to the second scenario (LTIP) with similar results. Figures 5.11 and 5.12 show corresponding histograms for the minimum TTC and maximum AR.

The low level of simulated errors provides convincing results, whereas especially for the minimum TTC, the medium-level error case gives a noise level in the computed metrics that is comparable to the signal amplitude. The same conclusion applies: displacement errors should be no more than approximately 20 cm in RMS terms for conflict metrics to be computed with reasonable confidence, at least after postprocessing filters have been applied.

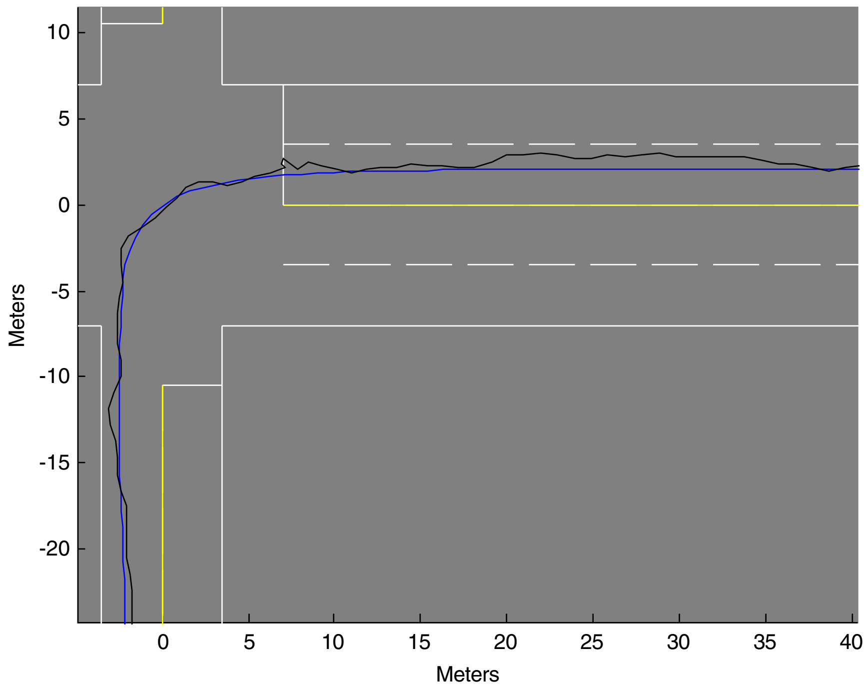


Figure 5.8. Effect of medium level spatial errors on the trajectory (blue: reference; black: perturbed).

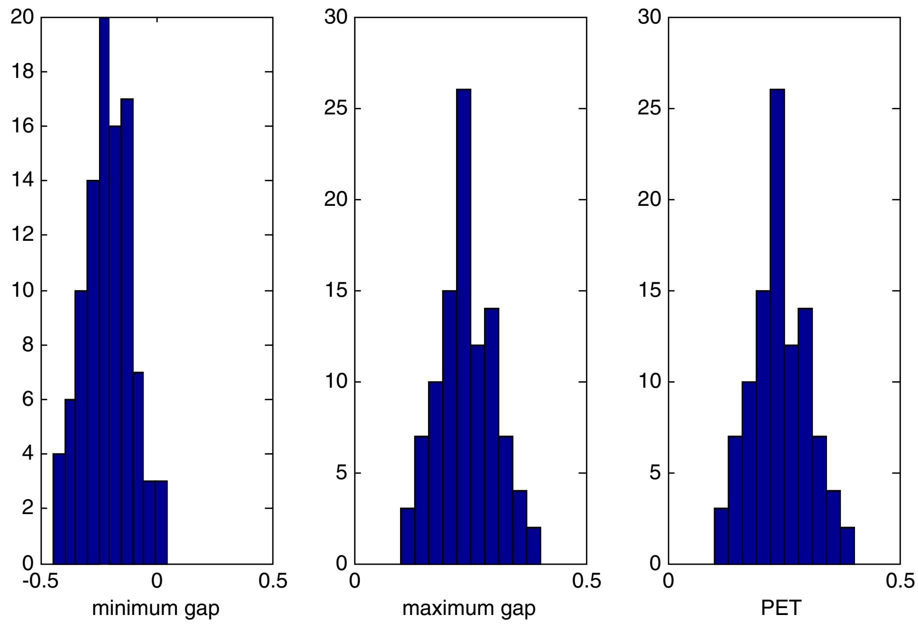


Figure 5.9. Frequency distribution of summary metrics (LTAP/OD: 100 simulations, low-level error).

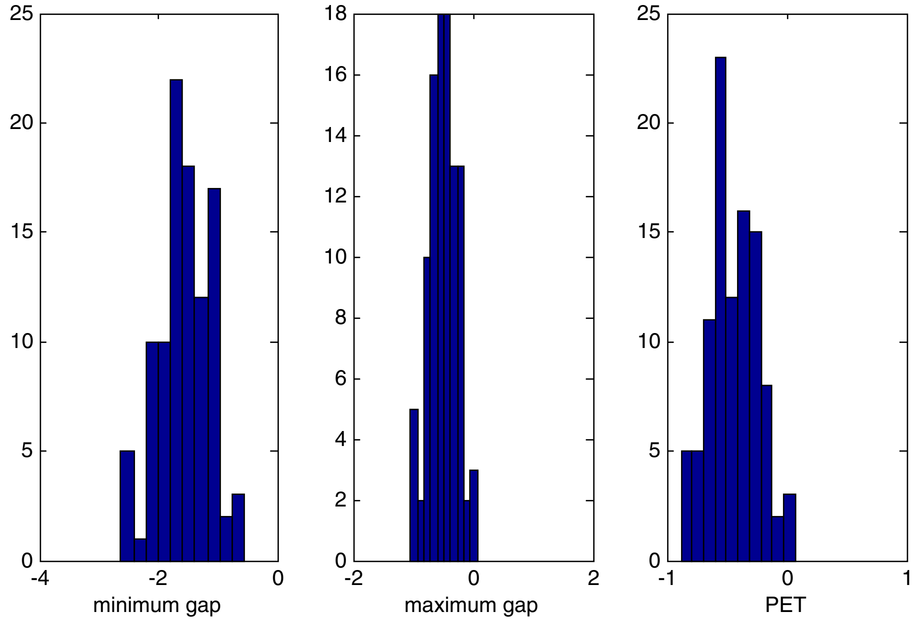


Figure 5.10. Frequency distribution of summary metrics (LTAP/OD: 100 simulations, medium-level error).

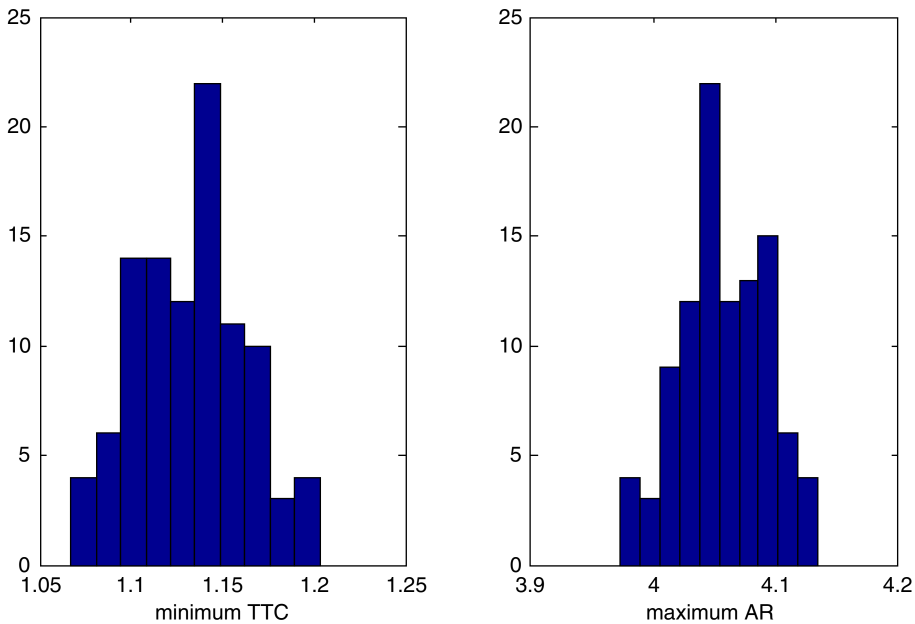


Figure 5.11. Frequency distribution of summary metrics (LTIP: 100 simulations, low-level error).

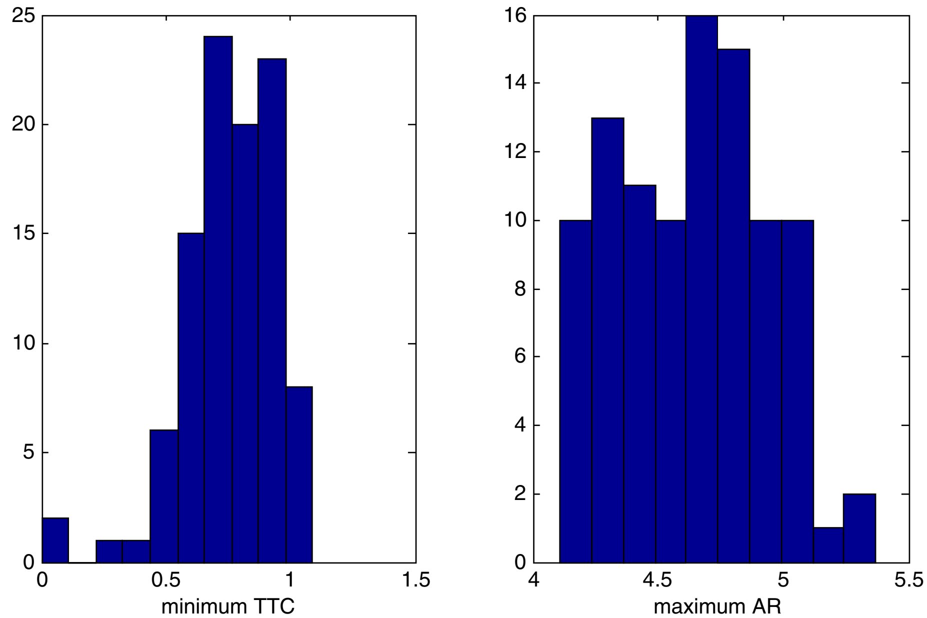


Figure 5.12. Frequency distribution of summary metrics (LTIP: 100 simulations, medium-level error).

CHAPTER 6

Site Observer Design

The video system design needs to be robust and accurate, which normally implies the need for sensor redundancy, which in this case means using multiple cameras. Accuracy is still dependent on the limitations of the individual seniors—in this case the video camera. Although the error process is complex, depending on the detail and quality of tracked visual features and whether there are unknown latencies in the video capture, it is instructive to estimate fundamental limits based on the simple assumption that errors arise purely from camera image pixilation; this is the ideal case that errors caused by lack of definition in the vehicle image, lens distortion, inaccurate calibration, blurring through motion, loss of images through occlusion, and so forth can be adequately controlled through the use of synchronized shutter operation, image redundancy of multiple cameras, careful camera calibration and checking, and error reduction through Kalman filtering. In Figure 6.1, we assume that the particular vehicle comes within distance $D = 10$ m of the camera and that the camera is mounted at a height of $d = 3$ m above the feature being tracked. The pixelation error θ is assumed to result from a field of view of 70° in the vertical sense, and with an image resolution of 640 × 480, the corresponding angle is 0.0025 rad. The resulting uncertainty in horizontal position is then:

$$x = \theta d\left(1 + \frac{D^2}{d^2}\right)$$

which turns out to be just a little less than 10 cm.

This suggests that the required tracking accuracy is feasible but only if suitable steps are taken to control errors at each step in the data capture and analysis process. This places considerable importance on the need for taking an essentially 3-D approach, for which feature heights are estimated as part of the estimation process.

Further from the intersection center, approach behavior is likely to be of some importance, although not beyond a distance of approximately 100 m. If the cameras are all based close to the intersection, the lateral and longitudinal positions of vehicles approaching are again dependent on field of view, image resolution, and baseline distance *d*, but in this case it may be horizontal rather than vertical (longitudinal position cannot be satisfactorily resolved if the camera is directly aligned with the approach lane). Then, a 10-m baseline, or offset from the vehicle approach path, assumed straight, leads through a similar calculation to a longitudinal error of approximately 2.5 m. Lateral position theoretically can be resolved to within approximately 1 m, and overall we would expect that the vehicle can be located with a precision of 2–5 m at that kind of range, without requiring additional cameras or other sensors positioned remotely from the intersection.

Design Concept

The design concept is not strictly derived from the earlier tasks, but it is strongly motivated by the positive capabilities and limitations of current approaches, as well as the most challenging system requirements, which can be loosely summarized:

- Automation: good for commercial systems but lacking in research-based systems;
- Accuracy: especially lacking in commercial systems and likely to be a continuing challenge for refinement; and
- System integration: again a weakness of all existing research-based systems.

Automation and system integration will emerge naturally within the design, so the focus here is on the accuracy problems and the shortfalls found in current system architectures. Basic video processing methods are heavily reliant on background subtraction, which easily provides a deceptively (to the human observer) elegant method to derive a trajectory:

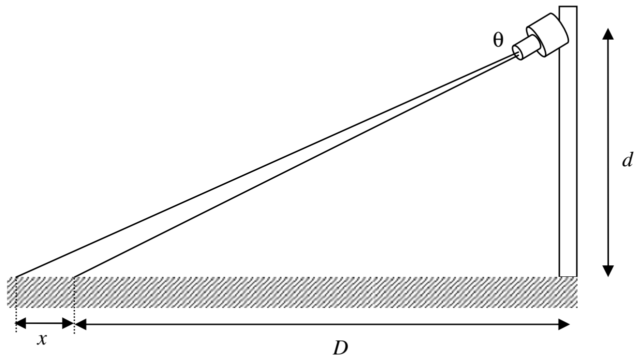


Figure 6.1. Basic camera geometry and resulting pixelation error.

Once the background is recognized, it is subtracted pixel by pixel from the immediate scene, and we can see that only the vehicles and other transient objects remain. Following the centroid of any extracted shape, we have an approximate trajectory in the camera image, and we are surely 90% of the way to a solution—the remaining 10% cannot be so hard? Unfortunately, this is not so. Used by SAVME and most other commercial systems, it supplies approximate data of transitory object locations; this is adequate to determine whether a preidentified road section is filled (e.g., for a virtual loop detector) but gives tracking results that are highly sensitive to factors such as shadows, occlusions, and the effects of a viewing angle (Figure 6.2). Here C_1 has a high-level viewpoint, whereas C_2 has a much lower viewing angle. The camera shadows shown are indicative of the parts of the visual scene that may be confounded with the actual object when viewed from that particular angle; the high viewing angle makes the image insensitive to the vertical dimension in the object (the vehicle to be tracked), and a reasonably accurate plan view of the object is captured. On the other hand, C_2 produces a stretched image of the solid object, and although this is obvious to the human observer, the background subtraction algorithm is not so smart, and it cannot be used directly to infer the 2-D (plan) view of the object from the perspective projection observed. As described earlier, SAVME countered this problem by using very high camera mounting towers, whereas NGSIM also had the luxury of high camera positions and made improvements using 3-D rigid body vehicle shapes to replicate the effect of the vertical dimension. Even the IVSS study used high camera positions although with problems of camera motion in windy conditions. In general, high camera locations are not available, and non-ideal camera angles are forced by site access. In addition, irregular or unpredictable vehicle shapes, rapid changes in perspectives, the effects of glare, and so forth make the solid models unreliable estimators.

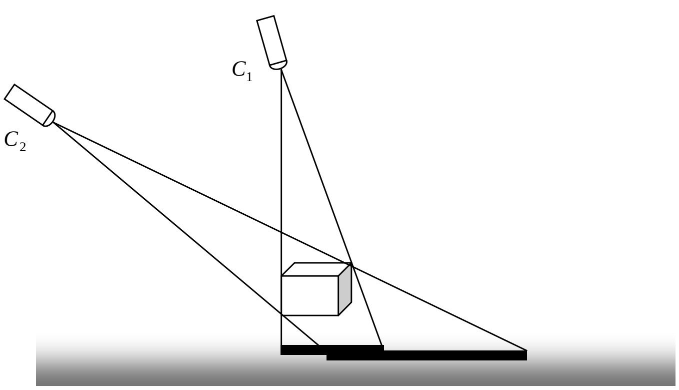


Figure 6.2. Effect of camera viewing (altitude) angle on boundary estimation.

Multicamera Feature Tracking

Based on the above information, a basic architecture for the system is presented. Enhancements are to be made by combining feature tracking and grouping with the background subtraction and doing this simultaneously from multiple cameras with overlapping fields of view. This is expected to provide greater precision in motion estimation, though with the same 2-D projection problem limits the accuracy of absolute position estimation. Another significant problem, one that is never easily solved, is the need to cluster and separate features, with each cluster uniquely associated with a rigid vehicle object in the world. The perspective of the single camera again makes this harder, and although any image-based tracker is going to face this challenge at some point, it is expected that multiple views can help reduce the problem.

It is worth comparing to the camera-domain image processing techniques, typical of most commercial and research lab systems, the architecture of which is illustrated in Figure 6.3. When information derived from video processing is incomplete or indeterminate, iteration is needed to improve and correct estimates, as represented by the closed loop. This may require manual user intervention in the estimation, and a video archive is also needed to support this iterative loop. To

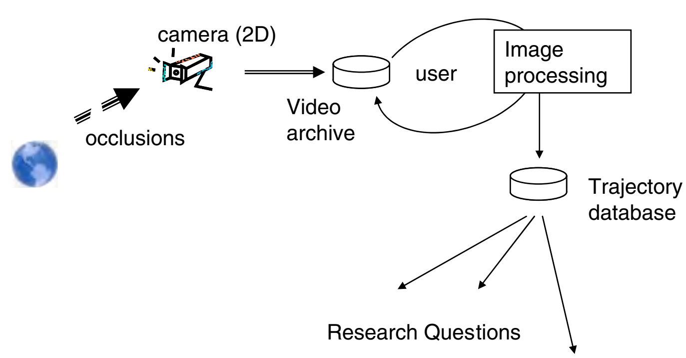


Figure 6.3. Camera-domain object tracking.

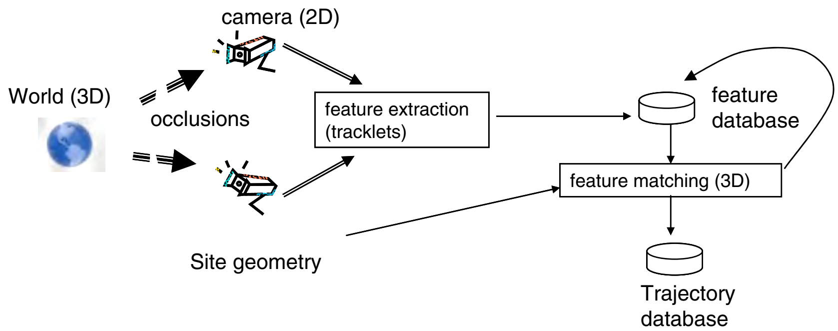


Figure 6.4. Site Observer concept.

avoid losing image quality, high volumes of video data must be recorded and retained for analysis. In the multicamera design concept, the known properties and existing algorithms of 2-D camera-domain video image processing are retained (Figure 6.4). The uncertainty in grouping and segmentation is removed from the image domain so that once processed for features, the video stream is discarded. As much basic information as necessary is extracted from the individual camera video stream, and additional compressed video is recorded to enable operator review. 3-D bounding boxes will be derived from the features and subsequently projected over the compressed video stream for visual confirmation.

The advantages of this general approach are summarized as follows:

- Only existing video image processing techniques are used.
- No uncompressed video archive is needed.
- Automated processing is built into the architecture.
- System development comprises three parallel strands of activity (see below).
- The parallel architecture is inherently expandable. There is no computational constraint on using large numbers of cameras for large and complex intersections or even networks of intersections.
- Large baseline stereographic information is directly incorporated, increasing the potential accuracy of the measurement system many times compared with that of the simple single-camera systems.

System Design

The basic system comprises multiple cameras (four are shown: C_1, C_2, C_3, C_4) that are simultaneously triggered via a real-time data acquisition system residing on a host computer that houses storage for both a database and compressed video. The compressed video is not part of the automatic system, being reserved for development and review purposes. In Figure 6.5, we mention MPEG-4 compression, but any form of compression, including subsampling of image frames or pixels, as well as JPEG compression of individual frames, could be used.

Preprocessing of the video image streams uses proven algorithms to extract features from the video input. The synchronous shutter trigger is significant, ensuring that features are extracted from different cameras at the same time, to allow feature registration at a later stage. The absolute time reference from GPS allows multiple systems of this type to be integrated at a later stage as feature databases are uploaded to a remote server. At the intersection, each camera station has a local networked computer to perform feature extraction and image compression. The results are sent via network to a host, which is also responsible for basic control functions, such as starting and stopping the local data collection and uploading the feature sets.

A large-scale system might comprise several CDAPS subsystems (Figure 6.6). In any case, the expectation is that a separate server, based at a research lab or transportation facility, will host the data, and it is there that final postprocessing into trajectories and other analysis will take place at this data center. Ideally, high-speed Ethernet is available to communicate to the data center, but if not, data upload can be performed periodically in manual fashion; in any case, without the need to capture raw video, data volumes should not be onerous. Assuming a rather generous 50 bytes per feature, and 50 features per frame, and a frame rate of 20 Hz, a system with four cameras will produce a data store that grows at less than 1 Gb per hour. This is not insignificant, but it is orders of magnitude less than the corresponding figure of 88 Gb per hour for raw video images (assuming only 1 byte per pixel monochrome images and a modest 640 × 480 camera resolution). The feature storage also does not greatly expand with camera resolution, or with the use of color, and at times of sparse traffic the storage growth drops to zero. Thus, even without a dedicated network connection, the system can be

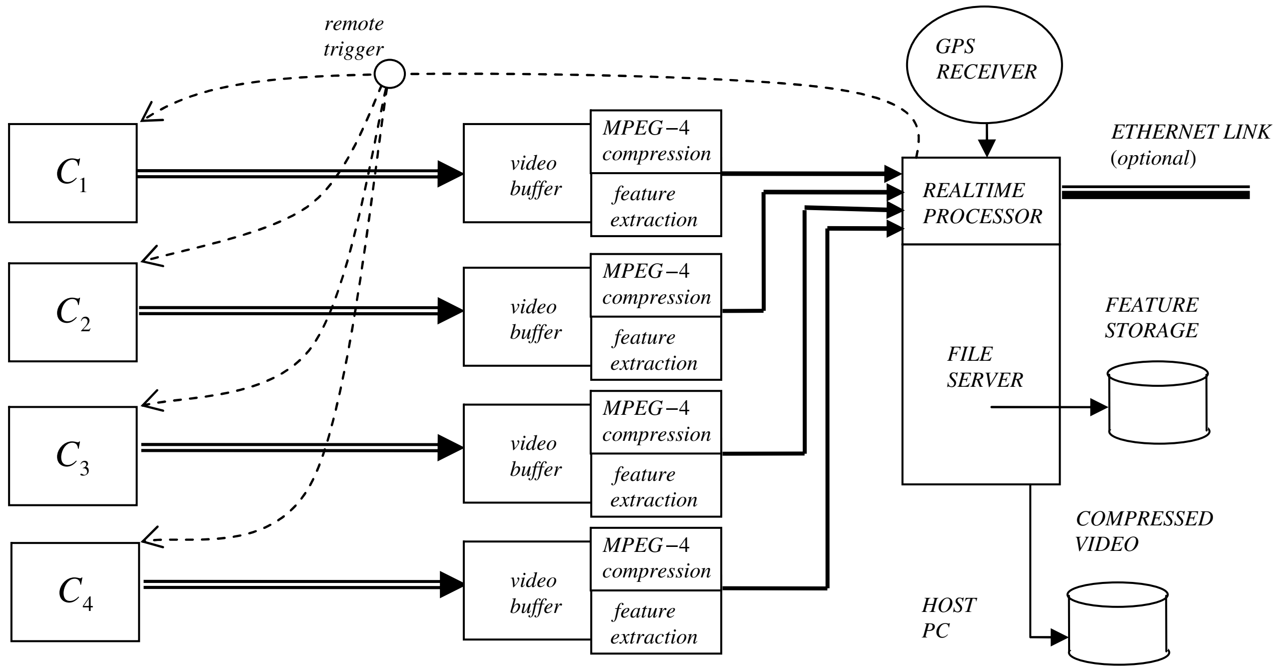


Figure 6.5. Camera, data acquisition and preprocessing system (CDAPS).

made to run for many weeks without the need to offload data, and the advantages are very clear.

The design concept has other potential advantages, not least that it builds on previous work and expertise. Challenges do exist in the design concept, but these are almost all concentrated at the level of algorithm performance at the integration stage: how best to carry out feature segmentation and grouping using stored features and how to ensure vehicle bodies are correctly located, with features merged or separated appropriately. These are not trivial problems, but they are familiar ones and depend on geometry much more than on low-level image processing, and the multicamera

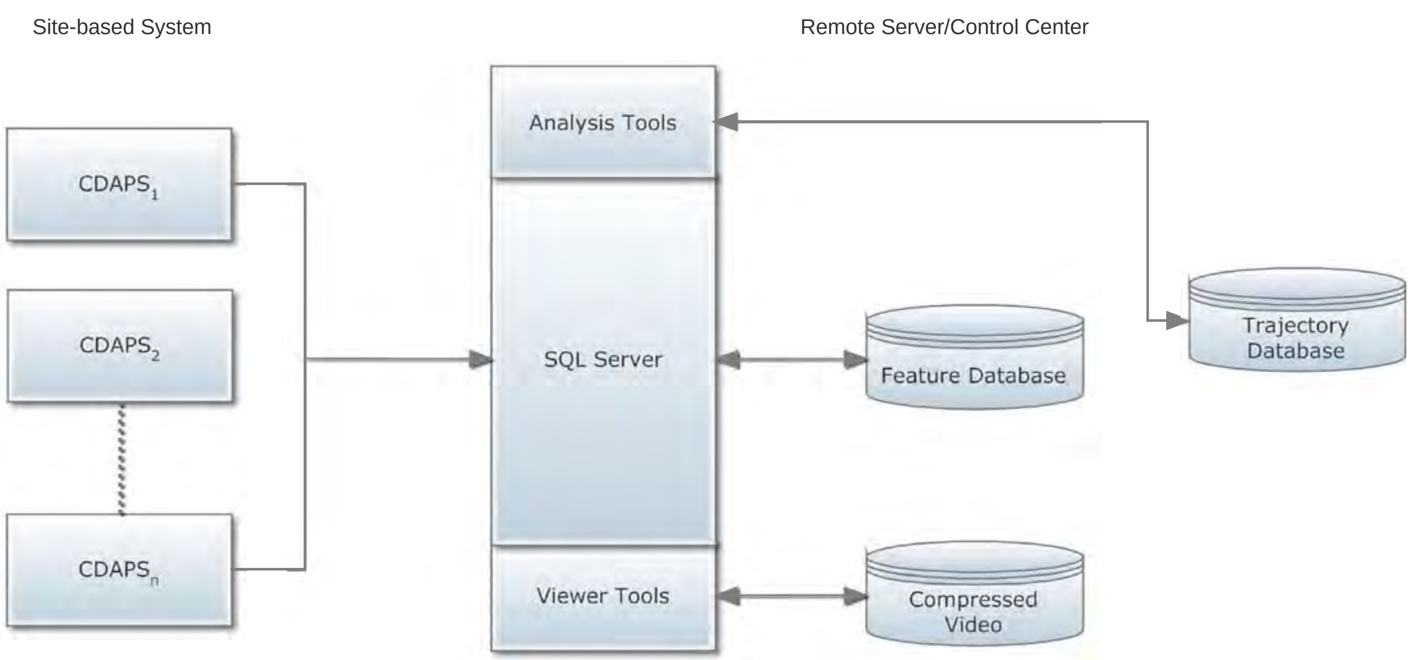


Figure 6.6. Integrated vehicle tracking system—Site Observer.

synthesis approach provides redundant information to resolve uncertainties not available for a single-camera system. Note however that, to ensure adequate coverage, the expectation is not to use stereo imaging per se; even when the same vehicle is seen in two cameras, the features recorded are normally expected to be from different parts of the vehicle. Expanding the concept to stereo vision may be possible in the future, but in this case it would be necessary to double the number of cameras.

One further point about the architecture is its suitability for migration to future hardware. New video products are becoming available all the time, and in the future it may be preferred to integrate the video preprocessing function within the camera body itself. Cameras are already available with powerful digital signal processors built in, and in many commercial vehicle-sensing systems, there is a clear trend toward locating the heavy computation within the camera. In this case, feature information may be broadcast over wireless Ethernet, greatly increasing installation flexibility. Although this is not the best choice for the development of the current robust prototype system, the option for future development in this direction is a major strength of the design concept.

CHAPTER 7

System Hardware and Site Installation

The modular CDAPS design was implemented at an intersection in Ann Arbor, Michigan; it is a suburban intersection between Lohr Road and Oak Valley Drive. Figure 7.1 shows an aerial view of the site. This chapter gives details about the site, including background reasons for selecting the particular location, and describes the component hardware and its installation.

Choice of Intersection

The choice of intersection was to provide a mix of realistic real-world challenges but without overburdening the prototype Site Observer installation with excessive complexity of traffic flow or logistical problems. As discussed, an intersection is the preferred type of location for the purpose of the pilot study. Because system installation would require many site visits, it was natural to use a location in the immediate vicinity of Ann Arbor. It should have reasonable peak traffic volumes and some history of traffic accidents and offer the possibility of conflicts taking place during pilot data collection. Ideally, the intersection should be signalized with rigid signal poles and mast arms; this helps with electrical power and camera mounts and may lead to traffic conflicts because of signal violations or late braking at a red light. Ideally, the turning priorities should include permissive turns. For installation, the site should offer acceptable accessibility, with conduits available for connecting cables. It is important that the intersection be owned by a cooperative and supportive agency, allowing access by the research team in a flexible way.

Washtenaw County Road Commission (WCRC) was contacted in the first instance; the road commission was very receptive to the requests of the research team, so no other agencies were contacted. Based on the general requirements for the study, as well as the available intersections near Ann Arbor, three locations of varying size and complexity were considered in detail. After review, including site visits to each location, the intersection of Lohr Road and Oak Valley Drive was chosen. It presented a broad central area—roughly 30 m^2 (~100 ft^2), moderate traffic volumes, easy access, electrical power, rigid signal poles, conduits for pulling cables, and convenient locations for additional cabinets to house the Site Observer computer and communications hardware. The intersection was also sufficiently typical that lessons learned could be applied to other locations.

Traffic Flow and Crash History

The traffic signal system at the intersection uses cameras as virtual loop detectors, and these were used by the WCRC to count traffic movements over a 24-h period. The count shown in Table 7.1 yielded a total of 12,000 vehicle movements, which indicated that this location could provide rich data without pushing the limits of the automated system. The aim of the pilot study was to collect data for at least 24 h total, and by limiting the collection time to daytime only, it was anticipated that approximately 15,000 vehicle trajectories would be captured during the pilot study.

A review of the crash history of the test intersection found that during the preceding 5 years, 10 police-reported crashes occurred within 100 m of the intersection. Figure 7.2 shows the locations and crash types. Most were rear-end crashes; two involved impacts during a turn; and one crash was a head-on. The posted speed limits are as follows: 40 mph on the north and west legs, 35 mph on the east leg, and 45 mph on the south leg. As seen in Figure 7.1, all left turns have dedicated lanes; the approaches from the north and south have protected left turns, whereas the east and west legs have permissive left turns, with the turning vehicles giving way to oncoming through traffic. Right turn on red is permitted from all approach directions.

Site Survey and Terrain Map

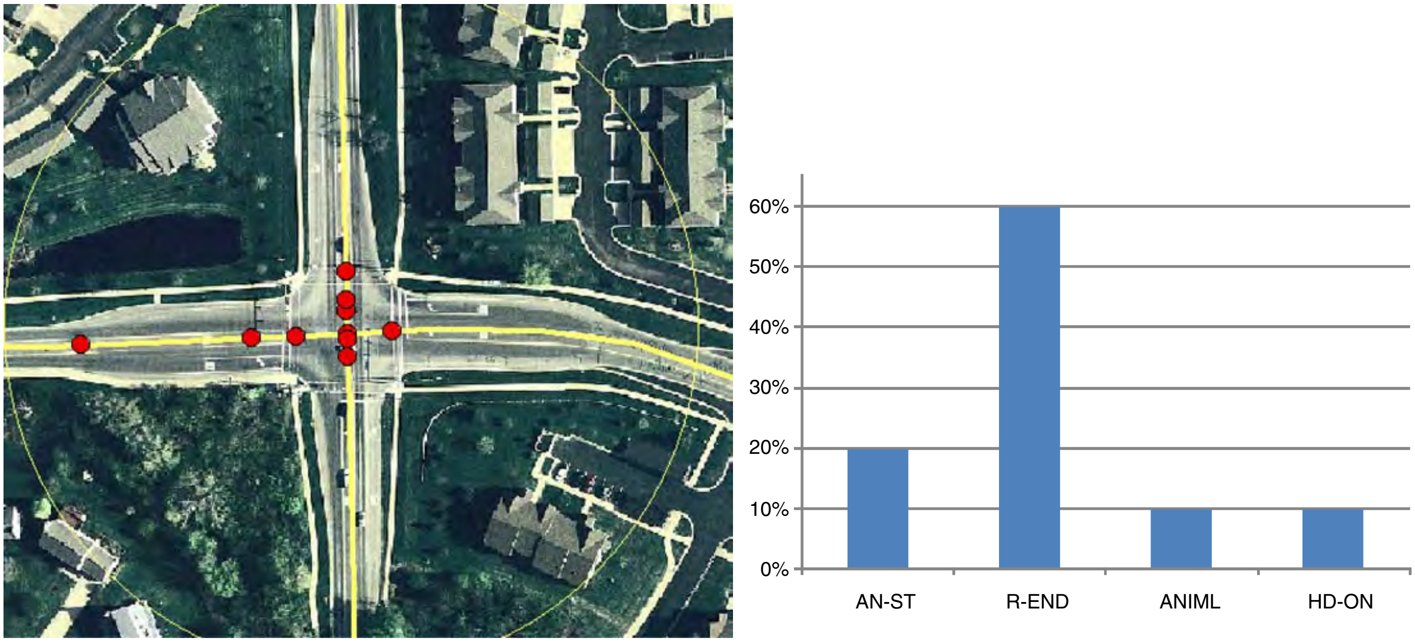
Figure 7.3, a ground level photo of the site looking out along the northern leg, clearly shows that the terrain is not flat. This

Sources: Imagery © 2012 Geo Eye; map data © 2012 Google.

Figure 7.1. Intersection satellite image.

Table 7.1. Turning Counts at the Pilot Intersection, May 4–5, 2009

Turning	South Leg			North Leg			West Leg			East Leg		
Counts	**Left**	**Through**	**Right**	**Left**	**Through**	**Right**	**Left**	**Through**	**Right**	**Left**	**Through**	**Right**
24-hr total	1,790	2,757	120	679	2,306	173	248	689	2,144	163	719	351



Note: AN-ST = angle-straight; R-END = rear end; ANIML = Involving an animal; HD-ON = head-on.

Figure 7.2. Crash locations (left, red dots) and types (right).

presents obvious problems for tracking accuracy; if the camera tracks a vehicle at some distance (for instance, 50 m) from the intersection and the ground is assumed flat, the distance error will be potentially very large, especially if the camera does not have a high mounting position. Thus, it was considered essential to model the terrain and use the model to establish coordinate transformations between ground and 3-D world coordinates. To achieve this, a high-resolution lidar survey was commissioned. Not only could the road height be found and included in the various coordinate transformations, but there also were two other important advantages resulting from the survey:

- Any static features and reference points visible in a camera image that are also found in the lidar survey can be used for camera calibration, either immediately after installation or later during data collection in case the camera position shifts.

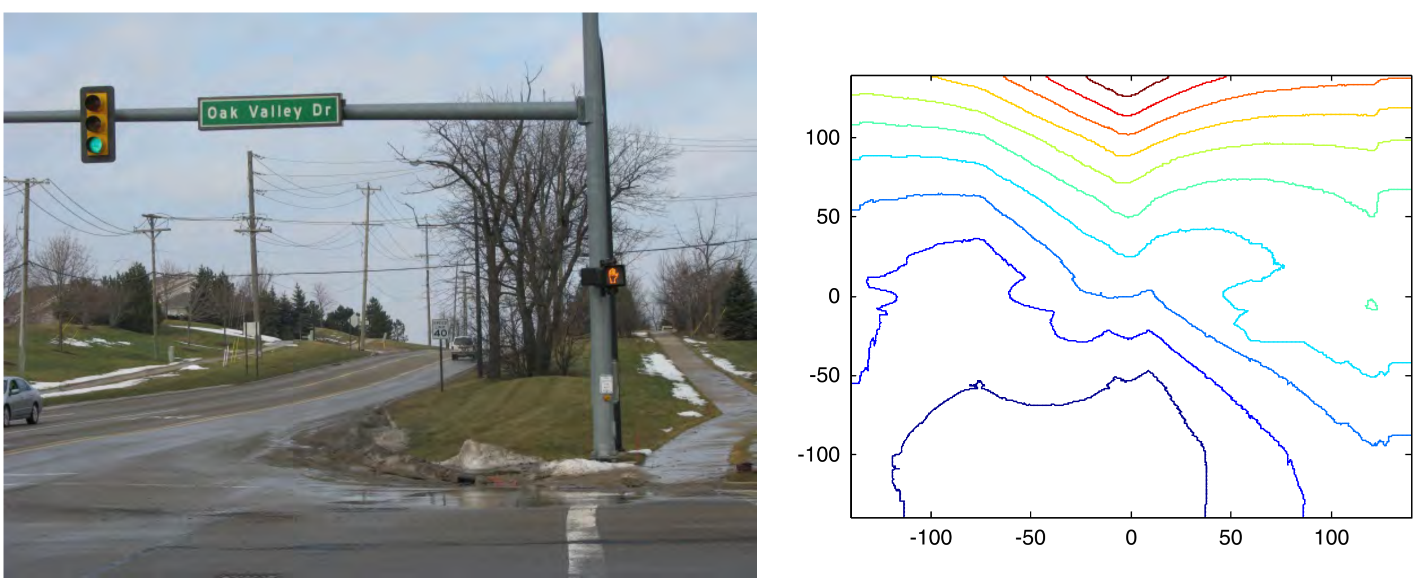


Figure 7.3. View of the pilot intersection and contour map of surface heights.

- Even before installation, virtual camera views can be used to optimize the camera installation by experimenting with different mounting positions, direction of camera view (camera pose), field of view, and camera aspect ratio.

Figure 7.3 includes a contour map of the terrain data extracted from the lidar survey. Because the survey was restricted to the road and its immediate surroundings, the terrain map is only approximate when considering points far from the paved road. This limitation, however, does not affect the geometric effect of road height on feature location in camera images.

The lidar survey was conducted by Midwestern Consulting, based in Ann Arbor, Michigan. It provided more than 100 million data points as a "point cloud." MATLAB routines were written to interrogate the point cloud, explore virtual camera locations, and generate virtual camera views before installation. Figure 7.4 shows an overall virtual view provided by Midwestern Consulting. The left image in Figure 7.5 shows a virtual scene from supplied software, Leica TruView, operating as a plug-in for Microsoft Internet Explorer. The right image shows zoomed-in detail of a particular point of interest, in this case one of the traffic signals. Figure 7.9 presents one of many virtual camera views generated from the lidar survey, typical of those created before cameras were installed.

Many key geometric points were found, even before cameras were installed. Coordinates used in the survey are shown for a test point on the top of the mast arm in Figure 7.5; coordinates are given in meters and based on a projected coordinate system

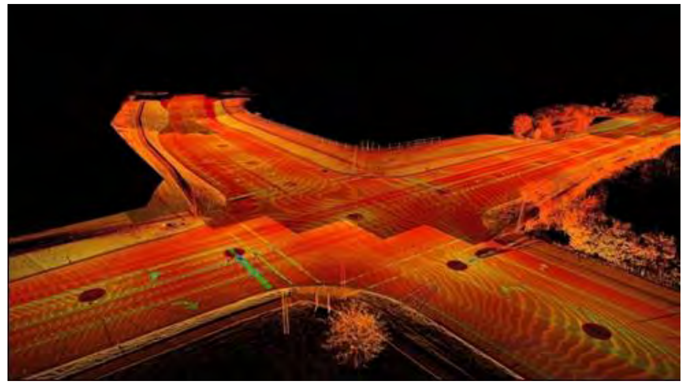


Figure 7.4. Virtual lidar survey image of the test intersection.

(NAD 1983 Michigan GeoRef, centered at latitude 45.30916667, longitude −86.00000000). An arbitrary local origin was selected for the intersection at 4050129, 81956, 269, which is close to the geometric center of the intersection, and local positions were calculated relative to this. All 3-D world positions are measured as (*x*, *y*, *z*) in meters in directions (east, north, vertically upward), respectively, relative to the local origin.

Hardware and Site Installation

A four-camera CDAPS system was installed at the intersection. The cameras chosen were Firewire units (DragonFly2) from Point Grey Research. At each corner of the intersection, a new

Figure 7.5. Virtual view from lidar survey using Leica TruView software. Right: detail of reference point on traffic signal used for camera calibration.

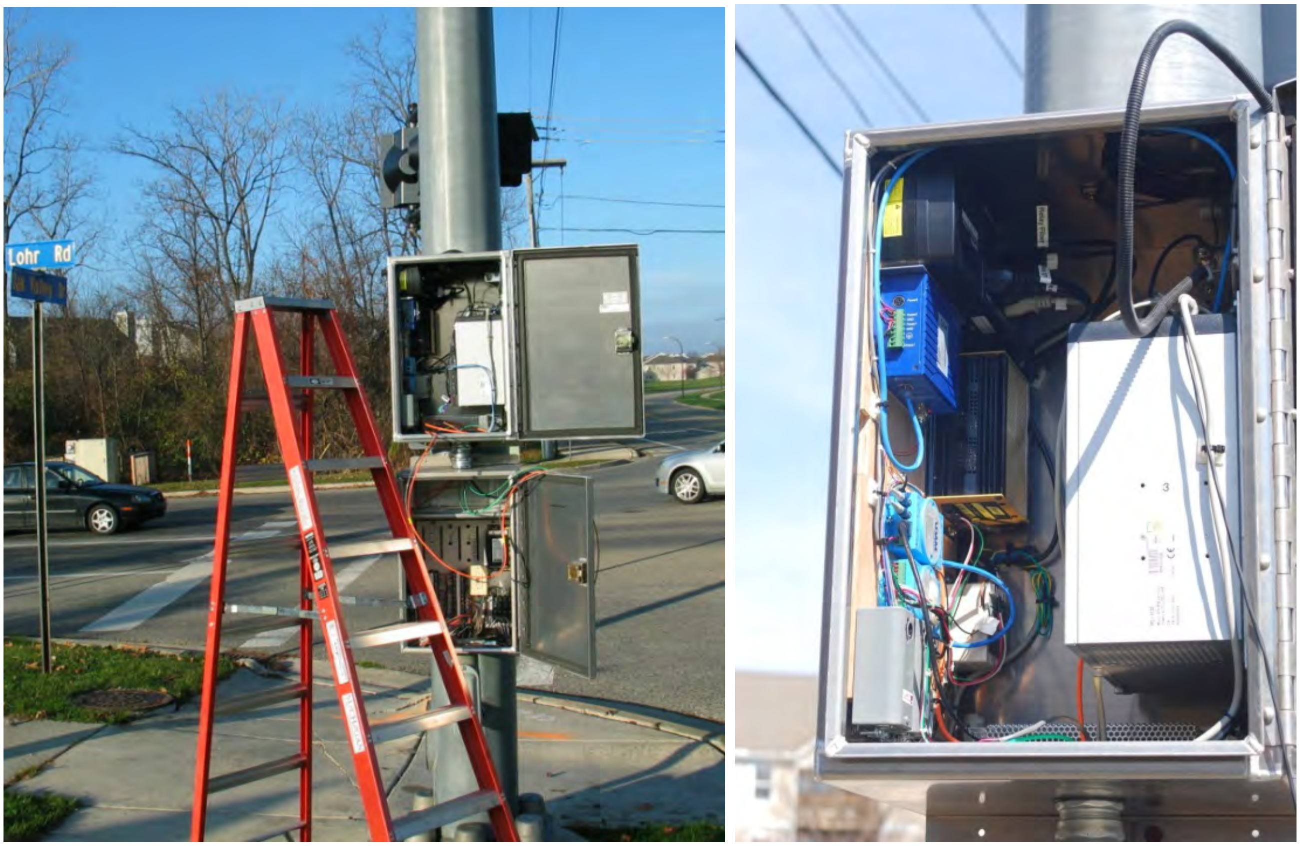


Figure 7.6. Local camera station cabinet.

cabinet was fitted to the signal mast immediately above an existing cabinet, which provided 110 V electrical power (Figure 7.6). The hardware for each of the camera stations is summarized in Table 7.2. The GPS transceiver was used to trigger the camera at accurately timed 20-Hz intervals, and the capability for external triggering was one reason this camera was selected. Another useful attribute of this camera is the ability to select a subrectangle inside of the 1024 × 768 full-frame image. The four installed cabinets also contained battery backup (UPS) in the event of power failure and heating/cooling fans. The central (host) computer was mounted in the main control cabinet, shown in Figure 7.7, before installation.

In Figure 7.7, the signal load switches are seen at the rear of the cabinet. This cabinet offered ample room for installing the

Table 7.2. Camera Station Major Hardware Components

Camera	**DragonFly2**
Camera enclosure with heater/fan	
Firewire to fiber converters (2)	Firenex-MX
TTL converters (2)	SITech 2817-R
GPS antenna	Laird GPST821
12 V power supply (2)	
TCP/IP relay module	ADAM6066
Ethernet switch	StarTech SV11071PEXT
Ethernet to fiber converter	Elinx EIR-G-SFP-T
Camera station computer	

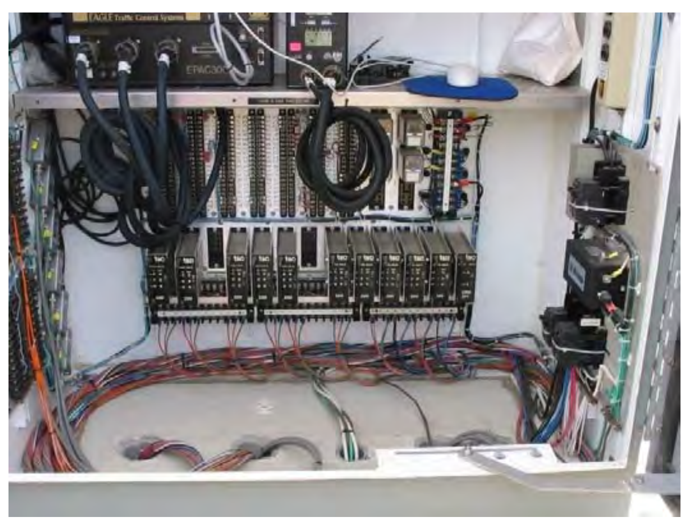


Figure 7.7. Main cabinet before installation of host computer system.

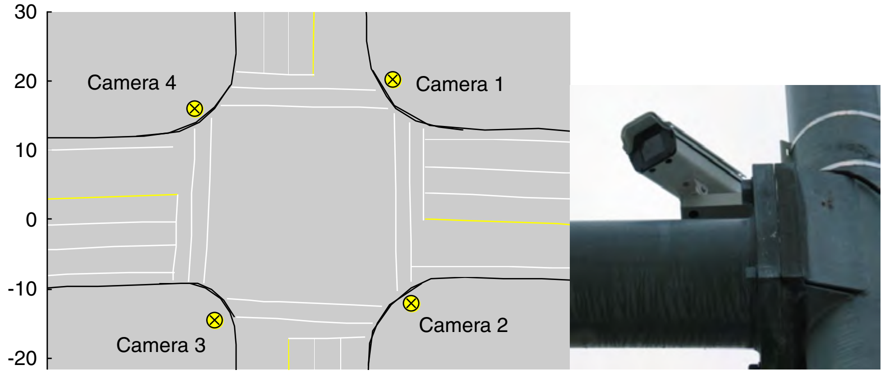


Figure 7.8. Camera positions. Left: Plan view. Right: Mounting positions are on traffic signal mast arms, close to the signal mast (Camera 3 is shown here).

main host computer. Camera image collection included the addition of frame-by-frame time stamps—the GPS trigger time in milliseconds. Each camera was connected to its local computer using optical fiber; fiber was also run under the road in conduits to the host computer. Optical fiber was preferred because of the need for electrical isolation in the event of lightning strikes and for the camera connections it avoided the need to add Firewire transceivers to extend beyond maximum Firewire cable lengths.

Initially, the plan was to use pan and tilt units for camera control and thus fine-tuning of the camera views. However, the motorized units tested were insufficiently rigid and could not be locked mechanically; when power was turned off, the unit could not accurately maintain a controlled position, and even modest loads caused camera motion. Given the importance of repeatable calibrations, it was decided to use fixed mounts. Then, to avoid expending time and effort making manual adjustments to the installed cameras, the lidar virtual views were used to adjust all the main geometric parameters of the cameras before installation: location, pose, focal length (zoom level), and image size were all selected in this way.

The starting point was the camera location and especially its height. As can be seen in Figure 7.5, the signal masts have a main mast arm for the signals and a higher arm for lighting and supporting the existing traffic signal control cameras. The lower arm has a height of approximately 6 m (19.7 ft), whereas the upper arm allows a camera to be mounted at 9.2 m (30.2 ft) above the ground. The lower mounting was chosen to be more representative of what could be expected at a typical installation and not to avoid the geometric challenges anticipated for the Site Observer in general. The extra rigidity of the lower mount was also preferred, although some camera motion (mainly side-to-side) was noticed in strong winds as variable wind loads on the mast arm caused the main mast to twist slightly. The locations of the cameras are shown in Figure 7.8.

Once locations had been set, the ground coverage became the main concern. Figure 7.9 shows the full (1,024 × 768) image view of the north leg, taken from the southwest camera

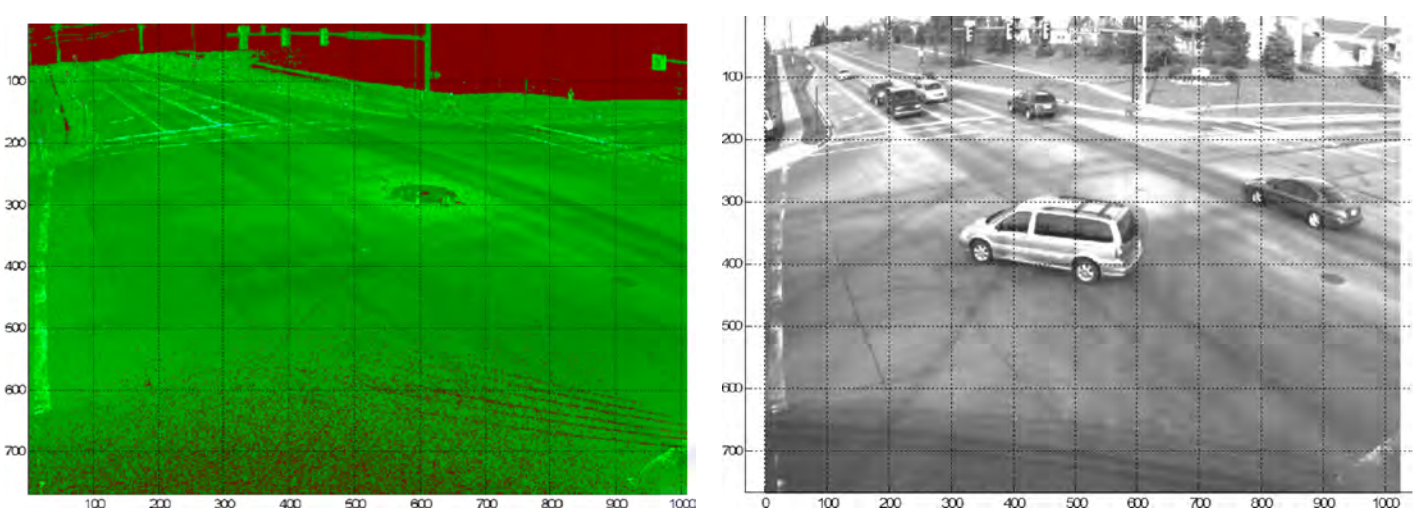


Figure 7.9. Virtual and actual camera views (SW camera).

Figure 7.10. Cropped image (992 × 304 subrectangle, SW camera).

location. Early tests with a 640 × 480 image size indicated that this likely presents a realistic upper limit on the number of pixels to be processed in something approaching real time, but there was still free choice of the aspect ratio. The key requirement was to maximize the useful part of the image (seeing more road and vehicles and less sky and foliage) and also maximize coverage, especially multicamera coverage within the intersection. This led to the choice of a 992 × 304 subrectangle, giving the camera view shown in Figure 7.10. Roughly speaking, each camera covered the opposite leg, most of the intersection center (but omitting the immediate foreground), and the intersection entry on a second leg toward the right field of view. Final adjustment could be carried out using the virtual pan and tilt offered by the choice of subrectangle in the camera; although most of the available width was used up, the tilt control could be used to capture distant points on the opposite leg, without wasting pixels showing areas of sky.

The overall coverage is mapped in Figure 7.11. It is seen that approximately 50% of the central square of the intersection has four-camera coverage. Only right-turning vehicles will avoid being seen simultaneously by at least three cameras at some time when passing through the intersection. Ideally, to improve the multicamera views, it would have been preferred to increase this so the minimum was three cameras for all paths through the intersection, but given the limited number of observation points and pixels for image analysis, this was thought to be the best compromise available.

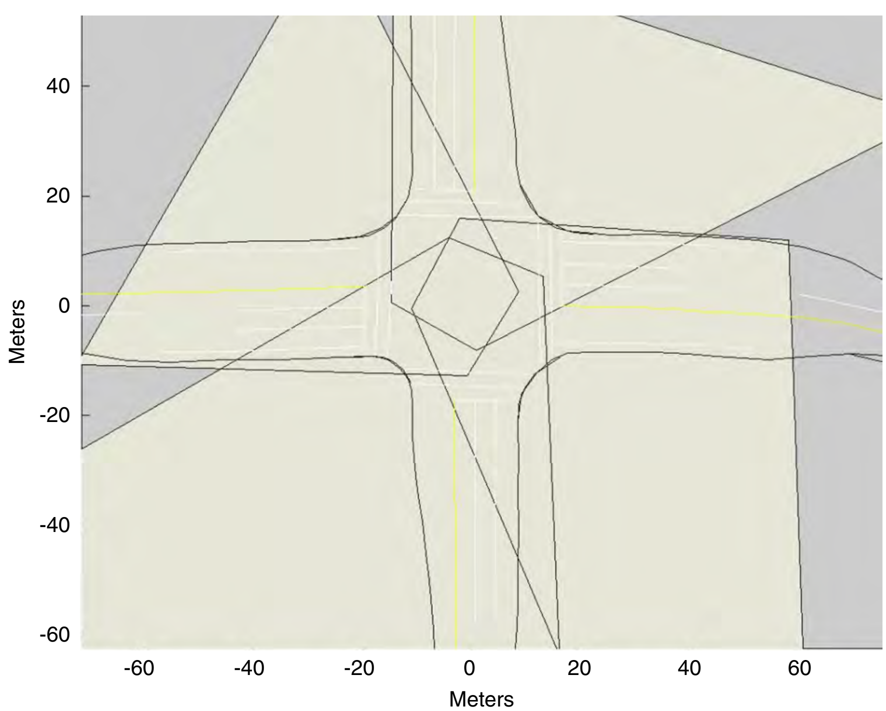


Figure 7.11. Overall camera coverage.

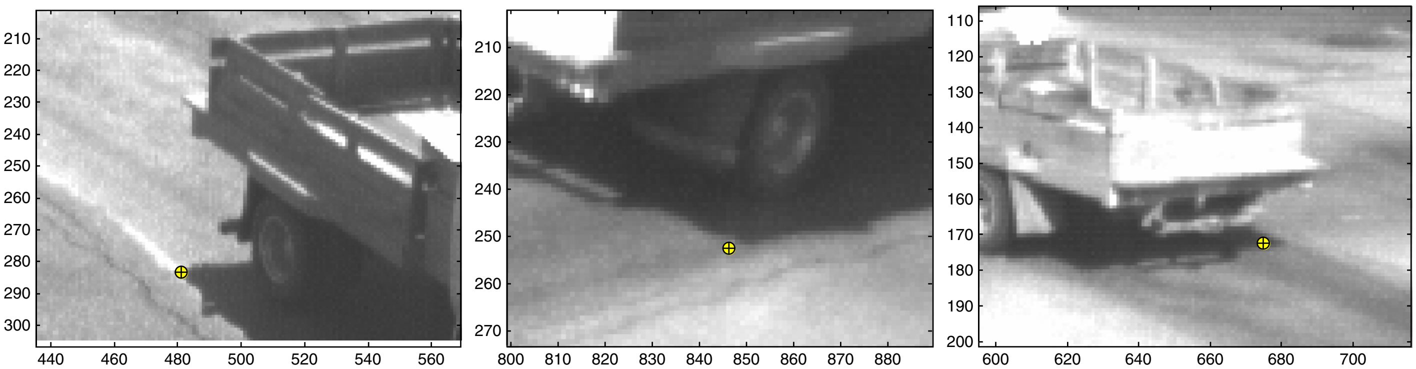


Figure 7.12. Simultaneous images (zoomed in) from the NE, SE, and SW cameras. The marked point is coincident in time and in 3-D world coordinates. Scales are pixel coordinates.

From Figure 7.11, it is expected that most through vehicles would be seen by four cameras, and this was indeed the case. Figure 7.12 shows detail of another vehicle, moving at approximately 40 mph through the intersection, as seen by three of the cameras. In fact, the NW camera also captured the vehicle, but the point of interest here is the corner of the shadow; the small yellow marker is a single world point, projected into the three camera views (for the NW camera this feature is obscured by the vehicle body).

Although the match is not absolutely perfect, the overall localization of the shadow corner confirms the validity of a number of parameters included in the transformations: camera intrinsic calibration (the way the camera captures and distorts points in the world), the extrinsic calibrations (precise location of where the camera is mounted and the angles of pan, tilt, and roll), and the validity of the height map and shutter synchronization. This set of images confirms the basic correctness of the coordinate transformations, as well as the multicamera synchronization and area coverage.

Additional hardware is shown in Figures 7.13 and 7.14. The main host computer collected signal states at 20 Hz using photodetector modules mounted to the load switch LEDs. WCRC personnel could remove these connectors if required, and sufficient LEDs were left unobscured to allow visual inspection of individual switches in the event of any reported fault. Importantly, there was no electrical connection between the Site Observer and the traffic control system.

Figure 7.13. Optical sensors attached to load switches in the main signal cabinet.

The radar subsystem (Figure 7.14) was connected during only one data collection experiment, lasting approximately 1 h. The CDAPS system includes a CAN interface allowing automotive (adaptive cruise control) radar to be connected. The left image shows the temporary nature of the installation, and the right image shows the field of view for each of the four radar units. The purpose of the radar installation was to provide an independent check on kinematic variables (range and range rate). It is easy to see why these types of radar are wholly inadequate for tracking; their narrow field of view makes it impossible to achieve adequate spatial coverage. Again, it was useful to have lidar survey data to plan the installation and also take account of the height variations in the road; the required azimuth and elevation angles were set beforehand, making installation relatively quick and easy. In fact, the validation experiment was completed in approximately half a day.

Finally two other sets of equipment were used for test and validation. First, two instrumented vehicles, Honda Accords from the Integrated Vehicle-Based Safety System (IVBSS) project (UMTRI, 2010), were used during a 30-min experiment. During this time, a small number of vehicle interactions were orchestrated (for example, with one vehicle turning right into the same path as the other vehicle, causing the driver of the second vehicle to brake). The purpose was to test motion capture under representative conditions, not to create artificial conflicts. This data set was for trajectory validation and is not included in any statistical data analysis.

The second set of equipment was the traffic signal camera system owned by WCRC and used for making turning count estimates (as well as supplying signal state information via the load switches). The system estimated vehicle numbers by path taken at 5-min intervals, to allow high-level comparison with the Site Observer. When differences are found, compressed video can be used to assess whether the Site Observer is missing vehicles or possibly duplicating them.

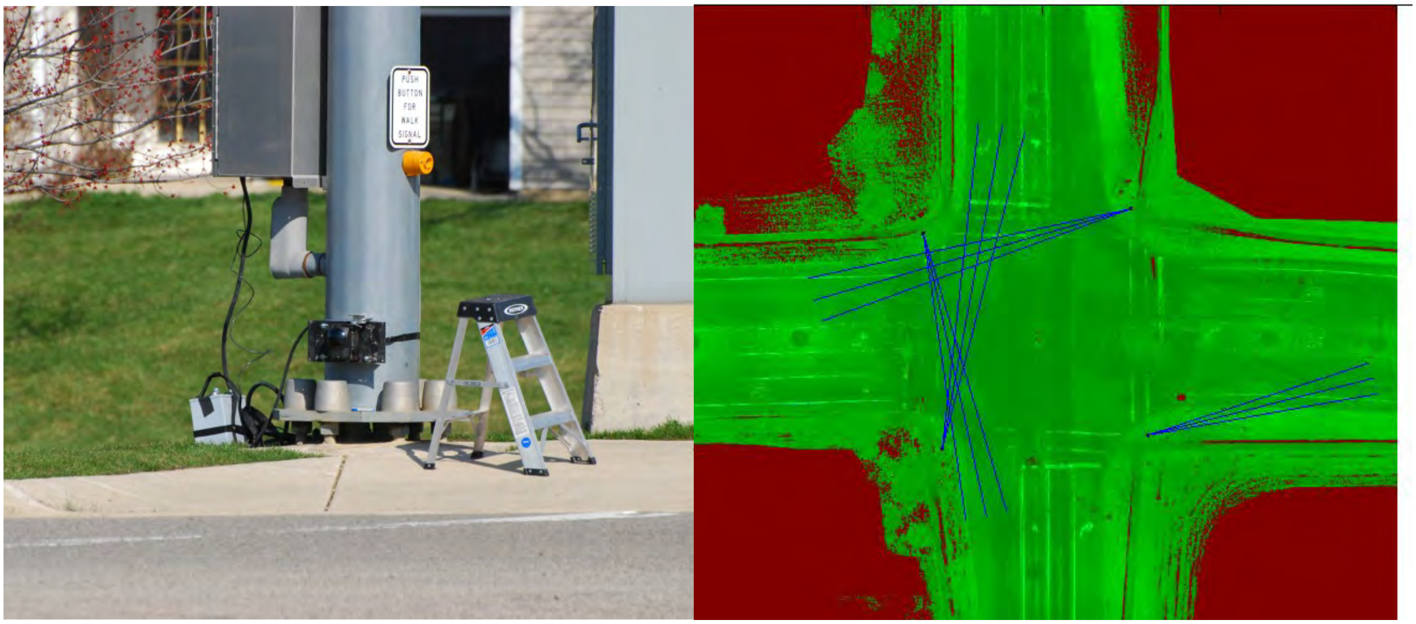


Figure 7.14. Radar installation for test and validation.

CHAPTER 8

Image Processing and Feature Extraction

Image processing generates a rich set of features that are later to be interrogated to identify vehicles and their kinematics. The image processing for feature extraction takes place at the level of an individual camera system, and there is no intention to fuse information from raw images. For this process, the input is a sequence of camera images, and the output is a set of geometric features in camera coordinates. Two types of feature were extracted, one based on corner features and one based on background subtraction. The processing steps are indicated in Figure 8.1. The extracted features are numerical arrays of camera coordinate points and velocities, as well as associated times and object identifiers. Compressed video is also extracted for review but is not used in any further numerical analysis; all image analysis takes place with the uncompressed video images, but once the features are extracted the raw images can be discarded.

Once extracted, feature data are stored locally and then sent over the Ethernet link to the main intersection host computer. Again, once they are stored centrally, there is no need to retain the features on the local computer. In this way, the local camera system can run in extended data collecting sessions without storage limits. The image processing algorithms operate at a low level and are fully automated. The key requirement for these algorithms is to extract sufficient information to isolate and track individual vehicles but to avoid storing excessive unwanted features.

In the process, high-volume image information is removed, and the result is a distilled version of what was visible in the image streams. The core image processing algorithms, developed by the UC Berkeley team, employ a number of important attributes that facilitate postprocessing into vehicle trajectories.

Background Subtraction

The background subtraction algorithm detects objects based on the intensity change of each pixel. Most existing vehicle detection and tracking systems are based on this algorithm, but here it plays more of a supporting role. The algorithm requires relatively low computation time and shows robust detection in good illumination conditions. In Figure 8.2, the right image shows how background (black) is separated from foreground objects (blue); the moving vehicle and the bicycle are sufficiently different in pixel intensity from the stored background image so that the foreground region is identified. This works reliably only if the background image is identified correctly, despite lighting changes and passing objects. (Note that Figure 8.2 was obtained from an intersection in California; it was not part of the Michigan data collection.)

There is a well-known problem with background subtraction, in that parked or otherwise stationary vehicles in the early images become part of the background image, and it may take several minutes before this aberration can be removed. In Figure 8.2, the parked white vehicle is part of the background, so when it moves there will be a false foreground object for some time. In the current analysis, this type of error is reduced by the interaction between background subtraction and feature tracking. In general, background subtraction can suffer from problems caused by occlusions, shadows, glare, and other sudden illumination changes, and such aberrations make background subtraction unsuitable for direct use in tracking applications.

Clusters and Cluster Tracking

Corner features may be detected all over any complex image. To limit the search, only foreground regions are analyzed, speeding up image processing. This is the first part of the interaction between background subtraction and feature tracking. As illustrated in Figure 8.3, nearby corner features have similar velocities and are grouped into clusters. As the object moves across the image, individual corner features may be lost or new ones detected, but for extended periods of time the cluster may be tracked. The mean position of the cluster may jump slightly as the number of component features changes, but the associated mean velocity tends to remain stable.

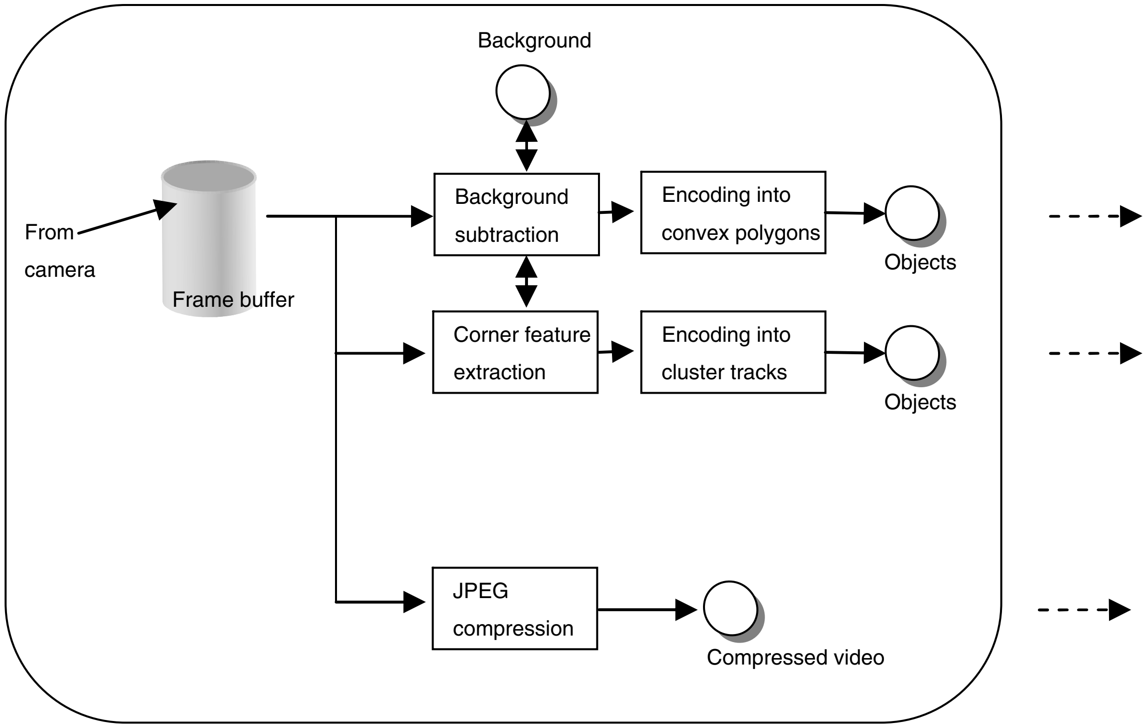


Figure 8.1. Image processing steps at an individual camera station.

The second part of the interaction is that regions with tracked clusters are not used for updating the background image, so even if a vehicle stops for a long time, provided the clusters continue to be recognized there is no danger that the background will be corrupted by its presence. The key point is that clusters are detected only in foreground regions, and as they are tracked across the image they are excluded from the background; the result is stable and tends to reduce errors in both cluster detection and updating of the background image.

Occasional errors do occur. For example, when a shadow or a glare line moves across another line (such as a fence rail) in an image, a resulting moving corner feature may be detected. It normally disappears soon after, but it can happen that the corner becomes confused with a stationary corner feature (e.g., the intersection between a fence rail and a fence post) and then remains in the image for some time. The use of clusters, rather than individual corner features, tends to exclude such aberrations, but certainly some erroneous clusters can find their way into the recorded data set. This provides a challenge for postprocessing; however, the cluster sets are largely free of corrupt data, which makes the whole approach feasible.

The corner feature grouping and tracking algorithm is implemented in a dynamic way, operating frame by frame, thus making it suitable for the real-time processing. In Figure 8.3, grouping into clusters is shown. The algorithm includes a cluster track generation algorithm to connect fragmented corner trajectories into a continuous cluster track. Note that the paths seen in Figures 8.2 and 8.3 are only determined in the camera frame; the human observer is deceived into seeing these as tracks in the world, but in reality *the feature height*

Figure 8.2. An example video image (left), a corner feature detection and tracking result (center), and a background subtraction result (right).

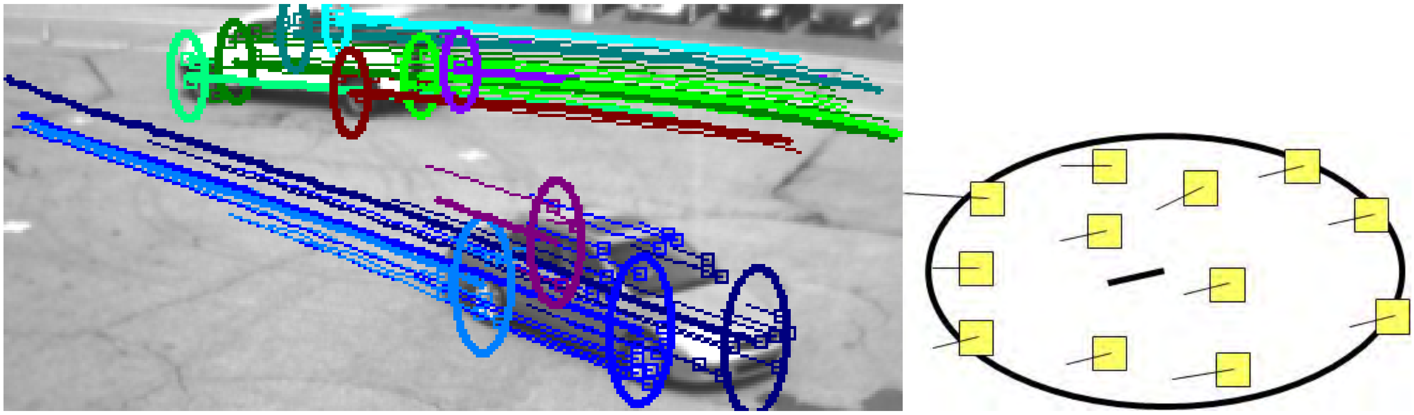


Figure 8.3. The clustering algorithm uses corner feature grouping and is sensitive to distances in the image. Left: Feature tracks and clusters on two vehicles (parking lot test at UMTRI). Right: Dispersion and mean velocities are recorded as part of the cluster properties.

must be known before that inference can be made. The cluster tracks are not located on a vehicle bounding box; again, the human observer interprets what is not known to the algorithms. At this stage in the process, there is no vehicle trajectory, only data that can be used to construct one.

In summary, a combination of corner feature detection, dynamic grouping and tracking, background estimation, and interaction between these processes provides a robust foundation for extracting essential information from the video stream.

Feature Extraction and Data Representation

Cluster tracks are immediately available for export to the site host computer; Table 8.1 shows the format of the data tables. The *site* refers to the camera used (NE corner in this case), *RunId* is the data collection run, *ProcessId* identifies the parameters used during the feature extraction, *FrameTime* is GPS time in milliseconds, and *ClusterId* is the identifier for the tracked cluster. The principal data elements are then (*X, Y*) coordinates of the cluster center, whereas (*XVelocity, YVelocity*) are the velocities determined by averaging the component feature motions; these last four columns are in units of camera pixels. Additional fields exist, in particular the size and orientation of the bounding ellipse (as seen in Figure 8.3) and flags for whether the cluster was actually present and whether the cluster was deemed stable at that time. Note that, during a short occlusion, clusters are allowed to persist by extrapolation based on their velocity in the image.

Unlike the cluster tracks, foreground regions ("blobs") are image based; certain pixels are filled, others are empty. To turn this pixelated information into something more compact, a polygon was fitted to each blob. In general, the blobs are convex in shape (any two points inside can be joined by a straight line that stays within the blob) and therefore a *convex hull* was determined; this is a minimum sized bounding polygon that is also convex. Thus, the blob is represented by a series of coordinate values, together with an index to identify the blob and another index to identify the particular vertex. Blobs can then be recreated in the image by joining consecutive vertices for a fixed *BlobId* (see Table 8.2).

Figure 8.4 shows three camera views taken at the same time, with cluster positions overlaid and blobs shown alongside. All except one cluster is attached to a vehicle, the erroneous case being attributable to glare. Several blobs are also erroneous, and generally the variable quality of blobs makes them suitable for use only with extreme caution. In the next chapter we will see that despite this, retaining blobs for data fusion offers a major advantage for vehicle localization and fusion between cameras. Figure 8.5 displays an image of pedestrian-generated cluster tracks.

A significant part of the software development was to take the core image processing routines and encapsulate them in real-time code for implementation in the CDAPS environment. The real-time code can be run in one of four modes:

- Full streaming: Image data are taken from the camera frame buffer and processed sequentially. This is the full real-term version of the system and is sensitive to image size and numbers of clusters that need to be processed.
- Data streaming: Full-resolution video images stored on a server are used as input to the processor; the processor controls the speed of the image stream, but of course this requires an image library. Apart from image input control, the process is identical to the full streaming mode.
- Image capture: In this mode, the system operates as a video image recorder; uncompressed images are stored on a hard

Table 8.1. Sample Cluster Data Extracted from Image Processing

Site	RunId	ProcessId	FrameTime (ms)	ClusterId	X	Y	XVelocity	YVelocity
0	122	7	242,064,800	374916	157.2	14.2	–12.00012	–3.999996
0	122	7	242,064,800	376791	162.6079	20.00808	–26.37817	–6.5411
0	122	7	242,064,800	380833	412.8333	162.6667	90	60
0	122	7	242,064,800	381780	144.3333	6.333333	–13.33344	0
0	122	7	242,064,800	383261	358	152.2727	87.27295	47.27264
0	122	7	242,064,800	388346	256.5	97	20	11.66672
0	122	7	242,064,800	388407	303	96.4	27.99988	6.666718
0	122	7	242,064,800	394415	386.6	161	84.00024	56.00006
0	122	7	242,064,800	396188	435.6667	179	117.7777	57.77771
0	122	7	242,064,800	396386	394	72	80	13.33328

disk, to be uploaded later to a server, and feed the data streaming process.

- Dual mode: In this mode, the camera spools image data to the local hard drive; once sufficient data have been obtained, or the hard drive in nearly full, the system switches to data streaming mode. In this way, the system swaps between waking and sleeping cycles and is just a combination of the previous two modes of operation.

The real-time system software was based on the UMTRI DAS (data acquisition system) platform, developed over several years during previous projects involving real-time data acquisition and control. For this pilot project, with the need for multiple tests using a fixed set of images, it was most convenient to run in the combined data streaming/image capture modes. Thus, uncompressed images were stored on a server at UMTRI, and the various processes ran from there. All hardware and software is compatible with the computer infrastructure at the intersection. Benchmarking showed that the system could run in full streaming mode if the number of pixels were effectively reduced by decimating by a factor of 2 in the horizontal direction; reducing the number of pixels speeds up processing. At the time of operation it was not possible to fully run in this mode because images eventually are lost when the rate of image processing falls below the capture rate of 20 Hz.

As mentioned above, with accurate transformations defined to map between world and image coordinates, it is possible

Table 8.2. Sample of Blob Feature Data Values

Site	RunId	ProcessId	FrameTime (ms)	BlobId	Vertex	X	Y
0	122	7	242,441,850	13	0	228	103
0	122	7	242,441,850	13	1	228	107
0	122	7	242,441,850	13	2	235	114
0	122	7	242,441,850	13	3	237	115
0	122	7	242,441,850	13	4	249	118
0	122	7	242,441,850	13	5	255	118
0	122	7	242,441,850	13	6	279	116
0	122	7	242,441,850	13	7	283	114
0	122	7	242,441,850	13	8	285	112
0	122	7	242,441,850	13	9	286	110
0	122	7	242,441,850	13	10	287	105
0	122	7	242,441,850	13	11	287	99

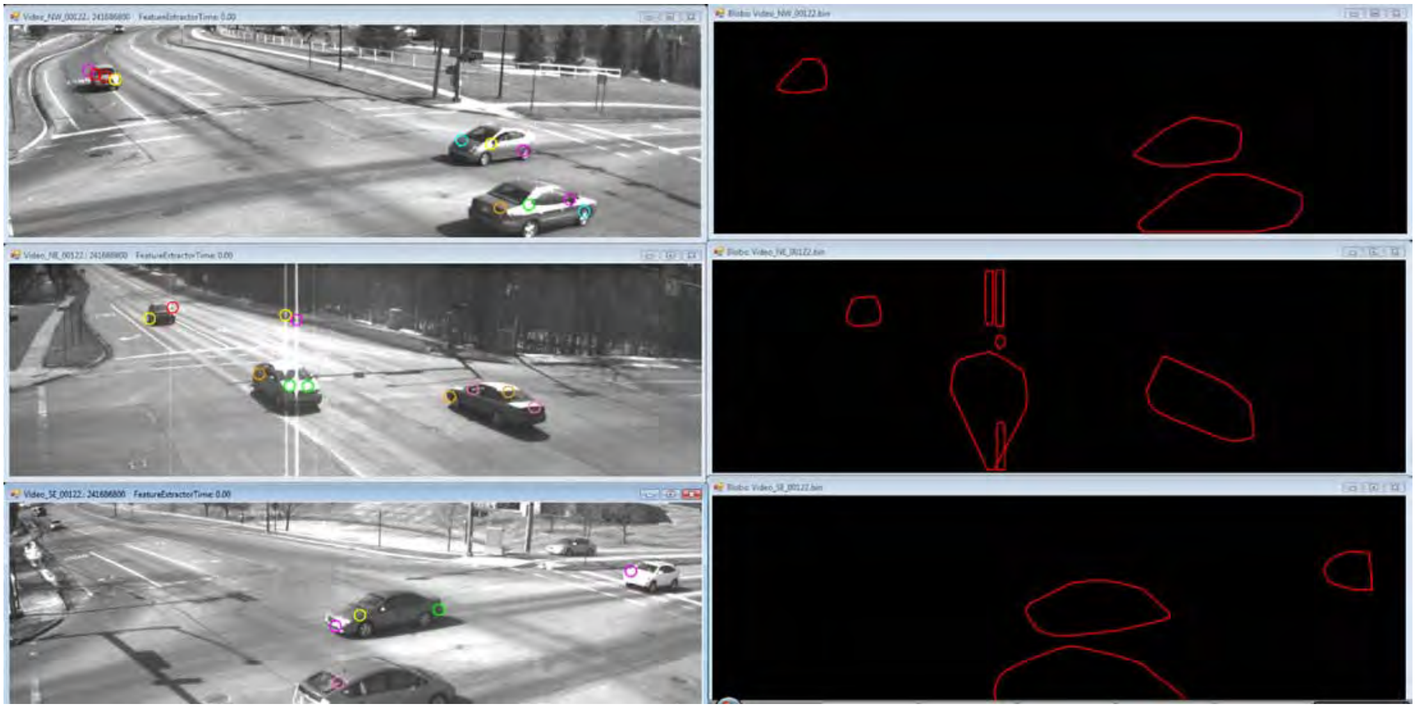


Figure 8.4. Clusters and blobs, including effects of glare and erroneous features.

to take features recorded on the camera and transform them into a world view. Figure 8.6 shows an example; on the left, five cluster tracks obtained from the SW camera location are mapped onto a world view. In this case there was no prior information about the height of the clusters on the vehicle and an assumed height of 0.5 m was used. At the time of the image, only three of the clusters were in existence, and these are marked as blue crosses. The blob for this vehicle is particularly well behaved, giving hope that the vehicle location can be precisely identified, assuming it can be combined

Figure 8.5. Image detail showing pedestrian-generated cluster tracks.

accurately with other camera views. The challenge then is to fuse the feature data sets in a reliable and automatic way that does the following three actions:

- Attaches features (blobs and clusters) to vehicles in a unique way;
- Locates the vehicle in 3-D space (i.e., allows a 3-D bounding box to be co-located with the vehicle); and
- Attaches features to the bounding box and thus estimates their heights, with sufficient feature numbers to track the vehicle all the way through the intersection.

In the case of Figure 8.6, clearly only the exit leg is populated with features; to do more than this requires data fusion between the four cameras.

Effects of Environmental Conditions

The Site Observer was tested during late winter and early spring, including test periods of light snow and low sun angles. The presence of glare can be seen in Figures 8.4 and 8.5, including corruption to blobs. Glare was not found to generate cluster tracks, although occasionally, as mentioned, the intersection of a moving shadow with a background object such as a fence can give the illusion of a moving feature. However, clusters are rarely created from such stray effects, and when they are, the cluster-triggering process (see Chapter 9)

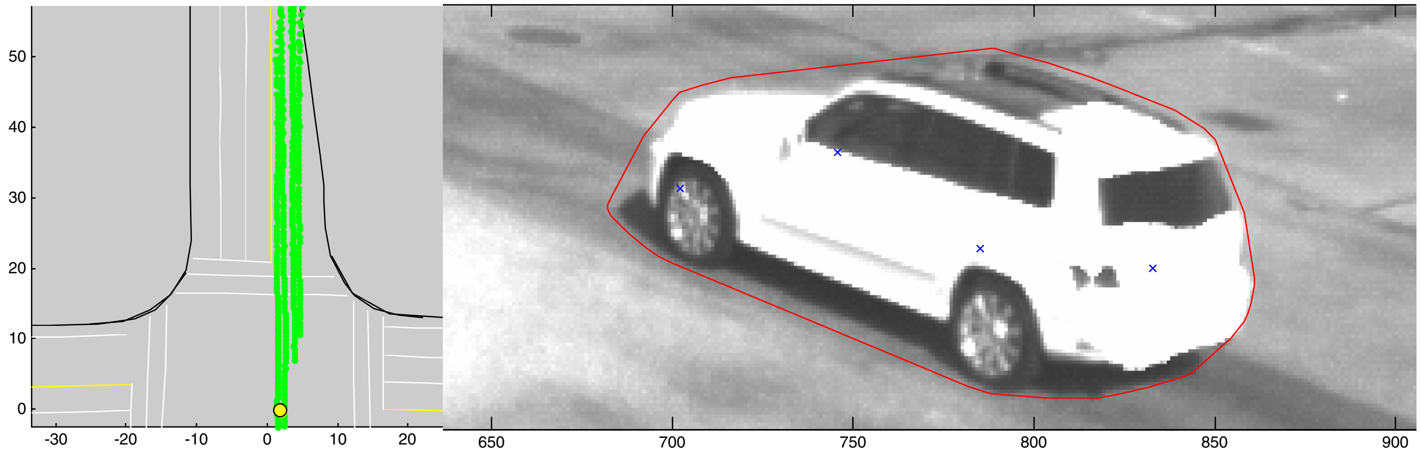


Figure 8.6. World view and camera view of cluster tracks on a single vehicle (through vehicle traveling north).

means that such stray phenomena cannot result in the detection of a false vehicle trajectory.

Shadows typically add to uncertainty in the location of vehicle boundaries, and the effects of shadows are included in the analysis of the following chapters. Again, the use of multiple cameras and 3-D projections tends to reduce the effects of shadows, but certainly their effects are not completely removed.

There was found to be no systematic influence arising from adverse weather conditions such as in light rain or snow; these conditions generate random patterns of corruption at the pixel level, and provided the resulting noise levels are not overwhelming, no effect is seen. On the other hand, in dense rain, snow, dust, or fog, it would not be possible to track vehicles. Because of the limited range of weather conditions under which data were captured, no particular benchmarking was possible for such effects. However, it is clear that when a human observer finds it difficult to see features on vehicles in captured images, the automated system likewise is challenged. It is expected that with deteriorating weather conditions, the number of clusters detected will reduce until the point that vehicle trajectories become incomplete, an error condition that is easily detected. Under extreme conditions, such as blizzard or dense fog, a human observer may need to recognize that the system is blind.

High winds can cause cameras to shake, although the location of the installed cameras meant that the effects were limited to twisting of the signal mast due to wind load on the mast arm. Rotations of as much as 1° were seen, and these tended to be at sufficiently low frequency not to disturb the feature tracking. Other parts of the trajectory estimation (Kalman filtering, see Chapter 10) ensure that these perturbations will have minimal effect on trajectories. On the other hand, loose or highly flexible camera mounts causing large or high frequency vibrations of the camera could not be tolerated. Improved mounts or camera-level shake removal hardware and software is needed in such cases.

At night, with the current camera hardware, there is no expectation that satisfactory tracking can be achieved; no specific tests were performed. With street lighting, CMOS cameras and customized control of iris and shutter, it is possible that features other than headlights and tail lights can be found and adequately tracked; a lot depends on the conditions at any particular site. If the camera can display a sufficiently rich set of corner features to obtain at least two or three clusters per vehicle, and there is adequate contrast between foreground and background, tracking may take place using the algorithms develop in this project.

The installed intersection hardware was located in cabinets that were heated or cooled as appropriate. It is not expected that any normal variations in air temperature or solar heating will cause the system to fail during operation.

CHAPTER 9

Vehicle Localization and Trajectory Estimation

In the previous chapter, it was noted that localization in the camera image is not the same as localization in the 3-D world, and that it is easy to confuse what is seen (by a human observer) with what is known or estimated quantitatively (by the machine vision tracking system). In this chapter the research team proceeds with the quantitative analysis. The first step is to convert camera coordinates into world coordinates, which are of course common to all four camera systems. Cluster tracks are transformed using a pair of assumed heights, at 0.5 m and 1.5 m. Of course, vehicles can display features well in excess of this, up to approximately 5 m for heavy commercial vehicles. Figure 9.1 shows a sample cluster track, with two corresponding projections into world coordinates based on these assumed heights. Here the vehicle is traveling toward the camera. The uncertainty in feature height amounts to more than half a lane width in location, so clearly some refinement is needed to improve precision. The quality of the projected cluster track is not that good either. It suffers random variations attributable to an effect noted in Chapter 8; as component features are gained or lost from the cluster, the mean position is displaced, and this can happen between consecutive frames. In the right plot, this is corrected by using a modification of a method proposed by Kim (2008): velocities are integrated and linear regression is applied to remove drift from the original position estimates. Of course, the result is no less sensitive to feature height because this is a basic geometrical effect, but certainly the reduction in high frequency variations represents a worthwhile improvement.

Multicamera Cluster Tracks

Any individual vehicle may have many clusters attached, including from different camera positions. So the next step is to pick out a single cluster as a seed for vehicle identification and build on conditions of consistency with rigid body vehicle motions to increase the information available but without overlapping features with other vehicles. This is the aim. The basic strategy is to set a trigger on the outgoing leg of the intersection, where there is minimal chance of a queue forming, and where vehicles are mainly well separated in the camera frame. Informally, a best case trigger location was set at 30 m from the center of the intersection, where longitudinal separation of vehicles is likely and where the camera height is still sufficient to avoid most occlusions. In addition, on exit, vehicles are squeezed into a single lane at this distance, which provides initial help with localization. Given that cluster height affects lateral position (because of the lateral offset in camera position relative to the exit lane) the condition that the cluster track should normally be within the lane boundaries offers a simple condition to select an approximate feature height. Thus, for features moving away from the intersection—and having a greater lateral offset than for the tracks shown in Figure 9.1—the preference for being within the single lane provides a simple means to select between the two assumed heights.

Other triggers on the same exit leg that match the first triggered cluster track along their whole common length are sought. Of those found, priority is given to the one that is generated first in time, so it has best visibility within the intersection area; the selected cluster track is then used as the reference cluster for the vehicle in question. In addition, it is easy to add clusters from all camera locations as long as they match the reference cluster based on rigid body conditions. Note that cluster height above ground is assumed to remain constant, so the rigid motion is well approximated by the condition that distances in the (x, y) plane remain constant, at least within some tolerance. In fact this is an important step because it often connects a cluster track entering the intersection with the reference one that exits. From this, the time range over which the maximum number of cameras are recognizing at least one of this group of clusters is determined; the central time within this range, rounded to the nearest frame time, is called the *reference time* t_{ref} for the detected vehicle. If only one camera view can be found, the set of cluster tracks is considered incomplete and is rejected.

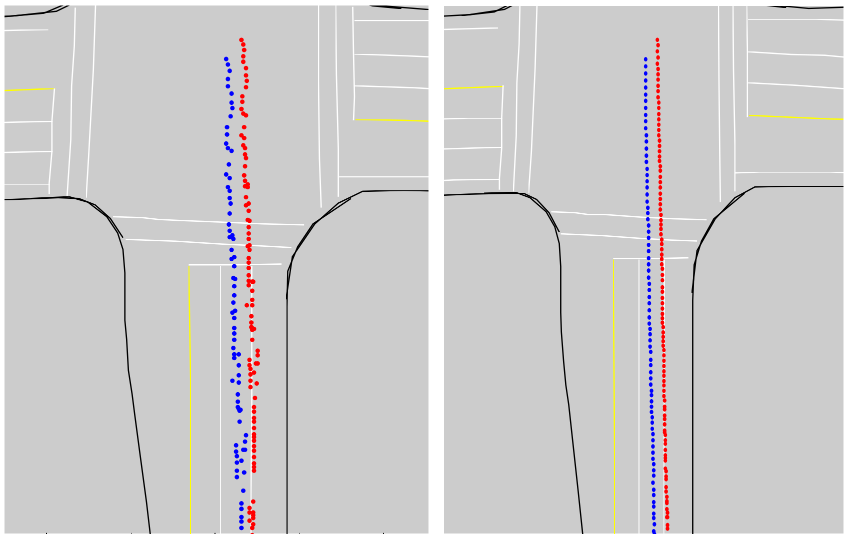


Figure 9.1. A single-cluster track resolved in world coordinates at an assumed height of 0.5 m (blue) and 1.5 m (red), with vehicle traveling to the north, viewed from NE camera.

Figure 9.2 shows all triggers obtained from a 30-min test run (Run 00122). Of the triggers obtained, each was referred to the two reference heights (0.5 m and 1.5 m), and the condition applied that at least one of the trigger points should be within the lane boundaries. If both are within the boundaries, the one nearest the lane center is selected. Of 6,599 cluster tracks, 1,427 triggers were found in this way. Given an approximate estimate of 500 to 1,000 vehicle movements per hour (see Chapter 7), it seems reasonable to expect that most vehicles, if not all, would have provided an exit trigger. It is worth noting that approximately 5 min of start-up delay was necessary; approximately 3 min elapsed before the background image had converged in all cameras, and an additional 2 min was allowed so that triggers of exiting vehicles would also show a corresponding entry track.

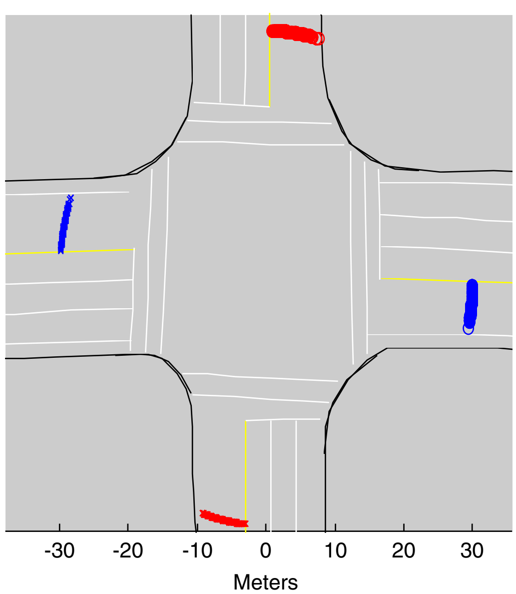


Figure 9.2. The 30-m triggers set on exiting cluster tracks (Run 122).

For illustration, an example trigger was selected and back-propagated into the intersection; compatible triggers were included, giving three matching cluster tracks, all three being outgoing on the east leg of the intersection. Searching for compatible cluster tracks across all cameras yielded a total of 18 clusters. Despite the lack of precise localization, given the uncertainty in cluster height, informal video review of many cases suggests there is a high probability that all are from the same vehicle. The results are seen in Figure 9.3, which shows the four camera angles, complete with blobs and identified clusters (shown here only in the relevant image). Note that not all 18 clusters are in view in this instant, which is the reference time mentioned above. It is clearly seen that in this case the clusters are all unique to the one vehicle, which is turning left from the north leg and exiting toward the east.

Figure 9.4 shows the scatter in the clusters when resolved at a nominal 0.5-m height above ground. Note that exceptions here are the original three triggering clusters, which have some initial height refinement based on position in the trigger zone. It might be thought that the yellow marker (placed on the reference cluster track) makes an acceptable vehicle center and that this would be adequate for localization. This would give an advantage through averaging over multiple views, but it is not making best use of the information available. In fact, this particular vehicle, by executing a left turn, is seen by all four

Figure 9.3. Simultaneous images of a turning vehicle identified with 18 cluster tracks. From the top, the cameras are located at the NE, SE, SW, and NW corners, respectively. Scales are in pixel coordinates.

cameras, but as was noted in Chapter 7, right-turning vehicles normally are covered by only two cameras, so any type of simple position average is likely to induce bias for such cases.

Given the need for precision when determining conflict metrics, it is worth seeking an improved method. Note that there is no way to associate clusters between different cameras (and typically there is no commonality), so stereographic analysis is not feasible.

The polygonal blobs are used to provide new information to localize the vehicle when it is maximally visible to all cameras. Figure 9.5 shows a projection of blobs (all those existing at a common time) onto the road surface, whereas Figure 9.6 shows a close-up view. These images are determined at the same time and using the same data from the turning vehicle in Figure 9.3. The key point here is that with a zero-height ground plane, the projection from 2-D camera to 3-D world is precisely known. In fact, although a ground plane is mentioned, the actual mapped surface heights are used. The projection of the blobs onto the road surface takes full account of any height variations in the surface geometry.

It can be seen that some of the blobs are projected far from the intersection center; for example, in the upper plot of

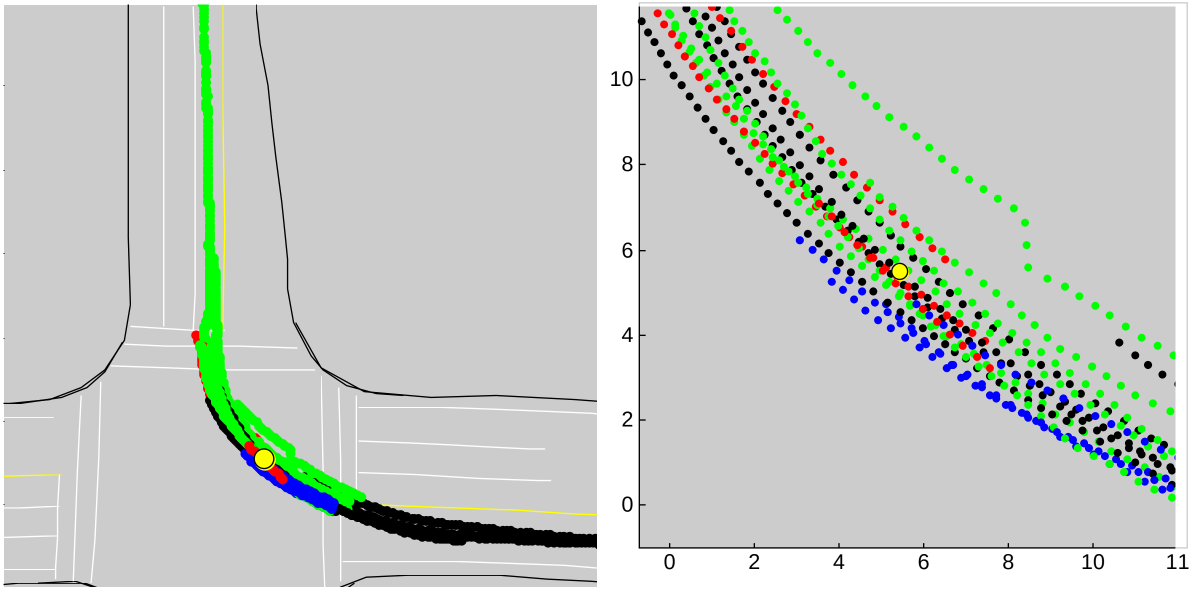


Figure 9.4. Multicamera cluster tracks using a nominal 0.5-m assumed height above ground. Left plot shows detail. Yellow dot is reference track at reference time. Tracks are seen from cameras at the NE (blue), SE (red), SW (green), and NW (black) corners. Scales are in meters.

Figure 9.3, a blob is cut off by the image frame, and its 3-D counterpart may extend well beyond the projected line. In this case, the projected polygon is extended well beyond the limits found from the visible points in the camera.

To localize further, the centroid of the clusters is used to select the nearest projected blob from each camera, and then the projected polygons are intersected to give a localized *bounding polygon* (BP) for the vehicle. This is shown in magenta in Figure 9.6.

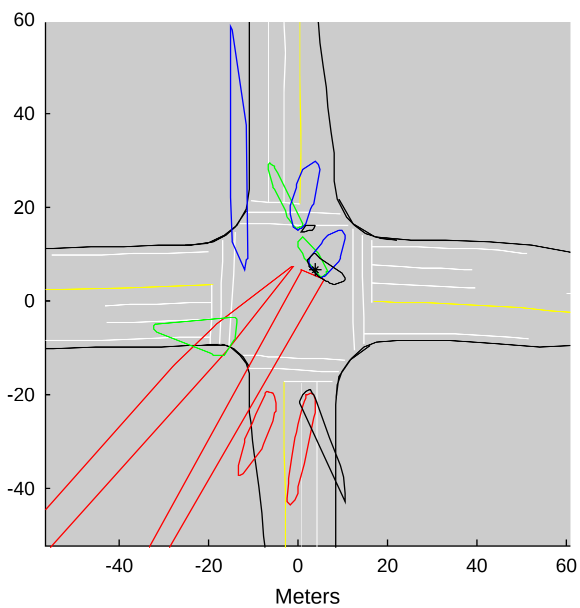


Figure 9.5. Multicamera blob projections at time t_{ref}. The black star shows the location of the centroid of associated clusters. Colors denote camera source: NE, blue; SE, red; SW, green; and NW, black.

To complete the basic vehicle localization at the reference time, a rectangle is fitted. This is not a unique process because many rectangles can fitted to a polygon, and given the uncertainty over the exact limits of the BP, the rectangle is allowed to protrude slightly beyond the BP. Rectangle fitting is simplified by first estimating the direction of motion, and this is easily done by tracking the cluster set between adjacent frames: essentially the cluster velocities are averaged to provide a direction for the orientation of the rectangle. Multiple lines are intersected with the BP in the directions parallel and perpendicular to the direction of motion, and the median lengths and widths obtained are used to estimate the size and positioning of the fitted rectangle. This process is found to be normally robust, especially provided the vehicle is visible to at least three cameras.

Cluster height estimation is now considered again; with the vehicle boundary now determined (at time t_{ref}), the localization of the clusters can be improved. This is carried out in local vehicle coordinates (Figure 9.7) based on an origin G at the center of the rectangle and GX_V, GY_V axes aligned with the vehicle rectangle; the GX_V axis points to the left of the direction of motion as shown. In Figure 9.7, O, X, and Y are the intersection coordinates, with O at the nominal center of the intersection, OX pointing east, and OY pointing north. The vehicles' axes are to move with the vehicle and are especially useful for projecting clusters into the local vehicle geometry and thus for estimating the unknown cluster

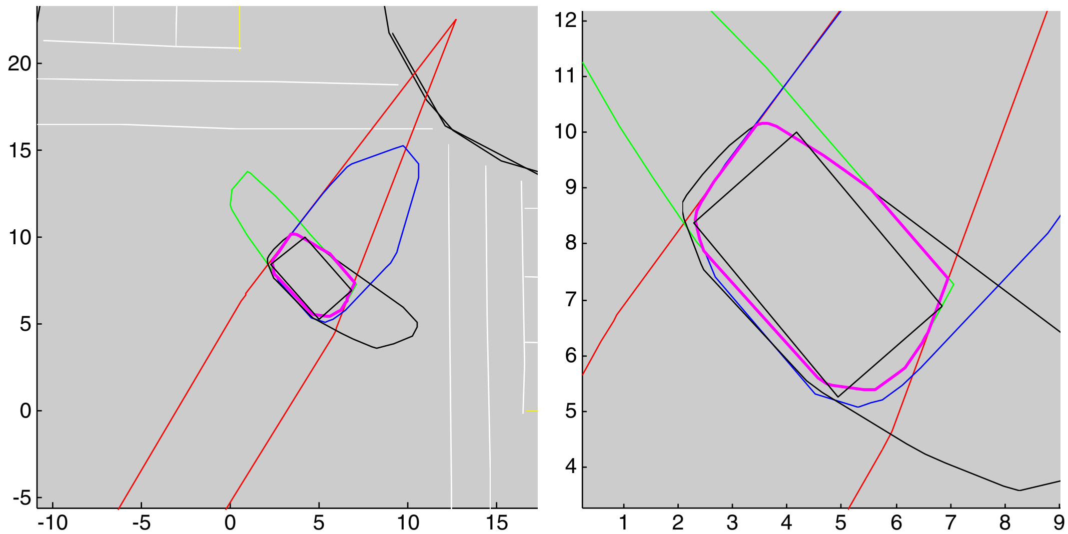


Figure 9.6. Detail and vehicle localization by intersecting blobs. Blob color codes are the same as in Figure 9.5, with the addition of a magenta bounding polygon and fitted rectangle (black). Distances are in meters.

heights. Cluster localization is shown in Figure 9.8, in which the blue dashed rectangle is the vehicle boundary (at ground height) and each red line represents the projection of a single cluster between the upper standard height ($h = 1.5$ m, marked with a red star) and the lower standard height ($h = 0.5$ m). Clearly the upper height indicates the point nearer to the relevant camera. Numeric values indicate the source camera, the directions of which are rotated because of the transformation to vehicle coordinates.

If the assumed height of any cluster between these reference heights is varied, it assumes a different position on its corresponding (red) cluster line. Of course, the cluster height may be outside of this range, in which case it should be extrapolated beyond the nominal endpoints of the cluster line. Intersecting each cluster line with the vehicle rectangle provides an estimation of cluster location and height. Although two intersection points normally are found, it is assumed that the vehicle boundary nearest the camera is the most probable location,

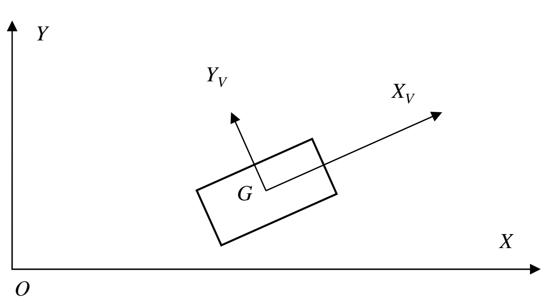


Figure 9.7. Vehicle local coordinates (ISO sign convention).

and this one is used. If no intersecting point is obtained, the point of nearest approach is used, unless it is further from the vehicle than a certain tolerance (1 m is assumed), in which case the cluster is rejected. The resulting cluster points are shown as blue squares in Figure 9.8. For comparison, the blue circles are the nominal heights used previously, mostly at the

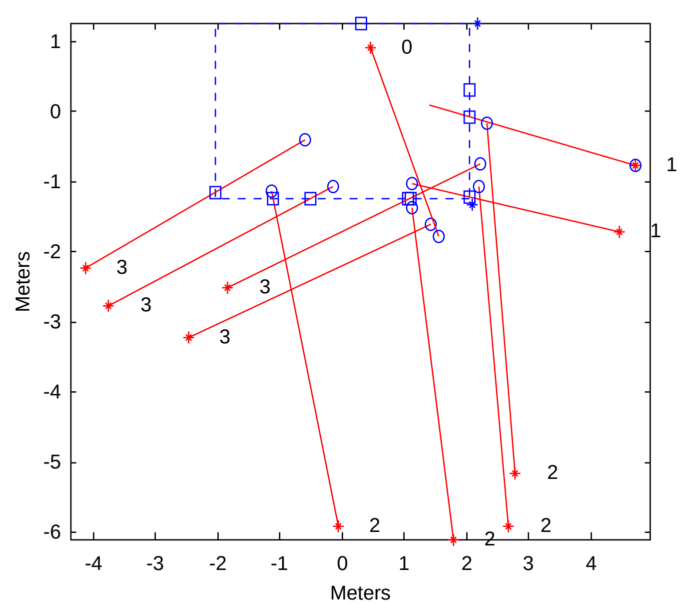


Figure 9.8. Cluster localization using a vehicle rectangle (vehicle coordinate system: X_V is horizontal, Y_V is vertical; units are in meters). Numeric values adjacent to cluster lines indicate the camera location: 0 = NE, 1 = SE, 2 = SW, and 3 = NW.

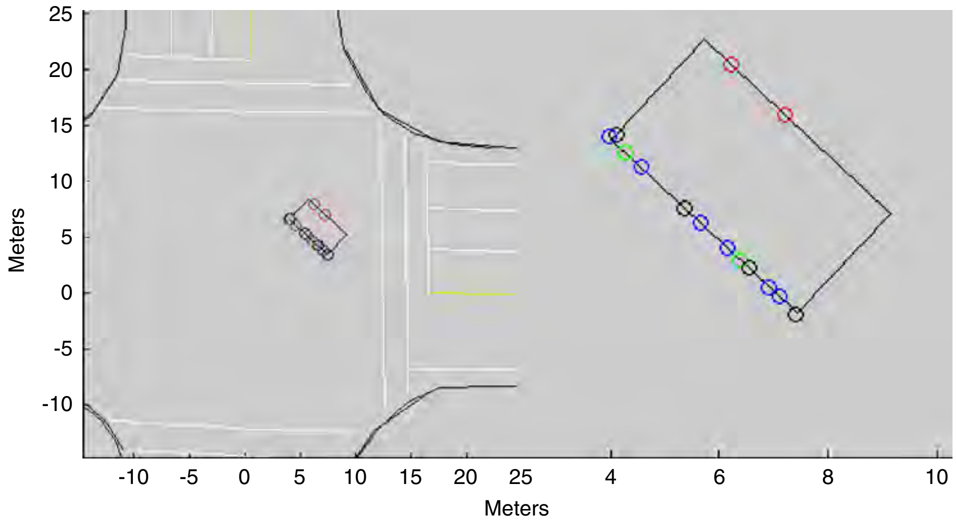


Figure 9.9. Fitted rectangle and associated cluster points, including close-up (right plot). Colors according to camera corner location: NE, red; SE, green; SW, blue; and NW, black.

lower (0.5 m) location. If negative heights are found, the cluster is rejected unless it is within a small tolerance, in which case its height is set to zero.

The resulting set of fitted clusters is shown in intersection coordinates in Figure 9.9, whereas a 3-D projection of the bounding box on one of the camera images is given in Figure 9.10. The height of the box was estimated as being twice the median height of the fitted clusters (which is more robust than choosing the maximum cluster height). It should be emphasized that all steps are fully automated and that the example was randomly selected.

The "nearest edge to camera" algorithm is not always accurate because clusters attached to the roof or other interior surfaces such as windshield or hood may exist further from the camera. This means that for tracking purposes, the clusters nearer the ground are preferred. If greater precision is required, it is possible to further refine locations of clusters using the coincidence of cluster lines from multiple frames, at the expense of additional computation and complexity. At this point, it is assumed that viable localization has been achieved with fully automated methods.

Figure 9.10. Fitted rectangular bounding box on camera image.

For any single-cluster track, there is now a simple way to estimate the motion of the vehicle center: follow the cluster in 3-D coordinates assuming fixed height above ground, estimate the velocity vector as part of the cluster tracking, and then apply known offsets (i.e., the vehicle-based coordinate of the fitted cluster). Averaging the results may be used to give a refined vehicle trajectory, but because cluster tracks may appear or disappear, a more systematic approach is preferred; this is considered in the next chapter. For now the research team includes the results of tracking a single outgoing cluster in this way, joining it with a single incoming cluster, where length and height preferences have been used in the cluster selection. The path shown has been truncated at 50 m because it is not expected that the single-cluster tracking is likely to be stable or sufficiently accurate beyond that distance. A similar limit will be imposed when performing velocity and acceleration estimation (see Chapter 10).

Figure 9.11 shows the resulting path of the example vehicle considered above. Although only a single cluster has been used for the path estimation, and clearly the lateral positioning is not as precise as it may be, this basic trajectory may be used for searching purposes. Using polynomial curve fitting, tangent directions also may be reliably determined, even when the precise vehicle location is uncertain. The tangent vector is stored so it is available when additional refinement requires the vehicle orientation. As a byproduct, the curve fitting gives estimates of speed and distance (see Figure 9.12), where the distance is measured along the curved path and has a nominal zero point at the reference time t_{ref}. The speed estimation is seen to be robust

Figure 9.11. Basic vehicle fitted path (red: using outgoing cluster; blue: using incoming cluster). The right plot shows a detail of the track near the stop bar; corresponding vehicle image is for the front car at the stop bar.

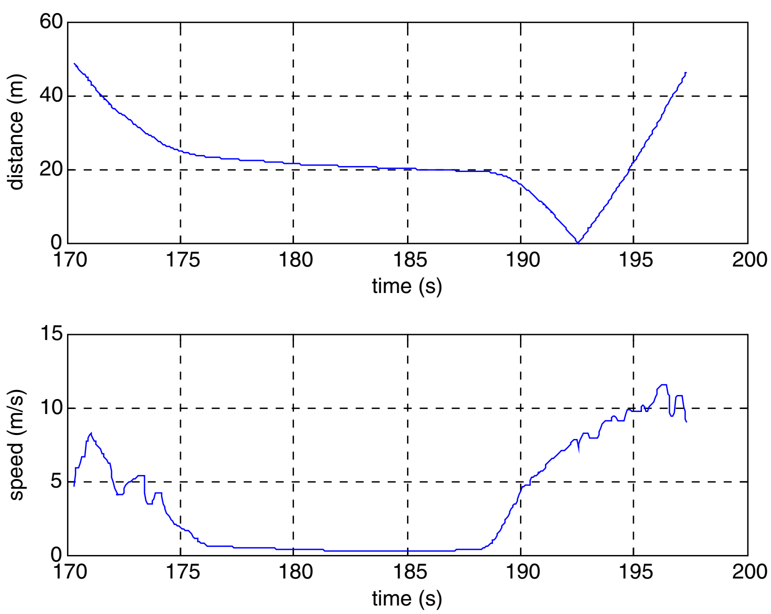


Figure 9.12. Speed and distance estimates from basic track fitting.

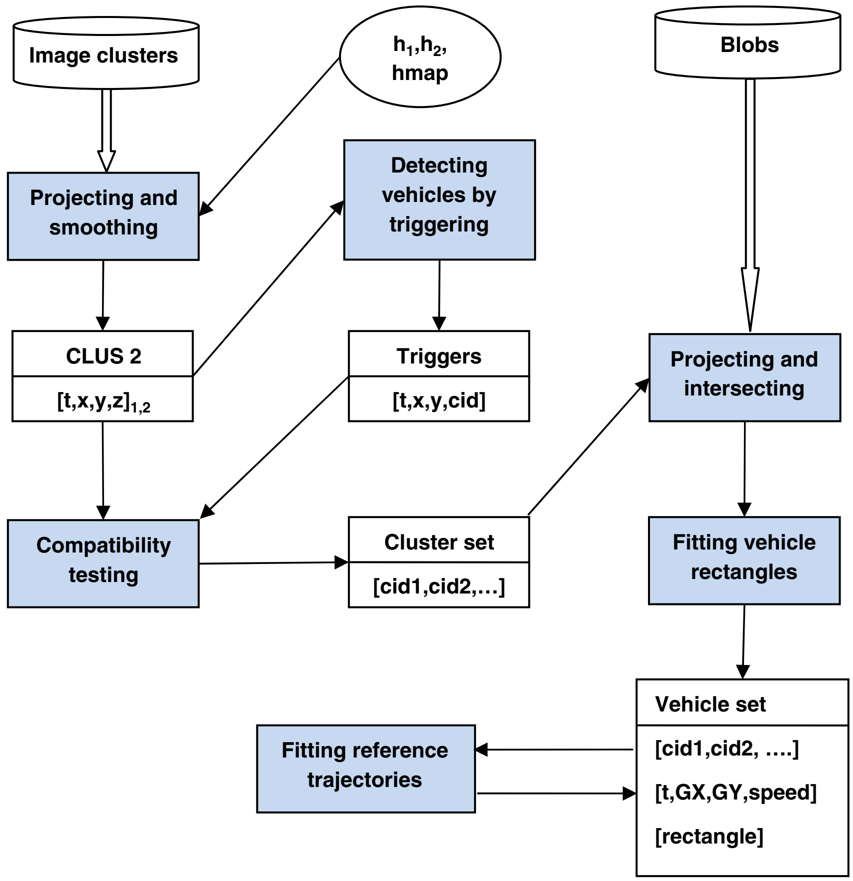


Figure 9.13. Block diagram summary of basic vehicle trajectory estimation.

to the fact that the vehicle actually stopped for a short period at the stop bar (between $t = 180$ and 188 s, as seen in video review), although this is captured as a very low drift speed of approximately 0.2 m/s.

Figure 9.13 summarizes in a block diagram the overall steps used in the foregoing basic vehicle trajectory estimation. The double arrows represent SQL-based data extraction from the feature database. White rectangles are tables constructed to store relevant data elements (where t, x, y, and z represent time and position coordinates, *cid* represents a cluster identification, and in CLUS 2 the dual reference heights h_1 and h_2 are indicated by the suffixes $[\ldots]_{1,2}$).

CHAPTER 10

Trajectory Refinement and Estimation of Motion Variables

Chapter 9 described a process for establishing basic vehicle trajectories, sufficient to localize vehicles and search for specific types of events (for example, to find potential conflicts based on timing, position and direction of motion, such as left turns across path). However, for many conflict metrics, such as those involving time to collision (TTC), the vehicle velocity is needed. For other purposes (e.g., hard braking events), estimates of vehicle acceleration information also may be required. Therefore, some degree of trajectory refinement is needed.

Here refinement of vehicle trajectories by two distinct methods is considered:

1. Position refinement: Tracking clusters in groups and allowing for the local position of each available cluster gives multiple estimate of the vehicle centroid path. Combining these into a mean path should reduce position errors and in particular reduce the tendency for position errors to drift over time. The main purpose is to apply something akin to cluster position averaging but take into account the known local coordinates of each cluster, as estimated at the reference time (see Chapter 9).
2. Motion estimation: To make use of commonly used tracking algorithms, specifically the model-based Kalman filter (KF) to estimate speed and acceleration. As was seen in Figure 9.12, such estimates are feasible, but the use of a KF can reduce noise and is essential when accelerations are estimated.

Multicluster Centroid Path

Figure 10.1 shows the method of centroid tracking. The dashed line is the projection into the ground plane of the cluster motion, using the height estimate obtained at the reference time $t = t_{\text{ref}}$. From the fitted rectangle and its orientation, the position of the cluster relative to the center of the rectangle $\hat{\mathbf{x}}_G$ is known. Assuming the cluster stays fixed relative to the vehicle body, its motion is used to infer the motion of $\hat{\mathbf{x}}_G$. With multiple clusters, multiple estimates of $\hat{\mathbf{x}}_G$ at later times are generated, and simple averaging can be used to improve the position accuracy.

Figure 10.2 shows a typical example: a left-turning vehicle, with $t = t_{\text{ref}}$ in the left plot and $t = t_{\text{ref}} + 1$ on the right. The red circles are the estimated locations of the clusters after projection into world coordinates, making use of the cluster heights estimated previously. The blue dots show the corresponding estimates of the centroid, which are averaged to fit the rectangle, centered at the cross marker, seen better in the zoomed figure, inset. The dashed line is the estimated path based on single clusters and without height or position refinement.

Figure 10.3 shows the same vehicle (at time $t = t_{\text{ref}} + 0.7$ to keep it within the image frame of the camera) where the left plot fits the bounding box based on the earlier path estimate, simply using a sample cluster track at a nominal height. Clearly there is significant positional bias; this path is also shown in Figure 10.2 as the dashed line. By comparison, the use of local coordinate offsets provides quite accurate positioning. This example was not specifically chosen and is quite typical of what is achieved when the bounding box fitting works well. Figure 10.4 shows a limitation of the current version of the localization method in that no attention has been devoted to countering the effects of shadows. Here the shadows clearly affect the location of the bounding box, and clearly the problem is inherited from the original blob and cluster fitting. There is ample scope to reduce such problems, especially by rejecting clusters at ground height and adjusting the box dimensions after checking the angle of the box relative to the sun. A more heuristic method based on typical vehicle rectangle shapes is also possible, but such heuristics have been avoided in the prototype analysis.

Motion Estimation: Kalman Filtering

A KF is similar to other filters used for smoothing data, except that it uses an underlying model of how the data are generated and includes estimates of the sources of disturbance (see,

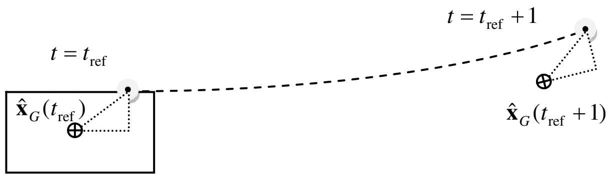


Figure 10.1. Centroid position estimation by local coordinate rotation and offset. The dashed triangle indicates the fixed distances of lateral and longitudinal offset relative to the vehicle center.

e.g., Bar-Shalom et al. 2001). This additional model-based information provides a way to not only smooth a noisy day but also to estimate other information contained within the model. For the current application, the purpose is to extract other vehicle motion variables, in particular the speed and (longitudinal) acceleration. In fact, to a large degree, the estimated trajectories are not especially noisy, mainly because it was possible, early on, to remove worst effects of fluctuations in the positions of the clusters.

Once we have estimated any vehicle centroid path, as above, we have data in the form of an estimated position vector $\hat{\mathbf{x}}(k) = [\hat{x}(k), \hat{y}(k)]$ equally spaced over time, where k is a time index corresponding to our chosen frame rate of 50 ms, and the hat denotes an estimate of the actual physical quantity. Between these time intervals, the KF model assumes the acceleration can vary linearly and in an unknown way; this is the classic Wiener process model (again see Bar-Shalom et al. 2001), for which the degree of *jerk* (rate of change in acceleration) varies unpredictably. Of course, here the source of disturbance is just the driver. The process model equations are presented briefly here, and the reader is referred to the many texts on the topic of how the KF is implemented, but suffice to say that compared with the many nonlinear transformations and image processing steps considered, the KF is an extraordinarily quick and simple method.

The underlying model equations are of the following form:

$$\mathbf{x}(k+1) = A\mathbf{x}(k) + Gv(k)$$

where the matrices of constant coefficients take the block form

$$A = \begin{bmatrix} A_3 & 0 \\ 0 & A_3 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & T & \frac{1}{2}T^2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix}, G = \begin{bmatrix} G_3 \\ G_3 \end{bmatrix}, G_3 = \begin{bmatrix} \frac{1}{2}T^2 \\ T \\ 1 \end{bmatrix}$$

and the measured output is

$$\mathbf{y} = C\mathbf{x},\ C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

The model has six states: two for translation ($x(k)$, $y(k)$), two for velocity ($\dot{x}(k)$, $\dot{y}(k)$), and two for acceleration ($\ddot{x}(k)$, $\ddot{y}(k)$). The input $v(k)$ represents errors between the model and the physical data resulting from unmodeled driver control actions, measurement noise, and so forth. In this model, there is no preferred sense of direction for the vehicle; a car that came into the intersection, turned two circles, and returned from where it came would be tracked as well as one taking a more normal path.

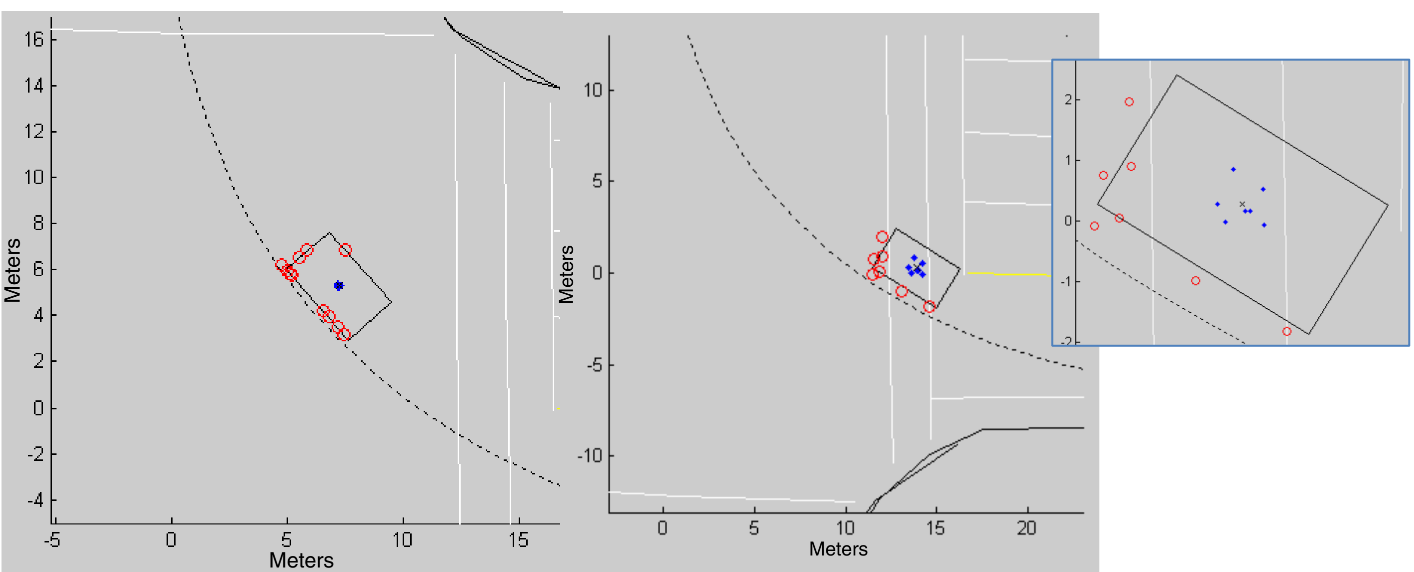


Figure 10.2. Example of position refinement. Red circles are cluster projections; blue dots are centroid estimates. Left: At reference time. Right: At time t_{ref} + 1. Zoomed image also is at time t_{ref} + 1.

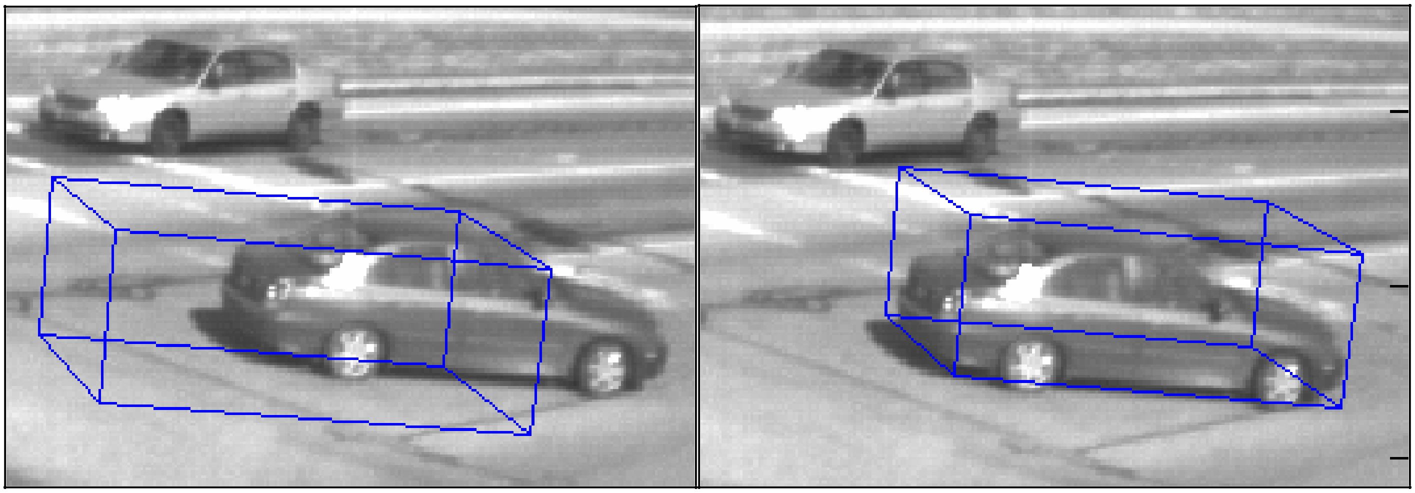


Figure 10.3. Effect of position refinement. Left: Initial approximation from single cluster and nominal height estimate. Right: Uses position refinement by coordinate offsets. Time: t_{ref} + 0.7.

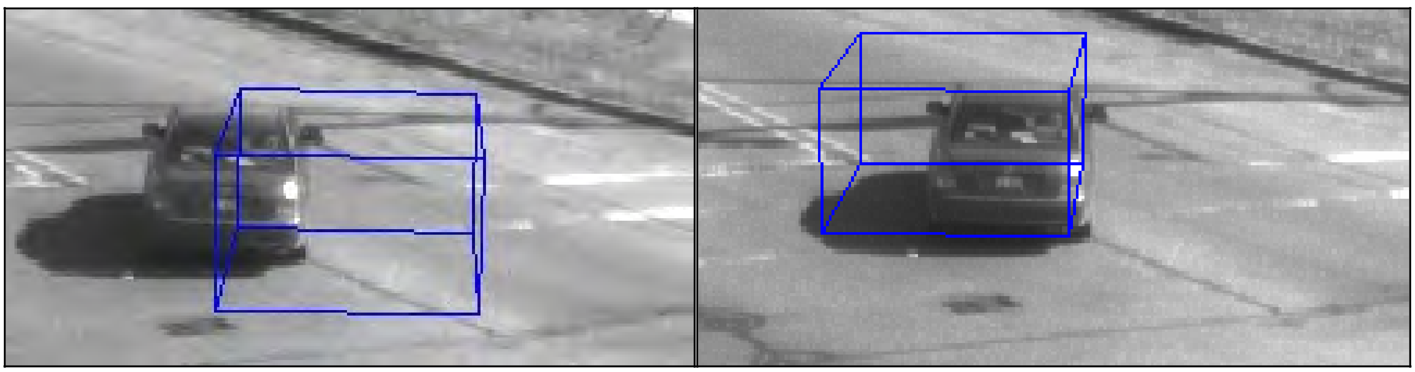


Figure 10.4. Lateral offset attributable to vehicle shadow. No shadow compensation has been applied in this analysis. Time: t_{ref} + 1.

One challenge for using the KF is to initialize the six states; some initial best guess of position, velocity, and acceleration is needed. Because position is best estimated within the intersection itself, the method adopted is as follows: (1) estimate the position at the reference time and for a few meters of travel near to that point; (2) fit a polynomial path to that segment and use this to estimate velocity and acceleration; (3) run the KF forward in time from this initial state until it reaches a distance 50 m from the intersection center (or runs out of data); (4) run the KF backward in time from the same reference position, again until 50 m is reached or the data run out; and (5) join the two paths into a single trajectory, which now includes velocities and accelerations. Afterward it is simple to resolve the motion variables in various directions to estimate speed, longitudinal acceleration, lateral acceleration, and yaw rate (rate of turn). Sample results for a left-turning vehicle are presented, and Kalman filter estimates are compared with the results of instrumented vehicles. In Figure 10.5,

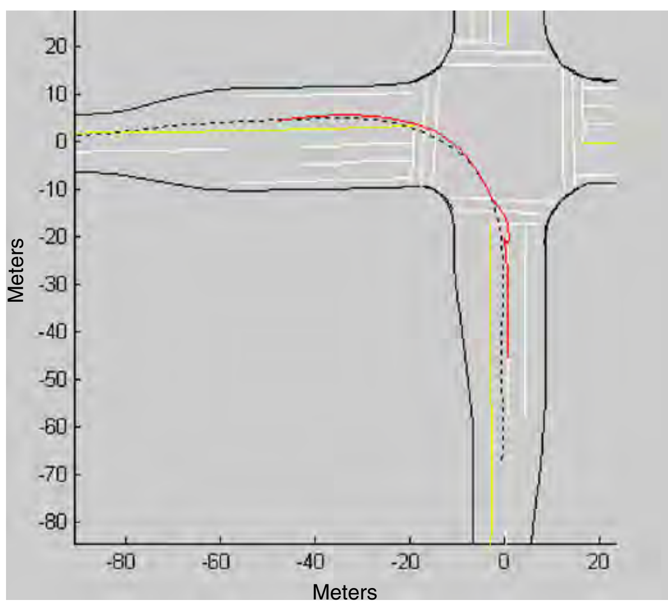


Figure 10.5. Kalman filter estimation of vehicle path.

the dashed line is the smoothed path based on single clusters, whereas the red line is the output of the KF position states. Some degree of wobble is seen in this plot, and overall it appears that the KF used, which is effective for speed and acceleration estimation, does little to improve the estimation of position. For this reason, the KF results are seen as *augmenting* the positioning refinement shown but not fully replacing it. It is possible in the future that the *extended* KF method can be used to improve the height estimation of cluster, but this is beyond the scope of this prototype development. Figure 10.6 shows the resulting speed and acceleration plots, similar to those shown in Figure 9.12; however, acceleration estimates are found, something not reliably obtained from polynomial curve fitting. Again, the fact that the vehicle temporarily stops does not cause any special problems for the KF.

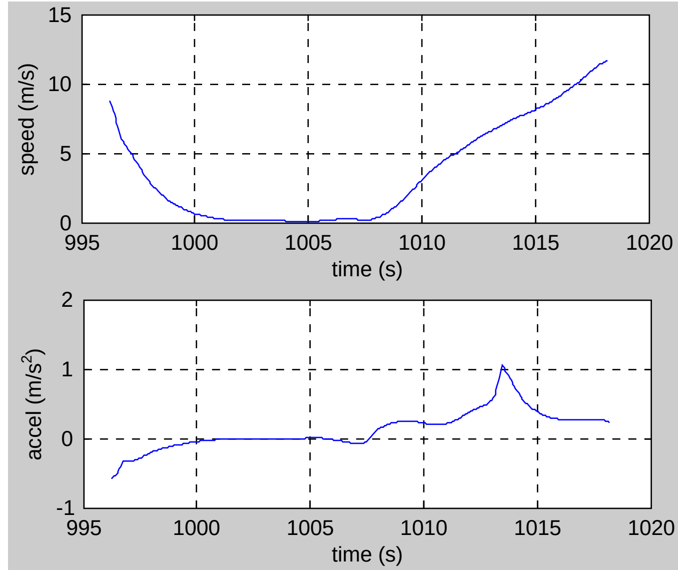


Figure 10.6. Speed and acceleration estimation for a left-turning vehicle.

CHAPTER 11

Performance Evaluation

The challenges of tracking vehicles using video imaging have been described and methods for overcoming those challenges explored. Now results of the prototype Site Observer are presented, and its feasible use as a fully automated motion and trajectory sensing device for robust use in traffic situations is considered. From a hardware standpoint, the system performed flawlessly during the latter part of a Michigan winter, even with a semi-permanent installation; camera and computer hardware were located at the intersection for several months. Not only were there no failures, the real-time system did not generate any dropped or missing image frames, and there were no times when the system froze or otherwise corrupted or lost source data. There is no doubt that the system runs and generates data, so the question reduces to whether it produces useful data, and if so, how complete and accurate are the data. This question is addressed in the following.

It is important to consider current performance in the context of a prototype for which additional refinement and improvement are possible and probably necessary. The key point to demonstrate is that the system operation—extracting features from individual cameras and combining them without further access to video images—provides valid data for safety analysis (for example, in conflict metric analysis). It is *not* expected that the system will resolve every single vehicle with the best possible precision; in its current functional form, some trajectories generated may be corrupted; an example is seen in Figure 11.4. Such corrupted trajectories are rare, and although a precise error rate is not objectively known, based on manual video review, it appears less than 1% of cases are prone to these kinds of problems. Such cases are easily screened out (for example, the KF automatically computes error levels) but are retained so the baseline performance of the system is not disguised. Although further analysis on the completeness of the data is possible, it is already clear that this is not an issue for the low traffic densities seen in the current study; minimal data loss means there is minimal bias in any conflict analysis.

The system was run for approximately 30 h during February and March 2010, at the installation site in Ann Arbor (see Chapter 7). During this time 17,593 vehicles passed through the intersection, with a breakdown by type shown in Figure 11.1. For comparison, Figure 11.2 shows corresponding numbers from a weekday 24-h period, in March 2010; these data were provided by Washtenaw County Road Commission, and obtained from the existing traffic signal system. The Site Observer was run over five separate days, capturing data for approximately 6 h in each case. A further 30-min run was added to capture data using instrumented vehicles. The two figures show similar trends, although the Site Observer data have greater mean counts per hour, as expected, because the system was run only during daytime hours.

Provided sufficient clusters are detected, the vehicles can be localized and tracked. Outgoing vehicles generally seem more consistent and well populated with clusters than do incoming vehicles, and it is possible that faster moving vehicles tend to have shorter incoming cluster tracks than do slower ones. This is tested with a sample of 200 trajectories (see Figure 11.3); all are through vehicles traveling north to south or south to north. Here the blue crosses correspond to incoming vehicles, and red circles are for outgoing vehicles. Both cases are shown because the road geometry is not symmetric; the ground rises to the north and falls to the south. It is clear that some very long tracks are seen in the outgoing vehicles, as great as 250 m, which is well beyond expectations (and nothing beyond 150 m should be deemed reliable). Interestingly, although the upper plot shows some speed dependence (the highest speed incoming tracks tend to be shorter), overall the only clear trend is for outgoing cluster tracks to be longer than incoming ones. For vehicles arriving from the north, there is a sharp cutoff at approximately 80 m, the reason for which is unknown.

Moving now onto more specific measures of performance, the team's focus is estimating (1) the accuracy of positioning and (2) the accuracy of derived motion variables.

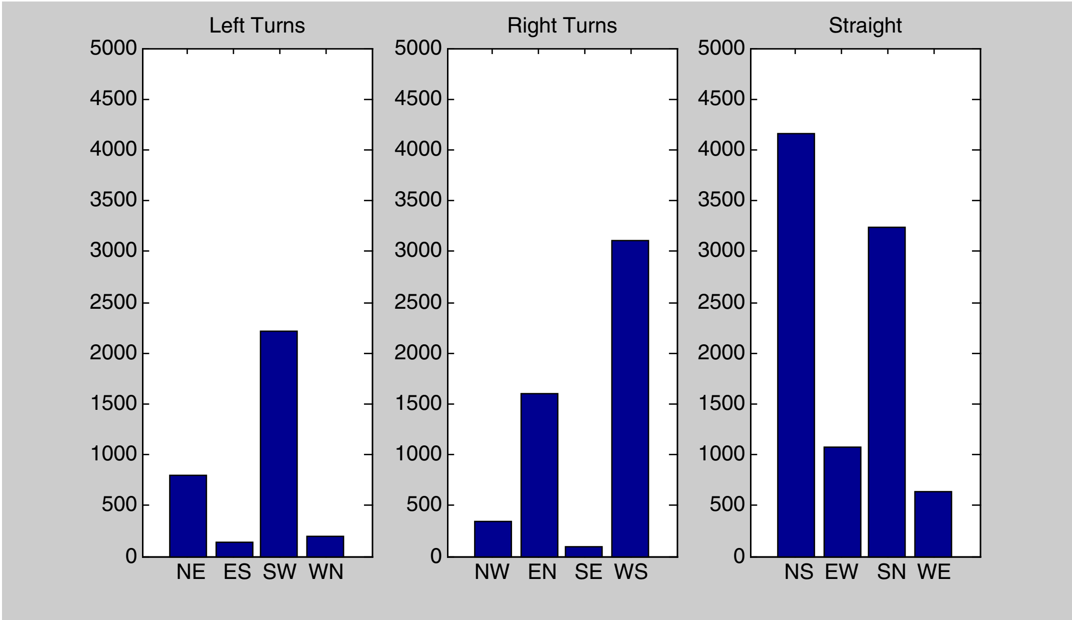


Figure 11.1. Traffic volumes during the 30-h study period (NE: enter from the north, exit toward the east, and so forth; 17,593 vehicle trajectories were recorded).

Positioning Accuracy

Unfortunately, a full and objective reference data set is not available for validation of vehicle positioning. Even the instrumented vehicles have limited precision in their positioning, with GPS position guaranteed only within a few meters. It is possible to use manual digitization of images to provide this reference, but this is tedious and would tell little beyond what has been seen informally: that position accuracy depends critically on camera coverage near the center of the intersection and that subject to this, the position errors are no worse than approximately 0.5 m within the intersection, increasing to approximately 1 m at distances of 100 m. A way to confirm these estimates and make them more precise is to look at trajectory dispersion.

Figure 11.4 illustrates the relevant type of data available from the Site Observer; centroid plots are made after criteria were applied for intersection entry and exit leg choice. Apart

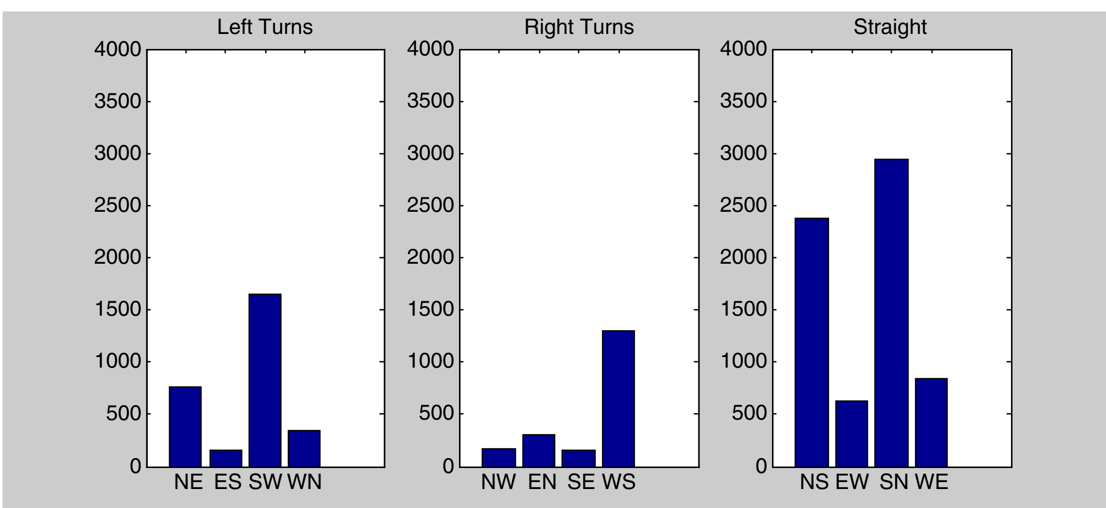
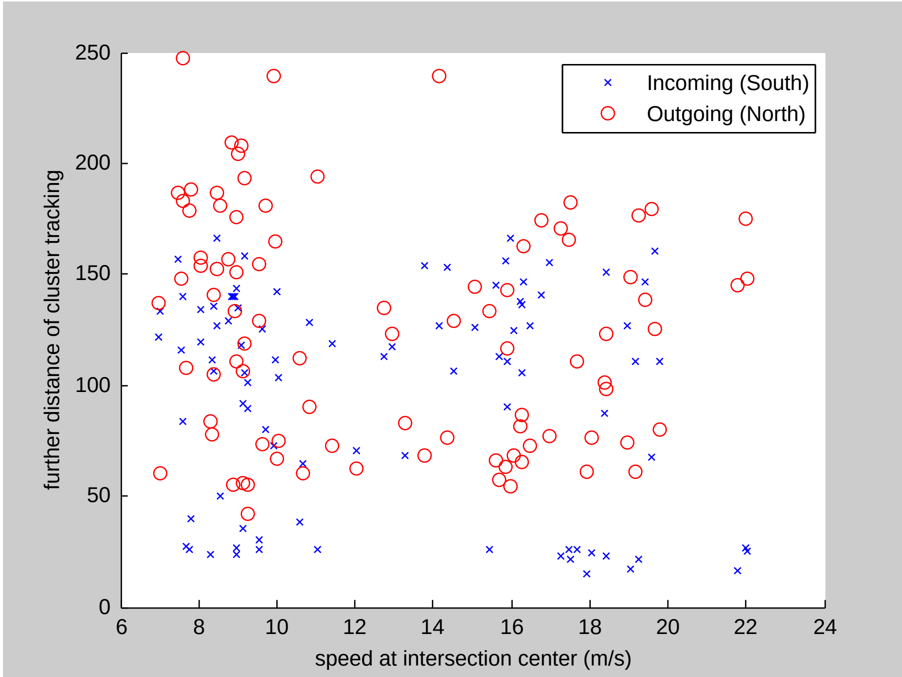


Figure 11.2. Traffic volumes during a comparable 24-h study period (from Traffic Signal Camera System, courtesy of Washtenaw County Road Commission; 11,622 vehicle movements were detected).



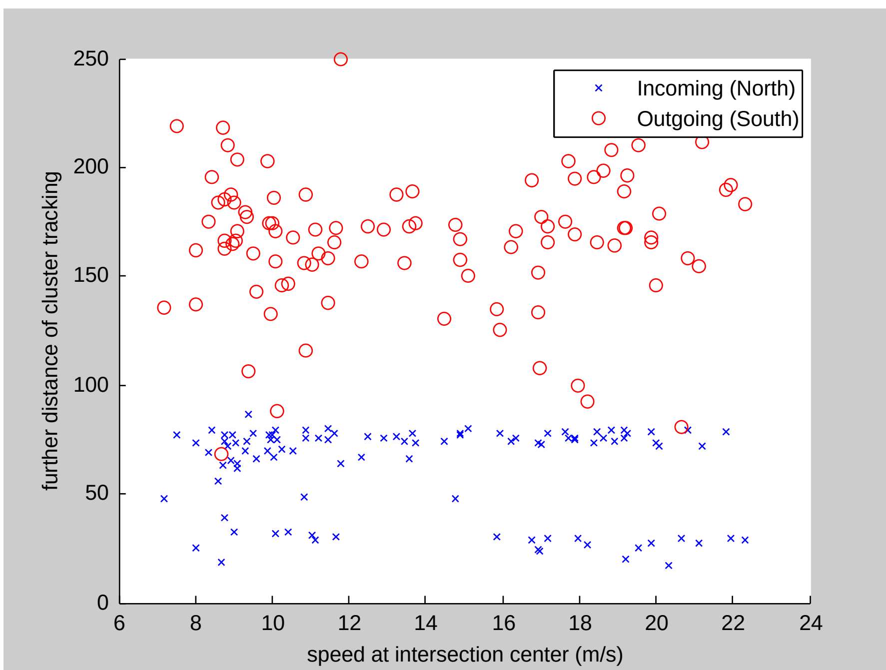


Figure 11.3. Lengths of cluster tracks: through vehicles, north and south directions.

from the one corrupt trajectory, the vehicle movements appear highly consistent. It undoubtedly is the case that some of the lateral dispersion seen is attributable to real-world variations in vehicle path, but some of the dispersion is caused by positioning errors. The two cannot be separated, but the overall result puts lower limits on the system accuracy. The dispersion appears greatest for left-turning vehicles (with only two cameras covering this movement) and possibly at distances further from the intersection.

Thus, lateral dispersion using samples of vehicles is analyzed. Five hundred trajectories of north-to-south vehicles were sampled and their lateral dispersion evaluated at distances of 50 and 25 m, both to the north and to the south of the intersection. Results are shown in Figure 11.5 and

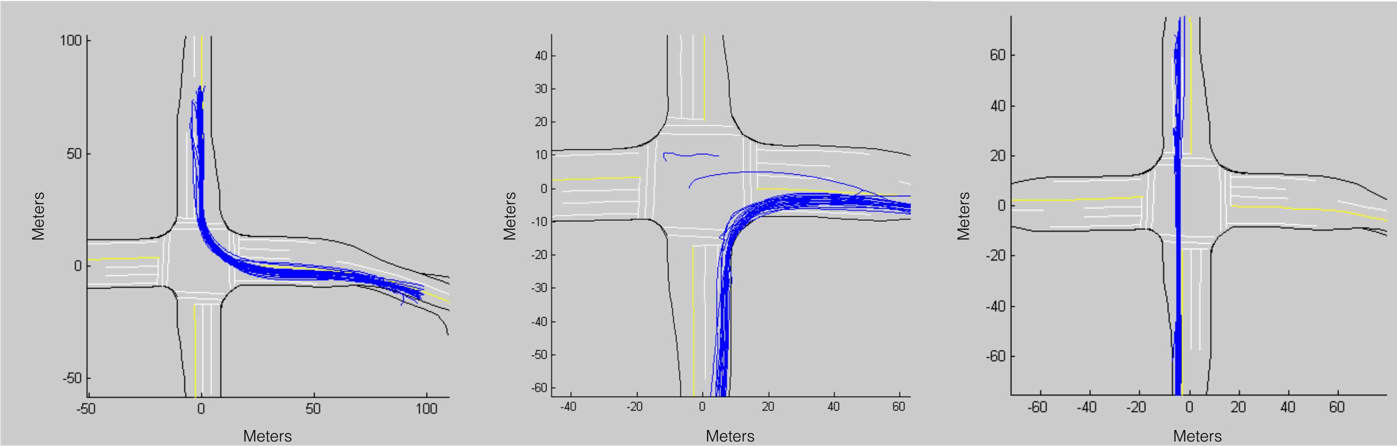


Figure 11.4. Samples of detected trajectories in selected directions. The central plot shows a single corrupted trajectory. Less than 1% of trajectories are this type and can be removed by screening.

Table 11.1. The mean values are biased to negative values, but this likely is because only an approximate value was used for the *x*-coordinate of the lane center, or it could be that drivers tend to follow the right lane marker more when approaching the intersection (additional analysis is easily possible; the power of this type of data set is that with additional effort, any number of detailed investigations are made feasible). The team's focus is on the dispersion values, which are very constant indeed—out to 50 m, there appears to be no disadvantage caused by pixilation errors. From this, it seems that the RMS errors are slightly higher than those suggested in Chapter 5. The ideal requirements for 20-cm RMS accuracy for the prototype system have not been met; rather, these results indicate values approximately twice that of the target, with RMS errors of approximately 40 cm. Although slightly disappointing, these are first generation results from

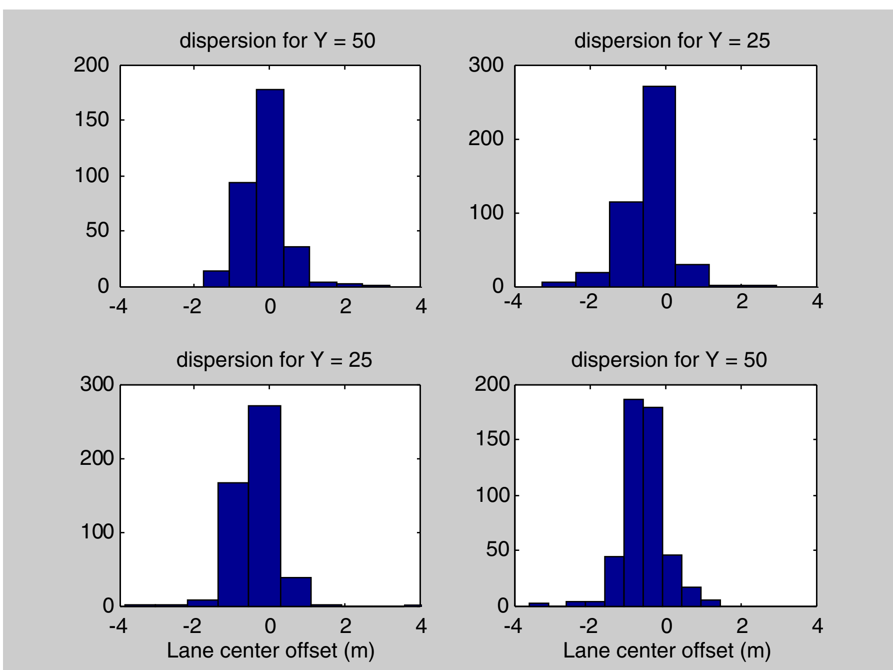


Figure 11.5. Frequency distributions for lateral dispersion within the lane for through vehicles. North-to-south movement: upper left, 50 m north; upper right, 25 m north; lower left, 25 m south; and lower right, 50 m south.

Table 11.1. Mean and Standard Deviations of Lateral Offset

Distance North from Intersection Center	Mean Offset (m)	Standard Deviation (m)
+50	–0.11	0.63
+25	–0.42	0.69
–25	–0.38	0.62
–50	–0.54	0.55

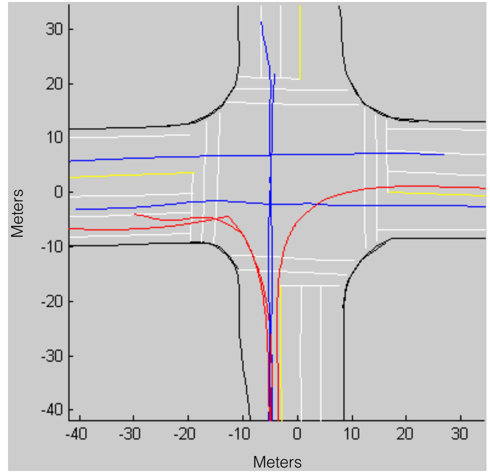


Figure 11.6. Simulated conflict paths using instrumented vehicles.

the prototype system, and there is every reason to expect improvement as various aspects are optimized (see Chapter 13 for additional discussion of this topic).

Speed and Acceleration Evaluation

For speed and acceleration, independent reference information does exist, albeit for a small number of cases. A number of passes were made through the intersection using instrumented vehicles. The onboard sensors measure a large number of variables, including speed, longitudinal acceleration, lateral acceleration, and yaw rate. Four particular (mild) conflict events were staged, two events with a legal right turn on red ahead of and into the path of the other vehicle going straight, and two events with a permissive left turn across path (LTAP/OD) in front of an oncoming through vehicle. Figure 11.6 shows the various paths.

The main interest here is not the conflict metrics (although these do show up in the next chapter) but on checking the performance of motion estimation from the KF. The results, shown below, are reasonably accurate, except around the initial points. The reason for this in not definitely known but seems likely to be associated with increased variation in cluster positions when the vehicle is first detected, approaching the intersection from distance. After this initial deviation, both velocity and acceleration values match the vehicle quite well, within approximately 1 ms^{-1} and 1 ms^{-2}, respectively. In Figure 11.7, the vehicle is approaching at a speed that will exacerbate the effect of early cluster detection, whereas for the left-turning vehicles in Figure 11.8 the challenge is more one of near-zero velocity (as well as reduced camera coverage); initial values are not badly compromised, and although

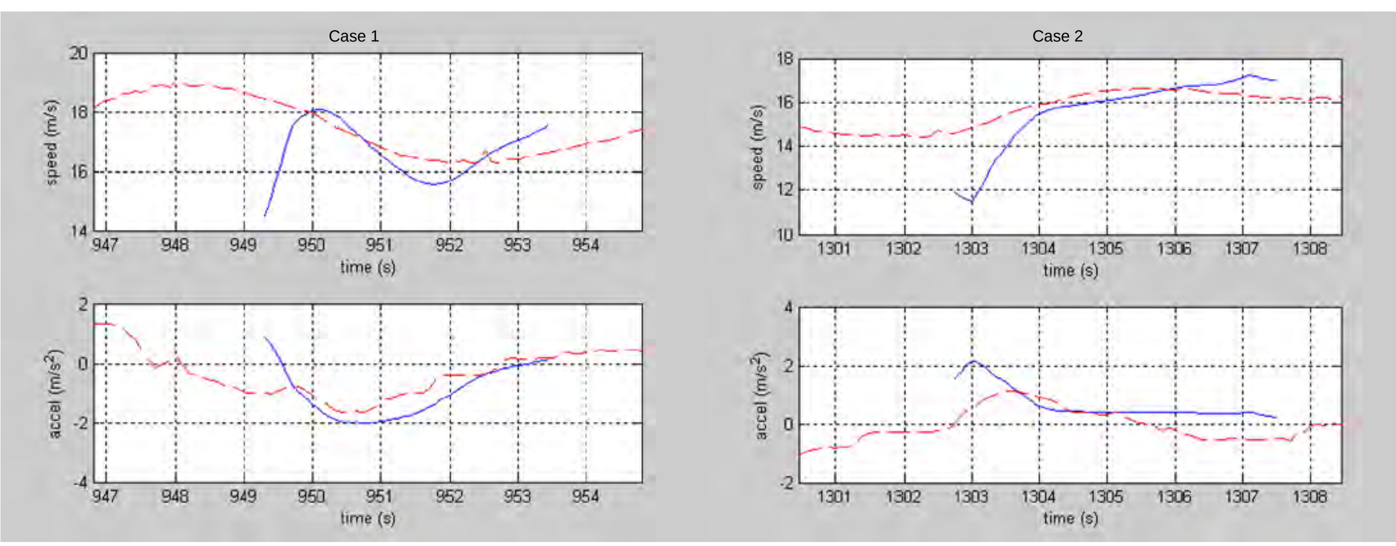


Figure 11.7. Speed and acceleration estimates for through vehicles, right turn into path. Blue: video measurement; red dashes: vehicle measurement.

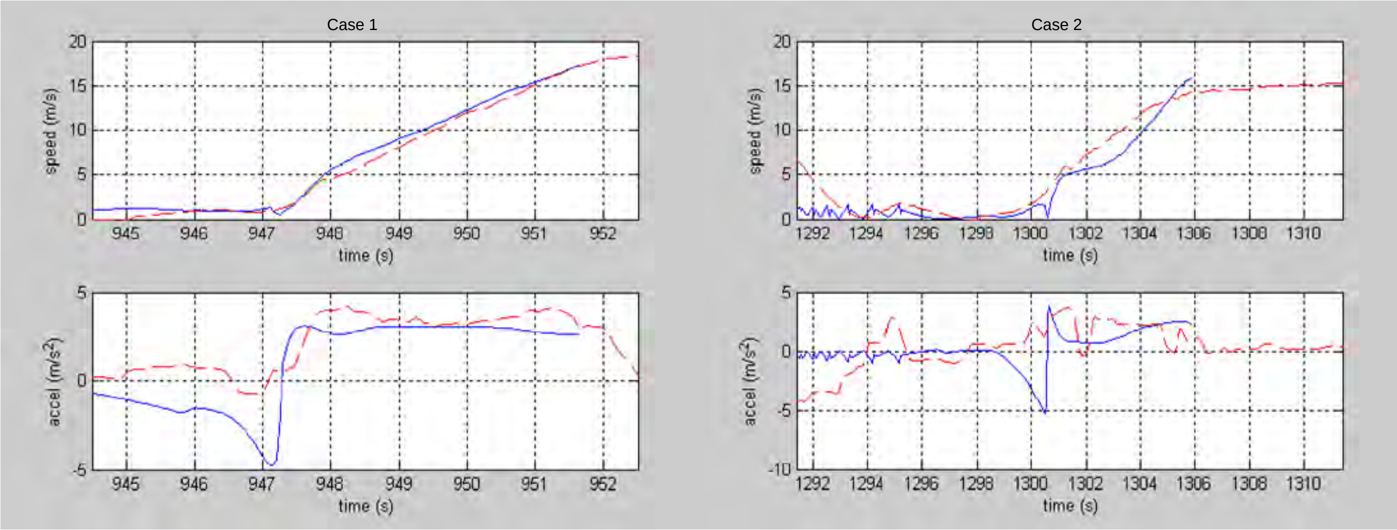


Figure 11.8. Speed and acceleration estimates for slow-turning vehicles, right turn into path. Blue: video measurement; red dashes: vehicle measurement.

accelerations are not especially well determined as the turning vehicle first starts to accelerate, the velocity estimates are stable and match well.

For the cases with left turn across path, the through vehicle again maintains its speed. In fact, for the case shown here (Figure 11.9), a period of acceleration is captured well by the camera system. The velocity and accelerations are not perfect, but they give usable quantitative information, sufficient for calculating time-to-collision values and acceleration or identifying braking events. In Figures 11.10 and 11.11, the KF results for the turning vehicle can be seen; the yaw rate (turning rate) is captured with a high degree of accuracy, and useful information on speed and acceleration is provided. Overall, accepting that there are imperfections because video images, rather than installed sensors, are being used, the Site Observer clearly achieves a highly useful degree of motion capture. The velocity estimates are used in Chapter 12 when TTC distributions are found.

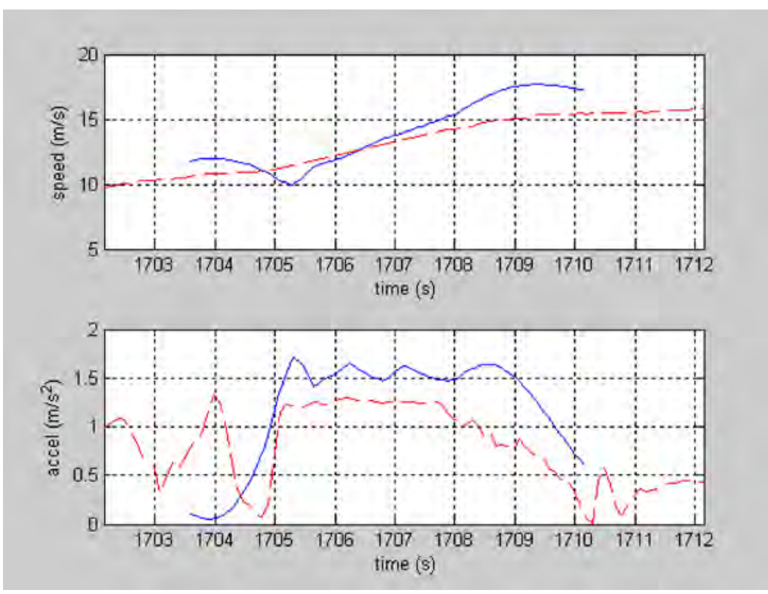


Figure 11.9. LTAP/OD conflict—through vehicle. Blue: video based; red dashes: vehicle based.

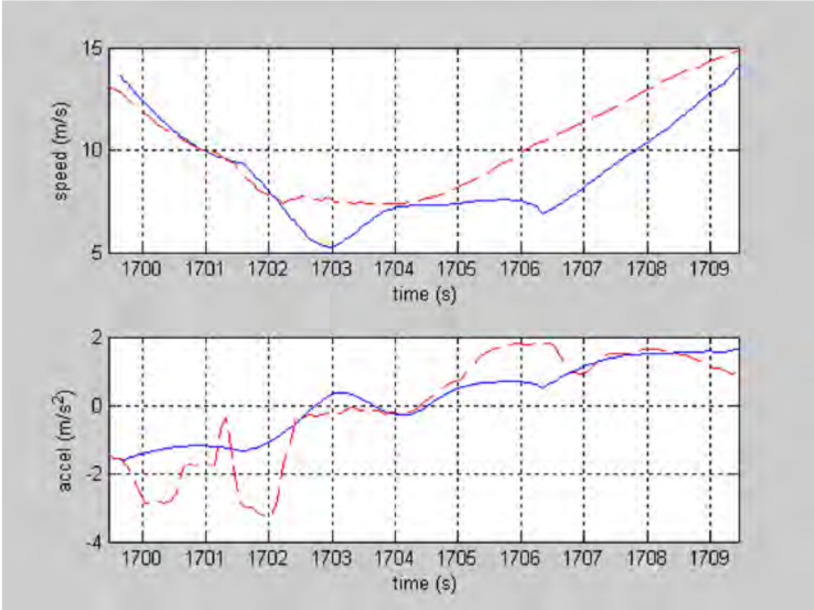


Figure 11.10. LTAP/OD conflict—turning vehicle. Blue: video based; red dashes: vehicle based.

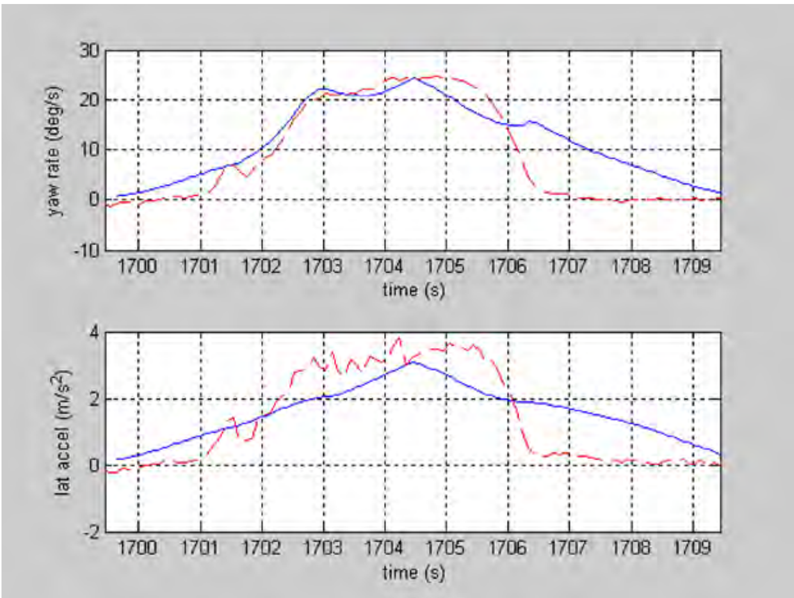


Figure 11.11. LTAP/OD conflict—yaw rate and lateral acceleration estimates for the turning vehicle. Blue: video based; red dashes: vehicle based.

CHAPTER 12

Conflict Analysis

In this chapter, two conflict types that were mentioned in Chapter 5 are considered: left turn across path/opposite direction (LTAP/OD), which is a legally possible maneuver at this intersection because of the presence of permissive left turns, and right turn into path (RTIP), which leads to the possibility of rear-end conflicts if the turning vehicle enters the path of a through vehicle traveling in the same direction.

The approach is to apply triggers to screen for relevant motion types and impose a constraint on the times that the vehicles are closest to the intersection; if two vehicles follow the selected turn paths and are at the intersection within 10 s of each other, they represent a candidate conflict pair. For each case, additional constraints (e.g., for RTIP the turning vehicle should exit the intersection before the through vehicle) can be applied and the appropriate conflict metrics evaluated. In fact, for LTAP/OD, both timing cases where the left turning vehicle turns in front or behind are included. It is also possible to relate turning events to traffic signal state because these data exist in the database, but for this small-scale study (focused on demonstrating the usability of the data) no such analysis is performed.

Left Turn Across Path

For LTAP/OD, postencroachment time (PET) is used as a metric (see Chapter 5). Recall that PET measures the time gap between the turning vehicle exiting the lane of the through vehicle and the time at which the through vehicle arrives. In Figure 12.1, the critical locations are shown with one vehicle trajectory shifted in time to the point of conflict (here for the turning vehicle ahead of the through vehicle); the size of the time shift equals the PET for the event. The extraction of these conflict events is completely automated, including vehicle positioning and size estimation; as described previously, length and width estimation is not fully mature in this analysis, so it is not all that surprising that the vehicle bounding boxes are somewhat distorted from reality. The actual vehicles are shown in Figure 12.2, in which the images shown are four frames apart (at 0.2-s intervals). The image is taken from the southwest camera, so we see the turning vehicle entering from the west and exiting to the north.

All LTAP/OD conflicts were extracted. *However, very few events actually took place.* For the full data set, only 67 permissive left turns were taken, with most (47 cases) with the turning vehicle entering from the east and exiting to the south. Figure 11.1 shows that permissive left-turn events (from the east and west) are relatively rare.

Of these events, only 15 cases included a conflict of the type shown above, with the turning vehicle generating an LTAP/OD conflict with PET <10 s. If other (mild) conflicts with negative PET values (where the turning vehicle waits for the oncoming vehicle to pass before turning) are included, another 47 events were found. Distributions are shown in Figure 12.3. Even with small data volumes it is clear that time separation is much less in the second case.

Right Turn into Path

As a second example of conflict metrics, RTIP events are considered. Starting with the full set of trajectories, all right turns are found, and for each case corresponding straight paths are sought. All pairs of trajectories meeting these requirements and with times nearest the intersection center within ±10 s are analyzed further. Table 12.1 shows that 2,764 such candidate events were found; in most cases, the turning vehicle arrives from the west and turns south and the through vehicle arrives from the north. Of 1,892 cases, in only 11 do vehicles exit to the east.

For these events three successive filters are applied: (1) the turning vehicle must exit the intersection first, (2) the through vehicle should the first in any group paired with the turning vehicle, and (3) the turning vehicle should be last of any

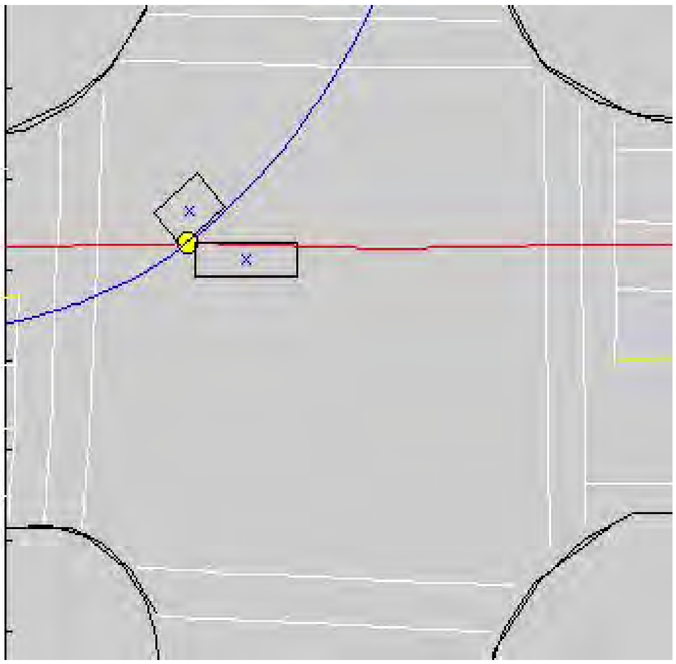


Figure 12.1. Conflict point for left turn across path (PET = 1.78 s).

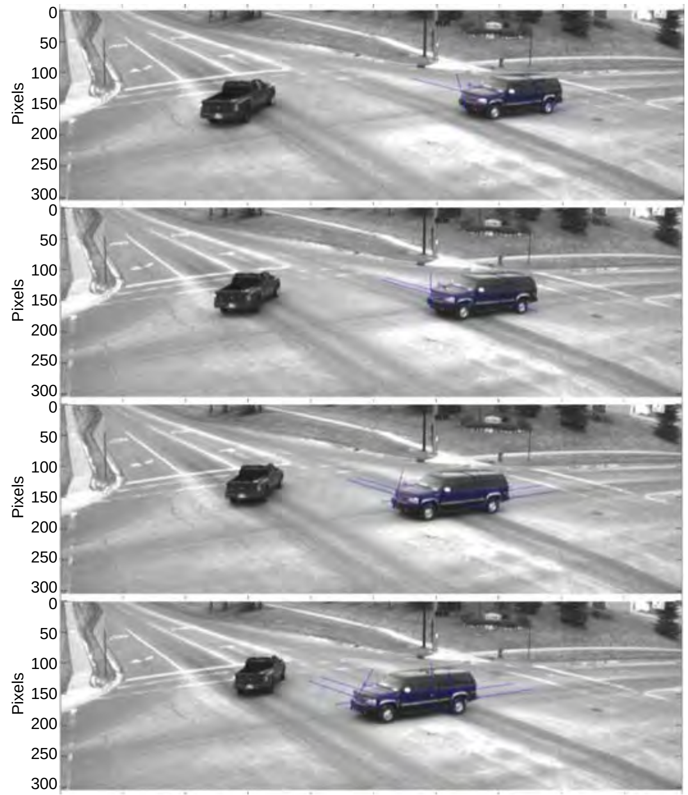


Figure 12.2. Path-crossing conflict (LTAP, PET = 1.78 s).

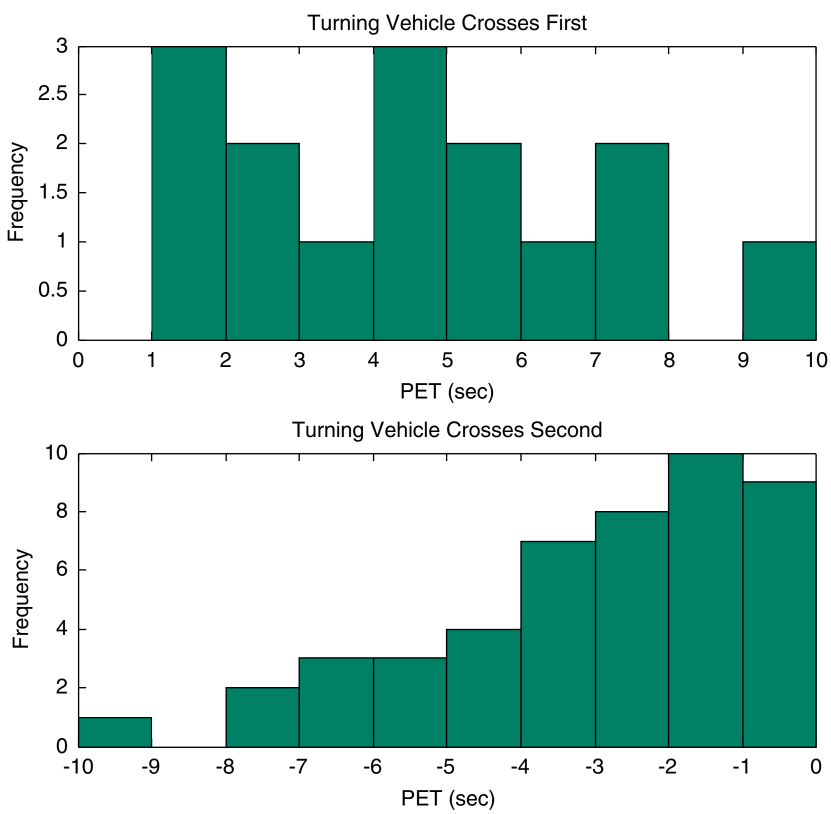


Figure 12.3. Distribution of postencroachment times for permissive left turns (62 cases).

Table 12.1. Counts of RTIP Events as Successive Filters Are Applied

Turning Vehicle Direction	WS	NW	EN	SE	
Through Vehicle Direction	**NS**	**EW**	**SN**	**WE**	**Total**
Coincide within 10 s	1,832	38	823	11	2,764
With turning vehicle leading	416	4	162	2	584
Filtered for nearest vehicle in group	253	4	109	1	367
TTC <10 s	135	3	62	0	200

turning vehicles paired with the lead through vehicle. These filters drastically reduce the number of cases, to 367, as seen in Table 12.1; however, the proportions of cases are roughly maintained. The plots in Figure 12.4 show trajectories of the 367 cases, where red indicates the through vehicle and black dashed lines indicate the turning vehicle. These are simple cluster tracks used for approximate timing and screening, so their positioning is not the most accurate available. Note that there is only one clearly corrupt track in the set, and this is screened out in the next stage, which is to apply the KF to the 367 cases and use speeds and distances to calculate TTC time histories. If the minimum TTC is less than 10 s, the event is retained.

Figure 12.5 shows a typical result. In this event, both vehicles exit to the north; distance is measured in the northerly direction to calculate the valid TTC, even when the turning vehicle is not yet moving due north; the black lines, representing the turning vehicle, are cut off at the beginning so values are found only for when both vehicles are in the through lane. In this case, tracking of the through vehicle started a little after the turning vehicle entered its path, so its position was extrapolated assuming constant speed; the extrapolation is shown in the thicker line. Figure 12.6 shows the time history of TTC during the event, for which the minimum value is the defined conflict metric. Table 12.1 shows there are 200 such events.

Interestingly, the through vehicle is seen to stop (Figure 12.5). This is not because of hard braking to avoid collision (in fact

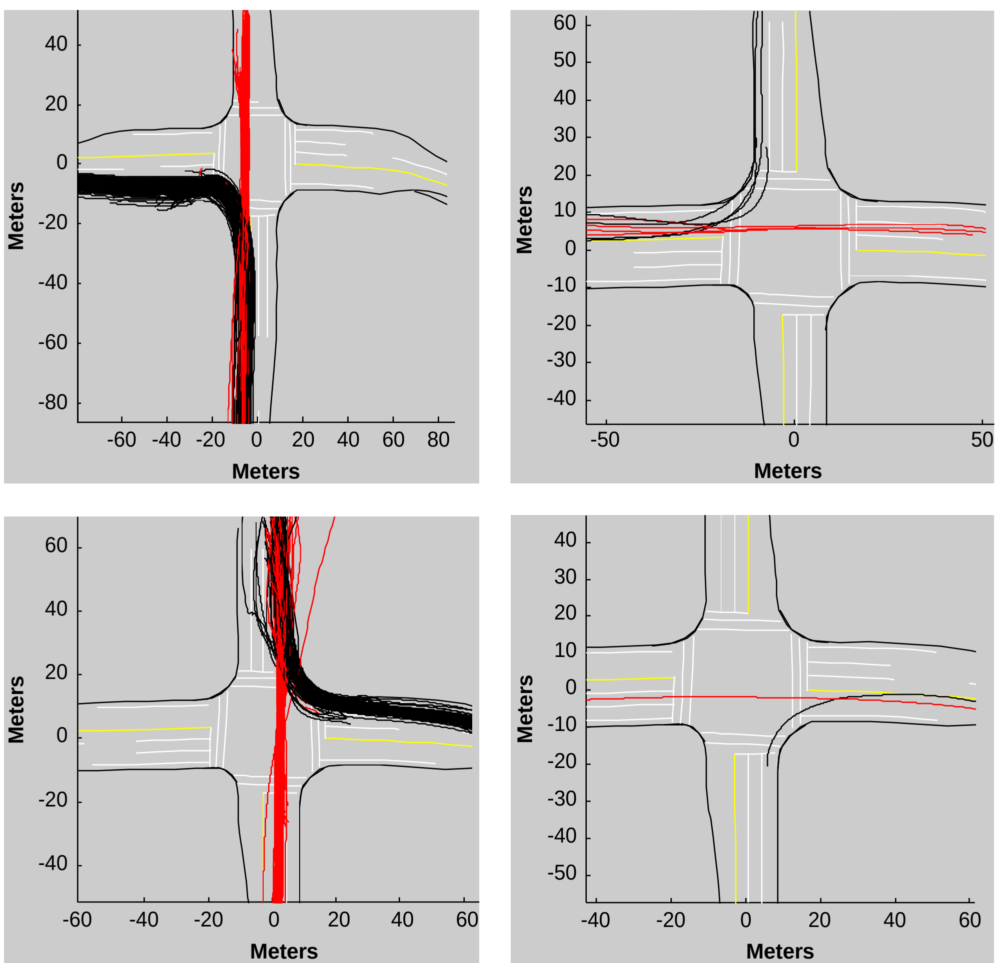


Figure 12.4. Vehicle trajectories used for TTC analysis.

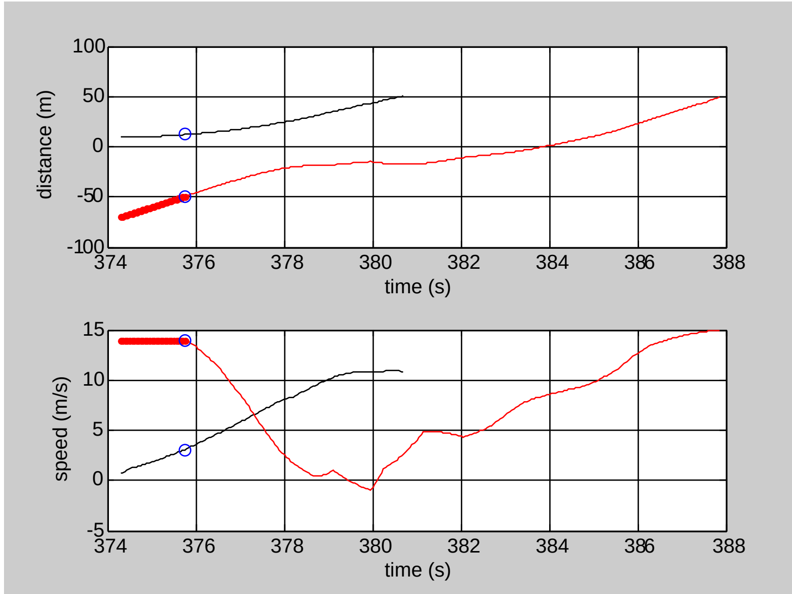


Figure 12.5. RTIP conflict. Red: through vehicle; black: turning vehicle; blue circle: point of minimum time to collision.

no such events took place in the field test). However, checking the traffic signal state and reviewing recorded video, this vehicle slows to a near-stop for a red traffic signal, then the signal turns green and the vehicle accelerates. The turning vehicle simply took advantage of this opportunity.

Finally, Figure 12.7 shows the results of the remaining 200 events. No event had TTC less than 1 s, and the minimum time was 1.4 s. There is a clear pattern of conflict frequencies increasing roughly linearly with increasing TTC value. The ability to draw conclusions or make comparisons between different turning directions, different signal states, traffic densities, weather conditions, and so forth is impossible with such a small event set. However, such analysis would be feasible if more extensive data collection takes place in the future.

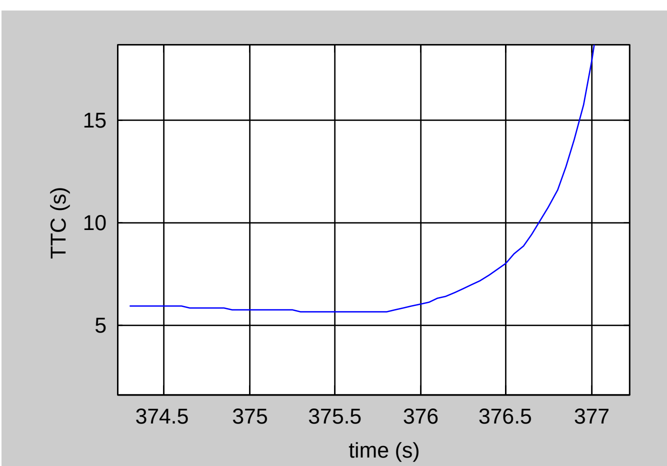


Figure 12.6. TTC time history during the conflict event of Figure 12.5 (TTC_{min} = 5.6 s).

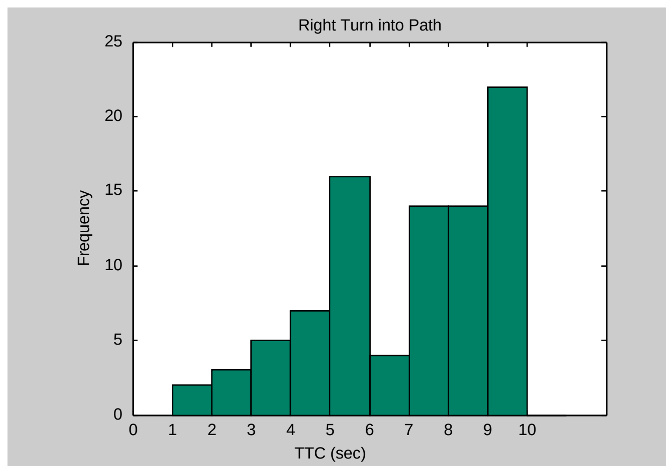


Figure 12.7. Distribution of time-to-collision times, right turn into path (200 events).

CHAPTER 13

Conclusions

The aim of this project was to develop and validate a new observation tool for vehicle safety research—an automated video tracking system, capable of wide-scale use, to captures vehicle trajectories, find traffic conflicts, and compute metrics such as time to collision. The concept design and system development took place with reference to the research agenda of SHRP 2. The system has been shown capable of high-fidelity tracking of vehicles for exposure and conflict analysis, and addressing research questions that are not amenable to other techniques. It achieves this by detection of vehicles using a small number of clusters, localizing the vehicle using overlapping polygons, and then using dense cluster sets to track each vehicle (see Figure 13.1). *The system is immediately capable of supporting new research into risk factors and countermeasure performance.* This is especially important in the areas where vehicle-based naturalistic data collection is silent, such as turning conflicts at intersections.

This report has considered in detail the design and development of the Site Observer. No comparable system currently exists, and the outcome of this project has been to develop a robust prototype system, together with supporting software and sample data from a small field trial; the system has been designed, built, and implemented and has captured representative data.

The system was designed to address key safety research questions posed by SHRP 2. In particular, it is capable of determining a variety of conflict metrics and provides a prototype working tool for evaluating safety performance of highways in the future. The project focused on turning conflicts at intersections, but the system is fully capable of being used at other locations and for other conflicts of particular interest. The system and the data it provides go beyond anything currently available, and the system is expected to become an invaluable tool for researchers and practitioners.

The system uses multiple machine vision cameras, networked together and accurately synchronized, and is organized in a hierarchical way with early stage processing taking place on site, local to where each camera is installed. In the future, it will be possible to integrate the image processing algorithms into the camera; such cameras with built-in digital signal processors are now available on the market. Coupled with wireless data transfer, the design has the scope to be easily installed and highly portable, reducing installation and setup times.

Another way that setup time can be reduced is to extend the process of on-site camera calibration to the point where the lidar survey is not required. This will also reduce cost on a per-installation basis, although for flexibility a pan-tilt-zoom unit should be considered. The whole calibration process may then be performed on site using a reference vehicle fitted with a highly visible marker or set of markers. Software exists to perform the calibrations automatically, provided the vehicle (and thus the markers) are located with sufficient accuracy and a sufficient number of traces are found. Real-time kinematics analysis can be performed after collecting raw GPS data to obtain centimeter-level positioning accuracy, and provided the vehicle is driven slowly to avoid excessive body roll or suspension travel, the necessary accuracy can be obtained. Additional survey points need to be located to determine an intersection map; fixed markers will also be needed to maintain the calibrated camera position or determine any small offsets during use. The requirement will again be met that vertical curvature of the roads is accounted for during vehicle tracking; the modified approach will be equivalent to the use of lidar survey data, avoiding the large data volumes and reducing setup time.

A major advantage of the Site Observer is its ability to reduce data volumes and offer full automation. The system does not require special high-level camera locations, although cameras do need to be mounted strategically to maximize overlapping fields of view. It has been found that covering all vehicle positions with several camera fields of view is important for robust and accurate positioning. Synthesis of extracted features has been carried out using database methods, followed by the

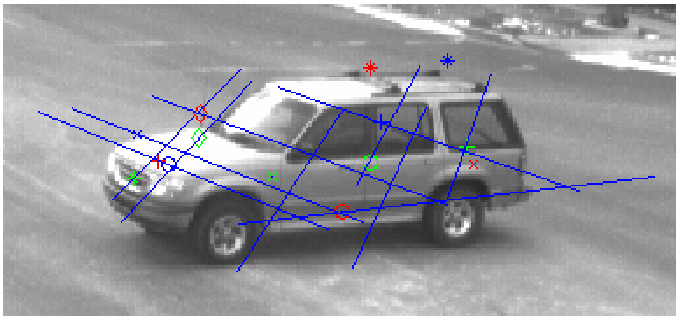


Figure 13.1. Motion capture by multiple cluster tracking.

application of 3-D mappings and dynamic analysis to create time histories of position, velocities, and accelerations.

Implementation took place at a suburban intersection in Ann Arbor, Michigan; approximately 17,000 vehicle trajectories were captured. A feature database was constructed, trajectories were extracted, and sample conflict analyses were undertaken for path crossing and rear-end conflict types (LTAP/OD and RTIP). The current system uses four cameras and five computers but is limited only by hardware costs; the design is fully scalable to larger installations.

An important contribution, beyond the actual development and use of the system, has been to link requirements and design principles directly to the needs of the SHRP 2 Safety program. The project has also validated the new design concept based on a modular and layered approach. While there remains scope for optimization and refinement, the system's tracking accuracy exceeds that of comparable research systems, while at the same time *decoupling the image analysis from the detection and tracking analysis.* To achieve this duality, two complementary feature types—clusters and blobs—have been required Neither feature type on its own is sufficient for the purpose.

The main purpose of the system is to capture research data. It is not essential that all trajectories are captured, only that loss of data does not bias results. Every time a system is switched on, the preceding data has been lost, and when it is switched off, more data are lost. The team samples only the population of interest. Data loss is not the vital issue; the goal is to collect sufficient unbiased data to answer safety research questions. It has been mentioned that *the proportion of corrupt vehicle trajectories at the test site was found to be around 1%,* which is certainly too small to adversely affect the validity of safety research results. The proportion of *missing vehicle trajectories* is a more difficult issue, and again no firm numbers were determined. In manual video review, from many tens of hours of viewing vehicles, together with overlays of triggered cluster tracks, *no cases were found in which a vehicle failed to show at least one such track.* This was not part of a formal analysis or controlled experiment, so it is possible that certain (unknown) conditions can lead to missed vehicles and that these were simply not observed. However, it seems reasonable to conclude that again the proportion of missing vehicles typically is very low. It may be possible in future work to make this part of a formal analysis by using a physical trigger, such as a loop detector, for comparison. Even if both the video and reference system miss a small proportion of actual vehicles, provided the sources of missed trajectories are independent, it will be possible to put a firm estimate of the proportion of missed trajectories.

Several aspects of the system allow scope for optimization and further refinement: the low-level image processing code can be made more efficient for real-time use, camera calibration can be automated using identifiable moving point targets on site (e.g., on the roof of a calibration vehicle), and greater use of wireless networks can allow researchers to control and monitor the system remotely. Also, the current version of the Site Observer has not fully exploited all information available in the feature sets. For example, the use of overlapping blobs projected onto the ground plane has been applied only at selected time instances, and their use can be increased. Overall, there is a tradeoff between computational efficiency and accuracy of results, and through additional field trials, at sites with a greater density of traffic, available tradeoffs can be exploited.

The current version of the Site Observer only sees cars and trucks, and this is based on the design intent within the project; for example, it is blind to bicycles and pedestrians. But the methods are more widely applicable; tracking and localization are feasible provided the initializing detection (triggering) mechanisms are set up for detecting these other road users. The same basic feature sets can be used, although camera positioning should take these extended requirements into account.

Motion variables were compared with those obtained from test vehicles. It was shown that velocities and accelerations can be estimated and provide useful information. However, it is not feasible to use the Site Observer to replicate what is observed in a naturalistic vehicle-based field study, quite apart from the fact that drivers are anonymous to the system. Site-based data are complementary to vehicle-based data, and there are opportunities to join these kinds of data in the future; when instrumented vehicles drive through an observed site, the data can be joined retrospectively by relating time and position information. This approach may offer new opportunities for addressing complex safety questions when driver factors are considered crucial.

The report has shown what the system currently does and mentions future visions for a more portable and packaged system, in effect a consumer product for professional researchers and practitioners. In the meantime, the system as it exists can be usefully deployed in a number ways. With current software and hardware, it can be installed at new locations where traffic densities and conflict levels are more challenging; this could be at a signalized intersection or really at any location of particular interest. The purpose will be to relate conflicts to crashes, either directly or with reference to other variables such as detailed turning counts. Useful data can be collected in a matter of weeks; it is not necessary to wait for crashes to occur. This makes for a very powerful research and evaluation tool.

Future Research

At the field test site used, the number of historical crashes was seen to be very low, but when the many thousands of such sites are considered, the safety problem is far from trivial. In general, relating conflicts to crashes at a single site may have limited scope, unless crashes are sufficiently frequent there. To overcome this limitation, one research option involves recording conflict data at a single location or at a small number of locations. Then, relating conflicts to turning counts (and other factors such as time of day), it will be possible to impute conflict rates at many other comparable locations for which turning count data are available. The resulting conflict and exposure rates can then be directly related to a larger pool of crash numbers and types. Such an analysis will be of immediate benefit in improving the understanding of risk factors and the relationship between crashes and conflicts.

References

Agent Video Intelligence Ltd. 2012. Agent Vi. www.agentvi.com. Accessed September 24, 2012.

Archer, J. 2001. Traffic Conflict Technique: Historical to Current State of the Art. www.ctr.kth.se/publications/ctr2001_05.pdf. Accessed September 10, 2012.

Bar-Shalom, Y., X. R. Li, and T. Kirubarajan. 2001. *Estimation with Applications to Tracking and Navigation.* John Wiley and Sons, Inc., New York.

Bauer, K. M., and D. W. Harwood. 1996. *Statistical Models of At-grade Intersection Accidents.* Report no. FHWA-RD-96-125. U.S. Department of Transportation Federal Highway Administration, McLean, Va.

Bauer, K. M., and D. W. Harwood. 1998. *Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes.* Report no. FHWA-RD-97-106. U.S. Department of Transportation, Washington, D.C.

Bauer, K. M., and D. W. Harwood. 2000. *Statistical Models of At-grade Intersection Accidents: Addendum.* Report no. FHWA-RD-99-094. U.S. Department of Transportation, Federal Highway Administration, McLean, Va.

Bitzel, I. F. 1957. Accident Rates on German Expressways in Relation to Traffic Volumes and Geometric Design. *Roads and Road Construction,* Vol. 35, No. 409, pp. 18–20.

Brodsky, H., and A. S. Hakkert. 1988. Risk of a Road Accident in Rainy Weather. *Accident Analysis & Prevention,* Vol. 20, pp. 161–176.

Cleveland, D. E., L. P. Kostyniuk, and K. L. Ting. 1984. Geometric Design Element Groups and High-volume Two-lane Rural Highway Safety. *Transportation Research Record,* Vol. 960, pp. 1–13.

Cleveland, D. E., L. P. Kostyniuk, and K. L. Ting. 1985. Design and Safety on Moderate-volume Two-lane Rural Roads Safety. *Transportation Research Record,* Vol. 1026, pp. 51–61.

David, N. A., and J. R. Norman. 1975. Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersections. Report no. FHWA-RD-76-129. U.S. Department of Transportation, Washington, D.C.

Eby, D. W., and L. P. Kostyniuk. 2004. Distracted-Driving Scenarios: An Analysis of 2001 Crashworthiness Data System (CDS) Data (SAVE-IT). University of Michigan Transportation Research Institute, Ann Arbor, Mich.

Ervin, R., C. MacAdam, J. Walker, S. Bogard, M. Hagan, A. Vayda, and E. Anderson. 2000, August. System for Assessment of the Vehicle Motion Environment (SAVME). Final Report no. UMTRI-00-21 to the National Highway Traffic Safety Administration, USDOT.

Ervin, R., C. MacAdam, A. Vayda, and E. Anderson. 2001. Applying the SAVME Database of Inter-Vehicle Kinematics to Explore the Natural Driving Environment. In *Proceedings of the Transportation Research Board (TRB) 80th Annual Meeting,* Washington, D.C., January 7–11, 2001.

Federal Highway Administration. 2003. *Manual of Uniform Traffic Control Devices.* U.S. Department of Transportation, Washington, D.C.

Federal Highway Administration. 2006. *Interactive Highway Safety Design Model (2006 release).* www.tfhrc.gov/safety/ihsdm/ihsdm.htm. Accessed September 10, 2012.

Garber, N. J., and R. Gadiraju. 1990. Factors Influencing Speed Variance and its Influence on Accidents. *Transportation Research Record,* Vol. 1213, pp. 64–71.

Gettman, D., and L. Head. 2003. Surrogate Safety Measures from Traffic Simulation Model. Report no. FHWA-RD-03-050. FHWA, Washington, D.C.

Glennon, J. C., T. R. Neuman, and J. E. Leisch. 1986. *Safety and Operational Considerations for Design of Rural Highway Curves.* Report FHWA-RD-86/035. U.S. Department of Transportation, Washington, D.C.

Gloyer, B., H. K. Aghajan, K.-Y. Siu, and T. Kailath. 1995. Video-based Freeway-monitoring System Using Recursive Vehicle Tracking. *Proc SPIE Int Soc Opt Eng,* Vol. 2421, p. 173.

Golob, T. F., and W. W. Recker. 1987. An Analysis of Truck-involved Freeway Accidents Using Log-linear Modeling. *Journal of Safety Research,* Vol. 18, pp. 121–136.

Gordon, T. J., L. P. Kostyniuk, P. E. Green, M. A. Barnes, D. F. Blower, S. E. Bogard, A. D. Blankespoor, D. J. LeBlanc, B. R. Cannon, and S. B. McLaughlin. Forthcoming. *SHRP 2 Report S2-S01C-RW-1: Development of Analysis Methods Using Recent Data: A Multivariate Analysis of Crash and Naturalistic Event Data in Relation to Highway Factors Using the GIS Framework.* Transportation Research Board of the National Academies, Washington, D.C.

Hall, J. W., and O. J. Pendleton. 1989. *Relationships Between V/C Ratios and Accident Rates.* Report FHWA-HR-NM-88-02. U.S. Department of Transportation, Washington, D.C.

Hannah J. T., T. E. Flynn, and W. L. Webb. 1976. Characteristics of Intersection Accidents in Rural Municipalities (Abridgement). *Transportation Research Record: Journal of the Transportation Research Board, No. 160,* Transportation Research Board of the National Academies, Washington, D.C., pp. 79–82.

Harwood, D. W., K. M. Bauer, I. B. Potts, D. J. Torbic, K. R. Richard, E. R. Kohlman-Rabboni, E. Hauer, and L. Elefteriandou. 2002. Safety Effectiveness of Intersection Left- and Right-Turn Lanes. Report no. FHWA-RD-02-089. U.S. Department of Transportation, Federal Highway Administration, McLean, Va.

Harwood, D. W., F. M. Council, E. Hauer, W. E. Hughes, and A. Vogt. 2000. *Prediction of the Expected Safety Performance of Rural Two-Lane Highways.* Report no. FHWA-RD-99-207. U.S. Department of Transportation, Washington, D.C.

Image Sensing. 2008. www.imagesensing.com. Accessed September 11, 2012.

Insurance Institute for Highway Safety. 2007, March 19. Older Drivers. *Status Report,* Vol. 42, No. 3.

Iteris. 2008. *Video Detection Product Overview.* www.iteris.com/vvd.aspx. Accessed September 10, 2012.

Ivey, D., L. Griffin, T. Newton, and R. Laden. 1981. Predicting wet weather accidents. *Accident Analysis and Prevention,* Vol. 13, pp. 83–89.

Joksch, H. C., and L. P. Kostyniuk. 1997. *Modeling Intersection Crashes and Traffic Volume.* Report no. UMTRI-97-37. University of Michigan Transportation Research Institute, Ann Arbor, Mich.

Jones, B. L., L. Janssen, and F. Mannering. 1991. Analysis of the frequency and duration of freeway accidents in Seattle. *Accident Analysis & Prevention,* Vol. 23, pp. 239–255.

Jovanis, P., and J. Delleur. 1981. Exposure-based analysis of motor vehicle accidents. *Accident Analysis & Prevention,* Vol. 13, pp. 83–99.

Juergens, W. R. 1972. *Construction Zone Detour and Temporary Connection Accidents.* Business and Transportation Agency, California Division of Highways.

Kanhere, N. K., S. T. Birchfield, W. A. Sarasua, and S. Khoeini. 2010. *Traffic Monitoring of Motorcycles during Special Events Using Video Detection,* TRB Paper 10-3933. Transportation Research Board, 89th Annual Meeting, Washington, D.C.

Kanhere, N. K., S. T. Birchfield, W. A. Sarasua, and S. Khoeini. 2011. *Evaluation of a Computer-Vision Tracking System for Collecting Traffic Data,* TRB Paper 11-4162. Transportation Research Board, 90th Annual Meeting, Washington, D.C.

Khattak, A. J., P. Kantor, F. M. Council. 1998. *The Role of Adverse Weather in Key Crash Types on Limited Access Roadways: Implications for Advanced Weather Systems.* Paper presented at the 77th Annual Meeting of the Transportation Research Board, Washington, D.C.

Khattak, A. J., A. J. Khattak, and F. M. Council. 2002. Effects of Work Zone Presence on Injury and Non-injury Crashes. *Accident Analysis & Prevention,* Vol. 34, pp. 19–29.

Kihlberg, J. K., and K. J. Tharp. 1968. *Accident Rates as Related to Design Elements of Rural Highways.* Report no. NCHRP 47. Highway Research Board, Washington, D.C.

Kim, Z. 2008. Real Time Object Tracking Based on Dynamic Feature Grouping with Background Subtraction. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), pp. 1–8.

Kim, Z., and J. Malik. 2003. Fast Vehicle Detection with Probabilistic Feature Grouping and its Application to Vehicle Tracking. In *Proceedings of the IEEE International Conference on Computer Vision,* Vol. 1, pp. 524–531.

Kim, Z., G. Gomes, R. Hranac, and A. Skabardonis. 2005. A Machine Vision System for Generating Vehicle Trajectories Over Extended Freeway Segments. In *12th World Congress on Intelligent Transportation Systems* (CD-ROM), ITS America, Washington, D.C.

Liste, F., R. Bernard, and B. Melvin. 1976. *Evaluation of Timber Barricades and Precast Concrete Traffic Barriers in Highway Construction Areas.* Virginia Highway Department and Transportation Research Council, Washington, D.C.

Matthews, L. R., and J. W. Barnes. 1988. Relation Between Road Environment and Curve Accidents. In *Proceedings 14th ARRB Conference,* Part 4, ARRB Group, Melbourne, Australia.

McDonald, J. W. 1953. Relation Between Number of Accidents and Traffic Volume at Divided-highway Intersections. *Highway Research Bulletin,* Vol. 74, pp. 7–17.

Miovision Technologies. 2008. www.miovision.com/home/index.php. Accessed September 10, 2012.

Najm, W., J. Smith, and D. Smith. 2001. Analysis of Crossing Path Crashes. Report no. DOT HS 809 423. NHTSA, Washington, D.C.

Nemeth, Z. A., and D. J. Migletz. 1978. Accident characteristics before, during, and after safety upgrading projects on Ohio's rural interstate system. *Transportation Research Record,* Vol. 672, pp. 19–24.

Next Generation SIMulation (NGSIM). 2008. http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm. Accessed September 19, 2012.

Parkhurst, D. 2006. *Using Digital Video Analysis to Monitor Driver Behavior at Intersections.* CTRE Final Report 2006. Iowa State University, Center for Transportation Research and Education (Project 05-214).

Preusser, D. F., A. F. Williams, S. A. Ferguson, R. G. Ulmer, and H. B. Weinstein. 1998. Fatal Crash Risk for Older Drivers at Intersections. *Accident Analysis and Prevention,* Vol. 309, No. 2, pp. 151–159.

Raff, M. S. 1953. Interstate Highway-accident Study. *Highway Research Board Bulletin,* Vol. 74, pp. 18–43.

Rouphail, N. M., Z. S. Yang, and J. Fazio. 1988. Comparative Study of Short- and Long-term Urban Freeway Work Zones. *Transportation Research Record,* Vol. 1163, pp. 4–14.

Satterwaite, S. P. 1976. An Assessment of Seasonal and Weather Effects of Road Accidents in California. *Accident Analysis & Prevention,* Vol. 8, pp. 87–96.

Schoppert, D. W. 1957. Predicting Traffic Accidents from Roadway Elements of Rural Two-lane Highways with Gravel Shoulders. *Highway Research Bulletin,* Vol. 158, HRB, National Research Council, Washington, D.C., pp. 4–26.

Shankar, V., F. Mannering, and W. Barfield. 1995. Effects of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies. *Accident Analysis & Prevention,* Vol. 27, pp. 371–389.

Siemens USA. 2008. www.itssiemens.com/en/t_nav114.html. Accessed September 10, 2012.

Smith, K., A. Bjelkemyr, J. Bärgman, B. Johansson, and M. Lindman. 2009, June. IVSS Intersection Accidents: Analysis and Prevention, Final Report, Vol. 1 (main report). Ref. AL80 A 2008:73343.

Snyder, J. F. 1974. Environmental Determinants of Traffic Accidents: An Alternate Model. *Transportation Research Record,* Vol. 486, pp. 11–18.

Staplin, L., and A. D. Fisk. 1991. A Cognitive Engineering Approach to Improving Signalized Left Turn Intersections. *Human Factors,* Vol. 33, No. 5, pp. 559–571.

Stokes, R. W., and M. I. Mutabazi. 1996. Rate-quality control method of identifying hazardous road locations. *Transportation Research Record,* Vol. 1542, pp. 44–48.

Tanner, J. C. 1953. Accidents at Rural 3-way Junctions. *Journal of Institute of Highway Engineers,* Vol. 11, pp. 56–57.

Traficon USA. 2008. www.traficon.com/index.jsp. Accessed September 10, 2012.

Highway Capacity Manual. TRB, National Research Council, Washington, D.C., 2000.

University of Michigan Transportation Research Institute (UMTRI). 2010. University of Michigan, Transportation Research Institute, Integrated Vehicle-Based Safety Systems. http://umtri.umich.edu/divisionPage.php?pageID=249. Accessed September 10, 2012.

Versace, J. 1960. Factor analysis of roadway and accident data. *Highway Research Board Bulletin,* Vol. 240, pp. 24–32.

VideoTrak. 2008. Quixote Traffic Corporation (QTC). www.peek-traffic.com/products/video/videotrak.html. Accessed September 10, 2012.

Vogt, A. 1999. Crash Models for Rural Intersections: Four-lane by Two-lane Stop Controlled and Two-lane by Two-lane Signalized.

Report no. FHWA-RD-99-128. U.S. Department of Transportation, Federal Highway Administration, McLean, Va.

Vogt, A., and J. G. Bared. 1998. Accident Models for Two-lane Rural Segments and Intersections. *Transportation Research Record,* Vol. 1635, pp. 18–29.

Washington, S., B. Persaud, C. Lyon, and J. Oh. 2005. Validation of Accident Models for Intersections. Report no. FHWA-RD-03-037. U.S. Department of Transportation, Federal Highway Administration, McLean, Va.

Webb, G. W. 1955. The Relationship between Traffic Volumes at Signalized Intersections. In *ITE Proceedings,* Institute of Transportation Engineers, Washington, D.C., pp. 149–167.

Wierville, W. W., R. J. Hanowski, J. M. Hankey, D. A. Kieliszewski, S. E. Lee, A. Medina, A. S. Keisler, and T. A. Dingus. 2002. Identification and Evaluation of Driver Errors: Overview and Recommendations, FHWA-RD-02-003. Federal Highway Administration, Washington, D.C.

Williams, J. C., S. A. Ardekani, and S. A. Asante. 1992. Motorist Understanding of Left-turn Signal Indications and Auxiliary Signs. *Transportation Research Record,* Vol. 1375, pp. 57–63.

Yilmaz, A., O. Javed, and M. Shah. 2006, December. Object Tracking: A Survey. *ACM Computing Surveys,* Vol. 38, No. 4, Article 13.

Zegeer, C., R. Stewart, D. Reinfurt, F. Council, T. Newman, E. Hamilton, T. Miller, and W. Hunter. 1990. *Cost Effective Geometric Improvements for Safety Upgrading of Horizontal Curves.* Report no. FHWA-RD-90-021. U.S. Department of Transportation, Washington, D.C.

Zimmerman, K., and J. A. Bonneson. 2005. Investigation of the Time-into-red for Red-light-related crashes. 84th Annual Meeting of the Transportation Research Board (CD-ROM). Transportation Research Board, Washington, D.C.

*Membership as of October 2012.

Related SHRP 2 Research

Development of Analysis Methods Using Recent Data: A Multivariate Analysis of Crash and Naturalistic Event Data in Relation to Highway Factors Using the GIS Framework (S01C)