Large-Scale Automated Analysis of Vehicle Interactions and Collisions

Nicolas Saunier, Tarek Sayed, and Karim Ismail

Road collisions are a worldwide pandemic that can be addressed through the improvement of existing tools for safety analysis. A refined probabilistic framework is presented for the analysis of road-user interactions. In particular, the identification of potential collision points is used to estimate collision probabilities, and their spatial distribution can be visualized. A probabilistic time to collision is introduced, and interactions are grouped into four categories: head-on, rear-end, side, and parallel. The framework is applied to a large data set of video recordings collected in Kentucky that contains more than 300 severe interactions and collisions. The results demonstrate the usefulness of the approach for studying road-user behavior and mechanisms that may lead to collisions.

Various approaches have tried to improve understanding of the mechanisms that lead to road collisions. An important distinction is whether analysis relies on microscopic data collected from the field. Recent interest in the use of traffic simulation for safety analysis (*1–3*) can be attributed to the difficulty of collecting adequate microscopic data, as stated by Cunto:

> Ideally, it would be preferable to obtain measures of traffic turbulence [i.e., safety performance] directly from field studies. However, such an approach [is] still not feasible given that it would require real-time monitoring of vehicles in the traffic stream, including those rare combinations of events when a crash is observed and this type of information is not readily available. (*3*)

The present work attempts to tackle the challenge of automatically monitoring all road users, including pedestrians, and extracting their trajectories for safety purposes. The data are collected by using video sensors, and computer vision techniques are used to process the video data (*4*).

Saunier and Sayed proposed a probabilistic framework (*5*), which relies on the concept of the safety hierarchy, that is, there is a continuum of all road users' interactions with collisions at the top, undisturbed passages or safe interactions at the bottom, and traffic conflicts somewhere between the two (*6*). The safety hierarchy is matched by a severity hierarchy, which is based on severity indicators that measure the proximity of an interaction to a collision. It is thought that the observation of all interactions, and traffic conflicts in particular, can be an alternative or complementary approach for analyzing road safety from a broader perspective than that of collision statistics alone. The widely accepted definition of a conflict is "an observational situation in which two or more road users approach each other in space and time to such an extent that a collision is imminent if their movements remain unchanged" (*7*). The core concept of this definition is the collision course and that various chains of events may lead road users to collide. This observation led to the proposal of a probabilistic framework to predict road users' positions and evaluate their probability of collision. This paper refines and expands the previous framework, adding more severity indicators, and applies it to explore a large data set of video recordings of collisions and conflicts.

N. Saunier, Department of Civil, Geological, and Mining Engineering, École Polytechnique de Montréal, C.P. 6079, succ. Centre-Ville Montreal, Quebec H3C 3A7, Canada. T. Sayed and K. Ismail, Department of Civil Engineering, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada. Corresponding author: N. Saunier, nicolas.saunier@polymtl.ca.

Transportation Research Record: Journal of the Transportation Research Board, No. 2147, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 42–50.
DOI: 10.3141/2147-06

RELATED WORK

Road safety studies traditionally rely on historical collision data. As reported by Davis and Morris (*8*), significant effort has been put into developing the *Highway Safety Manual* (HSM): the main tools are statistical models of observational data, using generalized linear models to describe baseline associations between collision frequency and observable road features, and the effect of countermeasures is captured through empirically determined collision modification factors. Yet there are well-recognized problems of availability and quality associated with collision data. Collision data are also intrinsically ill suited for understanding the mechanisms that lead to collisions.

An important concept to model is exposure (*9*). Recent work has shown that elementary units of exposure can be developed on the basis of known aggregate measures, such as annual average daily traffic (*10*). The framework is interesting and supports choices made in this paper for the categories of interaction, but it is disconnected from microscopic data collected in the field.

Davis and Morris expect that the statistical models proposed in the HSM "will be replaced by models explicitly describing mechanisms underlying crash occurrence" and advocate simulation models, in particular because those that "capture underlying mechanisms are usually able to represent a richer and more detailed set of alternatives than are statistical models" (*8*). Use of microscopic traffic simulation for safety analysis is not a new idea (*11*), but it is receiving renewed interest. FHWA funded the Surrogate Safety Assessment Model project to develop a program to automate conflict analysis from the vehicle trajectory data generated by traffic simulations (*1*). Two theses address the topic (*2, 3*), demonstrating that microscopic traffic simulation may be used for the estimation of road safety and performance effects of changes in the transportation system.

The limitations of traffic simulation are related to the difficulty of calibrating models and the suspicion that the models are too simple to replicate complex behaviors. In-field collection of data, in particular of surrogate safety measures (*12*), is necessary for tackling the problem and diagnosing real-world situations such as existing black spots. Such data are collected with various levels of automation. Collection of traffic conflicts has been extensively studied since their conceptualization in the late 1960s, but it is still largely performed manually by on-site observers (*6, 13, 14*). The extreme value method was applied by Songchitruksa and Tarko to estimate the frequency of right-angle collisions at signalized intersections by relying on the postencroachment time, which limits the categories of interaction that can be characterized (*15*). Davis et al. outlined a causal theory and built a minimal model capable of rigorously representing traffic conflicts and crashes, relying on the description of the evasive action (*16*). Yet little work has been done that relies on automated collection of road users' trajectories with the primary goal of safety analysis (*17–22*). This work and the previous work on which it is built (*4, 5*) are unique in their attempt to develop automated systems supporting a general framework for the analysis of road users' interactions and their severity.

PROBABILISTIC FRAMEWORK FOR SAFETY ANALYSIS

Moving-Object Trajectories

Road users have a certain size but are represented for simplicity by a point, for example, their centroid. The measurement of their position in space at each instant constitutes a trajectory. A trajectory is a mapping from a finite set $I \subset \mathbb{R}$ (I is typically a finite set of time instants at regular intervals) to $\mathbb{R}^2$ (the two-dimensional plane) (*23*):

$$I \subset \mathbb{R} \rightarrow \mathbb{R}^2 : t \rightarrow T(t) = (x(t), y(t))$$

The trajectory T of each road user U is measured for the time of its existence in a region of interest. Let $U(t_0)$ represent the knowledge available about a road user U up to time t_0, for example, its past n observed positions $T(t_{-n+1}), \ldots, T(t_{-1}), T(t_0)$. Studying the probability of collision requires the ability to predict road users' future positions. Let $\widehat{T_{t_0}}(t)$ be the prediction made at t_0 for the position of U at $t \geq t_0$, that is, based on the knowledge available about U at t_0.

General Collision Probability

The probability at time t_0 for two road users U_i and U_j to collide is the probability of the event Collision (U_i, U_j) of the two objects being at the same place at the same time at a later time $t \geq t_0$. Let Proximity$_\epsilon(A, B)$ be a function mapping from $\mathbb{R}^2 \times \mathbb{R}^2$ to $\{0,1\}$, called the proximity function, defined in the following way for a given distance d and threshold ϵ:

$$\text{Proximity}_\epsilon(A, B) = \begin{cases} 1 & d(A, B) \leq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The probability at time t_0 of a future collision for two road users U_i and U_j, denoted $P(\text{Collision}(U_i, U_j) | U_i(t_0), U_j(t_0))$, is the probability that there exists an instant $t \geq t_0$ such that

$$\text{Proximity}_\epsilon\left(\widehat{T_{i,t_0}}(t), \widehat{T_{j,t_0}}(t)\right) = 1$$

It can be written by using the complementary event that no collision occurs at any instant $t \geq t_0$ (an upper limit t_1 is added to take into account a reasonable amount of time for the road users to collide as they pass through the scene):

$$P\left(\text{Collision}(U_i, U_j) \middle| U_i(t_0), U_j(t_0)\right) = 1 - \int_{t_n}^{t_2} P\left(\text{Proximity}_\epsilon\left(\widehat{T_{i,t_0}}(t), \widehat{T_{j,t_0}}(t)\right) = 0 \middle| U_i(t_0), U_j(t_0)\right) dt \tag{2}$$

In the following, the conditional part of the probabilities is dropped for simplicity. The probability of noncollision at t, that is, of the proximity function Proximity$_\epsilon$ to be 0 at t, is the joint probability of the road users being in positions at t further than the given distance ϵ, which, assuming that the road users move independently, can be computed by

$$P\left(\text{Proximity}_\epsilon\left(\widehat{T_{i,t_0}}(t), \widehat{T_{j,t_0}}(t)\right) = 0\right) = \int_{A_i \in \mathbb{R}^2} \int_{A_j \in \mathbb{R}^2} P\left(\widehat{T_{i,t_0}}(t) = A_i\right) P\left(\widehat{T_{j,t_0}}(t) = A_j\right) \left(1 - \text{Proximity}_\epsilon(A_i, A_j)\right) dA_i dA_j \tag{3}$$

The independence assumption relies on the definition of traffic conflicts as interactions in which a collision is imminent if the road users' movements remain unchanged, that is, if the road users do not react to each other. If it is possible to draw from the distribution of possible future positions for road users, the probability of collision may be computed by simulation, counting the number of situations in which the proximity function is equal to 0.

Movement Extrapolation

Road users have particular dynamics that can be described by prior knowledge of their motion model and empirical knowledge learned from observations. This work relies on learning the distribution of road users' trajectories from observations (*24*), which can be used to predict road users' future positions with associated probabilities. More precisely, an extrapolation hypothesis H is defined by a trajectory $I \subset \mathbb{R} \rightarrow \mathbb{R}^2 : t \rightarrow H(t) = (x(t), y(t))$, derived from an observed prototype trajectory by translation and resampling, with a probability $P(H)$ of the road user following the extrapolation hypothesis. Each road user U_i can be assigned at t_0 a finite set of M_i extrapolation hypotheses $\{H_{i,1}, \ldots, H_{i,M_i}\}$ such that $\sum_{1 \leq m \leq M_i} P(H_{i,m}) = 1$. For N_U road users existing at time t_0, the sample space is the Cartesian product $\{H_{1,1}, \ldots, H_{1,M_1}\} \times \cdots \times \{H_{N_U,1}, \ldots, H_{N_U,M_{N_U}}\}$. Each road user is assumed to follow an extrapolation hypothesis independently from other road users.

It is then possible to enumerate the road users' predicted positions at each future instant $t \geq t_0$ for a limited time horizon (see t_1 in Equation 2) and identify the instants at which the proximity function will be 1, called collision points. A collision point CP_n is defined for two extrapolation hypotheses H_{i,m_i} and H_{j,m_j} as the first instant $t_n \geq t_0$ at which Proximity$_\epsilon(H_{i,m_i}(t_n), H_{j,m_j}(t_n)) = 1$. Let f be the function that associates to each pair of extrapolation hypotheses a collision point if it exists, or the element NoCollision (f is symmetric with respect to its inputs), and let g_1 and g_2 be the inverse functions defined over the collision points that return the extrapolation hypotheses that lead to the collision point. If $f(H_1, H_2) = CP$, then $g_1(CP) = H_1$ and $g_2(CP) = H_2$ (or vice versa).

Case of Two Isolated Road Users

The first simple case of two isolated road users U_i and U_j is considered (as if there are only two road users in the region of interest). They may collide at any of N_{CP} potential collision points. The event of their collision at collision point CP is denoted Collision(U_i, U_j, CP), which can be written simply Collision(CP) since CP is associated with the two given extrapolation hypotheses of the two road users that lead to it. The probability of Collision(CP) is approximated by the product of the probabilities of following $g_1(CP)$ and $g_2(CP)$. This can be written as

$$P\left(\text{Collision}\left(U_i, U_j\right)\right) = \sum_{1 \le n \le N_{CP}} P\left(\text{Collision}\left(CP_n\right)\right) = \sum_{1 \le n \le N_{CP}} P\left(g_1\left(CP_n\right)\right) P\left(g_2\left(CP_n\right)\right) \quad (4)$$

The formula presented by Saunier and Sayed (5) as a collision probability is very close to Equation 4 but cannot be properly considered as a probability and is referred in this paper as the severity index. The formula to compute the severity index for two road users U_i and U_j at t_0 is

$$\text{SeverityIndex}\left(U_i, U_j, t_0\right) = \sum_{1 \le n \le N_{CP}} P\left(g_1\left(CP_n\right)\right) P\left(g_2\left(CP_n\right)\right) e^{-\frac{(t_n - t_0)^2}{2\sigma^2}} \quad (5)$$

where σ is a normalizing constant, equal to an average user reaction time (chosen as 1.5 s in this paper).

Collision Probability and Other Indicators for Any Number of Road Users

However, analysis of the situation is much more complex if all road users are considered (and if there are three or more road users). The probability of collision at a collision point CP is the probability of the road users following the extrapolation hypotheses leading to the collision point and not having previously collided with other road users. As in the previous case, the N_{CP} collision points are enumerated for all road users and are now ordered by their predicted instant of occurrence so that $t_n \le t_{n+1}$ $\forall$ $1 \le n \le N_{CP} - 1$.

The probability of collision at the collision point CP_n is the probability of the corresponding road users to follow $g_1(CP_n)$ and $g_2(CP_n)$ and that there is no collision point CP_m with another road user occurring before CP_n (i.e., with $m < n$) involving one of the previous extrapolation hypotheses (i.e., such that there exists another extrapolation hypothesis H and $f(g_1(CP_n), H) = CP_m$ or $f(g_2(CP_n), H) = CP_m$). This can be computed recursively as

$$P\left(\text{Collision}\left(CP_1\right)\right) = P\left(g_1\left(CP_1\right)\right) P\left(g_2\left(CP_1\right)\right)$$

$$\forall\ 1 \le n \le N_{CP} - 1$$

$$P\left(\text{Collision}\left(CP_{n+1}\right)\right) = P\left(g_1\left(CP_{n+1}\right)\right) P\left(g_2\left(CP_{n+1}\right)\right) \prod_{\substack{1 \le m < n \text{ such that } CP_m \text{ involves} \\ g_1(CP_{n+1}) \text{ or } g_2(CP_{n+1})}} \left(1 - \text{Collision}\left(CP_m\right)\right) \quad (6)$$

Obtaining the individual collision probability for a single road user and a pair of road users at t_0, as well as the severity index for a pair of road users, is then a matter of summing over the corresponding collision points:

$$P\left(\text{Collision}\left(U_i, U_j\right)\right) = \sum_{\substack{1 \le n \le N_{CP} \text{ such that} \\ CP_n \text{ involves } U_i \text{ and } U_j}} P\left(\text{Collision}\left(CP_n\right)\right) \quad (7)$$

$$P\left(\text{Collision}\left(U\right)\right) = \sum_{\substack{1 \le n \le N_{CP} \text{ such that} \\ CP_n \text{ involves } U}} P\left(\text{Collision}\left(CP_n\right)\right) \quad (8)$$

$$\text{SeverityIndex}\left(U_i, U_j, t_0\right) = \sum_{\substack{1 \le n \le N_{CP} \text{ such that} \\ CP_n \text{ involves } U_i \text{ and } U_j}} P\left(\text{Collision}\left(CP_n\right)\right) e^{-\frac{(t_n - t_0)^2}{2\sigma^2}} \quad (9)$$

The expected time to collision (TTC) for two road users U_i and U_j can also be computed in this framework if $P(\text{Collision}(U_i, U_j)) > 0$, that is, if there is at least one collision point:

$$TTC\left(U_i, U_j, t_0\right) = \frac{\sum_{\substack{1 \le n \le N_{CP} \text{ such that} \\ CP_n \text{ involves } U_i \text{ and } U_j}} P\left(\text{Collision}\left(CP_n\right)\right) t_n}{P\left(\text{Collision}\left(U_i, U_j\right)\right)} \quad (10)$$

In a simple example with three road users and four collision points, presented in Figure 1, the resulting probabilities of collision at the collision points are

$$P\left(\text{Collision}\left(CP_1\right)\right) = P\left(H_{1,1}\right) P\left(H_{3,1}\right)$$

$$P\left(\text{Collision}\left(CP_2\right)\right) = P\left(H_{1,2}\right) P\left(H_{3,1}\right)$$

$$P\left(\text{Collision}\left(CP_3\right)\right) = P\left(H_{1,1}\right) P\left(H_{2,1}\right)\left(1 - P\left(\text{Collision}\left(CP_1\right)\right)\right)$$

$$P\left(\text{Collision}\left(CP_4\right)\right) = P\left(H_{1,1}\right) P\left(H_{2,2}\right)\left(1 - P\left(\text{Collision}\left(CP_1\right)\right)\right)$$

The probabilities of collision for the pairs of road users are the sum of the probabilities of collision at the corresponding collision points, CP_1 and CP_2 for U_1 and U_3, CP_3, and CP_4 for U_1 and U_2. These are also respectively the individual collision probabilities of U_3 and U_2 since they are each involved in a potential collision with only one other road user, U_2. Finally, the individual probability of collision for U_1 is the sum of the probabilities of collision at all collision points.

Interactions and Categories

The elementary traffic events considered in the analysis are road user interactions. An interaction is defined as a situation in which

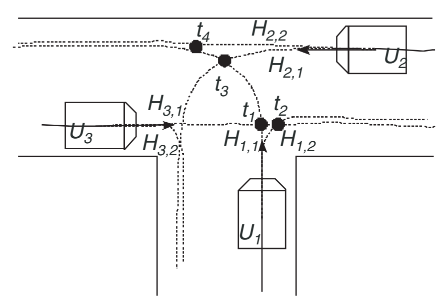

FIGURE 1 Simple example of three road users in interaction at three-leg intersection: four collision points, CP_1, CP_2, CP_3, and CP_4, at respective times t_1, t_2, t_3, and t_4, ordered temporally. (U_3 is too early at point of intersection for hypotheses $H_{2,1}$ with $H_{3,1}$ and $H_{3,2}$ for collision point to be considered.)

two or more road users are close enough in space and time, and their distance is decreasing. This is a necessary condition for road users to collide, that is, some form of exposure to the risk of collision following the definition by Elvik et al. of "an elementary unit of exposure . . . as any clearly defined and countable event that generates an opportunity for an accident to occur" (*10*). This is implemented as a test over the distance between road users, and over the cosine of the angle θ between the relative velocity, that is, the difference of the road users' velocities, and the vector that links the vehicle positions (see Figure 2). The actual condition is $\cos(\theta) \geq 0$, and a value close to 1 means that vehicles are heading almost straight toward each other. Simple measurements are made for all interactions at all instants, namely, the distance, cosine of the velocities, and speed differential (norm of the difference of the velocities). The collision probability, severity index, and TTC are also computed at each time instant: if there is no collision point between the interaction road users, the collision probability and severity index are 0, and TTC is undefined.

Interactions are classified according to the relative trajectories of the road users. Four categories are proposed similarly to those of Elvik et al. (*10*):

- Head-on: road users moving in opposite directions;
- Rear-end: road users following each other, potentially on different lanes;
- Side: road users originating from potentially conflicting directions, for example, at intersections; and
- Parallel: road users traveling in parallel in the same direction in different lanes.

Categories are identified by counting the number of instants at which the angle $\varphi(t)$ between the road users' velocities is within some intervals:

- If $\varphi(t) \in [-30°, 30°]$, the instant t counts for rear-end or parallel interactions.
- If $\varphi(t) \in [-180°, -150°]$ or $\varphi(t) \in [150°, 180°]$, the instant t counts for head-on interactions.
- If $\varphi(t) \in [30°, 150°]$ or $\varphi(t) \in [-150°, -30°]$, the instant t counts for side interactions.

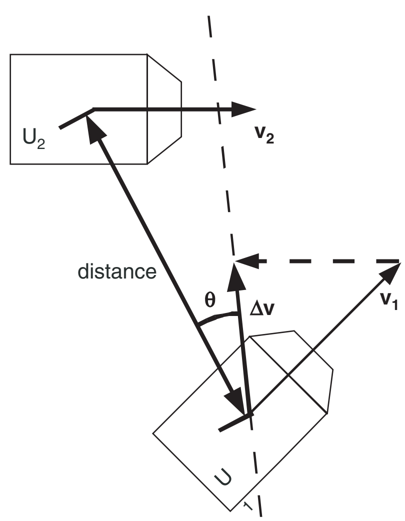

FIGURE 2 Interaction measurements at each instant. Two road users are defined to be in interaction if distance is below given threshold and $\cos(\theta) \geq 0$ (θ is angle between relative velocity $\Delta\vec{V} = \vec{v_1} - \vec{v_2}$ and vector that links road users' positions).

Rear-end and parallel interactions are further differentiated by the relative position of the road users for their common direction of movement. These rules may be crude and arbitrary, especially because interactions may last some time and really belong to different categories at different intervals. But they help to broadly categorize interactions. All the conditions and measures used to characterize interactions are symmetric with respect to the two road users, as they should be.

EXPERIMENTAL STUDY OF COLLISIONS AND CONFLICTS

Data Set and System Description

The framework proposed in this paper is used to explore a large data set of video recordings of traffic conflicts and collisions in Kentucky. The data set was first mentioned by Kim (*22*), who used it to test the video-based tracking system. All the analysis reported here was carried out with the video recordings as the only source of information. Another report provides details about the origin of the data (*25*). There are two subsets of video recordings, miss and incident, corresponding, respectively, to traffic conflicts of mild to high severity and collisions. It is not clear from the work of Green et al. how the severity was estimated to identify the subset of traffic conflicts (*25*). Each short video recording is composed of two subsequences with opposite viewpoints taken from two different cameras, each less than 10 s long. For this study, only one of these is used. Each recording contains, or should contain, one clear interaction, that is, a traffic conflict or a collision. From the original set of 238 traffic conflicts, nine were removed because of the poor quality of the video or because no obvious relevant traffic event was recorded. For similar reasons, 15 collisions were not considered from the original set of 116 collisions. In some cases, it appears that the recording started after the event of interest. It is not always clear if a collision occurred for interactions in the collision subset. Although the framework proposed in this paper is generic and can deal with all road users, the interactions contained in this data set involved only motorized vehicles.

The calibration of the scene is necessary to recover the real-world positions of the road users from their coordinates in image space. A robust method was developed to integrate various pieces of geometric information found in urban traffic scenes and address situations where little is known about the location, which is common when data are obtained from traffic cameras installed previously (*26*); see Figure 3. The quality of the video data makes detection and tracking of road users a challenge. The video recordings have a resolution of 352 pixels wide by 240 pixels high, and varying levels of compression, color aberrations, and so forth affect image quality. All the challenging conditions are covered, with various times of recording (day and night) and weather conditions: sunny days, which have strong shadows, and snow, fog, and rain (sometimes at night, when the reflection of vehicle headlights causes significant glare). Although these issues made some recordings impossible to analyze, detection and tracking of road users was possible in most recordings (see Figure 4) by using a video-based system developed previously (*4*). The parameters for tracking were taken from

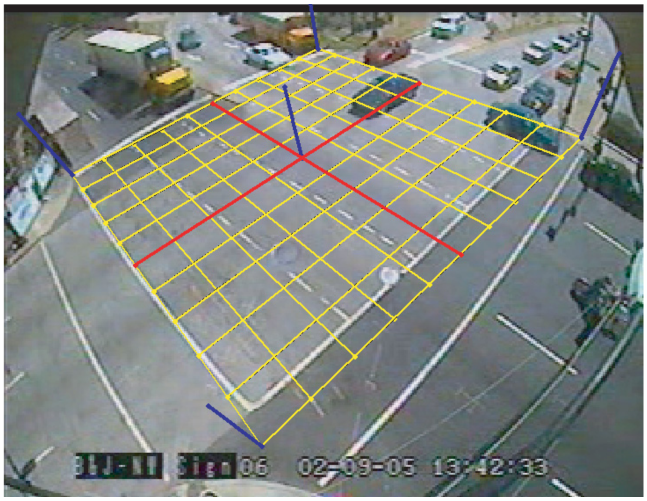

FIGURE 3 Reference grid in world coordinates projected in image space and overlaid on video frame. Grid spacing is 2 m, height of displayed vertical line segment (in blue) is 4 m.

a previous validation study done on a separate data set in which an automated search for the best parameters was conducted (*27*).

The distribution of road user trajectories, in the form of prototype trajectories, was determined by using a distance of 4 m (*24*). All pairs of road users existing simultaneously are considered. If they satisfy the conditions of interaction, their positions are extrapolated with the prototypes, the collision points are identified, and the severity indicators computed automatically. The distance threshold for the proximity function is set to 1.7 m, which corresponds to a typical minimum width of current cars. Assuming that road users' estimated positions are close to their centers, a location at a distance of less than this threshold means a collision occurred or was barely avoided.

Example of Interactions

Most interactions of interest in the data set are categorized as either side or parallel interactions, on which the rest of this study will focus. The collision probability and TTC as a function of time are represented for a few interactions of the two subsets in Figure 5. It was found that the severity index did not carry much additional information (it combines the collision probability and TTC) and is therefore not displayed. A very distinctive feature is that overall, the TTC exhibits a decreasing trend as time goes by for collisions, as one would expect. On the contrary, the TTC reaches a minimum, then increases again for the traffic conflicts, as the road users manage to avoid the collision. It may be surprising that extrema for the two indicators may be reached at different instants. Yet the gap is typically limited and is related to difficult tracking and less-accurate movement extrapolation when the road users become very close.

DISTRIBUTION OF INDICATORS AND COLLISION POINTS

To plot the distributions of indicators for all interactions, an aggregated measure is needed to characterize each interaction. A method similar to the one adopted by Saunier and Sayed is used (*5*), namely, to average the *n* maximum (respectively, minimum) values for the collision probability (respectively, for the TTC). The distributions are drawn for interactions with an aggregated collision probability above a given threshold, set to 0.1 in these results. The distributions are plotted for the two subsets of traffic conflicts and collisions in Figure 6. It would be expected that collisions would exhibit more-severe indicator values. This is not obvious on the plots, although collisions reach higher collision probabilities, and there is a high proportion of them with TTC around 0.5 s and 1 s. The two-sample Kolmogorov–Smirnov test was used to compare the sample distributions of each of the two indicators for the two subsets: the distributions were found to differ significantly at 4% and 1%, respectively, for the collision probability and the TTC.

It is also possible to study the spatial distribution of the interactions and in particular their potential collision points. For the same interactions as previously, the maps of all the collision points are plotted in Figure 7. It can be seen that the distribution is quite different, although conclusions are difficult to draw since the conditions of the data collection are unclear. This type of visualization should be useful for exploring large amounts of microscopic road safety data.

DISCUSSION OF RESULTS

An important aspect of this work is the level of automation of the data processing. Except for the sample of manually chosen interactions, all the processing can be done automatically. Yet, it would be difficult to rely on the tool without verifying its output. Only the side and parallel interactions were studied in detail, because the data set contains very little actual rear-end interactions and no head-on interactions. The interactions detected by the system as rear-end or head-on cover a lot of normal interactions, or at least interactions not as severe as some computed indicators could imply. These limits are first and foremost the limits of the current video-based systems for road user detection and tracking in urban intersections. The second source of errors in this analysis was the challenging data quality and the lack of information.

However, this system can be useful in the exploration of road safety data. A particular focus is the development of methods robust to errors and noise characteristics of real data, to produce aggregated results such as distributions that can be used for road safety diagnosis.

Another limitation of this study is the available data set with its high proportion of severe interactions. Conclusions are difficult, unlike, for example, studies of data collected before and after a countermeasure is implemented. The lack of normal traffic makes comparisons difficult and has an influence on the distribution of the prototype trajectories used for the prediction of road users' positions. Evasive actions may therefore have been picked up as prototype trajectories. It is believed that the impact is limited because these trajectories should not be common and therefore have low probabilities.

CONCLUSION AND FUTURE WORK

This paper presented a refined probabilistic framework for the analysis of road user interactions. In particular, the identification of potential collision points was used to estimate collision probabili-

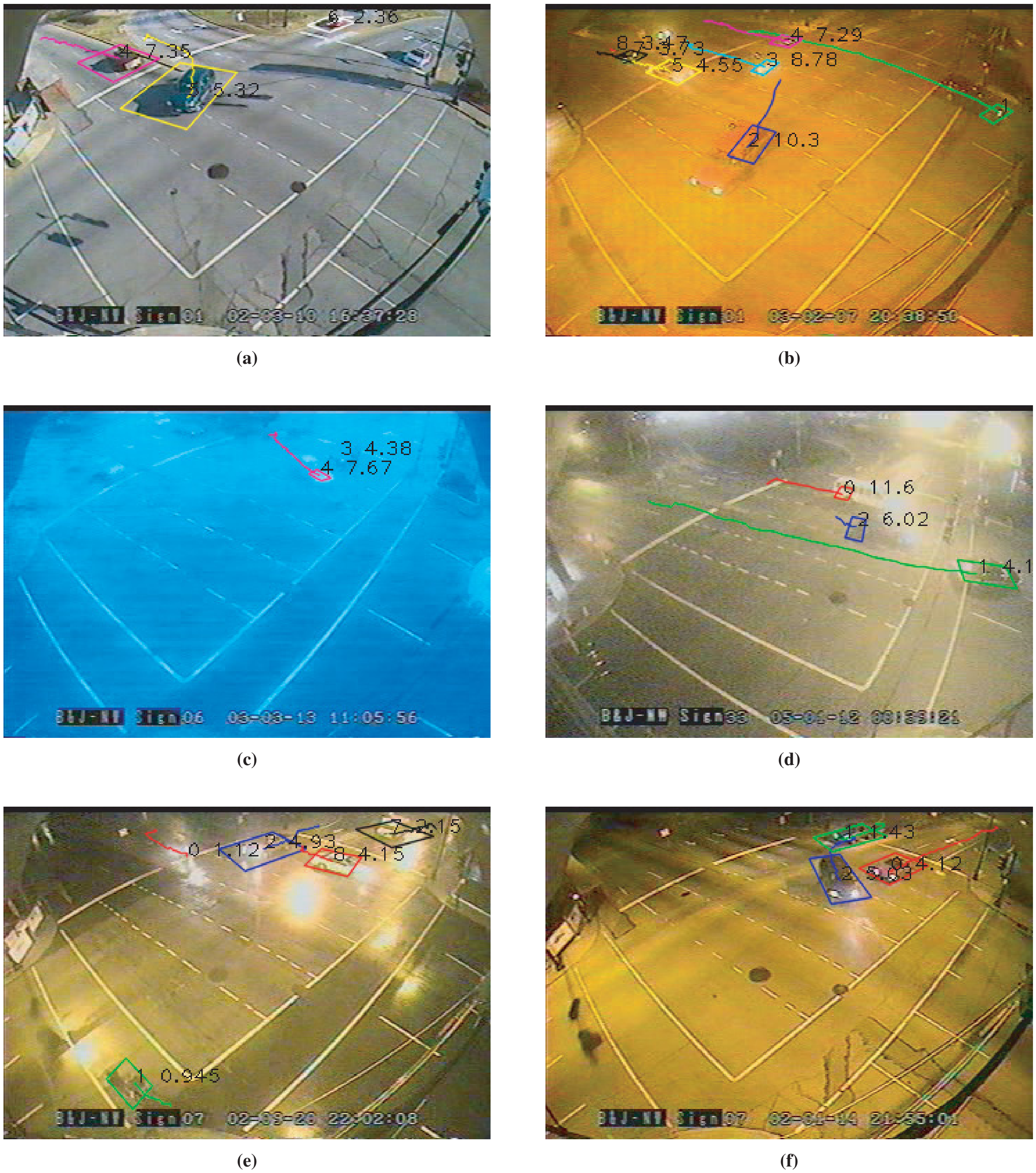

FIGURE 4 Examples of video recording conditions with road users' tracks, identification number, and speed overlaid.

ties, and their spatial distribution can be visualized. The framework was applied to a large data set of more than 300 severe interactions and collisions collected in Kentucky. Despite the quality of the data, the road users could be tracked and their interactions studied, including the computation of the proposed severity indicators. This demonstrates the usefulness of the approach in studying road user behavior and mechanisms that may lead to collisions.

Future work will explore the possibility of simulating future positions to generate more-varied outcomes and improve the robustness of the computation of the probability of collision. It will also focus on the validation of the proposed measurements and severity indicators with respect to other methods for road safety analysis, in particular as based on historical collision data.

ACKNOWLEDGMENTS

The authors thank Zu Kim of California Partners for Advanced Transit and Highways (PATH) and Ann Stansel of the Kentucky Transportation Cabinet for providing the video data set.

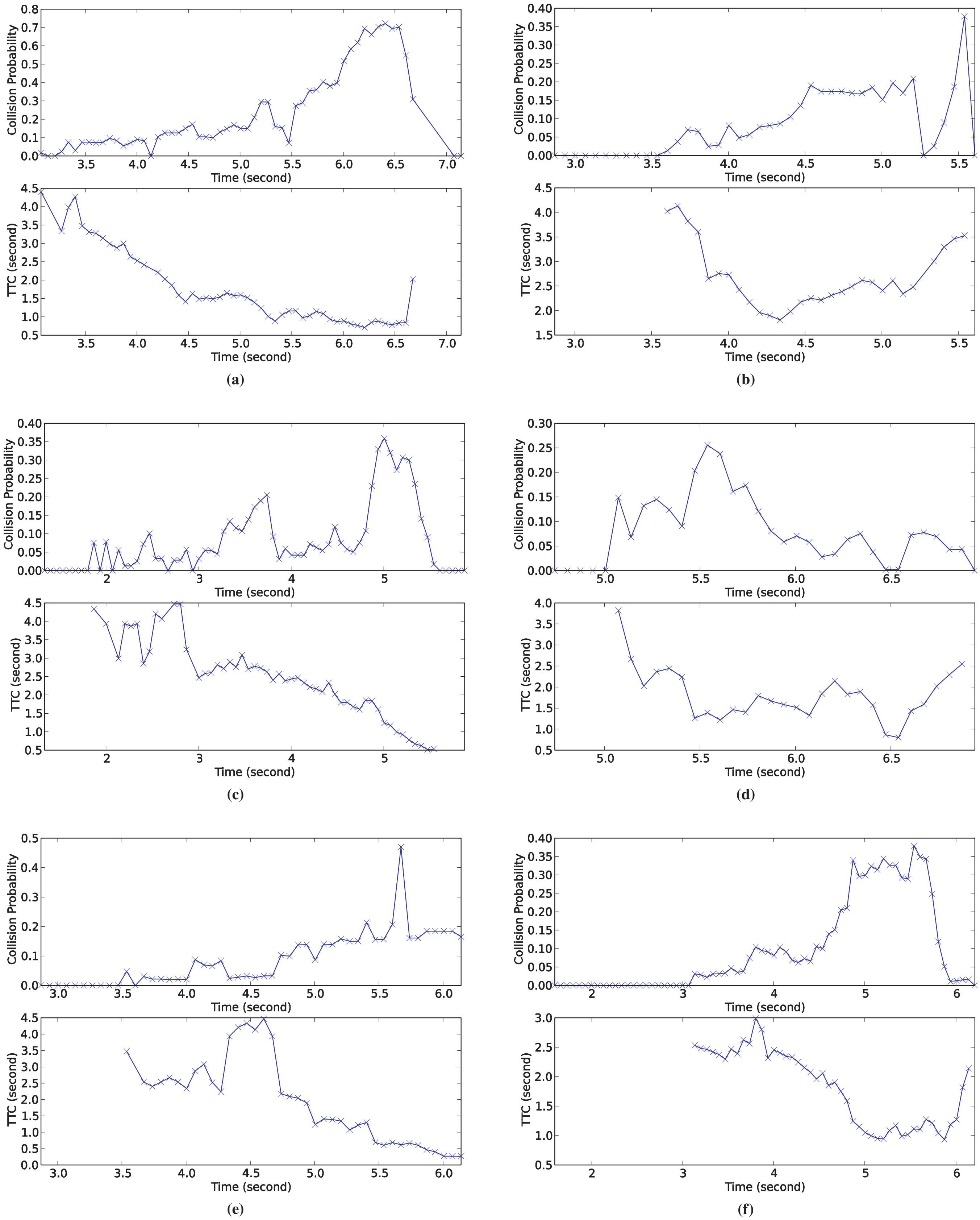

FIGURE 5 Plots of collision probability and TTC at each instant for small sample of interactions (without any postprocessing) from two subsets of collisions (*a*, *c*, *e*) and traffic conflicts (*b*, *d*, *f*). Interactions plotted in (*a*)–(*d*) are categorized as side interactions; those on last row (*e* and *f*) are categorized as parallel interactions.

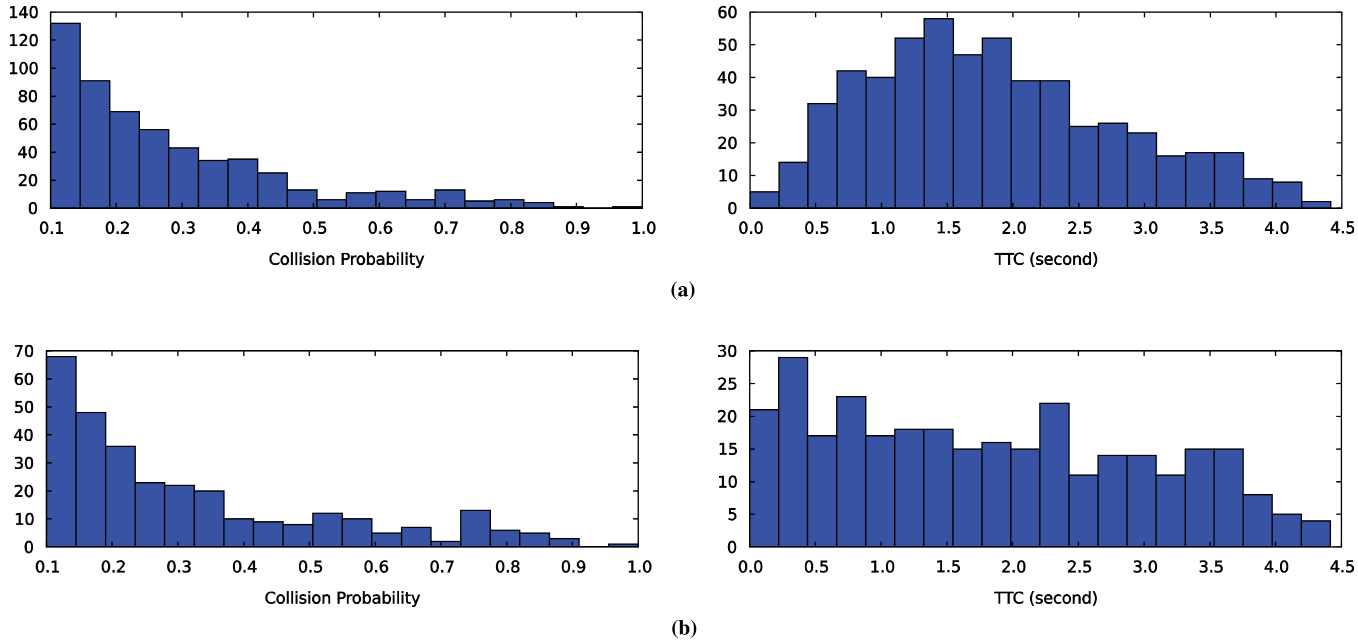

FIGURE 6 Distribution of maximum collision probability and TTC for interactions in subsets of (*a*) traffic conflicts and (*b*) collisions.

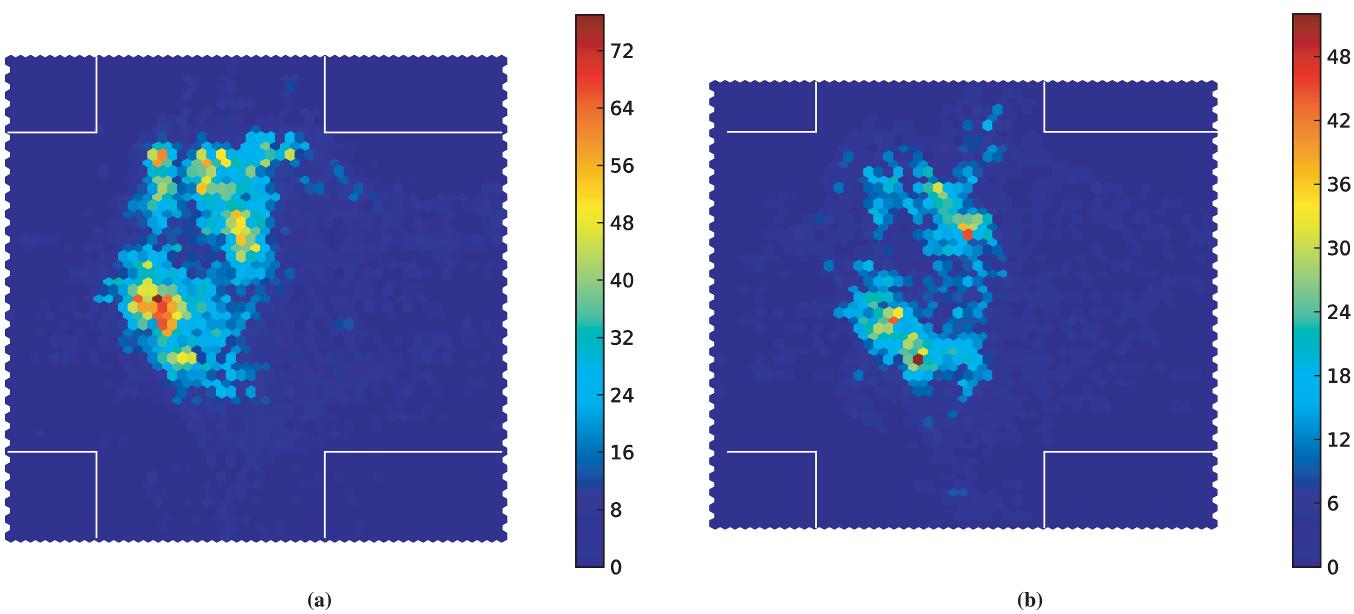

FIGURE 7 Maps of distributions of collision points (two-dimensional hexagonal binning plot) for two subsets of sequences: (*a*) traffic conflicts and (*b*) collisions.

REFERENCES

1. Gettman, D., and Head, L. *Surrogate Safety Measures from Traffic Simulation Models,* FHWA, U.S. Department of Transportation, 2003.
2. Archer, J. *Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and Their Application in Micro-simulation Modelling.* Thesis. Royal Institute of Technology, Stockholm, Sweden, 2004.
3. Cunto, F. *Assessing Safety Performance of Transportation Systems Using Microscopic Simulation.* University of Waterloo, Ontario, Canada, 2008.
4. Saunier, N., and T. A. Sayed. A Feature-Based Tracking Algorithm for Vehicles in Intersections. Presented at 3rd Canadian Conference on Computer and Robot Vision, IEEE, Quebec City, Canada, 2006.
5. Saunier, N., and T. Sayed. Probabilistic Framework for Automated Analysis of Exposure to Road Collisions. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2083,* Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 96–104.
6. Svensson, Å., and C. Hydén. Estimating the Severity of Safety Related Behaviour. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 379–385.
7. Amundsen, F., and C. Hydén. *Proceedings of the First Workshop on Traffic Conflicts.* Lund Institute of Technology, Sweden, 1977.
8. Davis, G. A., and P. M. Morris. Statistical Versus Simulation Models in Safety: Steps Toward a Synthesis Using Median Crossing Crashes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2102,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 93–100.
9. Hakkert, A. S., and L. Braimaister. *The Uses of Exposure and Risk in Road Safety Studies.* SWOV Institute for Road Safety Research, Leidschendam, Netherlands, 2002.
10. Elvik, R., A. Erke, and P. Christensen. Elementary Units of Exposure. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2103,* Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 25–31.
11. Sayed, T., G. R. Brown, and F. Navin. Simulation of Traffic Conflicts at Unsignalised Intersections with TSC-Sim. *Accident Analysis and Prevention,* Vol. 26, 1994, pp. 593–607.
12. Tarko, A., G. Davis, N. Saunier, T. Sayed, and S. Washington. *Surrogate Measures of Safety.* Subcommittee on Surrogate Measures of Safety, Transportation Research Board of the National Academies, Washington, D.C., 2009.
13. van der Horst, R. *A Time-Based Analysis of Road User Behavior in Normal and Critical Encounter.* Ph.D. thesis. Delft University of Technology, Netherlands, 1990.
14. Sayed, T., and S. Zein. Traffic Conflict Standards for Intersections. *Transportation Planning and Technology,* Vol. 22, 1999, pp. 309–323.
15. Songchitruksa, P., and A. Tarko. The Extreme Value Theory Approach to Safety Estimation. *Accident Analysis and Prevention,* Vol. 38, 2006, pp. 811–822.
16. Davis, G. A., J. Hourdos, and H. Xiong. Outline of Causal Theory of Traffic Conflicts and Collisions. Presented at 87th Annual Meeting of the Transportation Research Board, Washington, D.C., 2008.
17. Kamijo, S., Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic Monitoring and Accident Detection at Intersections. *Proc., IEEE Conference on Intelligence Transportation Systems,* Vol. 1, 2000, pp. 108–118.
18. Hu, W., X. Xiao, D. Xie, and T. Tan. Traffic Accident Prediction Using 3D Model Based Vehicle Tracking. *IEEE Transactions on Vehicular Technology,* Vol. 53, 2004, pp. 677–694.
19. Messelodi, S., and C. M. Modena. A Computer Vision System for Traffic Accident Risk Measurement: A Case Study. *Advances in Transportation Studies,* Vol. 7, 2005, pp. 51–66.
20. Atev, S., H. Arumugam, O. Massoud, R. Janardan, and N. Papanikolopolous. A Vision-Based Approach to Collision Prediction at Traffic Intersections. *IEEE Transactions on Intelligence Transportation Systems,* Vol. 6, 2005, pp. 416–423.
21. Laureshyn, A., and H. Ardo. Automated Video Analysis as a Tool for Analysing Road Safety Behaviour. Presented at ITS World Congress, London, 2006.
22. Kim, Z. Real Time Object Tracking Based on Dynamic Feature Grouping with Background Subtraction. *Proc., IEEE Computer Vision and Pattern Recognition,* Anchorage, Alaska, 2008, pp. 1–8.
23. Macedo, J., C. Vangenot, W. Othman, N. Pelekis, E. Frentzos, B. Kuijpers, I. Ntoutsi, S. Spaccapietra, and Y. Theodoridis. Trajectory Data Models. In *Mobility, Data Mining, and Privacy,* Springer, Berlin, 2008, pp. 123–150.
24. Saunier, N., T. Sayed, and C. Lim. Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis. Presented at 10th International IEEE Conference on Intelligent Transportation Systems, Seattle, Wash., 2007.
25. Green, E. R., K. R. Agent, and J. G. Pigman. *Evaluation of Auto Incident Recording System.* Kentucky Transportation Cabinet, Frankfort, 2005.
26. Ismail, K., T. Sayed, and N. Saunier. Camera Calibration for Urban Traffic Scenes: Practical Issues and a Robust Approach. Presented at 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.
27. Saunier, N., T. Sayed, and K. Ismail. Object Assignment Algorithm for Tracking Performance Evaluation. *Proc., 11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance,* Miami, Fla., 2009, pp. 9–16.

The Safety Data, Analysis, and Evaluation Committee peer-reviewed this paper.