Mining Microscopic Data of Vehicle Conflicts and Collisions to Investigate Collision Factors

Nicolas Saunier, Nadia Mourji, and Bruno Agard

Road collisions lead to great human and financial costs for society. Although some progress has been made, this worldwide issue needs more attention, as the costs increase. Proactive methods for road safety analysis that do not depend on collision occurrences are needed. Collection and analysis of microscopic data (road user trajectories) about all traffic events with and without a collision are the only ways to gain insight into collision factors and processes; that is, the chains of events that lead to collisions. The first phase of the project reported in this paper used microscopic data extracted from video sensors and data mining techniques to identify patterns in the traffic event database. Decision trees, the *k*-means algorithm, and the hierarchical agglomerative clustering method were used to analyze the relationship between interaction attributes and outcome (collision or not) and identify groups of interactions with similar attributes. This approach was demonstrated on a data set collected in Kentucky of 295 traffic events and contained 213 conflicts and 82 collisions. The decision tree confirmed the importance of evasive action in the interaction outcome. Three clusters were found from speed indicators extracted from road users' trajectories: the cluster containing the fewest collisions had the lowest speeds of the three. This result hints at the existence of conflicts that are dissimilar from most collisions and may therefore not be suitable for surrogate safety analysis.

The social cost of road collisions is among the largest negative side effects of road transportation. Including the costs of fatalities, disabilities, injuries and property damage, as well as medical care, lost productivity and traffic delays, the social cost of road collisions is estimated at $63 billion for Canada in 2004 (2004 U.S.$ = $45 billion) (*1*). In many age groups, road collisions are among the leading causes of death: in particular, it is the first leading cause of death for people age 15 to 29 years (*2*). The World Health Organization predicts that road collisions will jump from the ninth leading cause of death in 2004 to the fifth in 2030 (*2*). Compared with other health problems, road collisions are more burdensome because the victims are overwhelmingly young and healthy before the collision.

Road safety improvements may be achieved within the three components of the road system through changes in

N. Saunier and N. Mourji, Department of Civil, Geological, and Mining Engineering, and B. Agard, Department of Mathematics and Industrial Engineering, École Polytechnique de Montréal, C.P. 6079, Succursale, Centre-Ville, Montreal, Quebec H3C 3A7, Canada. Corresponding author: N. Saunier, nicolas.saunier@polymtl.ca.

Transportation Research Record: Journal of the Transportation Research Board, No. 2237, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 41–50.
DOI: 10.3141/2237-05

1. Infrastructure design,
2. Vehicle safety, and
3. Road user behavior.

This work deals with methods for road safety analysis at a given location, which allows identification of contributing factors related to the three components and particularly to the road. Traditional road safety analysis methods rely on collision databases that are filled with collision data manually collected after the occurrence of the collision, typically in the form of insurance and police reports. These data suffer from the following issues (*3*):

1. Difficult attribution of collisions to a cause (reports are skewed toward the attribution of responsibility, not to the search for the causes that led to a collision);
2. Small quantity of data; and
3. Too few data reconstituted after the event (with a bias toward more damaging collisions).

Safety analysts need to wait for accidents to happen to prevent them. There is a need for new proactive methods for road safety analysis that rely on more frequent traffic events without a collision.

Few data are available on the context of collisions and the collision process, that is, the chain of events that lead to a collision. The solution is to record information continuously about all traffic events: this can be achieved by using video sensors and computer vision techniques to extract all road user trajectories. Such data aid the investigation of the safety hierarchy (*4*), that is, the framework that places all traffic events on a continuum with collisions at the top, undisturbed passages or safe traffic events at the bottom, and traffic conflicts between. The position of a traffic event in the safety hierarchy measures its proximity to a potential collision or severity. Significant effort has been invested to develop techniques to collect and link to collisions the specific class of the most severe traffic conflicts. It is believed that the observation of all traffic events can provide a complementary safety diagnosis, more complete than can be done with collision data alone. It is a way to gain more knowledge about the factors and processes that lead to collisions.

This work has implications for the development of proactive methods for road safety analysis. A better understanding of the characteristics of traffic events with and without a collision should help derive better relationships between them, to identify types of traffic events without a collision that can be used as surrogates for road safety analysis. This is critical, because the work of Davis et al. on a small set of traffic events suggests that evasive actions undertaken by road users involved in conflicts may be of a different nature than those attempted in collisions (*5*).

This work pursues a line of research started at the University of British Columbia into automated road safety analysis that uses video sensors. A probabilistic framework for the computation of the probability of collision for all road users in interaction was proposed by Saunier and Sayed (*6*) and refined by Saunier et al. (*7*). This paper reports on the first phase of a research project that seeks to better understand collision factors and processes by using a large set of traffic events composed of conflicts and collisions. Contextual information and microscopic data, that is, the trajectories of road users involved in the traffic events, are extracted automatically and mined for patterns with artificial intelligence techniques (data mining). By comparing traffic events with and without a collision, this work will help to identify surrogate measures of safety. To the authors' knowledge, this work is unique in the size of the analyzed data set and the observation of safety-related events.

RELATED WORK

Road Safety Analysis

Safety is defined as the number of collisions expected to occur at a given location per unit of time, where expected means "the average in the long run if it were possible to freeze all prevailing conditions that affect safety" (*8*). Considerable research has been done to estimate safety models as a function of explanatory variables describing the transportation system: the road, the vehicle, and the driver. These safety models, also called crash prediction models (CPMs) or safety performance functions, typically take the form of an equation linking safety to a set of variables and rely on historical collision data. These models are at the core of the *Highway Safety Manual* (HSM). Historical collision data obtained from insurance and police reports are ill-suited for the analysis of collision processes.

Two methods based solely on collisions stand out to help shed more light on collision factors and processes: in-depth accident analysis and naturalistic driving studies. In-depth accident analysis relies on detailed reconstitutions to investigate collision factors (*9*) and as such may provide some information on the chain of events that led to the collision. However, they share many shortcomings with methods based on historical collision data: they provide limited amounts of data at a higher cost, they rely on reconstitutions in which the collision processes may be only guessed at, and they still require a wait for collisions to occur. Naturalistic driving studies rely on the continuous collection of data on road users, driving behavior, the vehicle, and the environment, over extended periods (*10*). Very large projects, for example, those within the Strategic Highway Research Program (*11*), are in process and should provide unprecedented information. An advantage will be the observation of all traffic events, not only collisions. Nevertheless, naturalistic driving studies also have limitations: they typically provide detailed information only on one of the road users involved in a safety-related event; instrumenting vehicles is costly and requires access to the vehicle, whereas fixed video cameras provide external nonintrusive monitoring of all traffic events and their context at a lower cost.

Traffic conflict studies are the most common proactive methods for road safety analysis (*4*, *12*). Although mixed validation results, cost, and reliability have hindered their development, they have been integrated into traditional approaches, including the HSM, providing complementary information and alternative methods. The framework of the safety hierarchy was developed in the context of traffic conflict studies (*13*) and is the basis for more recent approaches that take into account all road user interactions, not only the most severe traffic conflicts, for more complete and robust diagnoses (*4*–*6*). However, traffic conflict data traditionally collected by observers in the field are limited and are subject to reliability issues that make it unsuitable to understand collision processes: objective microscopic data are required for this purpose.

Little road safety analysis has been based on microscopic road user data, which was not easily available until recently: computer vision techniques now allow microscopic data to be extracted automatically from video data. Road user trajectories are rarely collected with the primary goal of safety analysis (*14*–*20*). To the authors' knowledge, no work aiming to understand collision processes relies on automatically collected microscopic data. The present work is also unique in the size of the data set and that it contains traffic events with and without a collision.

Data Mining in Road Safety Analysis

Machine learning models, like artificial neural networks (ANN) and support vector machines (SVM) (*21*), have been widely applied to estimate CPMs. However, the goal of this project is an understanding of collision factors, which requires extracting patterns from data and can be achieved through data mining techniques (*22*). These include classification through use of, for example, decision trees that can be interpreted, as opposed to the "black box" nature of ANNs and SVMs, and clustering, that is, finding groups through some similarity measure by using, for example, the *k*-means algorithm. Data mining has been used for the analysis of databases made up only of collisions without any microscopic data.

Safety models for some collision attributes have been built to classify collisions and identify collision factors. Collision-prone locations and the relationship of collision factors to road and driver of vehicle were investigated by Sayed et al. (*23*) and Sayed and Abdelwahab (*24*), who used a fuzzy *k*-means algorithm, ANN, and fuzzy *k* nearest neighbors. Several attempts to model collision outcomes have been made. Sohn and Shin used ANNs, decision trees, and logistic regression to identify collision severity-related factors to predict one of three possible outcomes: bodily injury, death, and property damage (*25*). They found that the presence of a protective device (i.e., seat belt or helmet) is the most important factor in differences in crash severity. Further work by Sohn and Lee indicated that a clustering-based classification algorithm worked best for their data (*26*). ANNs and the fuzzy adaptive resonance theory ANN were used to show that gender, vehicle speed, use of seat belt, type of vehicle, point of impact, and area type (rural versus urban) affect the likelihood of injury severity levels (*27*). Classification and regression trees were applied to analysis of the risk factors that can influence the injury severity in traffic collisions (*28*); it was concluded that the most important variable associated with collision severity is vehicle type, whereas pedestrians, motorcycle riders, and bicycle riders have higher risks for being injured than other types of vehicle drivers in traffic collisions.

A latent class clustering technique was used by Depaire et al. to segment traffic collision data and to identify homogeneous collision types (*29*). Injury models were then developed with the multinomial logit model for each resulting cluster, based on features such as collision type, crossroad type, built-up area, road type, road user age, dynamics of road user (moving or stationary), and vehicle type. It was concluded that cluster models reveal new variables affecting injury outcome and provide a more complete interpretation of the relationship causal variables and injury outcome.

Despite significant use of data mining techniques to analyze collision data, the lack of microscopic data describing traffic events with and without a collision limits the scope of the collision factors that can be identified and the analysis of similarities between traffic events of different severities.

DESCRIPTIVE ANALYSIS OF DATA SET

Context

This work relies on a unique data set of video recordings of traffic conflicts and collisions collected at one signalized intersection in Kentucky between August 16, 2001, and May 31, 2006 (*30*). All analyses reported in this paper used only the video recordings as the source of information, except for date and time of recording that can be derived from file names. Two subsets of video recordings, "miss" and "incident," correspond, respectively, to traffic conflicts of mild to high severity and collisions. It is not clear from Green et al. how severity was estimated to identify the subset of traffic conflicts (*30*). Each recording contains, or should contain, one clear safety-related traffic event, that is, a traffic conflict or a collision. From the original set of 238 traffic conflicts and 116 collisions, respectively, 213 and 82 were used in the analysis. The remainder of the recordings was not analyzed because of video quality, tracking issues for the video analysis tool, or absence of a relevant traffic event. In the remaining video recordings, all safety-related traffic events involve at least two road users: all traffic events are therefore referred to as interactions.

The quality of the video data makes road user detection and tracking challenging. The video recordings have a resolution of 352 pixels wide by 240 pixels high, varying levels of compression, color aberrations, and so forth, all affecting the image quality, and a frame rate of 15 frames per second. Many challenging conditions for automated video analysis are covered, with various times of recording (day and night) and weather conditions: sunny days cause strong shadows; there are many cases of snow, fog, and rain (sometimes at night, in which case the reflection of vehicle headlights causes particular glare). Although some recordings were impossible to analyze because of these issues, road user detection and tracking was possible in most recordings through use of a video-based system developed previously; some frames are shown in Figure 1 (*31*). The analysis relies on the trajectories used by Saunier and Sayed (*6*).

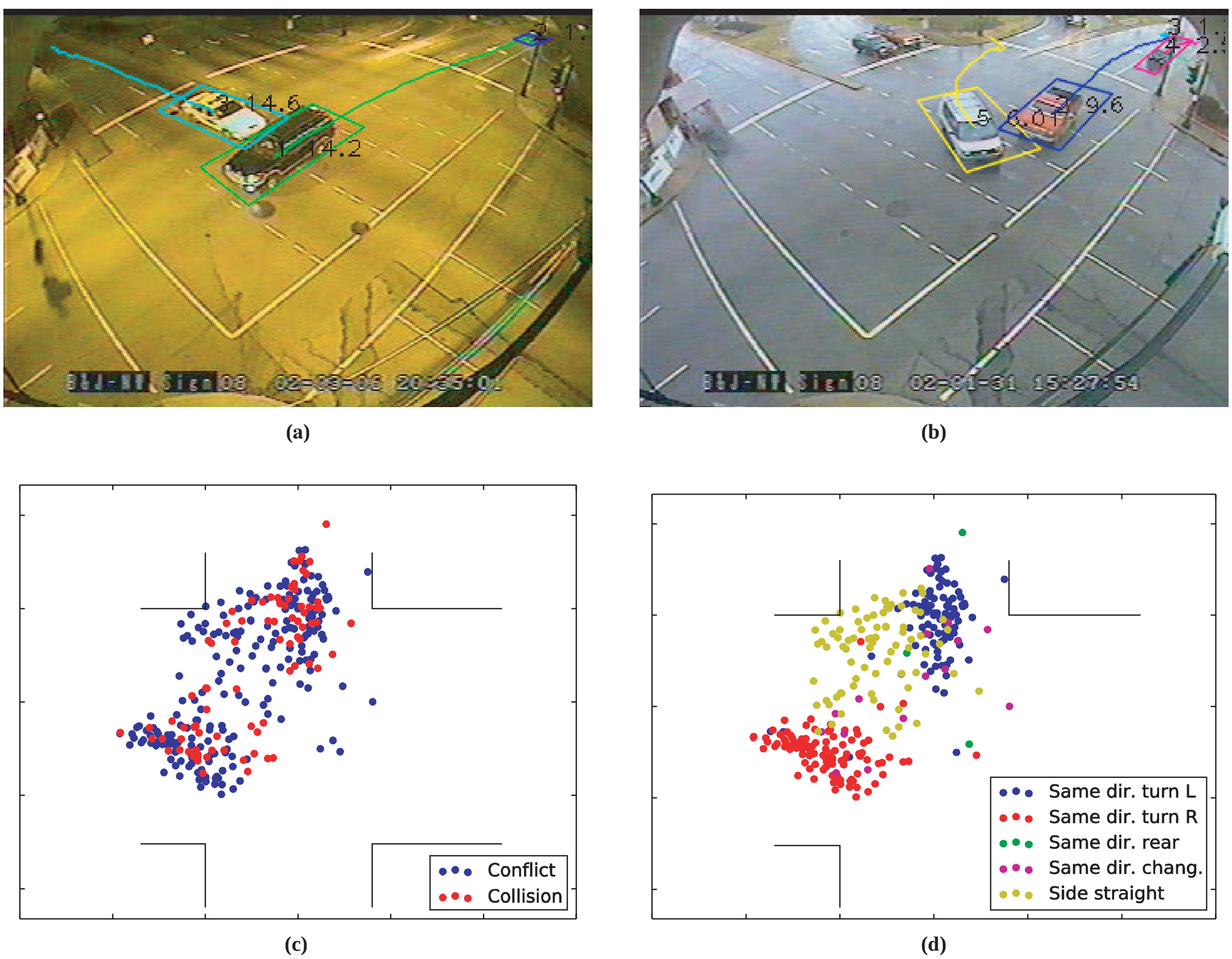


FIGURE 1 Sample interactions of categories and spatial plots of interaction outcomes and categories: (*a*) side straight collision, (*b*) parallel turning left collision, (*c*) interaction outcome, and (*d*) interaction category (upper left intersection corner in maps corresponds to farthest corner in video images).

TABLE 1 Attributes of Interactions

Attribute	Auto	Values
Categorical Attribute		
Type of day	×	Weekday, weekend
Lighting condition		Daytime, twilight, nighttime
Weather condition		Normal, rain, snow
Interaction category (see Figure 2)		Same direction (turning left and right, rear-end, lane change), opposite direction (turning left and right, head-on), side (turning left and right, straight)
Interaction outcome		Conflict, collision
Numerical Attribute	**Auto**	**Units**
Road user type		Number of road users per type
Passenger car		
Van, 4 × 4, SUV		
Bus		
Truck (all sizes)		
Motorcycle		
Bike		
Pedestrian		
Road user origin		Number of road users per origin
I-65 Jefferson Street exit		
I-65 Brook Street exit		
Brook Street		
Jefferson Street		
Type of evasive action		Number of evasive actions per evasive action
No evasive action		
Braking		
Swerving		
Acceleration		
3 attributes from Δv	×	km/h
6 values from s	×	km/h

NOTE: × in auto column indicates attributes were automatically extracted from data; Δv is time series of norm of difference of velocities; and s designates road users' speed time series.

Database

All 295 interactions are described by attributes listed in Table 1. Most categorical attributes were extracted manually though watching of the video recordings. The simple interaction categories used by Saunier and Sayed (*6*) were expanded in this work (Figure 2), covering most of the categories proposed by Hauer et al. (*8*). Road users' origin and type were also stored in the database; the number of road users involved in each interaction was counted by type. In addition to bikes and pedestrians, the types of motorized vehicles were taken from FHWA vehicle classes. There were no pedestrians, bikes, or buses in the database. In some cases, more than two road users could be considered to be involved in an interaction, but for simplicity in this first phase, only information on the two closest was included in the database. The numbers of evasive actions attempted by the road users were also counted by type.

So information from the trajectories of the road users involved in each interaction could be automatically extracted, they were manually identified among all road user trajectories in each video sequence. For this phase of the project, only speed data were extracted. From the time series of road user velocities (speed vectors), the following attributes were computed:

- Minimum, maximum, and mean values of the speeds of each road user, denoted, respectively, s_{min}, s_{max}, and $\overline{s}$. For a unique

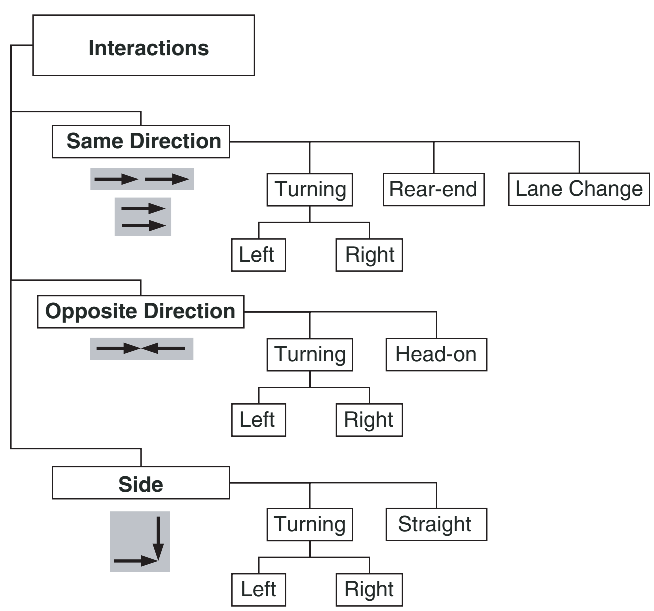


FIGURE 2 Hierarchy of interaction categories.

description of the interaction, that is, symmetric with respect to the involved road users, the attributes are ordered by increasing value (i.e., $s_{\mathrm{min}1} < s_{\mathrm{min}2}$, $s_{\mathrm{max}1} < s_{\mathrm{max}2}$, and $\bar{s}_1 < \bar{s}_2$). Six such attributes are therefore used to describe the road users' speeds during their interaction.

- Minimum, maximum, and mean values of the norm of the difference of the road users' velocities for the entire time interval during which they were tracked simultaneously, denoted, respectively, Δv_{min}, Δv_{max}, and $\overline{\Delta v}$.

Descriptive Analysis

As shown in Figure 3, there appears to be no relationship between the interaction outcome and type of day, lighting conditions, and weather conditions. Overall, weather conditions were mostly normal. There were slightly fewer collisions during twilight and at night than during the day, and there is no particular pattern in the interaction category. As can be expected, there is a strong relationship between the type of evasive action and the interaction outcome: in most collisions (62 of 82), no evasive action was attempted. There is also a sizable amount of conflicts in which at least one of the involved road users did not attempt an evasive action.

Speed attributes are displayed in Figure 4. The speed differences are quite similar for conflicts and collisions. However, speeds are systematically higher for collisions than for conflicts, although the difference is within the standard deviation of each category.

A rough localization of each interaction was obtained by averaging the positions of the involved road users during the period they were simultaneously tracked. No pattern was visible when the

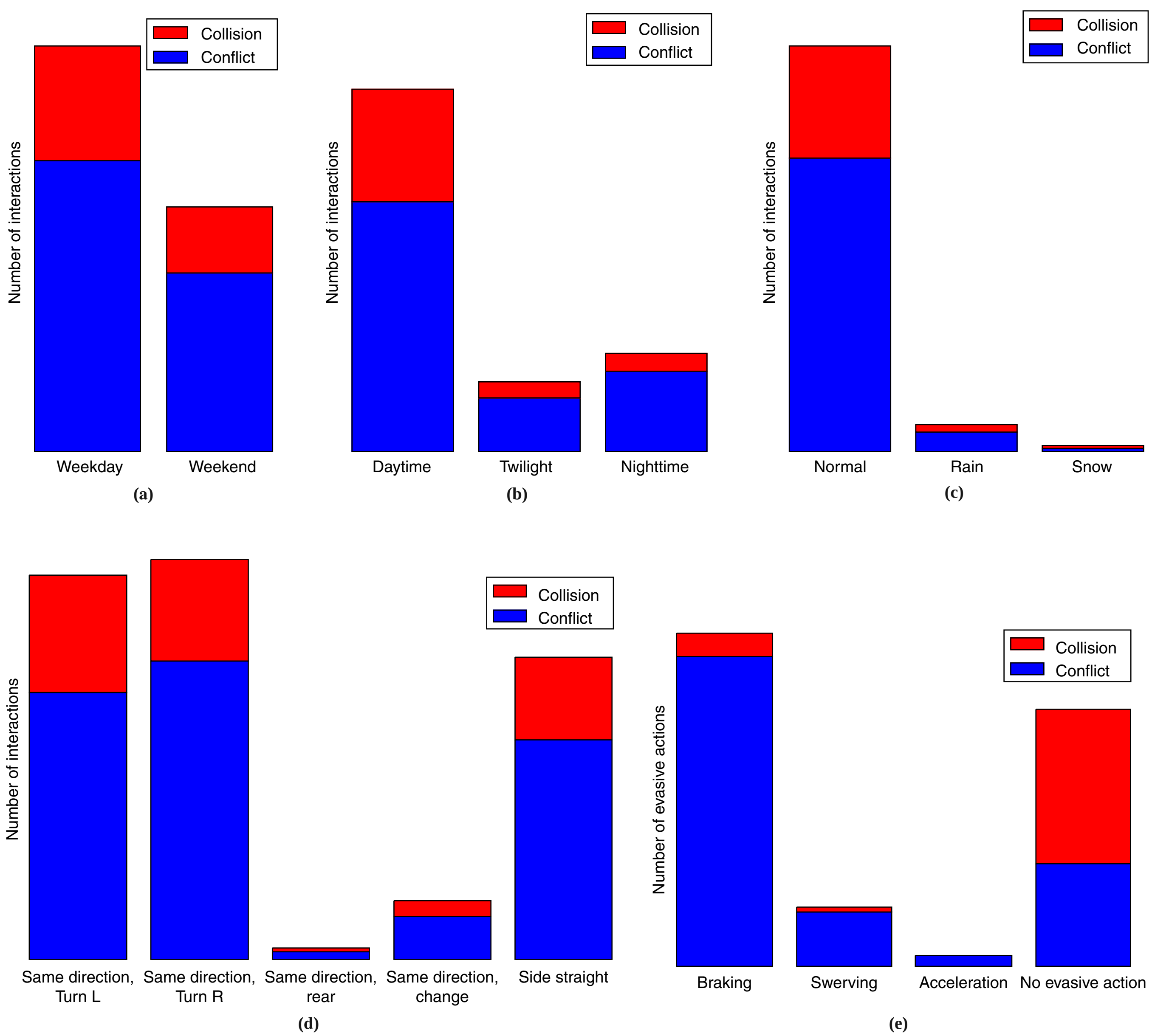
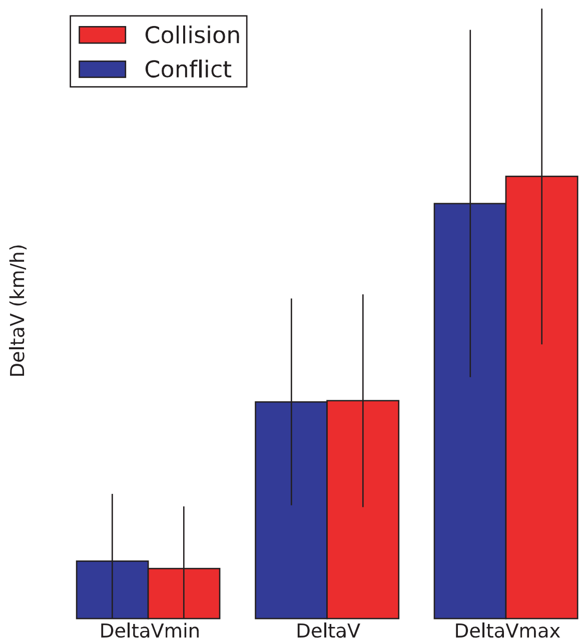
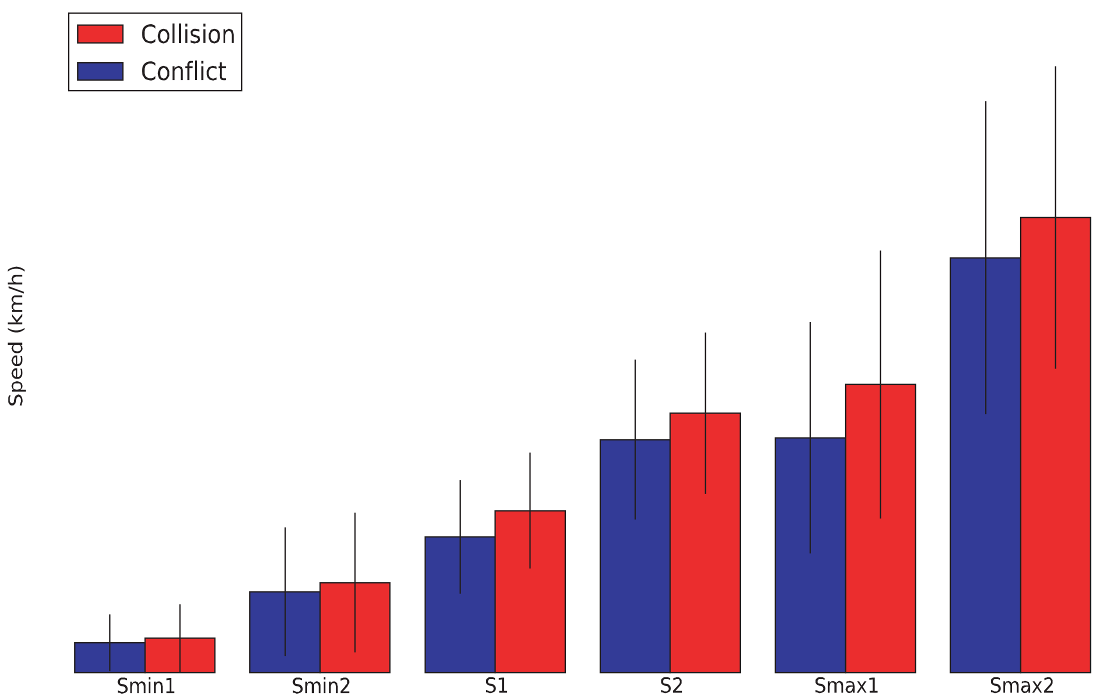


FIGURE 3 Distributions of database attributes per interaction outcome: (*a*) type of day, (*b*) lighting condition, (*c*) weather condition, (*d*) interaction category, and (*e*) type of evasive action.



(a)



(b)

FIGURE 4 Distributions of speed attributes of database per interaction outcome: (*a*) norm of velocity difference and (*b*) speed (error bars represent standard deviation).

various interaction attributes were plotted, in particular the interaction outcome, except for the interaction category (Figure 1). This makes sense because the side–straight interactions occur at the upper left corner of the intersection, whereas same-direction interaction happens between vehicles coming from similar origins on either street.

USE OF DATA MINING TECHNIQUES FOR EXPLORATORY ANALYSIS

Two well-known data mining techniques, decision trees and the *k*-means algorithm, were used to mine the interaction database for patterns (*22*). Association rules were tried but did not yield a strong result. The free and open source software TANAGRA was used for this analysis (*32*).

Classification

The C4.5 decision tree algorithm was used to predict the interaction outcome from the attributes. The goal was not to classify interactions but to identify rules in the conditions generated by the algorithm at each node to predict the interaction outcome. The following numerical attributes are transformed and grouped as categorical attributes so that they are treated as one attribute by the decision tree: road user types, road user origins, and types of evasive actions. The new attribute can be any pair of two of the original attributes, for example, passenger car or truck for road user type or braking or swerving for types of evasive actions. There is no need to normalize numerical attributes for decision trees.

A decision tree was thus learned by use of all the interaction attributes, with the parameters minimum size of leaves set to six and the confidence level to 0.25. The decision tree had 13 nodes and 10 leaves and an error rate of 0.0746 (which was computed on the training data set and therefore overestimates the classification performance). The resulting rules are presented in Figure 3, with the characterization of the corresponding leave. The first split done by the tree addresses the evasive actions, which is expected (Figure 5); in particular, the absence of an evasive action led in 91.18% of cases to a collision. The presence of at least an evasive action is associated with conflicts, except for swerving or no evasive action and braking or no evasive action. In this latter case, it is possible to refine the rules on the basis of the speed attributes $\overline{\Delta v}$ and $\overline{s}_1$. Conflicts were associated with larger mean velocity differences, or, if the average velocity difference is low, lower mean individual speeds; the first case is probably related to successful braking to avoid collision, whereas the second is more difficult to interpret. Collisions occur in the remainder of the cases, that is, for low mean velocity differences and higher individual mean speed, which is logical, although this pertained to only five of six interactions.

This analysis provides interpretable knowledge about the interactions and their attributes and confirms the obvious link between collision avoidance and the presence of at least an evasive action. In particular, mean velocity differences are higher when collision is avoided because of braking by one of the road users.

Clustering

Given the descriptive analysis of the database as presented in the previous section and the goal of this project to study collision processes, the choice was made to use only the nine speed attributes for the clustering. It is hoped that the resulting clusters can help identify relationships between conflicts and collisions that can be used for surrogate safety analysis, as well as the lack of such relationships. Three types of clusters can be produced by the method, depending on the proportion of conflicts and collisions in the cluster, with the following potential implications:

- A mixed cluster of similar conflicts and collisions, which could indicate that the conflicts in the cluster can be used as surrogates for the collisions in the same cluster;

- Evasive actions in [braking/braking] then Interaction outcome = **conflict** (94.62% of 93 examples)
- Evasive actions in [braking/no evasive action]
  - $\overline{\Delta v} < 12.6183$
    - $\overline{s_1} < 13.4022$ then Interaction outcome = **conflict** (83.33% of 12 examples)
    - $\overline{s_1} \geq 13.4022$ then Interaction outcome = **collision** (83.33% of 6 examples)
  - $\overline{\Delta v} \geq 12.6183$ then Interaction outcome = **conflict** (95.31% of 64 examples)
- Evasive actions in [no evasive action/no evasive action] then Interaction outcome = **collision** (91.18% of 68 examples)
- Evasive actions in [braking/swerving] then Interaction outcome = **conflict** (96.55% of 29 examples)
- Evasive actions in [no evasive action/swerving] then Interaction outcome = **conflict** (55.56% of 9 examples)
- Evasive actions in [swerving/swerving] then Interaction outcome = **conflict** (100.00% of 5 examples)
- Evasive actions in [braking/acceleration] then Interaction outcome = **conflict** (100.00% of 7 examples)
- Evasive actions in [no evasive action/acceleration] then Interaction outcome = **conflict** (100.00% of 2 examples)

FIGURE 5 Rule generated by decision tree.

- A pure cluster with no or few conflicts in the cluster, which could indicate that it is not possible to use conflicts as surrogates to the collisions in the cluster; and
- A pure cluster with no or very few collisions in the cluster, which could indicate that the conflicts in the cluster cannot be used as surrogates to any type of collision.

Before analysis can be done, a required preliminary step is to normalize all the attributes: the objective is to standardize the scale of effects of each variable on the results. The Euclidean distance is chosen as the distance used by the *k*-means algorithm to compare the interaction instances.

The partitioning algorithm *k*-means is a very effective method used to identify homogeneous groups assuming a number of classes known at the beginning. The number of clusters depends on the depth of analysis desired, and its determination can be a challenge. Following the approach used by Le et al. (*33*), the data were first clustered in a large number of groups, larger than what can be reasonably expected for this analysis (25 in this case). A hierarchical agglomerative clustering method was then used to merge the groups. At each iteration, the ratio of between-cluster sum-of-squares (BSS ratio) and the gap were calculated and used to determine the number of clusters by the hierarchical algorithm. The compactness of the data was measured by the gap value, and the dissimilarity between them was considered by the BSS ratio. A good clustering yields clusters where there are high BSS ratio and gap value. The dendogram shows that a division into three clusters is appropriate (Figure 6).

After an appropriate number of clusters were identified with the previous method, the *k*-means algorithm was applied again for three groups. Some distributions of the clusters are presented in Figure 7. Clusters 1 and 3 are characterized by higher proportions of collisions (respectively, 40.8% and 44.4%), whereas Cluster 2 contains few collisions (7.8%). The distribution of evasive actions per cluster is not shown but was consistent with the previous characterization and the proportions of conflict and collisions in the clusters. The distribution of interaction categories in Cluster 2 was homogeneous to the whole database, whereas it was skewed in Cluster 1 (overrepresentation of side–straight and same-direction turning right) and Cluster 3 (overrepresentation of same-direction turning left and right and same-direction changing lanes). No relationship to the other categorical attributes could be discovered.

Since clustering is based on speed attributes, the clusters should show differences for these attributes. As shown in Figure 7, the clusters comprehended different mean speed and velocity difference values: the first cluster had the highest speeds for all attributes, followed by Cluster 3 and Cluster 2 on all attributes but one. Similarly to Figure 1, the spatial distribution of the clusters was investigated but showed no discernible pattern. The final characterization of the clusters, that is, of the overrepresentations of some attributes with respect to the whole database, can thus be made:

Cluster 1. Collisions, highest speeds, categories side–straight and same-direction turning right;
Cluster 2. Almost pure conflicts, lowest speeds; and
Cluster 3. Collisions, medium speeds, categories same-direction turning left and right and same-direction changing lanes.

Observation of a clustering of the interactions only on speed attributes indicates that the three clusters show some contrasted characteristics. Clusters 1 and 3 are mixed, with overrepresentation of collisions with respect to the whole database, and represent two classes of interactions differing in speeds and categories: the conflicts in these clusters are good candidates for surrogates for the collisions in their respective clusters. Cluster 2 is overwhelmingly made up of conflicts and is characterized by lower speeds: it could characterize a group of conflicts of lower severity that cannot be used as surrogates or used only for some specific collisions.

CONCLUSION

This paper presented the first results of a larger project undertaken to gain better understanding of collision factors and processes through use of microscopic data. A large data set of 295 interactions, made up of conflicts and collisions, was characterized through mining of their attributes with decision trees and the *k*-means partitioning algorithm. Obvious relationships such as the link of evasive actions and their absence to the interaction outcome were confirmed. The clustering analysis yielded evidence that not all conflicts should be used as surrogates for all collisions and showed how groups of similar conflicts and collisions can be identified.

Further investigation is necessary to confirm these findings. The conditions of the data collection are not exactly known, in particular, the sampling conditions; is the set of interactions available in the data set representative of all interactions of similar severity that occurred during the period of the data collection? Exposure data are not available, and thus no conclusion should be drawn about the risk of collision in the conditions under study.

This work paves the way for larger efforts and increases knowledge of collision processes. The next phases of this work will make use of the whole road user trajectories, develop temporal indicators to characterize the interactions, and use better similarities, which

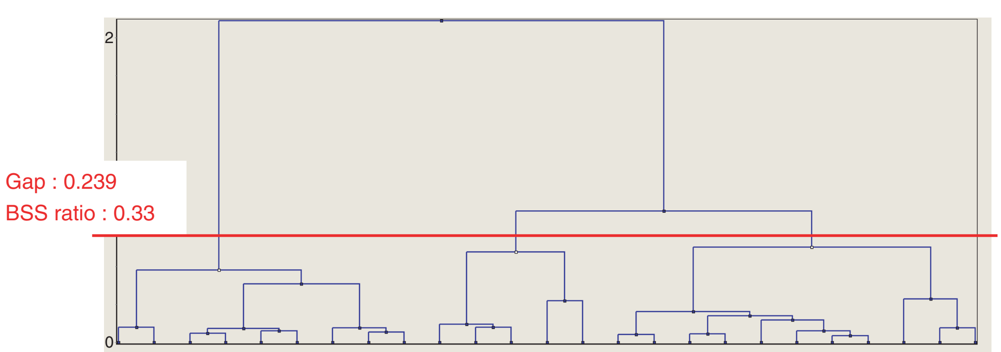


FIGURE 6 Dendogram of 25 clusters to obtain number of clusters in data.

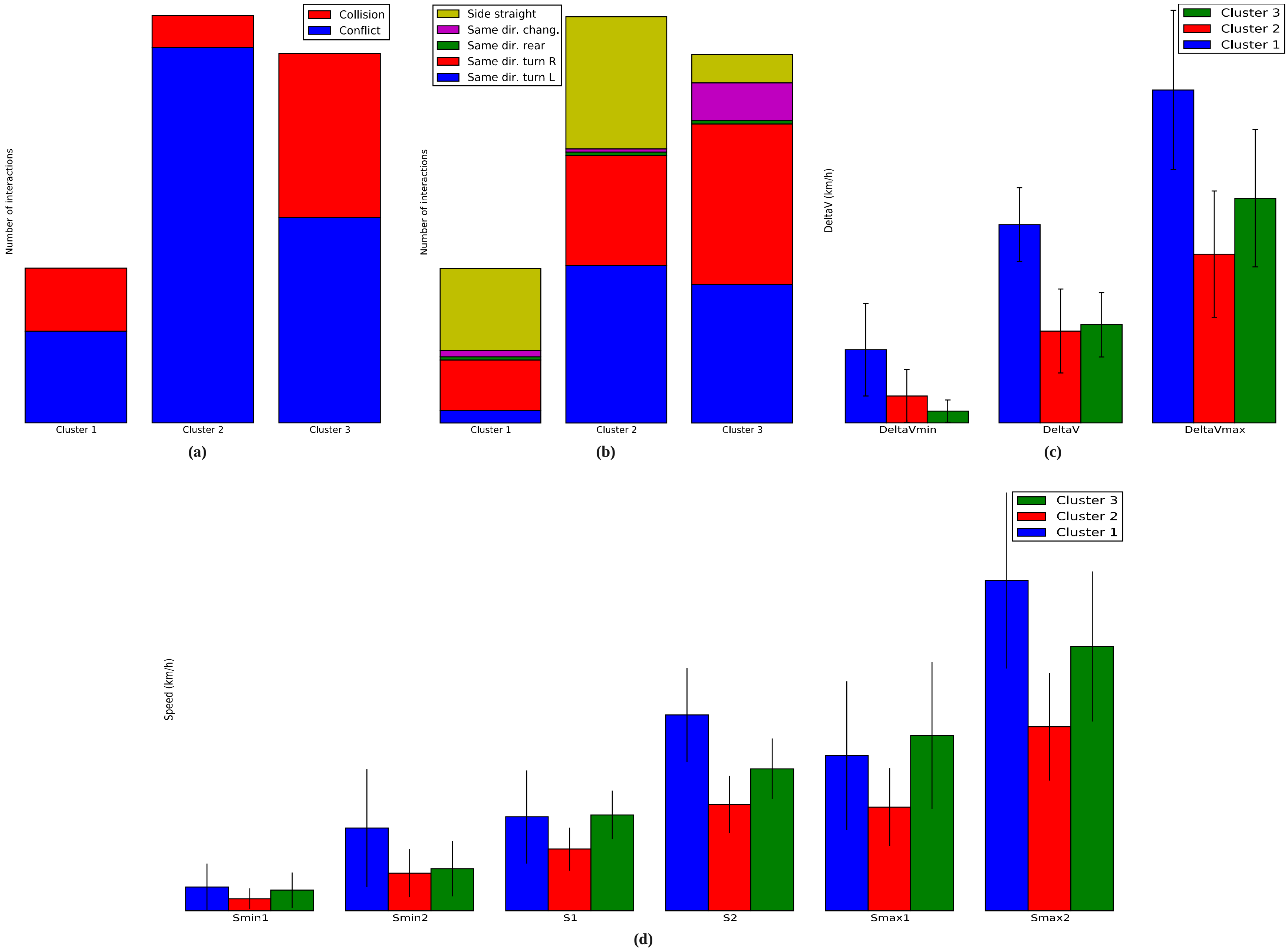


FIGURE 7 Distributions of (*a*) interaction outcome, (*b*) interaction category, (*c*) norm of velocity difference, and (*d*) speed attributes for the three clusters (error bars represent standard deviation).

can accommodate multidimensional vectors of varying length. Future projects will include the collection of large-scale data sets of all road user interactions in known conditions, which will allow stronger conclusions to be drawn.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the Foundation Polytechnique and the Research and Innovation Directorate of École Polytechnique of Montréal. The authors thank Zu Kim of California Partners for Advanced Transportation Technology and Ann Stansel of the Kentucky Transportation Cabinet for providing the video data set.

REFERENCES

1. Vodden, K., D. Smith, F. Eaton, and D. Mayhew. *Analysis and Estimation of the Social Cost of Motor Vehicle Collisions in Ontario.* Transport Canada, Ottawa, Ontario, Canada, 2007.
2. *Global Status Report on Road Safety: Time for Action.* World Health Organization, Geneva, 2009.
3. Ismail, K., T. Sayed, and N. Saunier. Automated Analysis of Pedestrian-Vehicle Conflicts: Context for Before-and-After Studies. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2198,* Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 52–64.
4. Svensson, A., and C. Hydén. Estimating the Severity of Safety Related Behaviour. *Accident Analysis and Prevention,* Vol. 38, No. 2, 2006, pp. 379–385.
5. 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.
6. Saunier, N., and T. A. 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.
7. Saunier, N., T. Sayed, and K. Ismail. Large-Scale Automated Analysis of Vehicle Interactions and Collisions. In *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.
8. Hauer, E., J. C. N. Ng, C. N. Morris, and J. Lovell. *Estimation of Safety at Signalized Intersections.* In *Transportation Research Record 1185,* TRB, National Research Council, Washington, D.C., 1988, pp. 48–61.
9. Fleury, D., and T. Brenac. Accident Prototypical Scenarios, A Tool for Road Safety Research and Diagnostic Studies. *Accident Analysis and Prevention,* Vol. 2, No. 33, 2001, pp. 267–276.
10. Neale, V. L., T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman. *An Overview of the 100 Car Naturalistic Study and Findings. Proc., 19th ESV Conference,* Washington, D.C., 2005.
11. *Special Report 296: Implementing the Results of the Second Strategic Highway Research Program: Saving Lives, Reducing Congestion, Improving Quality of Life.* Transportation Research Board of the National Academies, Washington, D.C., 2009.
12. Sayed, T., and S. Zein. *Traffic Conflict Standards for Intersections.* Transportation Planning and Technology, Vol. 22, 1999, pp. 309–323.
13. Hydén, C. *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique.* PhD dissertation. Lund University of Technology, Lund, Sweden, 1987.
14. Kamijo, S., Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic Monitoring and Accident Detection at Intersections. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 2, No. 2, 2000, pp. 108–118.
15. Hu, W., X. Xiao, D. Xie, T. Tan, and S. Maybank. Traffic Accident Prediction Using 3-D Model-Based Vehicle Tracking. *IEEE Transactions on Vehicular Technology,* Vol. 53, No. 3, 2004, pp. 677–694.
16. Messelodi, S., and C. M. Modena. *A Computer Vision System for Traffic Accident Risk Measurement: A Case Study.* Conferenza Italiana sui Sistemi Intelligenti, Perugia, Italy, 2005.
17. Atev, S., H. Arumugam, O. Masoud, R. Janardan, and N. P. Papanikolopoulos. A Vision-Based Approach to Collision Prediction at Traffic Intersections. *IEEE Transactions on Intelligent Transportation Systems,* Vol. 6, No. 4, 2005, pp. 416–423.
18. Laureshyn, A., and H. Ardö. *Automated Video Analysis as a Tool for Analysing Road Safety Behaviour.* ITS World Congress, London, 2006.
19. Laureshyn, A. *Application of Automated Video Analysis to Road User Behaviour.* PhD dissertation. Lund University, Lund, Sweden, 2010.
20. Kim, Z. Real Time Object Tracking Based on Dynamic Feature Grouping with Background Subtraction. Presented at IEEE International Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, 2008.
21. Xie, Y., D. Lord, and Y. Zhang. Predicting Motor Vehicle Collisions Using Bayesian Neural Network Models: An Empirical Analysis. *Accident Analysis and Prevention,* Vol. 39, No. 5, 2007, pp. 922–933.
22. Duda, R. O., and P. E. Hart. *Pattern Classification.* Wiley Interscience, New York, 2000.
23. Sayed, T., W. Abdelwahab, and F. Navin. Identifying Accident-Prone Locations Using Fuzzy Pattern Recognition. *ASCE Journal of Transportation Engineering,* Vol. 121, No. 4, 1995, pp. 352–358.
24. Sayed, T., and W. Abdelwahab. Comparison of Fuzzy and Neural Classifiers for Road Accidents Analysis. *ASCE Journal of Computing in Civil Engineering*, Vol. 12, No. 1, 1998, pp. 42–47.
25. Sohn, S. Y., and H. Shin. Pattern Recognition for Road Traffic Accident Severity in Korea. *Ergonomics,* Vol. 44, No. 1, 2001, pp. 107–117.
26. Sohn, S. Y., and S. H. Lee. Data Fusion, Ensemble and Clustering to Improve the Classification Accuracy for the Severity of Road Traffic Accidents in Korea. *Safety Science,* Vol. 41, No. 1, 2003, pp. 1–14.
27. Abdel-Aty, M. A., and H. T. Abdelwahab. Predicting Injury Severity Levels in Traffic Crashes: A Modeling Comparison. *ASCE Journal of Transportation Engineering*, Vol. 130, No. 2, 2004, pp. 204–210.
28. Chang, L.-Y., and H.-W. Wang. Analysis of Traffic Injury Severity: An Application of Non-Parametric Classification Tree Techniques. *Accident Analysis and Prevention*, Vol. 38, No. 5, 2006, pp. 1019–1027.
29. Depaire, B., G. Wets, and K. Vanhoof. Traffic Accident Segmentation by Means of Latent Class Clustering. *Accident Analysis and Prevention,* Vol. 40, No. 4, 2008, pp. 1257–1266.
30. Green, E. R., K. R. Agent, and J. G. Pigman. *Evaluation of Auto Incident Recording System.* Kentucky Transportation Cabinet, Frankfort, 2005.
31. Saunier, N., and T. Sayed, T. *A Feature-Based Tracking Algorithm for Vehicles in Intersections.* IEEE, New York, 2006.
32. Rakotomalala, R. TANAGRA: Un Logiciel Gratuit pour l'Enseignement et la Recherche. *Actes de EGC '2005,* Vol. 2, 2005, pp. 697–702.
33. Le, T. T. H., B. Agard, and S. Deveault. Decision Support Based on Socio-Demographic Segmentation and Distribution Channel Analysis in the U.S. Furniture Market. Presented at International Conference on Industrial Engineering and Systems Management, Montreal, Canada, 2009.

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