



Large-scale automated proactive road safety analysis using video data



Paul St-Aubin^{a,*}, Nicolas Saunier^a, Luis Miranda-Moreno^b

^a Polytechnique Montreal, 2500, chemin de Polytechnique, Montreal, Quebec H3T 1J4, Canada

^b McGill University, 817 Sherbrooke Street West, Montreal, Quebec H3A 0C3, Canada

ARTICLE INFO

Article history:

Received 2 June 2014

Received in revised form 18 February 2015

Accepted 6 April 2015

Available online 22 April 2015

Keywords:

Big data

Transportation

Traffic

Road safety

Computer vision

Trajectories

Surrogate safety

Conflicts

Road users

Time-to-collision

Road user behaviour

Roundabouts

ABSTRACT

Due to the complexity and pervasiveness of transportation in daily life, the use and combination of larger data sets and data streams promises smarter roads and a better understanding of our transportation needs and environment. For this purpose, ITS systems are steadily being rolled out, providing a wealth of information, and transitionary technologies, such as computer vision applied to low-cost surveillance or consumer cameras, are already leading the way.

This paper presents, in detail, a practical framework for implementation of an automated, high-resolution, video-based traffic-analysis system, particularly geared towards researchers for behavioural studies and road safety analysis, or practitioners for traffic flow model validation. This system collects large amounts of microscopic traffic flow data from ordinary traffic using CCTV and consumer-grade video cameras and provides the tools for conducting basic traffic flow analyses as well as more advanced, pro-active safety and behaviour studies. This paper demonstrates the process step-by-step, illustrated with examples, and applies the methodology to a case study of a large and detailed study of roundabouts (nearly 80,000 motor vehicles tracked up to 30 times per second driving through a roundabout).

In addition to providing a rich set of behavioural data about Time-to-Collision and gap times at nearly 40 roundabout weaving zones, some data validation is performed using the standard Measure of Tracking Accuracy with results in the 85–95% range.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Affordable computing and flexible and inexpensive sensor technology are transforming current practice and methods for traffic data collection, monitoring, and analysis: big data is changing how we interact with our environment and how we approach problem solving tasks in the field of transportation. This should come as no surprise as the complexity and pervasiveness in daily life of urban mobility lends itself naturally to large amounts of data. In this context, the use of mobile and/or fixed video sensors for traffic monitoring and data collection is becoming a common practice not only for freeways but also for urban streets. Early and notable examples include the NGSIM project which included a dataset of extracted trajectories from video data of four corridors (freeways and urban arterials) (Kim et al., 2005) and the SAVEME project which fielded a small but early implementation of video tracking for surrogate safety analysis (Ervin et al., 2000; Gordon et al.,

* Corresponding author.

E-mail addresses: paul.st-aubin@polymtl.ca (P. St-Aubin), nicolas.saunier@polymtl.ca (N. Saunier), luis.miranda-moreno@mcgill.com (L. Miranda-Moreno).

2012). The availability of such large data sets opens up possibilities for more dynamic traffic load balancing and congestion easing of road networks and in return provides researchers with participatory network usage data collection. This new situation in which traffic data is being collected intensively demands more intelligent and automated methods for traffic data analysis; it is then not surprising that computer vision techniques have gained popularity given their potential for transforming the existing CCTV infrastructure (or inexpensive consumer-grade video sensors (Jackson et al., 2013)) into a highly detailed traffic data collection tool to identify and study traffic behaviours.

One of the most prominent behavioural study application is in proactive road safety diagnosis using surrogate safety methods. This has been a long-standing goal in the field of transportation safety, as traditional statistical methods using accident data require long observation periods (years of crash data): one must wait for (enough) accidents to occur in this time. Beginning in the 1960s, attempts were made to predict the number of collisions based on observations without a collision rather than historical accident records (Perkins and Harris, 1968). The Traffic Conflict Technique (TCT) (Hydén and Linderholm, 1984; Parker and Zegeer, 1989) was one of the earliest methods proposed which entailed the observation of qualitatively-defined quasi-collision events: situations in which road users were exposed to some recognizable risk (probability) of collision, e.g. a “near-miss”. However, several problems limited their adoption: manual data collection is costly and may not be reliable, and the definition and objective measurement of these events were lacking (Hauer, 1978; Williams, 1981; Kruysse, 1991; Chin and Quek, 1997).

Today, with technological improvements in computing power, data storage, ubiquitous sensor technologies, and advances in artificial intelligence, these issues are rapidly being addressed. This research presents the application of a video-based automated trajectory analysis solution which combines the latest advances in high-resolution traffic data acquisition (Saunier et al., 2010) and machine learning methods to model and predict collision potential (Mohamed and Saunier, 2013; St-Aubin et al., 2014) from relatively short, but extremely rich traffic data. This data is typically obtained from ordinary video data via computer vision from a camera situated at 10 m or more above the road surface (Jackson et al., 2013), although some of the early footage was taken at 5 m above the road surface. This trajectory data consists of position and velocity measurements of road users captured 15–30 times per second to a relatively high degree of accuracy. This amounts to several million individual instantaneous measurements over a period of one day at a typical site (for each camera).

One of the limitations of past studies involving surrogate analysis is the use of few sites or small datasets. This paper presents, step-by-step, a complete automated system for proactive road safety analysis using large amounts of video data. To the authors' knowledge, the presented system is the most comprehensive to be applied to such a large amount of data collected in the field for a real world traffic engineering study. A large video data set was collected at nearly 40 roundabout weaving zones in Québec across 20 different roundabouts for the specific purpose of studying road user behaviour and corresponding safety using surrogate safety analysis. Roundabout weaving zones are defined as the area within the roundabout delimited by an approach and the next following exit. Each camera recorded 12–16 h of video on a typical workday, constituting a dataset of over 470 h of video data. Applying the proposed method to this large dataset yielded a considerable number of indicators, from individual road user measurements, e.g. speed, to individual interaction measurements, e.g. time to collision (TTC), to aggregated indicators per road user or interaction, to aggregated indicators per site over time and space.

This paper is organized as follows: the next section outlines the methodology, briefly reviews surrogate safety theory, and is followed by a step-by-step review of the methodology with practical examples drawn from the roundabout dataset; the results section provides early video data calibration results, safety indicator aggregation and prediction comparisons, and initial results of the complete roundabout study.

2. Methodology

2.1. Overview

Fig. 1 outlines the general data collection and analysis methodology. Video data is collected for the study: for a cross-sectional study, at a number of sites with adequate representation of contributing factors and controlled external factors; for a before-after study, at one or more sites before and after a change in contributing factors. With scene data and camera calibration parameters, feature tracking is performed to extract road user trajectories. The trajectories are in the form of series of positions of moving objects over time within the scene. This positional data is then processed to obtain derived measures such as speed, heading, and acceleration. Finally, data is analysed and interpreted in a variety of ways: (i) simple summaries such as average speed and volume counts; (ii) generalized spatial relationship analysis, such as surrogate safety analysis; or (iii) high-level interpretation of behaviour relative to elements of the scene (typically specific to the study) such as gap times and motor vehicle infractions. With a large number of potential contributing factors, it may be beneficial to apply site clustering techniques before initiating behavioural measure correlation.

2.2. Video data processing

2.2.1. Trajectories: positions in space and time

Road user trajectories are extracted from video data using a feature-based tracking algorithm described in Saunier and Sayed (2006). Trajectories are a series of points in Cartesian space representing the centre position of a moving object (road

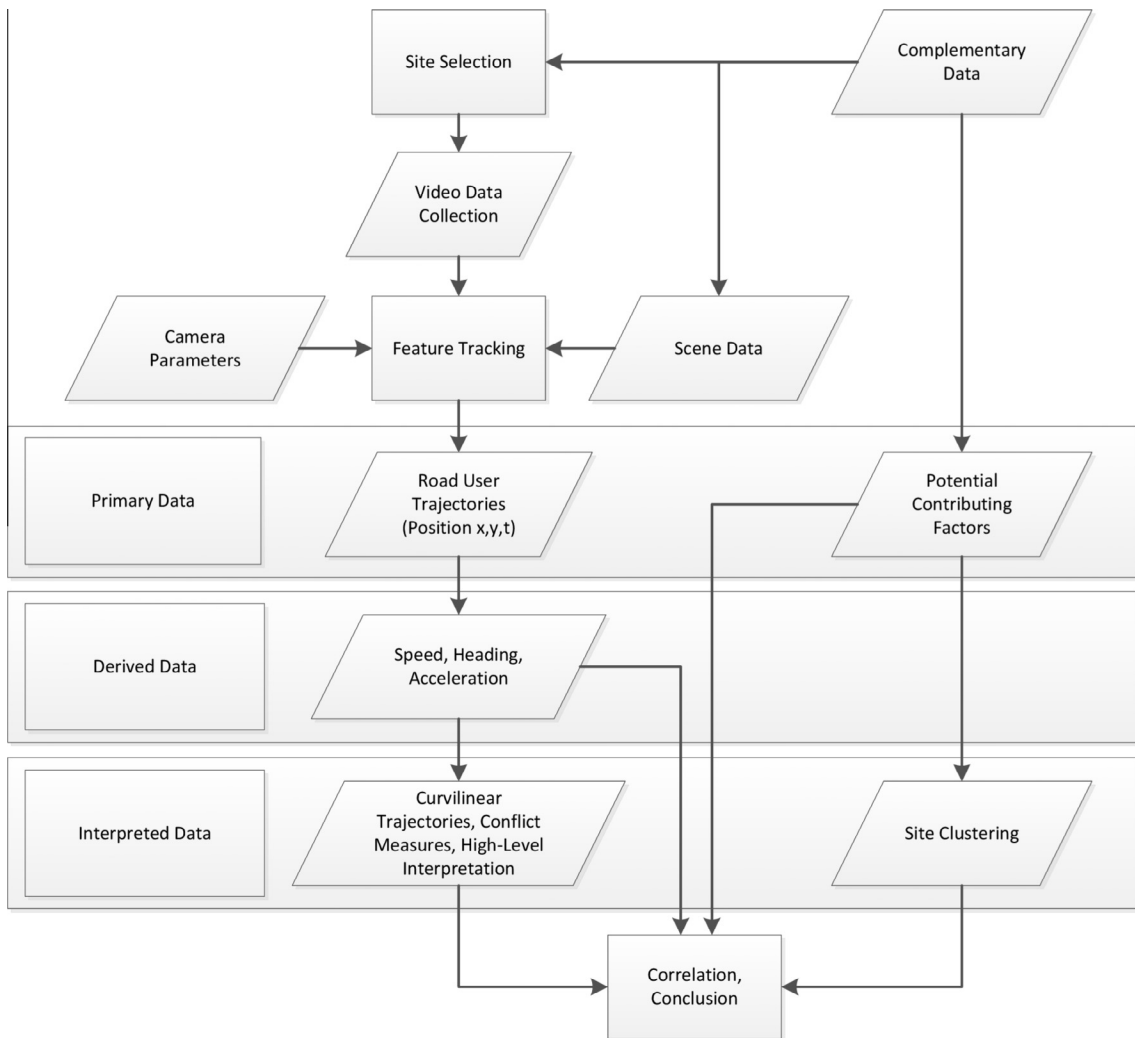


Fig. 1. Data flow diagram showing the overview of the system.



Fig. 2. Vehicle #304 is shown approaching vehicle #303 which is engaging the roundabout in the wrong direction, demonstrating a frequent violation leading to a traffic conflict.

user) at time t on a planar surface. Points are evenly spaced in time with a consistent Δt equivalent to the inverse of the frame rate of the video (typically 15–30 frames per second), i.e. a measurement is done for each frame. The object (road user) itself is represented by a group of characteristic features spread over the object and moving in unison. A sample of tracked road user trajectories is presented in image space in Fig. 2. The computer vision software is covered in greater detail in Section 2.6.3.

Three potential sources of error exist: parallax, pixel resolution, and tracking:

- **Parallax error** is mitigated by maximizing the subtending angle between the camera and the height of tracked objects. In practical terms this requires a high angle of view, ideally a bird's eye view, and tracking objects with a small height-to-base ratio. Passenger cars are more forgiving in this respect than heavy vehicles or pedestrians.
- **Pixel resolution** determines measurement precision. Objects further away from the camera experience lower tracking precision than objects near the camera. Error due to pixel resolution is mitigated by placing the camera as close to the study area as possible and using high-resolution cameras, though increases in resolution offer diminishing returns in tracking accuracy.
- Finally, **tracking errors** may occur with scene visibility issues or due to limits to current computer vision techniques. These erroneous observations have to be rejected or reviewed manually. Attempts have been made in recent literature and for this work to measure and optimize tracking accuracy (Ettehadieh et al., 2015) using the Measure Of Tracking Accuracy (MOTA) methodology (Bernardin and Stiefelhagen, 2008). MOTA is a measure of accuracy that combines multiple sources of tracking error simultaneously, including false positive and false negative detection (Measure of Tracking Precision, MOTP, may be used for the spatial accuracy of measurements). The tracking optimization is replicated for this study using a genetic algorithm that searches heuristically for optimal tracking parameters based on MOTA as a measure of fitness from manually annotated video data. See Section 3.1 for more details and results of this process.

Depending on the steps taken to minimize tracking errors, feature-based tracking functions best over study areas of up to 50–100 m in length with high-to-medium speed, low-to-medium density traffic flows.

2.2.2. Derived data: velocity and acceleration

Velocity and acceleration vectors are derived by simple differentiation of position and subsequently velocity over time, after undergoing smoothing with a moving average window over several frames. The heading of the velocity vector may be used to determine the orientation of the vehicle.

2.2.3. Size of data

Feature tracking provides a microscopic level of detail. Individual observations measured at a single site over the course of a normal day typically register in the tens of millions. The sample size (number) of individual tracking measurements (positions, velocities, etc.) per hour n can be estimated with the equation

$$n = fQd \quad (1)$$

where f is the number of frames per second of the video, Q is the average hourly flow-rate, and d is the average dwell-time of each vehicle in the scene (excluding full stops). Dwell-time is affected by the size of the analysis area in the scene and the average speed. As such, the size of the analysis area needs to be carefully selected.

2.3. Complementary data

With the exception of speed and aggregated traffic volume counts, vehicle trajectories offer little insight without context. Complementary data about the scene is collected and added in order to perform further analysis. This data includes a wide variety of design geometry and environment attributes characterizing the factors under study.

2.3.1. Analysis area

The analysis area is a bounding polygon which confines analysis to a particular region of the scene. This serves to (i) reject regions of the image with unsatisfactory feature tracking (particularly at the edges of camera space), and (ii) confine analysis to a particular region. An example of the analysis area is demonstrated in Fig. 3.

2.3.2. Alignments

Trajectory clustering is the first step in behavioural analysis. Trajectory clustering is an abstract representation of movements along prototypical paths through a scene, called alignments. This is the foundation for relating spatial position of trajectories with elements of road geometry and, in particular, the position of moving objects in relation to traffic lanes, bike paths, and side walks.

Many approaches to trajectory clustering have been explored. While some methods are supervised (Schreck et al., 2008), many more are unsupervised (e.g. k-means (MacQueen, 1967) or hidden Markov models (Rodríguez-Serrano and Singh, 2012)). Manual trajectory clustering is labour intensive and may be a source of bias, but it allows for tight control of scene description and analysis oversight. Unsupervised clustering is systematic but naive, as this form of clustering can only make use of trajectory data to infer spatial relationship. Manual clustering along a series of splines, called alignments, is chosen for its simple implementation and tight control over interpretation. A hybrid approach, which automatically refines the spatial positioning of manually defined alignments through traditional unsupervised clustering approaches, is considered for future improvements (Morris and Trivedi, 2008; Schreck et al., 2008).

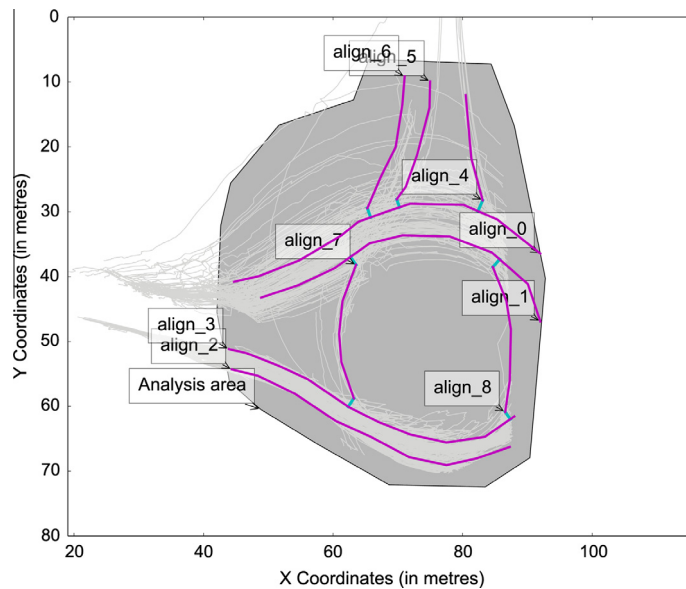


Fig. 3. The partial trajectories and scene of a multi-lane roundabout with a complex configuration of lanes (the south and east approaches are not visible). The analysis area is the grey area inside the black polygon, the alignments are the pink lines, while the connectors are in cyan. Some sample trajectories are highlighted in light grey. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The alignment is represented as a simple series of points with a beginning and an end, in the same direction of travel as the majority of movements along this path. This process introduces a new coordinate system which maps a position of a moving object in Cartesian space to a position in curvilinear space:

$$(x, y) \rightarrow (l, s, \gamma). \quad (2)$$

where a point located at (x, y) in Cartesian space is snapped orthogonally to the nearest position on the nearest alignment l , and is represented by the curvilinear distance s along this alignment from its beginning and the offset γ (positive to the right of the vector), orthogonal to this alignment, measuring the distance between the original point and its position snapped to the alignment. A second pass may be performed over a moving window in less than the time users take to perform real lane changes to correct any localized lane “jumping” errors which frequently appear near converging or diverging alignments. These coordinates are useful for studying following behaviour, lane changes, and lane deflection.

2.3.3. Network topology

Once trajectories are clustered, a network topology is constructed in order to be able to intelligently propagate future possible positions of moving objects through the network. In simple networks (i.e. two alignments), these movements are implicitly defined simply by observing lane change ratios, but in more complex networks, such as the network shown in Fig. 3, movements may involve multiple lane changes and therefore may require a more generalized approach. A recursive tree model is employed.

Alignment extremities are linked to other nearby alignments, creating diverging or converging branches, as are momentarily adjacent alignments. In addition, alignments which run parallel over a distance of more than 15 m are instead grouped into corridors over which lane changes may occur freely. This creates a series of links and nodes with implicit direction which can be searched to determine all possible future positions of a moving object inside this network. This serves to reduce processing times of spatial relationship calculations between objects (triage) and provides more intelligent interpretation of spatial relationships.

2.3.4. Geometric data/inventory

Finally, a traditional inventory of contextual factors that may be related to the behaviours under study needs to be constructed and associated with each site. These typically include:

- **Geometric characteristics** including lane width, curvature, approach angles, and number of lanes, presence of slip lanes and/or lane configuration;
- Presence of **horizontal and vertical signalization**: if proper placement is under evaluation, location may be recorded, or a typology of signage quality may be used instead;
- **Pedestrian facilities** such as crosswalks, and pedestrian refuges;
- **Built environment**: school zones, land usage, clearance;
- **Upstream/downstream distances to features** such as intersections, speed limit changes, and rail-road crossings.

2.4. Measurement definitions

2.4.1. General versus specific analysis measures

Some measures are generalizable for all traffic studies using alignments, while others are not (Gettman and Head, 2003). General traffic measures include speed profiles, counts, lane changes, origin–destination matrices, and basic spatial relationships including conflicts. Other measures may be specific to the study and generally require high-level interpretation (HLI). This interpretation makes use of study-specific geometric information to generate custom measures. As such, this section will not cover these custom measures, but will instead focus on generalizable measures. However, an application of HLI calculations will be briefly presented in Section 3.4.

2.4.2. Interactions

An interaction quantifies the spatial relationship between moving objects in a scene, as is depicted in Fig. 2. At the most fundamental level, an interaction is defined as a pair of moving objects simultaneously present in a scene over a common time interval (also referred to as a user pair). We further define an instantaneous observation (i.e. in a given video frame) within this time interval as an interaction instant (St-Aubin et al., 2015).

This interaction definition is generic, if not naive, as the quality depends largely on how the scene is constructed. For example, the significance of an interaction between two vehicles separated from each other physically (e.g. via a median or a building) may not be comparable to an interaction between two vehicles merely separated by lane markings because of the implication that they may cross the lane marking intentionally or inadvertently very easily. This may interfere with collision prediction attempts, particularly if scenes are not consistently selected and geometry is not controlled.

One solution is to perform a triage of user pairs based on physical access and proximity. A network topology coupled with a driving distance horizon is proposed. This is not a perfect solution, however, as physical access may not necessarily be a discrete choice. For the example of the median, it is still physically possible, although much less likely, for a vehicle to cross over into oncoming traffic.

2.4.3. Motion prediction

While vehicle trajectories offer a rich set of observed behavioural data, they do not provide much collision data. The proactive road safety approach requires that predicting collisions should be performed without observing them directly. To be studied, collisions must be extrapolated from traffic events with potential for collision. This potential is modelled by predicting future possible positions between each pair of road users at every instant in time. Several motion prediction models are proposed for study, including constant velocity (Laureshyn et al., 2010), normal adaptation (Mohamed and Saunier, 2013), and motion patterns (Saunier et al., 2007; Morris and Trivedi, 2008; Sivaraman and Trivedi, 2013), particularly the implementation of a discretized motion pattern (St-Aubin et al., 2014). Specific details of implementation of each method can be found in St-Aubin et al. (2015).

As illustrated in Fig. 4, motion prediction is performed for each user pair over each instant t_0 for a number of time steps of size Δt between t_0 and some chosen time horizon. Each motion prediction may generate for two road users a series or a matrix of collision points with a sum of probabilities inferior or equal to 1.

2.4.4. Time-to-collision

Time-to-collision (TTC), first proposed by Hayward (1971), is one of the most popular surrogate safety measures. It is a method of quantifying proximity to danger. Time-to-collision measures the time, at a given instant t_0 , until two road users

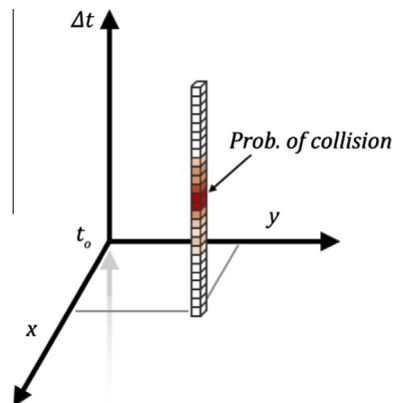


Fig. 4. Probability of collision in motion prediction space in (x, y, t) projected into the future from $t = t_0$ over Δt steps where TTC is the measured Δt in projected motion prediction space where and when a significant probability of collision occurs.

on a collision course collide, based on the motion prediction model. In its simplest form, e.g. motion prediction at constant velocity or in a car-following situation, time-to-collision is the ratio of differential velocity or speed and differential position. A TTC value of 0 s is, by definition, a collision. TTC is particularly useful as it has the same dimensions as some important traffic accident factors such as user perception and reaction time, typically set at the critical value of 1.5 s (Hydén, 1987; Green, 2000) in the literature, and breaking time. Larger values of observed TTC thus provide greater factors of safety for particular driving tasks.

Time-to-collision is measured instantaneously: a new value of TTC may be computed at every instant. Thus, a pair of users may have a time series of TTC observations evolving over time. Some efforts have been made to study these evolutions (Mohamed and Saunier, 2013) as a form of safety continuum (Zheng et al., 2014). Other approaches have focused counting discrete events using a threshold similar to the TCTs (Hydén, 1987; Svensson and Hydén, 2006), although care must be taken because results can be shown to vary with prediction methodology and threshold level (St-Aubin et al., 2015).

A sample pair of road user trajectories (#303 and #304, Fig. 2) and spatial relationships simultaneously existing over a time interval lasting 64 instants, or just over 4 s, is presented in Fig. 5. In this scenario, vehicle #304 is approaching vehicle #303, which is engaged in an illegal U-turn, at high velocity. The norm of the differential velocity Δv , relative distance d , and corresponding time t are measured for every instant. In a matter of just under 4 s, the differential velocity changes from 9.63 to 2.26 m/s while the relative distance changes from 28.57 to 9.57 m. For every interaction instant of this user pair, motion prediction is used to calculate resulting TTC under each motion prediction method. The time series of the predicted collisions and associated TTC measures for this pair of users and for each different motion prediction method is presented in Fig. 6. At each instant in time, normal adaptation and motion pattern prediction generate multiple possible collisions, while constant velocity can only ever produce one possible point of collision. The existence of more predicted collision instants using motion patterns demonstrates the potential for predicting many more complex situations. When several potential collision points are predicted, the expected time-to-collision, defined as TTC'_i , at time t_i (for instant i) is calculated as the probability-weighted TTC average of all possible collision points indexed $j = 1, \dots, n$ with $Prob(collision)_{ij}$ as defined and measured in St-Aubin et al. (2014).

$$TTC'_i = \frac{\sum_{j=1}^n TTC_{ij} Prob(collision)_{ij}}{n} \quad (3)$$

It is clear from both this figure and the trajectories themselves that constant velocity and normal adaptation motion predictions are inadequate for roundabout conflict analysis: the trajectories share the same destination yet their headings lead them to a potential collision point only for a brief period of time with these prediction methods.

2.5. Indicator aggregation over time and space

Instantaneous surrogate safety indicators such as speed and TTC are observed continuously. Currently, there is a lack of consensus in the literature on how to interpret continuous measures in terms of safety besides the event-based (Hydén, 1987; Svensson and Hydén, 2006) or safety continuum (Zheng et al., 2014) paradigms. One qualitative approach used in the literature has been to compare shifts in the probability distributions of these indicators when the magnitude of impact of individual indicators has no bearing on the overall direction of shift of safety risk (Ismail et al., 2010; Autey et al., 2012; St-Aubin et al., 2013). For example, as illustrated in Fig. 7(c) and (d) it's unclear if converting "high and low risk" events into "medium risk" events is beneficial, but converting "high risk" events into "low risk" events is always beneficial as illustrated in (a) and (b). Still, there may be more complex interpretations related to driver awareness and learning as a function of perceived danger as described in Svensson and Hydén (2006). Alternatively, changes in frequency of events meeting a given threshold may yield similar results. While traffic event conversion factors have been developed and used in the past (Hydén, 1987; Svensson and Hydén, 2006), transferability of validated conversion factors has been cited as problematic (Mohamed and Saunier, 2013). Indeed, while the shape of probability distributions of TTC has generally been characterized as Gamma-like in the literature (Ismail et al., 2010; Autey et al., 2012), variations or compound effects can be found, particularly when traffic streams become mixed (St-Aubin et al., 2013). One recent approach proposed a shifted gamma-generalized Pareto distribution model (Zheng et al., 2014).

Indicator interpretation may depend on the precise definition of what constitutes a "traffic interaction". The disaggregated interaction approach treats indicators as a representation of incremental risk which add up over time, but it tends to bias slower moving objects which dwell in the scene longer and complicates conditional probability calculations. The user-pair aggregated approach solves these problems but is sensitive to how the analysis area is defined and which users pairs are defined as interacting. Overall, a more detailed interaction exposure framework is still needed in the literature. Some simplifications have been made in the meantime. The traditional TCT methodology frequently represented user pair interactions using a single value of TTC, the minimum TTC (TTC_{min}). In practice, due to the imperfect nature of automated video data extraction, minimum values tend to oversample instantaneous tracking errors, and so a 15th percentile approach is preferred (St-Aubin et al., 2015).

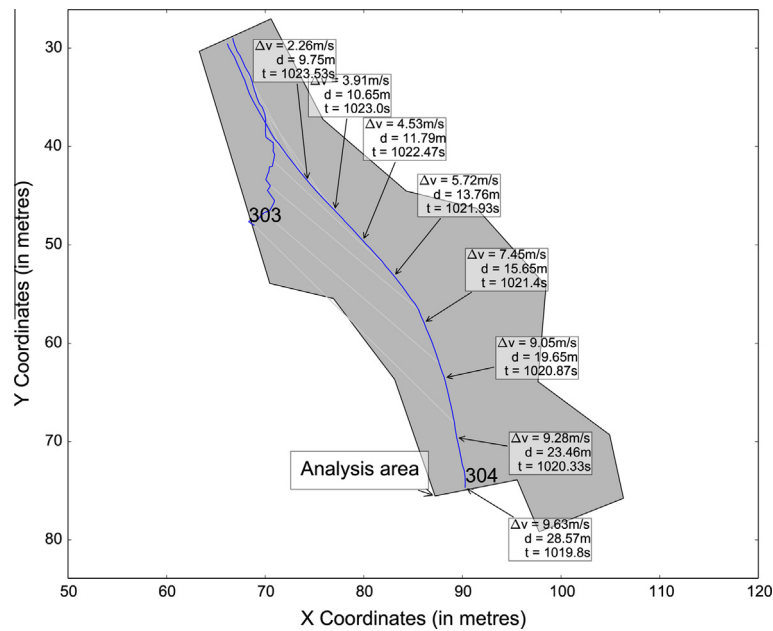


Fig. 5. Vehicle #304 is shown approaching vehicle #303 which is engaging the roundabout in the wrong direction. Spatial relationship measures Δv , relative distance d , and time stamp t are labelled along the time series every eight frames between the two trajectories. Light grey lines join the two trajectories at common time frames for visualization purposes.

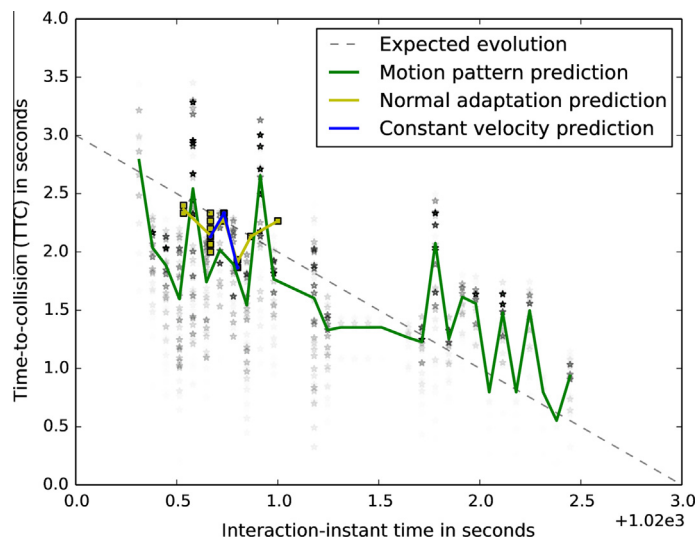


Fig. 6. Time series of TTC observations for different motion prediction methods for the interaction between vehicles #303 and #304. Markers correspond to TTC for a specific collision point at each instant, the probability of collision associated with each collision point is indicated by the opacity of the marker and lines are weighted average TTC observations per instant.

2.6. Tools and techniques

2.6.1. Data collection

Video data collection is primarily performed using a specially constructed mobile video-data collection system built for temporary high-angle video data collection and has tamper-proof, weather-proof, self-contained features presented in Jackson et al. (2013). Ordinary fixed traffic surveillance CCTV cameras can be used as well and are usually more reliable, but only where and when available.

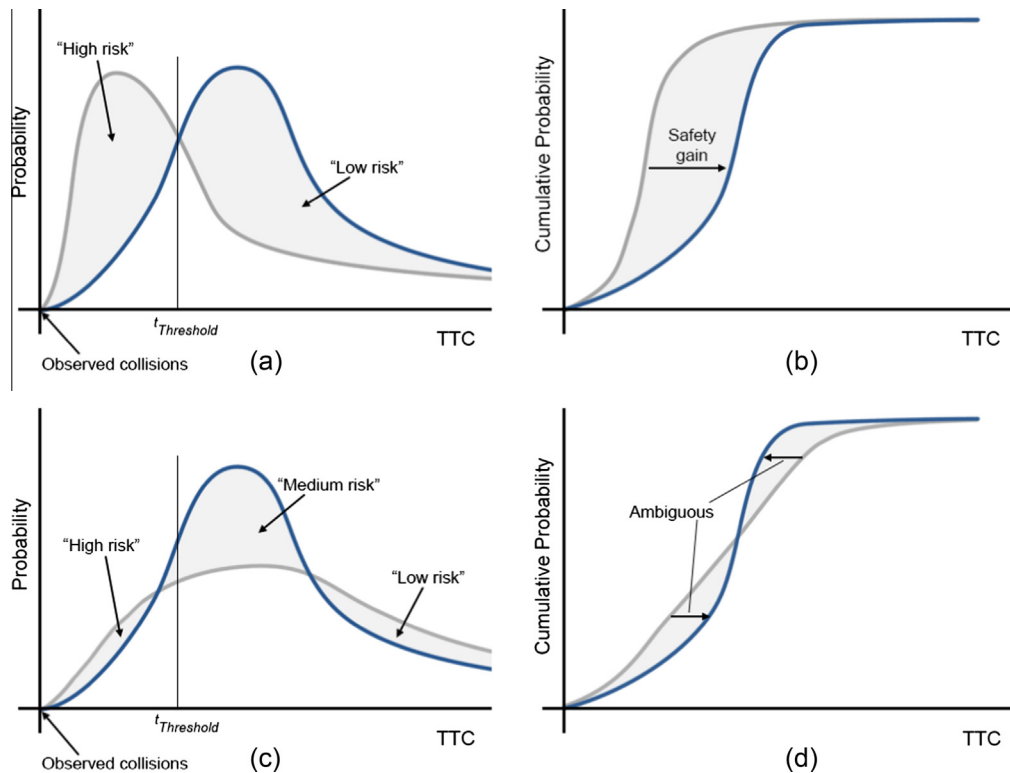


Fig. 7. TTC probability distribution function comparisons. Qualitative analysis of TTC shift is possible in (a), while it is not as straightforward in (b). An alternative approach is to perform a comparison on a low threshold.

2.6.2. Camera types

Two cameras were tested. The first was a Vivotek IP security camera with a narrow lens filming at 15 frames per second at a resolution of 800×600 , while the second camera was a GoPro 2 with a wide-angle lens filming at 30 frames per second at a resolution of 1280×960 . Both cameras encoded video in H.264, although the GoPro used a significantly larger bitrate. The consumer-grade GoPro provides higher-quality videos, however these gains seem to have minimal impact on vehicle tracking performance while the video files are much larger and more difficult to handle and process.

2.6.3. Software

The software used is the open-source Traffic Intelligence project (Saunier et al., 2010; Jackson et al., 2013), itself based on the computer vision library OpenCV (Brahmbhatt). This software provides the basic feature tracking (the algorithm presented in Saunier and Sayed (2006)), trajectory management and coordinate projection functionality, as well as a few usage-specific tools such as correction for lens distortion, trajectory clustering, and basic motion prediction functions. Some of the more advanced analysis tools and techniques presented in this paper are under development and will be made available as their functionality is completed and validated.

2.6.4. Data model

All this data, from the raw video data to the high level interpretation, need to be organized and managed. A high level conceptual data model is presented as an entity-association diagram in Fig. 8. It is a little simplified to avoid some implementation details which can be found in the Traffic Intelligence project. The diagram has two main parts:

- The entities (objects) resulting from the video analysis (upper half of Fig. 8): the raw data extracted from video analysis is stored as time series (indexed by frame number) of positions and velocities (with corresponding unique keys), and are grouped as road user objects which may have a type (passenger vehicle, truck, cyclist, etc.). Interactions are between two road users and may be characterized by several indicators such as TTC, and PET.
- The entities providing the data description or meta data (lower half of Fig. 8): sites, e.g. the different roundabouts studied in this work, are the corner stone of the meta-data. They may correspond to several video sequences (the actual video files), each being characterized by a camera view with camera parameters such as a homography matrix (but a camera

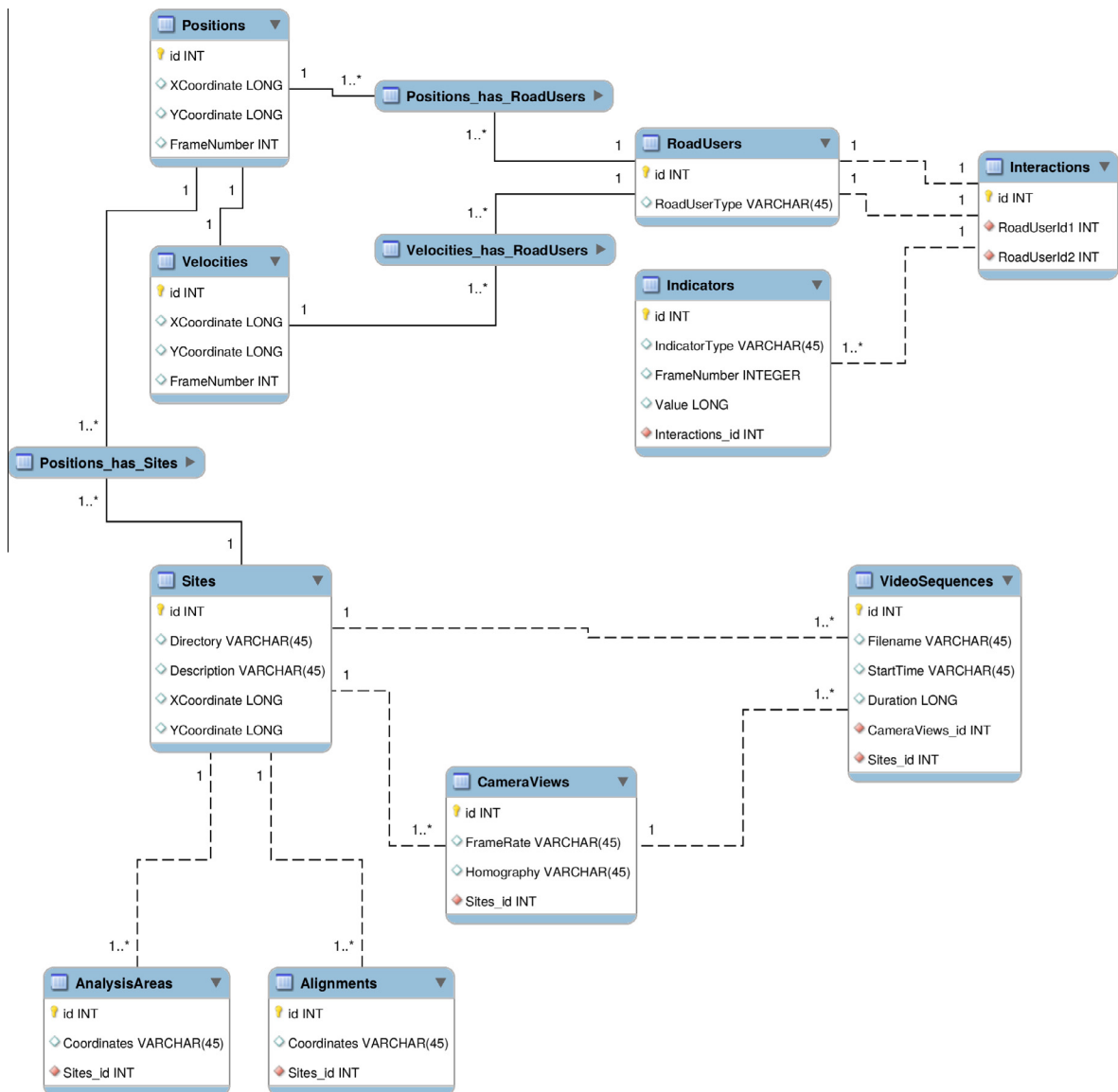


Fig. 8. High level data model as an entity-association diagram (using symbols of the Unified Modelling Language).

view can be the same for several files, e.g. when video sequences are split into several files for each hour). Various types of site features (complementary data) may be added, e.g. the alignments and analysis areas shown as examples in the diagram.

Positions and road users are obviously linked to a camera view, a video sequence, and a site (not through an actual key in the positions table as shown in Fig. 8, but through configuration files).

2.6.5. Processing

Real-time analysis is not an explicit goal of this technology, as its intended use is primarily the off-line analysis of data recorded in the field from non-permanent video cameras. However, performance is a serious consideration, if for no other reason than to ensure that processing remains affordable and does not fall behind data collection. In any case, some calculations may require pre-processing of as much data as possible, in particular machine learning tasks such as motion prediction (see Section 2.4.3).

In the current iteration of the software, and with today's multi-core processors, tasks are highly parallelizable. Feature tracking and trajectory analysis can be performed on multiple video sequences at a time, typically cut up into 20-min or one-hour segments, in parallel on a single mid-to-high-performance machine, or on a computer cluster. With parallel

processing of video sequences on a single computer, memory becomes the main bottleneck; 32 GB or more of memory are highly recommended on a multi-core machine, to take full advantage of up to 8 threads. Alternatively, the large majority of calculation tasks can be parallelized at the observation level, as they are independent events.

Feature tracking is written in C++ for performance, while the majority of trajectory analysis is written in Python for ease of development and extensibility. Where possible, expensive trajectory analysis calculations make use of Python wrappers for fast compiled libraries.

3. Experimental results

3.1. Tracking performance

Tracking accuracy was evaluated by way of the CLEAR MOT metric Measure Of Tracking Accuracy (MOTA) (Bernardin and Stiefelwagen, 2008) using as ground truth manually annotated trajectories with an open source annotation tool from the Urban Tracker project (Jodoin et al., 2014). Perfect tracking would yield MOTA of 100%, while it can become negative if more false alarms are done than there are ground truth objects. Nearly 40,000 observations (instants) of 371 motor vehicles were annotated manually across two prototypical sample sites: one site where objects were tracked at a distance of between 20 and 50 m away from the camera and one site where objects were tracked at a distance no greater than 20 m away. An initial MOTA was calculated for default tracking parameters and offered modest performance (in the neighbourhood of 70%). A tracking optimization was then performed using a genetic algorithm similarly to Etehadieh et al. (2015) to search for parameters that maximized MOTA. After optimization, accuracy increased to 94% (measured over the same ground truth used for optimization). These statistics are summarized in Table 1. Tracking parameter optimization converged in less than 24 h.

Of important note is the fact that this calibration was performed on the oldest and poorest quality video data (resolution of 800×600 , without software-assisted image stabilization, and before lens correction). Tracking results are expected to improve with increases in quality of video data and video pre-treatment. The optimized tracking parameters should be portable to other sites with similar view and camera characteristics, though this will be investigated further and be the subject of a future paper in depth.

3.2. Data size

In total, 473 h of video data were collected (see Table 2 for details). Data was typically collected for 12 h from 7 a.m. to 7 p.m. At 30 frames per second, a data collection at an intersection over a driving distance of 50 m and at an average driving speed of 30 km/h, and with an average hourly flowrate of 500 veh/h yields approximately 90,000 instantaneous moving object measurements per hour. Additionally, each of these observations can have anywhere between 3 and 100 feature tracks associated with it. For an adequate balance between performance and tracking accuracy, the recommended number of features to aim for is roughly 15–20 per object over time, depending on the typical time spent by a road user in the field of view. This yields manageable data sizes (roughly 500 MB of storage per hour of video) while maintaining an adequate level of data richness and object representation. Video storage needs will vary greatly by camera choice, resolution, frame rate, and video encoding settings.

Fig. 9 shows hourly number of user pairs observed versus traffic volume. Trends are evident, but explaining factors are not clear (probably a mix of several lane arrangement indicators). This will need further study. The number of user pairs per hour should be linearly correlated with the number of interaction instants as is demonstrated in Fig. 10. If they are not, it is possible that analysis areas across sites are not comparable, particularly for disaggregated TTCs.

Most of the analysis is conducted on a pair of dedicated consumer-grade high-performance machines (INTEL Core i7 3770k processor with 16–32 GB of memory), with some work offloaded to a computing cluster when acceleration of work is necessary. Feature tracking performance depends on video resolution and special post-processing requirements such as stabilization or lens correction (for distortion). A typical one hour 800×600 pixels video at 30 frames per second is processed with current consumer-grade hardware in about an hour. A typical one hour 1280×960 video with correction for distortion can be processed in about two hours. Basic analysis on one of these trajectory sequences takes between 5 min and 30 min, depending on traffic in the scene, while interaction analysis, particularly using motion patterns, can typically take anywhere between 1 and 48 h to complete. Interaction analysis processing times are very sensitive to the interaction complexity of the scene.

Table 1
Tracking performance.

| Site | Ground truth objects | Ground truth observations | Unoptimized MOTA (default parameters) | Optimized MOTA |
|------------------------|----------------------|---------------------------|---------------------------------------|----------------|
| #1-1-12 (far camera) | 146 | 12,887 | 0.685 | 0.944 |
| #22-3-16 (near camera) | 225 | 25,011 | 0.744 | 0.853 |

Table 2
Data details.

| | |
|--------------------------------------|-------------|
| Roundabouts | 20 |
| Analysis Areas | 36 |
| Hours of Video Data | 473 |
| Estimated Total Traffic Volume | 79,432 |
| Disk Space (Video + Data + Overhead) | 1.9 TB |
| Veh-km Traveled | 9505 veh km |

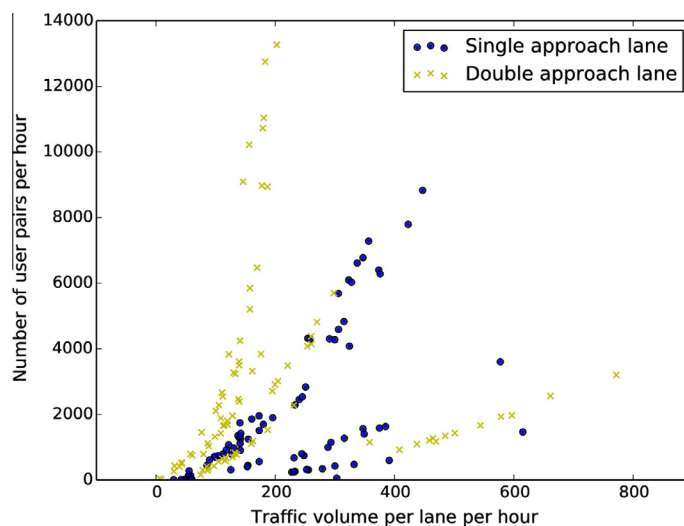


Fig. 9. Number of user pairs observed at each roundabout versus inflow per lane per hour.

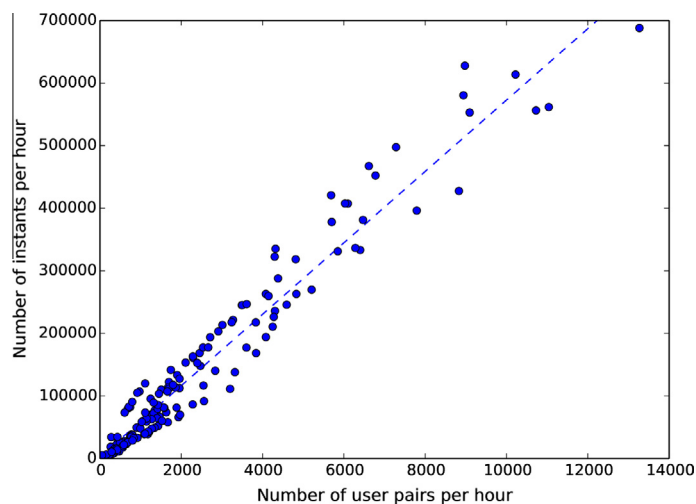


Fig. 10. Number of interaction instants observed versus number of user pairs observed.

3.3. Sample surrogate safety analysis

A sample surrogate safety analysis of three of the sites is demonstrated in Fig. 11. This shows road user trajectories projected in and with respect to the scene, mean velocity vectors and speed along these trajectories, and spatial distribution of collision points based on motion pattern prediction with instantaneous *probability* > $1e5$ and *TTC* < 1.5 s. Problematic weaving conflicts are highlighted in the first and third examples, particularly for multi-lane roundabouts. The second example, on the other hand, demonstrates car-following conflicts, particularly at the approach (not surprising, given that this single lane approach frequently experiences queuing as drivers yield to the high conflicting flow) and exit (more surprising).

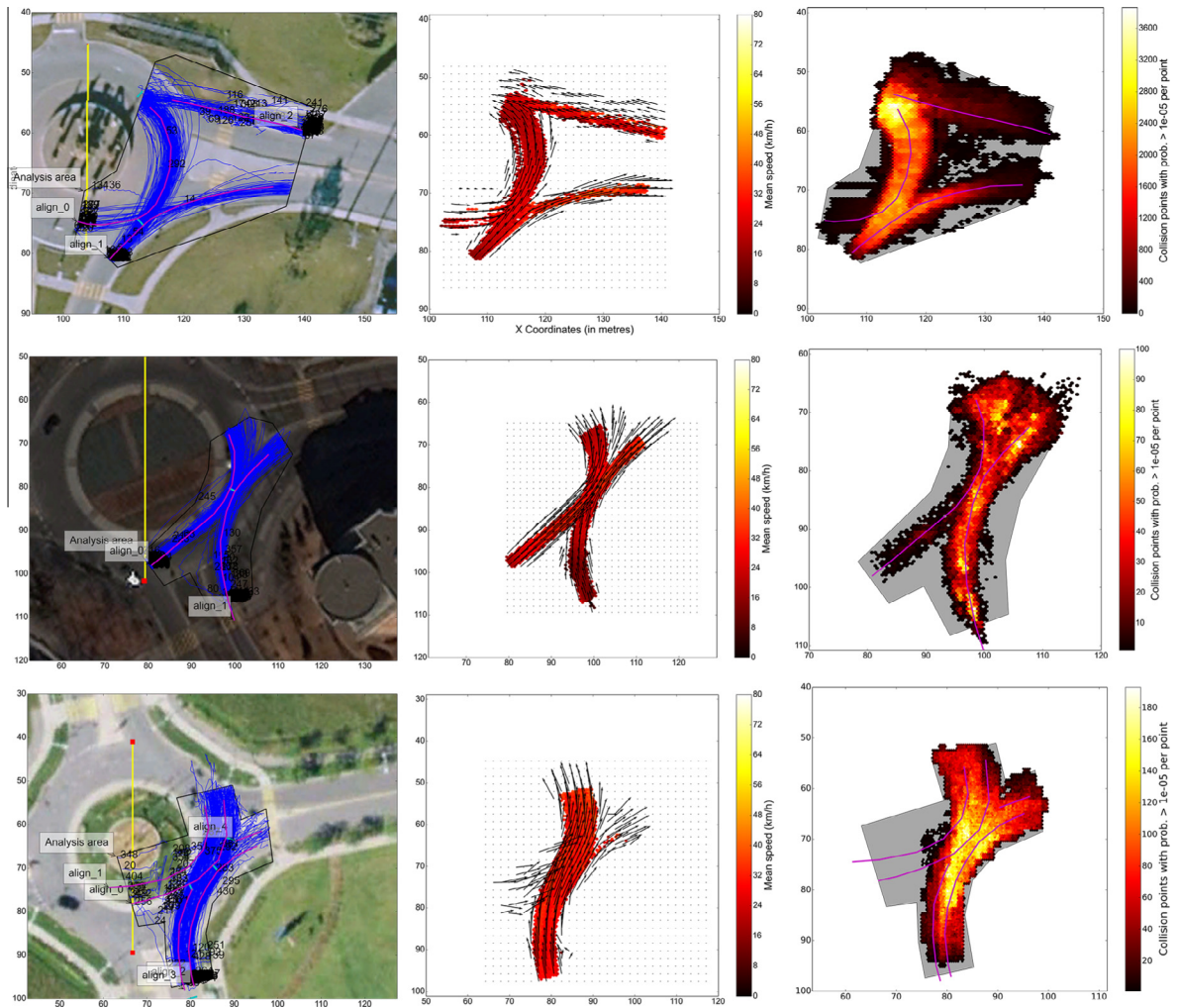


Fig. 11. Sample spatial data and analysis at 3 selected sites from top to bottom. From left to right, diagrams demonstrate trajectory tracks (positions) in analysis area and with descriptive alignments, mean speed and heading, and spatial distribution of collision points based on motion pattern prediction with instantaneous probability $> 1e5$ and TTC < 1.5 s. All coordinates in metres, north pointing upwards.

Table 3

K-means clusters for weaving zone sites.

| Cluster | Description | Group size | Observations |
|---------|--|------------|--------------|
| cl_1 | Small single and double lane residential collectors | 11 | 4200 |
| cl_2 | Single-lane regional highways and arterials with speed limits of 70–90 km/h and mostly polarized flow ratios | 16 | 26,243 |
| cl_3 | 2-lane arterials with highly polarized flow ratios | 5 | 13,307 |
| cl_4 | Hybrid lane 1 \rightarrow 2 \rightarrow 1 arterials with flow ratios favouring the conflicting flow | 3 | 4809 |
| cl_5 | Traffic circles converted to roundabouts (2 lanes, extremely large diameters, tangential approach angles) | 4 | 10,295 |
| cl_6 | Single-lane regional highways with large-angle quadrants (140 degrees) and even flow ratios | 2 | 2235 |

Due to the large number of potentially contributing factors (at least 20) and the relatively small number of corresponding sites, sites were clustered into 6 groups according to Table 3 using the k-means algorithm (St-Aubin et al., 2015).

The cumulative distribution of 15th percentile motion pattern TTCs is broken down by roundabout cluster as illustrated in Fig. 12. Cluster cl_3 (2-lane arterials with highly polarized flow ratios) offers the best and clearest safety benefits, while cluster cl_5 (traffic circles converted to roundabouts with 2 lanes, extremely large diameters, tangential approach angles) offers markedly decreased safety performance. cl_6 (single-lane regional highways with large-angle quadrants (140 degrees) and even flow ratios) is ambiguous in its interpretation: while there is a general left-shift in TTC distribution (towards lower values) there is also a decrease in frequency of the lowest (most severe) TTC indicators. The remaining clusters, representing

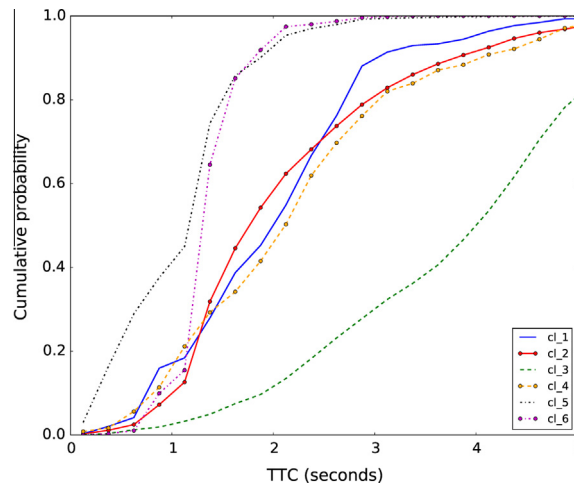


Fig. 12. Cumulative probability distribution of 15th percentile motion pattern TTCs for all roundabout weaving zones clustered into six groups.

31 sites, tend towards an average TTC distribution. This is still a large number of sites. Further investigation, regressing for specific factors, is warranted (St-Aubin et al., 2015).

3.4. Sample high-level interpretation analysis

Some high-level interpretation measures are also compiled using the data sample. Fig. 13 shows the mean speed profiles, with the interval at mean \pm one standard deviation, through the roundabout weaving zones. Speed profiles are mapped, not as units of distance, but rather as proportions of curvilinear location relative to the start and end of the merging zone, measured at the yield line of the approach and equivalent line at the corresponding exit. This is done to account for the large variability in diameter of roundabouts and angular size of weaving zones across sites. Mean speeds are generally consistent

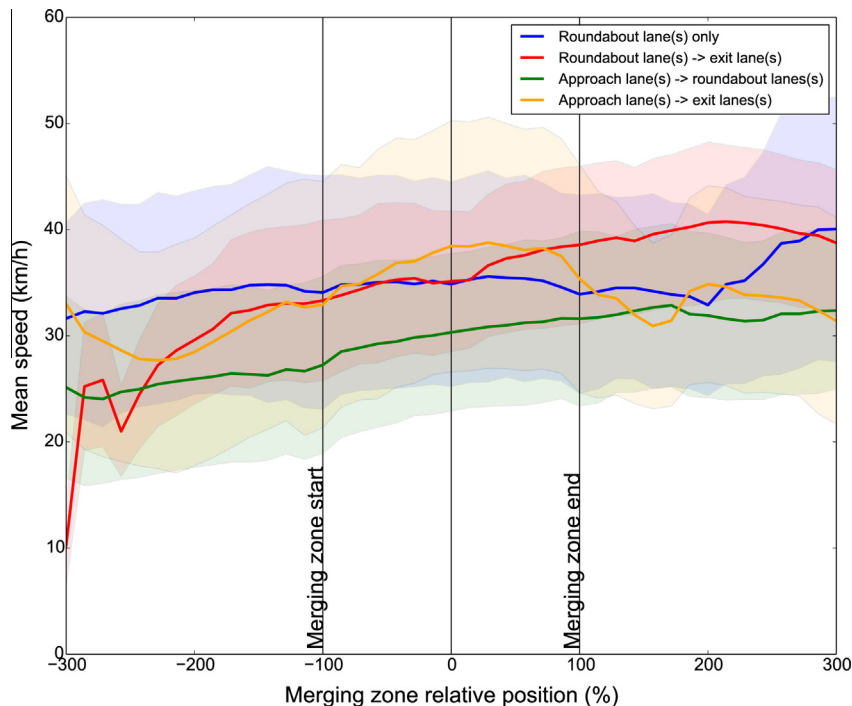


Fig. 13. Mean speed profiles and \pm one standard deviation relative to merging zone start and end according to movement type.

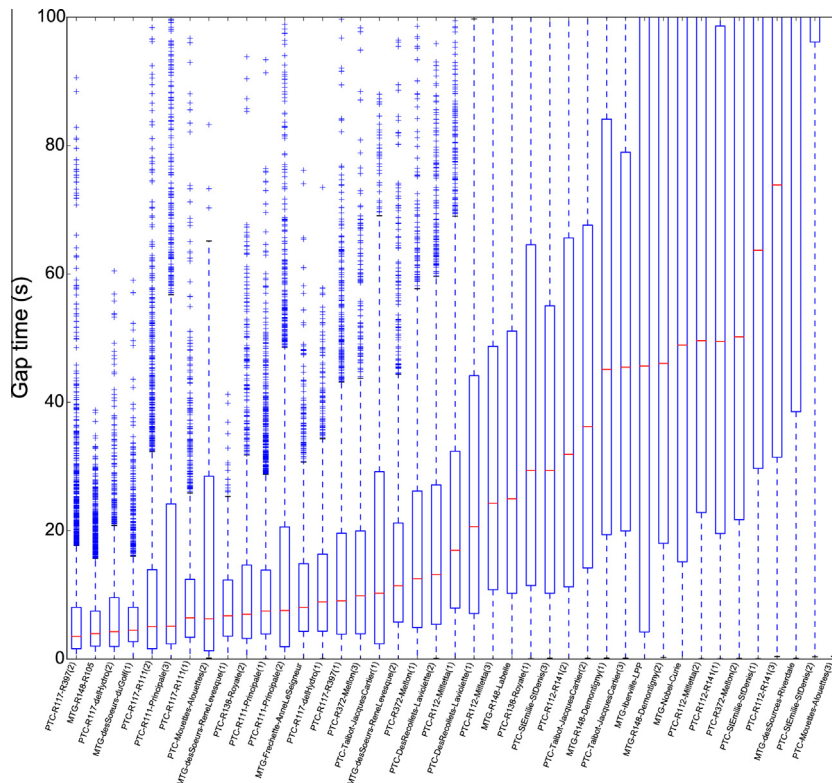


Fig. 14. Distributions of gap at merging instant of approaching vehicles (critical gap), presented as boxplots (the maximum represented gap is 100 s, but some sites see even larger values).

with those in the literature, but variation does occur with relative location and movement type (Hydén and Várhelyi, 2000). In addition, cross-sectional analysis uncovers even larger variations in mean speed profiles (St-Aubin et al., 2013).

Finally, Fig. 14 presents boxplots of lag gap acceptance of approaching vehicles and conflicting vehicles. In a cross-sectional analysis, this quantifies vehicle insertion aggressivity. Smaller accepted gaps might be explained by more impatient drivers, typically symptomatic of high-volumes of continuous flows inside the roundabout and long wait times at the approach.

4. Conclusion

Large-scale automated video data allows for larger surrogate analysis, in practice and research, for the same purpose as the traditional historical accident data approach: road network screening, evaluation of countermeasure, road safety diagnosis, etc. With this in mind, surrogate safety analysis is expected to play an important role as a complementary safety approach or as an approach that can potentially replace the traditional methods, particularly when historical accident data is limited or doubtful (given its poor quality in some cases).

This paper demonstrates the theoretical and practical application of a large-scale automated video data collection system using computer vision for highly detailed traffic studies, in particular for proactive road safety analysis using surrogate safety analysis. The reader is led step-by-step through the process of collecting, processing, and analysing video data, with examples and discussion of challenges along the way. This paper demonstrates an early implementation of the methodology in the form of a cross-sectional analysis of driver behaviour in the largest set of roundabout video data analysed to date.

Several technical challenges and their solutions were outlined, notably tracking errors (MOTA optimized to 94% and no less than 85%), analysis of TTC distributions, and aggregation and sampling considerations. It is expected that these issues will be further addressed as processing and analysis tools become more accessible, more collaborators contribute solutions to the open source software stack, and techniques applied to transportation issues become more sophisticated. Future work will examine camera lens, angle, and visibility considerations for effect on tracking accuracy.

The roundabout analysis concludes that 2-lane arterials with highly polarized flow ratios offer the best gains in safety, though this is likely due to the effect of polarized flows and not the multi-lane aspect, as it was also shown that multi-lane roundabouts generate increased concentration of conflict. The true effects of multi-lane roundabouts on safety may not be as simple as previously thought, and may depend on the specifics of lane configuration. Traffic circles converted to roundabouts

(change of signage only) yielded the poorest results and should be avoided. A future paper will be dedicated to examining regression models for individual contributing factors, in more detail as well as before-and-after studies of winter driving and roundabout conversion.

Acknowledgements

The authors would like to acknowledge the funding of the Québec road safety research program supported by the Fonds de recherche du Québec – Nature et technologies, the Ministère des Transports du Québec and the Fonds de recherche du Québec – Santé (proposals number 2012-SO-163493), as well as the various municipalities for their logistical support during data collection. Shaun Burns, a Masters student at McGill University was also instrumental in the collection of video data.

References

- Autey, J., Sayed, T., Zaki, M.H., 2012. Safety evaluation of right-turn smart channels using automated traffic conflict analysis. *Accid. Anal. Prev.* 45, 120–130. <http://dx.doi.org/10.1016/j.aap.2011.11.015>.
- Bernardin, K., Stiefel, R., 2008. Evaluating multiple object tracking performance: The CLEAR MOT Metrics. *EURASIP J. Image Video Process.* 2008, 1–10. <http://dx.doi.org/10.1155/2008/246309>.
- Brahmbhatt, S., Practical OpenCV. <http://dx.doi.org/10.1007/978-1-4302-6080-6>.
- Chin, H.-C., Quek, S.-T., 1997. Measurement of traffic conflicts. *Saf. Sci.* 26 (3), 169–185. [http://dx.doi.org/10.1016/S0925-7535\(97\)00041-6](http://dx.doi.org/10.1016/S0925-7535(97)00041-6).
- Ervin, R., MacAdam, C., Walker, J., Bogard, S., Hagan, M., Vayda, A., Anderson, E., 2000. System for Assessment of the Vehicle Motion Environment (SAVME), Tech. Rep. UMTRI-00-21.
- Ettehadieh, D., Farooq, B., Saunier, N., 2015. Systematic parameter optimization and application of automated tracking in pedestrian-dominant situations, in: *trb*, 15–2400.
- Gettman, D., Head, L., 2003. Surrogate Safety Measures From Traffic Simulation Models.
- Gordon, T., Bareket, Z., Kostyniuk, L., Barnes, M., Hagan, M., Kim, Z., Cody, D., Skabardonis, A., Vayda, A., 2012. Site-Based Video System Design and Development, Tech. Rep. S2-S09-RW-1, Strategic Highway Research Program (SHRP2).
- Green, M., 2000. How long does it take to stop? Methodological analysis of driver perception-brake times. *Transport. Human Fact.* 2 (3), 195–216. http://dx.doi.org/10.1207/s1070203_1.
- Hauer, E., 1978. Traffic Conflict Surveys: Some Study Design Considerations, Tech. Rep., Transport and Road Research Laboratory, Questions Concerning Reliability and Validity.
- Hayward, J., 1971. Near Misses as Ameasure of Safety at Urban Intersections, Ph.D. Thesis, Pensilvania State University, first use of TTC.
- Hydén, C., 1987. The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique, Ph.D. Thesis, University of Lund, Lund Institute of Technology Department of Traffic Planning and Engineering, tCT thresholds, conversion rates, and reliability tests.
- Hydén, C., Linderholm, L., 1984. The swedish traffic-conflicts technique. In: *International Calibration Study of Traffic Conflict Techniques*. Springer, Berlin Heidelberg, pp. 133–139. http://dx.doi.org/10.1007/978-3-642-82109-7_12.
- Hydén, C., Várhelyi, A., 2000. The effects on safety time consumption and environment of large scale use of roundabouts in an urban area: a case study. *Accident Anal. Prev.* 32 (1), 11–23. [http://dx.doi.org/10.1016/S0001-4575\(99\)00044-5](http://dx.doi.org/10.1016/S0001-4575(99)00044-5).
- Ismail, K., Sayed, T., Saunier, N., 2010. Automated analysis of pedestrian-vehicle conflicts: context for before-and-after studies. *Transport. Res. Rec.* 2198, 52–64.
- Jackson, S., Miranda-Moreno, L.F., St-Aubin, P., Saunier, N., 2013. Flexible, mobile video camera system and open source video analysis software for road safety and behavioral analysis. *Transport. Res. Rec.: J. Transport. Res. Board* 2365 (1), 90–98. <http://dx.doi.org/10.3141/2365-12>, <<http://dx.doi.org/10.3141/2365-12>>.
- Jodoin, J.-P., Bilodeau, G.-A., Saunier, N., 2014. Urban tracker: multiple object tracking in urban mixed traffic. In: *IEEE Winter Conference on Applications of Computer Vision*. IEEE. <http://dx.doi.org/10.1109/wacv.2014.6836010>.
- Kim, Z., Gomes, G., Hranac, R., Skabardonis, A., 2005. A machine vision system for generating vehicle trajectories over extended freeway segments. In: *12th World Congress on Intelligent Transportation Systems*, San Francisco, CA.
- Kruysse, H.W., 1991. The subjective evaluation of traffic conflicts based on an internal concept of dangerousness. *Accident Anal. Prev.* 23 (1), 53–65. [http://dx.doi.org/10.1016/0001-4575\(91\)90035-4](http://dx.doi.org/10.1016/0001-4575(91)90035-4).
- Laureshyn, A., Svensson, Å., Hydén, C., 2010. Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. *Accid. Anal. Prev.* 42 (6), 1637–1646. <http://dx.doi.org/10.1016/j.aap.2010.03.021>.
- MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: *5-th Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, pp. 1:281–297.
- Mohamed, M.G., Saunier, N., 2013. Motion prediction methods for surrogate safety analysis. *Transport. Res. Rec.: J. Transport. Res. Board* 2386 (1), 168–178. <http://dx.doi.org/10.3141/2386-19>.
- Morris, B., Trivedi, M., 2008. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Trans. Circuits Syst. Video Technol.* 18 (8), 1114–1127. <http://dx.doi.org/10.1109/TCSVT.2008.927109>.
- Parker, M., Zegeer, C., 1989. Traffic Conflict Techniques for Safety and Operations – Observers Manual.
- Perkins, S., Harris, J., 1968. Traffic conflicts characteristics: accident potential at intersections. *Highway Res. Rec.* 225, 35–43.
- Rodríguez-Serrano, J.A., Singh, S., 2012. Trajectory clustering in CCTV traffic videos using probability product kernels with hidden Markov models. *Pattern Anal. Appl.* 15 (4), 415–426. <http://dx.doi.org/10.1007/s10044-012-0269-7>.
- Saunier, N., Sayed, T., 2006. A feature-based tracking algorithm for vehicles in intersections. In: *Canadian Conference on Computer and Robot Vision*. IEEE, Québec. <http://dx.doi.org/10.1109/CCRV.2006.3>.
- Saunier, N., Sayed, T., Lim, C., 2007. Probabilistic collision prediction for vision-based automated road safety analysis. In: *The 10th International IEEE Conference on Intelligent Transportation Systems*. IEEE, Seattle, pp. 872–878. <http://dx.doi.org/10.1109/ITSC.2007.4357793>.
- Saunier, N., Sayed, T., Ismail, K., 2010. Large-scale automated analysis of vehicle interactions and collisions. *Transport. Res. Rec.: J. Transport. Res. Board* 2147 (1), 42–50. <http://dx.doi.org/10.3141/2147-06>.
- Schreck, T., Bernard, J., Tekusova, T., Kohlhammer, J., 2008. Visual cluster analysis of trajectory data with interactive Kohonen Maps. In: *2008 IEEE Symposium on Visual Analytics Science and Technology*. IEEE. <http://dx.doi.org/10.1109/vast.2008.4677350>.
- Sivaraman, S., Trivedi, M.M., 2013. Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis. *2013 Intelligent Vehicles Symposium (IV)*, vol. 14. IEEE, pp. 1773–1795 (large review of tracking tech).
- St-Aubin, P., Saunier, N., Miranda-Moreno, L.F., Ismail, K., 2013. Use of computer vision data for detailed driver behavior analysis and trajectory interpretation at roundabouts. *Transport. Res. Rec.: J. Transport. Res. Board* 2389 (1), 65–77. <http://dx.doi.org/10.3141/2389-07>.
- St-Aubin, P., Miranda-Moreno, L.F., Saunier, N., 2014. Road user collision prediction using motion patterns applied to surrogate safety analysis, in: *Transportation Research Board-XXII # STOC*, Washington, D.C.
- St-Aubin, P., Saunier, N., Miranda-Moreno, L.F., 2015. Comparison of various time-to-collision prediction and aggregation methods for surrogate safety analysis. In: *Board, T.R., (Ed.), Transportation Research Board (TRB) 94th Annual Meeting, National Academy Of Sciences, Washington, DC*.

- St-Aubin, P., Saunier, N., Miranda-Moreno, L.F. 2015. Large-scale microscopic traffic behaviour and safety analysis of Québec roundabout design. In: Board, T.R. (Ed.), *Transportation Research Board (TRB) 94th Annual Meeting*, National Academy Of Sciences, Washington, DC.
- Svensson, Å., Hydén, C., 2006. Estimating the severity of safety related behaviour. *Accid. Anal. Prev.* 38 (2), 379–385. <http://dx.doi.org/10.1016/j.aap.2005.10.009>.
- Williams, M., 1981. Validity of the traffic conflicts technique. *Accident Anal. Prev.* 13 (2), 133–145. [http://dx.doi.org/10.1016/0001-4575\(81\)90025-7](http://dx.doi.org/10.1016/0001-4575(81)90025-7).
- Zheng, L., Ismail, K., Meng, X., 2014. Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes. *Saf. Sci.* 64, 155–162. <http://dx.doi.org/10.1016/j.ssci.2013.12.003>.