

Big Brother is Watching You... To Predict Crashes

Paul St-Aubin^a, Nicolas Saunier^b, Luis Miranda-Moreno^c

^a*Ecole Polytechnique de Montréal*

^b*Polytechnique Montreal*

^c*McGill University*

Abstract

The age of Big Data is here and many industries have already started embracing it. The transportation industry stands much to gain from large-scale data analysis due to the complexity and pervasiveness of transportation in daily life, which promises smarter roads and a better understanding of our transportation needs and environment. But this inertia is also one of the greatest challenges to big data adoption initiatives. Transitional technologies may, however, provide the answer to kick-start this migration today. This paper presents, in detail, a practical framework for implementation of an automated, high-resolution, video-based traffic-analysis system, particularly geared towards traffic flow modelling, behavioural studies, and road safety analysis. This system collects large amounts of microscopic traffic flow data from ordinary video cameras and provides the tools for studying basic traffic flow measures as well as more advanced, pro-active safety measures. This paper demonstrates the process step-by-step illustrated with examples and applies it to a case study of a large set of roundabout data. In addition to providing a rich set of behavioural data, the analysis suggests a relationship between flow ratio and safety, between lane arrangement and safety, and is inconclusive about the relationship between approach distance and safety.

1. Introduction

Affordable computing and flexible and inexpensive sensor technology are transforming the current practice and methods for traffic data collection, monitoring and analysis: big data is changing how we interact with our environment and approach problem solving tasks in the field of transportation. This should come to no surprise as the complexity of urban mobility complexity and the pervasiveness of geo-location devices in daily life lend themselves naturally to large data sets. 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. This 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 advanced methods for traffic data analysis; it is then not surprising that computer vision techniques have gained popularity given the potential of transforming the existing video-based traffic monitoring infrastructure into a highly detailed traffic data collection tool to identify and study traffic behaviours.

One such behavioural study application is in proactive road safety diagnosis. This has been a long-standing goal in the field of transportation safety. Traditional statistical methods applied to accident data require long observation periods (years of crash data): one must wait for (enough) accidents to occur. Beginning in the 1960s, attempts were made to predict collision rates based on observations without a collision rather than historical accident

records [1]: these methods are now termed surrogate safety methods. The traffic conflict technique was one of the earliest methods proposed which entailed the observation of 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: the manual data collection method is costly and may not be reliable, and the definition and objective measurement of these events are difficult [2].

Today, with technological improvements in computing power, data storage, sensor technologies, and advances in artificial intelligence, these issues are quickly 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 [3] and machine learning methods to model and predict collision potential 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 roadway [4]. This trajectory data consists of position and velocity measurements of road users captured 15 to 30 times per second to a relatively high degree of accuracy. This amounts to several million individual instantaneous measurements over the period of one day at a typical site.

This high-resolution data permits the measurement of precisely defined instantaneous surrogate safety measures identifying collision probability. One such measure is time-to-collision (TTC) which measures the time remaining at any given instant to some collision point in the future defined by a collision course with another road user. This measure is useful as it provides the remaining time road users have to react to and avoid potential collisions.

Higher TTCs are generally considered safer, though the precise link has yet to be validated. However, this measure relies on motion prediction hypotheses to identify collision courses. The traditional approach is to use constant velocity projection [5] [6] (situations in which road users fail to correct their course or even turn), which is the motion prediction method most frequently used, without any justification. This approach does not natively provide a collision course probability, and it will not be suitable in situations where observed trajectories do not include constant velocity displacements: for example, turning lanes in an intersection and movements in a roundabout.

More advanced collision course modelling efforts are being developed, including motion patterns which represent naturalistic (expected) driving behaviour learnt from the same data set. This procedure provides several potential collision points and their probability as a function of both the characteristics of the specific site and the road users' behaviour. The motion patterns, or the distribution of trajectories at a site and their probabilities, may be described discretely over time and space [7] or with prototype trajectories [8]. The motion and collision predictions are computationally intensive as they explore, for each pair of road users, at each point in time, all future positions in time and space (typically subject to a time horizon). Furthermore, interaction complexity and exposure increase exponentially as the number of simultaneous road users in a scene increases. For example, over the course of one day, a typical intersection can experience between 100 thousands and 100 millions of these instantaneous interactions, depending on the intersection complexity.

This paper presents a complete automated system for proactive road

safety analysis that can deal with large amounts of video data. To the authors' knowledge, the presented system is the most comprehensive to be applied to such big data collected in the field for a real world traffic engineering study. A large video dataset was collected at more than 20 roundabouts in Québec to study road user behaviour and their safety. Camera views record data at more than 40 roundabout weaving zones, an area within the roundabout delimited by an entry and the next following exit. Each camera records 12 to 16 h of video on a given work day, which constitutes a dataset of over 600 hours of video data. Applying the proposed method to this large dataset yields considerable amounts of indicators, from individual road user measurements, e.g. speed, to individual interaction measurements, e.g. TTC, to aggregated indicators per road user or interaction, to aggregated indicators per site over time and space.

Analysing such big data is a challenge of a magnitude that has never been undertaken before in driver behaviour and road safety research. It holds the key to understanding the processes that lead road users to collide, and to design and validate safety indicators that do not require accidents to occur. The approach will be demonstrated on this video dataset to identify roundabout characteristics that influence road safety.

The paper is organized as follows: the next section presents the methodology, with practical examples drawn from the roundabout dataset, which is then applied to about half of the collected data and various system outputs are presented, before the conclusion and discussion of future work.

2. Methodology

2.1. Overview

Figure 1 outlines the general data collection and analysis framework. For a given research mandate, factors are selected for testing and a set of video data is collected at a sample of sites with adequate representation of these factors, while controlling for as many other factors as possible. With scene data and camera calibration parameters, feature tracking can be performed to extract trajectories [9]. The trajectories are raw spatial-temporal position data of moving objects within the scene. This positional data is processed to obtain derived measures such as speed, heading and acceleration. Finally, scene information can be added to obtain higher-level data, such as movements referenced by lane, conflict measures, and other high-level interpretation behavioural measures (specific to the study). With a large amount of potential contributing factors (e.g. site characteristics), it may be beneficial to apply site clustering techniques before initiating behavioural measure correlation.

2.2. Video Data

Road user trajectories are extracted from video data using a feature-based tracking algorithm described in [9] and implemented in the open source project Traffic Intelligence¹.

¹<https://bitbucket.org/Nicolas/trafficintelligence/>

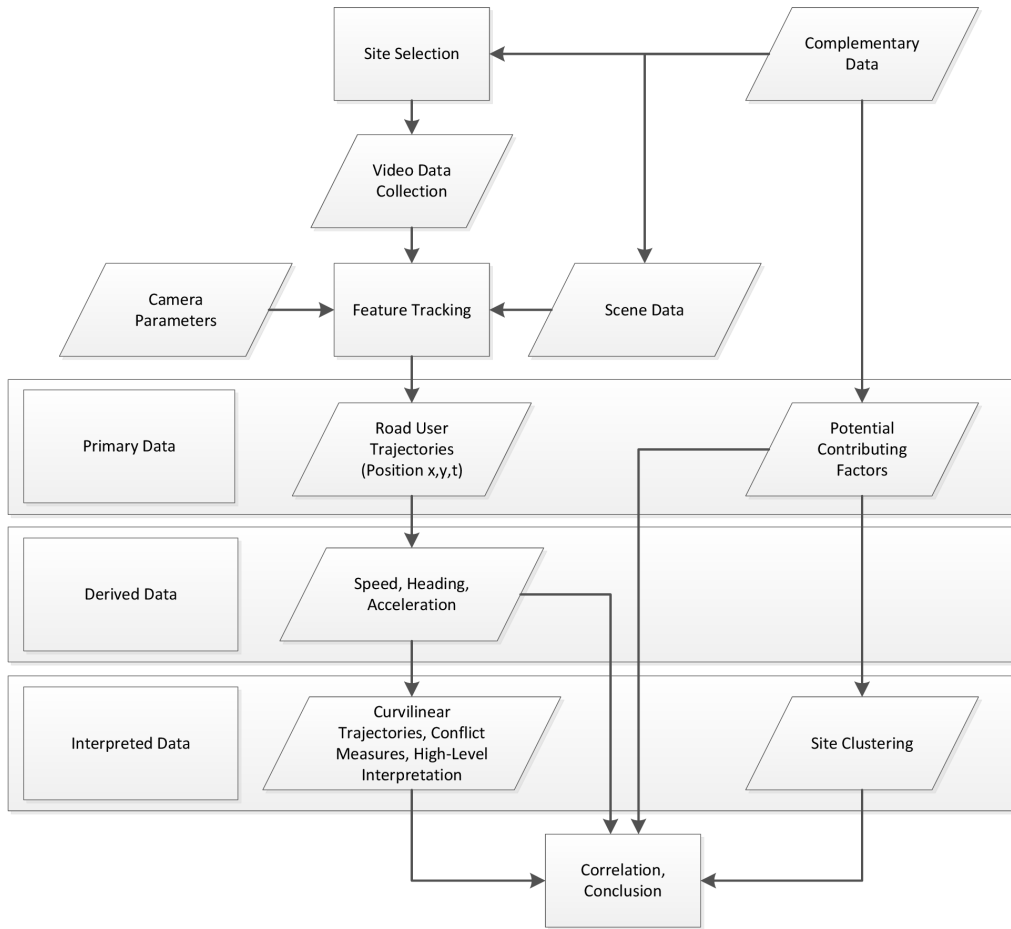


Figure 1: Data flow diagram showing the overview of the system.

2.2.1. Trajectories: Positions in Space and Time (x,y,t)

Trajectories are a series of points in Cartesian space representing the position of (the centroid of) a moving object (road user) at time t on a planar surface. Height z is usually not considered. Points are evenly spaced in time with a consistent Δt equivalent to the inverse of the framerate of the video, i.e. a measurement is done for each frame. Typical framerates for

video are between 15 to 30 frames per second, providing 15 to 30 observations per moving object per second. The object (road user) itself is represented by a group of characteristic features spread over the object and moving in unison.

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

- **Parallax error** is mitigated by maximising the subtending angle between the camera and the height of tracked objects. In practical terms this requires a high view or ideally a bird's eye view, tracking objects with a small height to base ratio. Passenger cars are generally more forgiving in this respect than trucks 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 study areas nearer to the camera and using high-resolution cameras, although increases in resolution offer diminishing returns of tracking distance.
- Finally, **tracking errors** may occur due to scene visibility issues or limits with current computer vision techniques, in particular to handle data association (e.g. attach the trajectories to the right objects when they occlude each other). These erroneous observations have to be rejected or reviewed manually.

Depending on the steps taken to minimize tracking errors, feature-based

tracking functions best over study areas of 50-100 m in length with high-to-medium speed, low-to-medium density flows.

A sample of road user trajectories is presented as they are tracked in image space in Figure 2. For more information on computer vision, see section ??.

2.2.2. Derived Data: Velocity & Acceleration

Velocity and acceleration measures are derived through differentiation from position and velocity over time respectively. These are 2-dimensional vectors with a magnitude (speed and acceleration) and a heading.

It should be noted however that each successive derivation increases pixel precession error for that measure. A velocity measure requires twice as many pixels as a position measurement. Similarly, an acceleration measurement requires three times as many pixels as a position measurement. This type of error can be compensated for with moving average smoothing over a short window (e.g. 5 frames). At this time, acceleration measurements are still too noisy to be useful for instantaneous observations. Higher camera resolutions should solve this problem in future applications.

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

$$n = fQd \tag{1}$$

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.

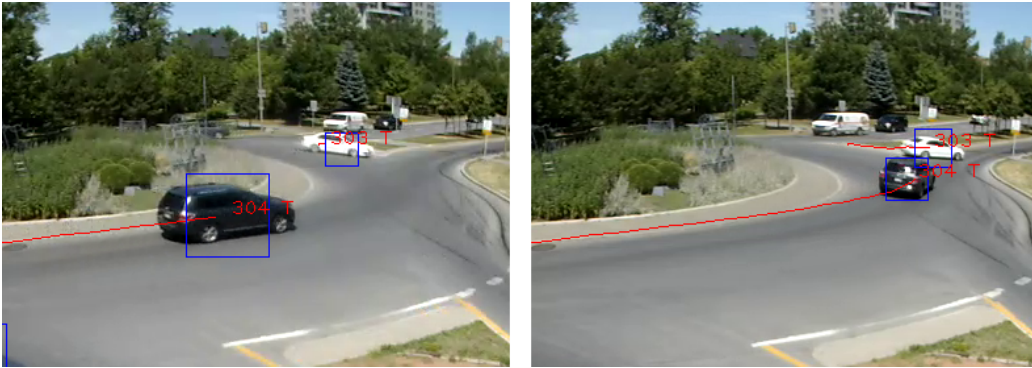


Figure 2: Vehicle #304 is shown approaching vehicle #303 which is engaging the roundabout in the wrong direction demonstrating a frequent violation leading to a traffic conflict.

2.3. Complementary Data

With the exception of speed and vehicle counts, vehicle trajectories offer little insight without context. Complementary data about the scene is collected in order to perform traffic studies and for higher-level interpretation. This data includes a wide variety of scene descriptors and design geometry attributes characterizing the factors under study. 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 include number of lanes, lane width, horizontal and vertical signalisation, pedes-

trian facilities, the built environment, and upstream/downstream distances to other intersections.

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 areas of the image with unsatisfactory feature tracking (particularly at the edges of the video), and ii) confine analysis to a particular region. For a cross-sectional or before-after study, analysis areas should conform to the same region of the roadway as much as possible. An example of the analysis area is demonstrated in Figure 3.

2.3.2. Alignments

Trajectory clustering is an important preliminary step in scene interpretation. Trajectory clustering is an abstract representation of movements along prototypical paths through a scene, called alignments. This is the foundation for relating spatial position with road geometry and, in particular, position of moving objects in relation to lanes and sidewalks. The alignment is represented as a simple series of points with a beginning and an end, typically in the same direction as the majority of flows 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

$$(x, y) \rightarrow (l, s, \gamma). \tag{2}$$

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 , 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 window of time less than the time users take to perform real lane changes to correct any localised lane "jumping" errors which frequently appear near converging or diverging alignments. These coordinates are useful for studying following behaviour, lane changes, and lane deflection.

Many approaches exist to trajectory clustering: while some methods are supervised, many more are unsupervised (e.g. k-means [10]). Manual trajectory clustering is labour intensive and potentially a source of bias, but 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 spatial positioning of the manually defined alignments through traditional unsupervised clustering approaches, is considered for future improvements.

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 Figure 3, movements may involve multiple lane changes and therefore may require a more general

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. Alternatively, alignments which run in parallel over a distance of more than 15 metres 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 of this network. This serves to reduce processing times of spatial relationship calculations between objects (triage) and provides more intelligent interpretation of spatial relationships.

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. 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, and will instead focus on generalizable measures. However, an application of HLI calculations will be briefly presented in section 3.3.

2.4.2. Interactions

An interaction quantifies the spatial relationship between moving objects in a scene, as is depicted in Figure 2. At the most fundamental level, an

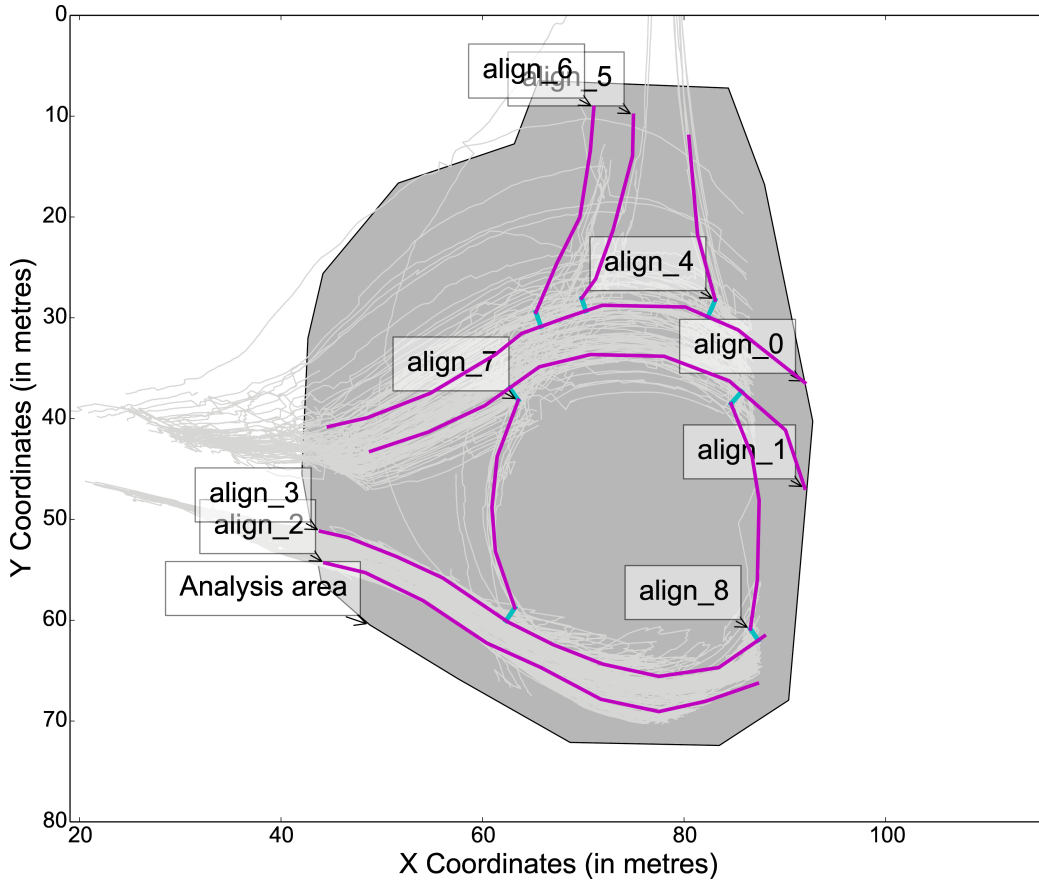


Figure 3: The partial trajectories and scene of a multi-lane roundabout with a complex configuration of lanes (the south and east approaches are not visible). The alignments are in pink, while the connectors are in cyan. Some sample trajectories are highlighted in light grey.

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.

This interaction definition is a generic precondition for any safety-related

event of interest. In many scenes, it will include events of widely varying relationship to safety. For example, the significance of an interaction between two vehicles separated from each other physically (e.g. via a median or a large building) may not be comparable to an interaction between two vehicles merely separated by a painted line because the probability that one of the vehicles comes into contact with the other vehicle is reduced in the case of the median. This may cause issues when comparing different scenes if the analysis areas are not drawn consistently and may increase the computational burden of with collision prediction.

One solution is to filter 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 isn't necessarily a binary option. In our median example, it is still physically possible, although less likely, for vehicles to cross-over into an opposing lane and cause a collision, although this is something that could be modelled.

2.4.3. Motion Prediction

Safety is evaluated from the observations of all vehicle interactions, by predicting future positions to determine if they are on a collision course and to characterize that collision course. The potential for collision of all interactions is measured by predicting future positions of vehicles at every instant in time and examining i) situations of particular probability of collision (i.e. threshold) or ii) evolution of the probability of collision over a time series. Several motion prediction methods are proposed for study [11]:

- **Constant velocity** is the classic motion prediction model, wherein

vehicles are projected along straight paths at a constant speed and heading using the velocity vector at that moment in time. This model is the simplest but also makes the most assumptions: only one movement is predicted at every instant, it does not depend on the context (road geometry or traffic), and the natural (non-reacting) motion of a moving object is a straight path (not always true). These assumptions may be adequate for specific applications of the methodology, e.g. highways [12]. The current implementation is based off of [6].

- **Normal adaptation** uses the initial velocity vector at the prediction moment to project trajectories, but modifies the velocity vector to account for normal variation. This model benefits from a wider range of possible outcome velocity vectors, but otherwise suffers the same problems and makes the same assumptions as constant velocity. The implementation of normal adaptation studied is based off of [11], using a acceleration maximum α of

$$\alpha = \pm \frac{2}{f^2} \tag{3}$$

and a maximum steering parameter σ of

$$\sigma = \frac{0.2}{f} \tag{4}$$

where f is the number of frames per second of the video.

- **Motion patterns** are a family of models which use machine learning to calculate future position likelihoods from past behaviour [8, 13]. This

type of model is the most promising as motion prediction is probabilistic in nature and inherently models naturalistic behaviour. However, motion patterns are complex to implement and expensive to process. The type of motion pattern being studied for implementation is a discretized motion pattern [7].

As illustrated in Figure 4, motion prediction is performed for each user pair over each interaction instant t_0 for a number of time steps of size Δt between t_0 and t_0 plus 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.

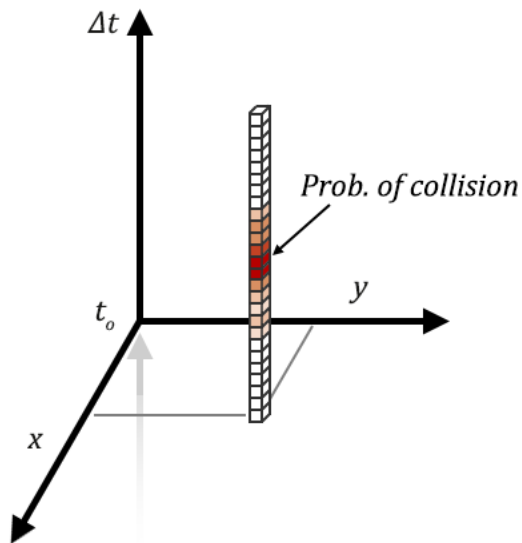


Figure 4: Collision prediction space in (x, y, t) over Δt steps based on the conditions at $t = t_0$.

2.4.4. *Time-to-collision*

Time-to-collision (TTC) 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 collide, if they collide, based on the motion prediction model. In the simplest form, e.g. constant velocity, time-to-collision is the ratio of differential velocity and differential position. A TTC value of 0 seconds 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 and breaking time. Larger values of observed TTC thus provide greater factors of safety for these driving tasks.

Time-to-collision is measured instantaneously: a new value of TTC may be computed for 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 [14]. Other approaches have focused on quantile or threshold observations (i.e. counting the number of interactions with minimum TTC below a threshold as in classical traffic conflict techniques [15]), or even to examine instantaneous risk and significance of TTC [12].

A sample pair of road user trajectories (#303 and #304, Figure 2) and spatial relationships simultaneously existing over a time interval lasting 64 instants or just over 4 seconds is presented in Figure 5. In this scenario, vehicle #304 is approaching at high velocity vehicle #303 which is engaged in an illegal U-turn (in a right-hand roundabout, users are supposed to travel counter-clockwise around the centre island at all times). The differential velocity Δv , relative distance d , and corresponding time t is measured for

every instant. In a matter of just under 4 seconds, 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. These predicted collisions and associated TTC measures are presented in Figure 6. Motion pattern prediction generates many more possible collision points than constant velocity prediction, though each of these points has a lower associated probability. When several potential collision points are predicted, the expected TTC $ETTC_i$ at time t_i is calculated as the probability-weighted TTC average

$$ETTC_i = \frac{\sum_{j=1}^m TTC_{ij} Prob(collision)_{ij}}{n} \quad (5)$$

of all possible collision points indexed $j = 1..m$ that could be reached with probability $Prob(collision)_{ij}$ [7].

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 they are on a collision course only for a brief period of time with these prediction methods.

2.4.5. Post-encroachment time

While prediction models and TTCs relate to the collision potential, other surrogate safety measures aim to measure collision proximity from crossing, but not necessarily colliding movements. Trajectory data is detailed enough to provide gap acceptance time (GT) and post-encroachment times (PET). These are measures that broadly characterise how aggressively and close in

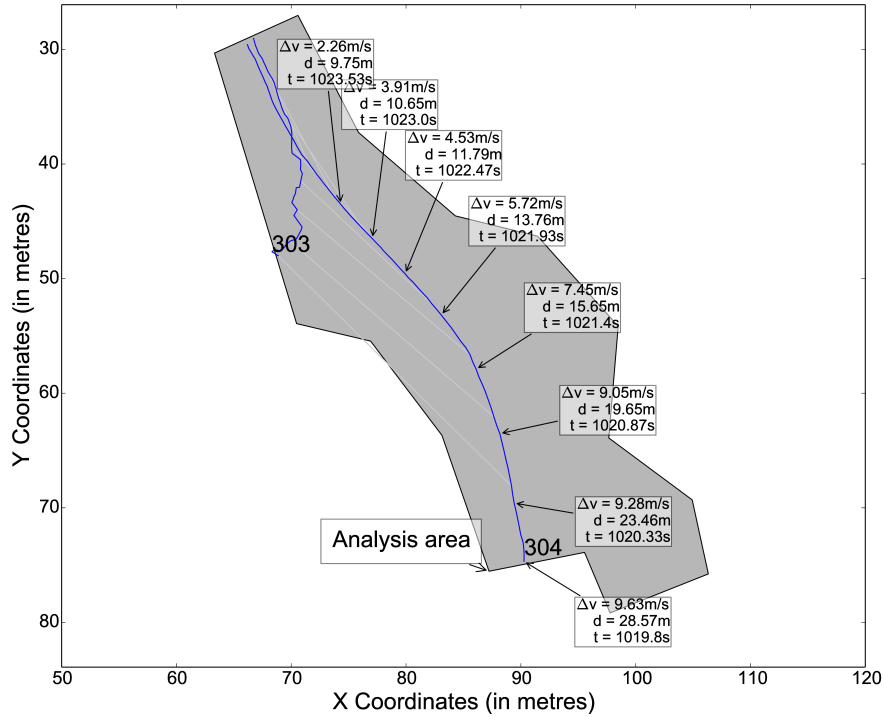


Figure 5: Vehicle #304 is shown approaching vehicle #303 which is engaging the roundabout in the wrong direction. Spatial relationship measures ΔV , relative distance d , and time stamp t are labelled along the time series every eight frames between the two trajectories. Light grey lines join the two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

space and time merging and crossing tasks, respectively, are performed. As such, there is generally only one of these measures for the entire common time interval of a pair of road users. Gap acceptance time and PET fall under the category of high-level interpretation measures as the calculation of these measures cannot be generalised for all traffic studies, in part because

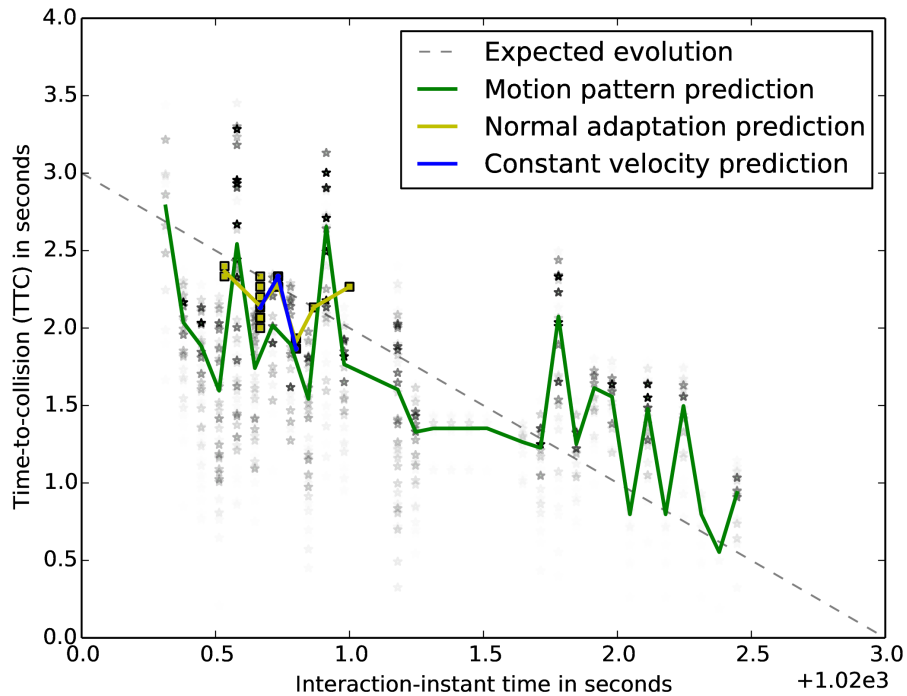


Figure 6: Time series of TTC observations for different motion prediction methods for the interaction between vehicles #303 and #304. Points correspond to TTC for a specific collision point and lines are weighted average observations per instant. The expected evolution of the timeseries occurs with a slope of one second to one second when the TTC observation at a given instant holds true, i.e. vehicles do not correct their collision course and a collision ensues.

the behaviour does not apply to all types of traffic interactions, and, in the case of gap acceptance time, because the measuring method may vary from one type of geometry to another.

For the crossing zone defined by the intersection of the two trajectories of a pair of road users, the post-encroachment time measures the time between

complete departure of the first arriving vehicle, and first arrival of the next arriving vehicle. If $PET = 0$, a collision has happened. As such, higher PETs should demonstrate safer behaviour, although not necessarily linearly. An alternative to PET is predicted PET (pPET) which is measured from motion prediction instead of direct observation [11].

Gap acceptance time similarly measures arrival and departure of a road user at a common crossing zone, but in this case, the crossing zone occurs in-line during a merging task, usually followed by following behaviour.

2.5. Indicator aggregation over time and space

Instantaneous surrogate safety indicators may be aggregated over time for each interaction (or user pair), over a given time interval for several road users and over space. Indicator distributions are generally shaped like Gamma distributions across the literature [16, 17, 12]. Quantifying collision risk based on any of the surrogate safety indicators is the remaining puzzle piece. Using a TTC threshold has been the traditional approach in traffic conflict techniques [15], correlating a number of interactions with minimum TTC below a threshold with an expected number of collisions, though this constitutes a significant loss of information [14] and this introduces assumptions in the model. One recent approach proposed a shifted gamma-generalised Pareto distribution model [18].

Nevertheless, some qualitative analysis is possible in some circumstances, for example with a continuous mass shift of a probability distribution function as demonstrated in Figure 7. This approach has been tried in some early applications of the methodology, e.g. in [16, 17, 12]. Figure 8 demonstrates three different TTC distribution aggregation methods as used to represent

nearly 3 million TTC observations over the course of one day at a single site: i) all instantaneous indicator values (subject to over-sampling of low severity values as well as over-sampling by slower road users and longer corridors), ii) minimum value of time series per user pair, or iii) 15th percentile value of time series per user pair. The 15th percentile is a practical solution to ignoring outliers that influence the maxima.

3. Experimental Results

3.1. Data Size

Video was collected at 20 roundabouts using two types of camera, a security camera VIVOTEK with a narrow lens filming at 15 frames per second at a resolution of 800*600 and a consumer camera GoPro 2 with a wide-angle lens filming at 30 frames per second at a resolution of 1280*960. The cameras are mounted on a specially constructed mobile video-data collection system built for temporary, high-angle video data collection, with tamper-proof, weather-proof, self-contained features presented in [4].

In these 20 roundabouts, video was recorded for 40 merging zones of varying lane configuration, geometry, land use, and traffic volumes across the province of Québec. The merging zone of the roundabout is defined as the portion of the ring intersected by an approach and an exit. There is generally one merging zone between every pair of adjacent branches. Video data at each site was taken on one mild summer workday from 6 AM to 7 PM or 10 PM and captures both peak traffic hours [12]. This yields a total of 600 hours of video data, 50 % of which has been fully processed at this time (see Table 1 for more information).

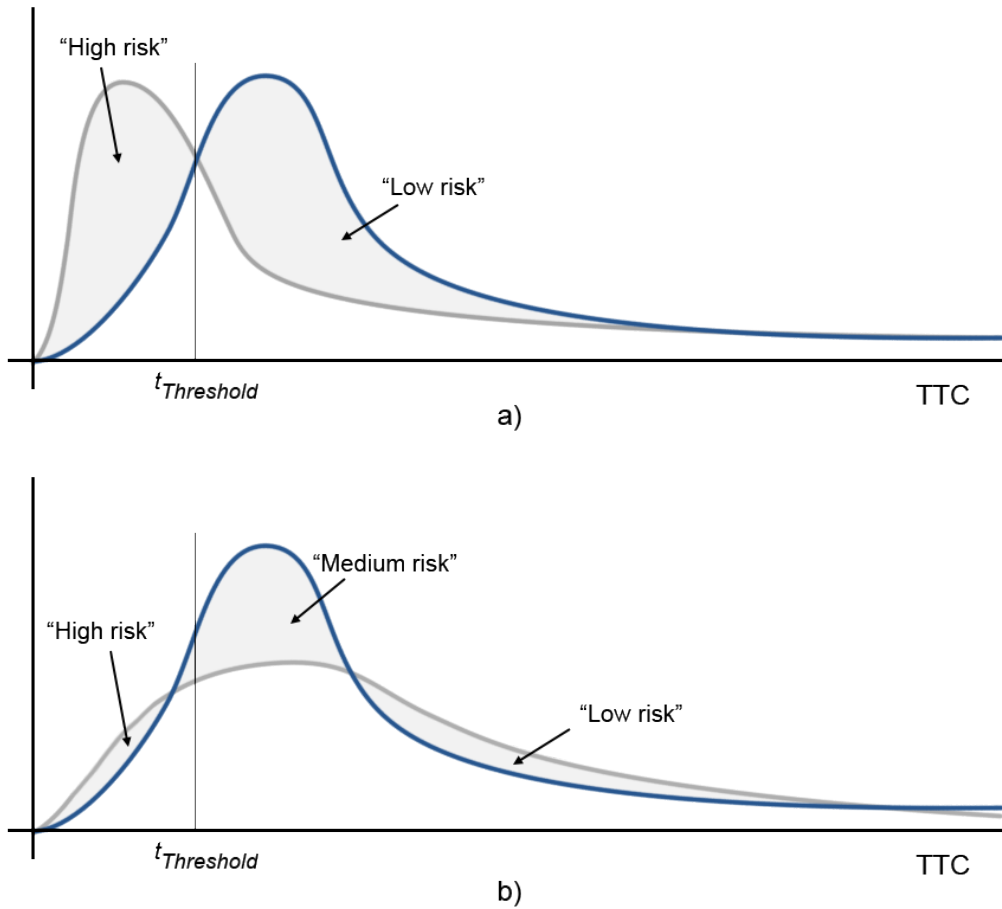


Figure 7: TTC probability distribution function comparisons. Qualitative analysis of TTC shift is possible in a), while it is not as straightforward in b). An alternative approach is to perform a comparison on a low threshold.

The software used is the open-source Traffic Intelligence project [3, 4], itself based on the computer vision platform OpenCV [19]. This software provides the basic feature tracking (the algorithm presented in [9]), trajectory management and coordinate projection functionality as well as a few usage-

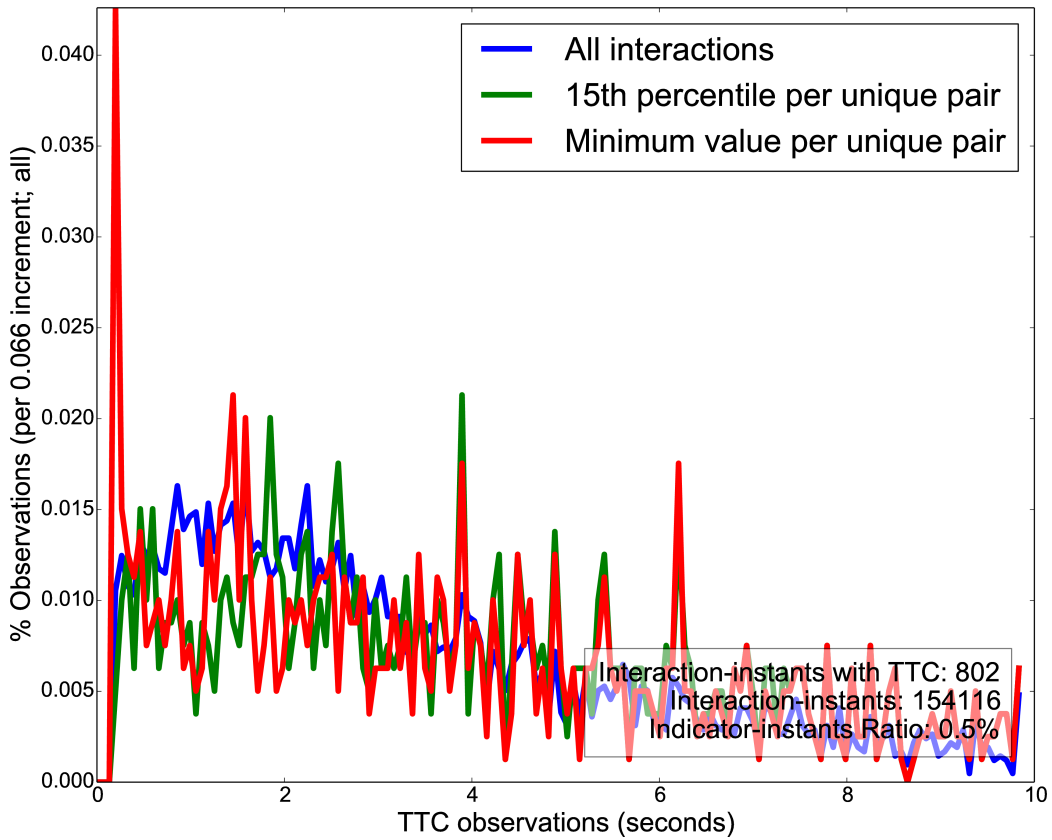


Figure 8: A sample distribution of the same TTC measures observed over the course of a day at a single site illustrating three different aggregation options.

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.

At 30 frames per second, a data collection at an intersection over a period of 12 hours (e.g. 7 AM to 7 PM), over a driving distance of 50 metres and at an average driving speed of 30 km/h, and with an average hourly vol-

Table 1: Data details

Roundabouts	20
Analysis Areas	41
Hours of Video Data	610
Estimated Total Traffic Volume	120,000
Disk Space (Video + Data + Overhead)	1.9 TB
Veh-km Traveled	8400 veh-km
Processed to Date	$\approx 50 \%$

ume of 500 veh/h yields approximately 90,000 instantaneous moving object measurements per hour. Additionally, each of these observations can have anywhere between 3 to 100 feature tracks associated with it. The recommended number of features to aim for is roughly 15-20 per object over time, depending on the typical duration of 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, framerate, and video encoding settings.

Figure 9 shows hourly number of user pairs observed versus traffic volume. Trends are evident, but contributing 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 Figure 10. If they are not, it is possible that analysis areas across sites are not comparable, particularly for time-series analysis and aggregated TTCs.

Most of the analysis is conducted on a pair of dedicated consumer-grade high-performance machines (INTEL Core i7 3770k processor with 16 to 32 GB of memory), with parallelisation of some tasks and work offloaded to a computing cluster when acceleration is necessary. Feature tracking performance depends on video resolution and special post-processing requirements such as stabilisation or lens correction (for distortion). A typical one hour 800×600 video 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 minutes and 30 minutes, depending on traffic in the scene, while surrogate safety analysis, particularly motion patterns, is very sensitive to the interaction complexity of the scene and can typically take anywhere between 1 to 48 hours to complete.

3.2. Sample Surrogate Safety Analysis

A sample surrogate safety analysis of three of the sites is demonstrated in Figure 11. This shows trajectory tracks projected in and with respect to the scene, mean speed and heading, and spatial distribution of motion-pattern-predicted collision points with instantaneous *probability* $> 10^{-5}$ and *TTC* < 1.5 s.

Figure 12 demonstrates a cross-sectional comparison of TTC distributions based on motion prediction at constant velocity for 20 merging zones for two contributing factors, each using all interaction instants. These distributions are aggregated directly from all TTC observations, they are not means of the distributions at each site. Kolmogorov-Smirnov tests are performed between the distributions to quantify non-parametric dissimilitude.

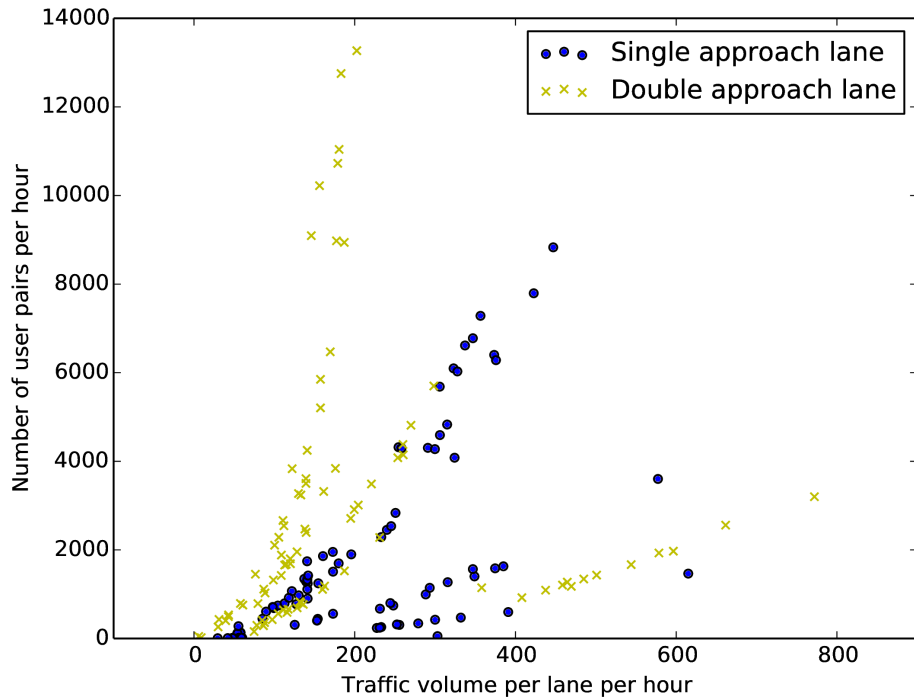


Figure 9: Number of user pairs observed at each roundabout versus inflow per lane per hour.

In the first diagram, a cross-sectional comparison is made for merging zones situated nearer or further than 300 metres upstream from another intersection. When this distance exceeds 300 metres, the distribution mass appears to shift left except for a sharp increase in small TTC below 0.5 seconds. It is so far unknown whether this small concentration of low-TTC conflicts offsets all other increases in TTC. This comparison remains therefore inconclusive. In the second diagram, a cross-sectional comparison is made between merging zones with high approach traffic volume ratios and low approach traffic

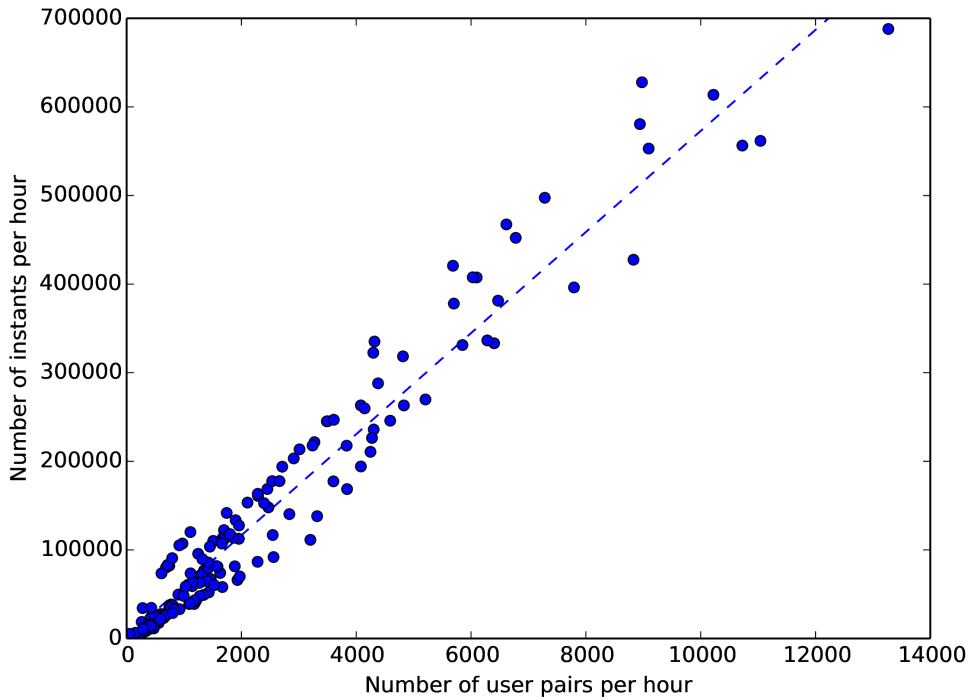


Figure 10: Number of interaction instants observed versus number of user pairs observed.

volume ratios, where R is the flow ratio between approach volumes and total volumes at the merging zone. In this comparison, a clear and consistent mass shift is observed, suggesting that high approach traffic volume ratios contribute to safer merging behaviour in a roundabout.

3.3. Sample High-Level Interpretation Analysis

Some high-level interpretation measures are also compiled using the data sample (for the same 20 merging zones). Figure 13 shows the mean speed profiles, with the interval at mean \pm one standard deviation, through the roundabout merging zone of the same 20 samples as previously used for sur-

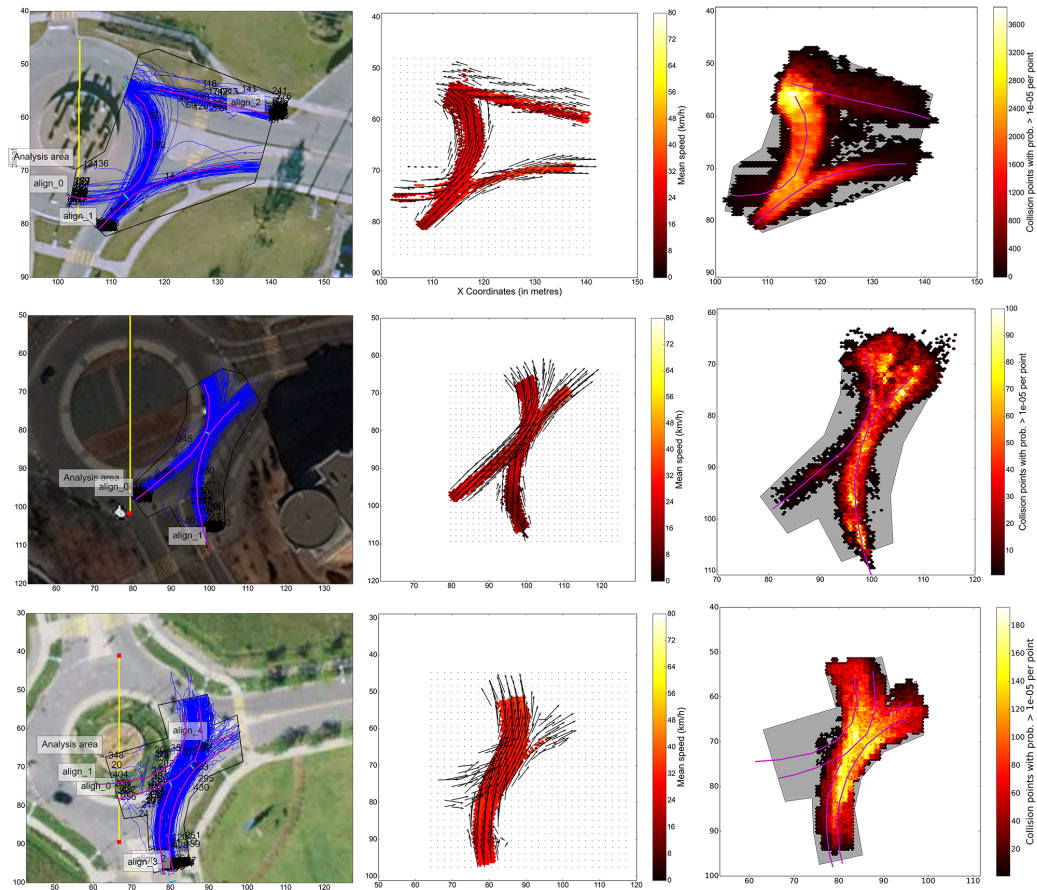


Figure 11: Sample spatial data and analysis at 3 selected sites from top to bottom. From left to right, diagrams demonstrate trajectory tracks (positions) in analysis area and with descriptive alignments, mean speed and heading on a regular grid, and spatial distribution of motion-pattern-predicted collision points with instantaneous *probability* $> 10^{-5}$ and *TTC* < 1.5 s. All coordinates in metres, north pointing upwards.

rogate safety analysis. Speed profiles are mapped, not as a unit of distance, but rather as a unit of curvilinear location relative to the start and end of the merging zone. This is done to account for the large variability in diameter

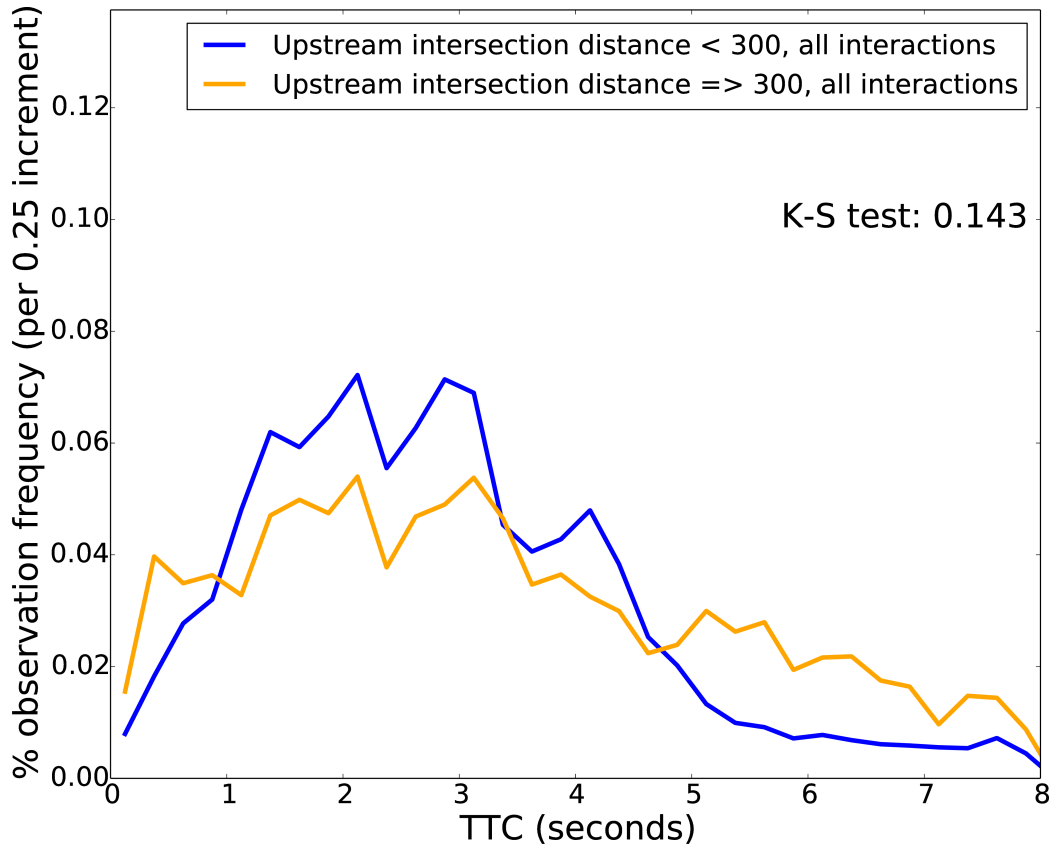


Figure 12: TTC distributions based on motion prediction at constant velocity across 20 merging zones for two testable factors, using all interaction instants: a) for the upstream merging distance, results are difficult to interpret in terms of safety, and b) for the flow ratio of approach/total flows R , results suggest that merging zones where the approach accounts for the majority of flows are safer.

of roundabouts and in the angle between successive approaches across the sites. The position measurement re-sampling method as described in section 2.2.3 is used here to correct for oversampling bias introduced from varying speed between road users. Mean speeds are generally consistent with those

in the literature, but variation does occur by relative location and movement type. In addition, cross-sectional analysis as in Figure 13 uncovers even larger variations in mean speed profiles (not shown).

Finally, Figure 14 shows distributions of accepted gap times of approaching vehicles and corresponding roundabout vehicles at the same sites. 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 flow inside the roundabout and long wait times at the approach. Figure 15 demonstrates platoon sizes (uninterrupted passage of sequential vehicles). Users already inside the roundabout are generally more clustered than users entering the roundabout from the approach.

4. Conclusion

This paper demonstrates the theory and practical application of large-scale, automated, proactive road safety analysis using computer vision. The reader is led step-by-step through the challenges and process of collecting, processing, and analysing video data with examples along the way. It demonstrates an early implementation in the form of a spatial and cross-sectional analysis using a large data set of roundabout video data to test several contributing factors. In addition to a rich set of behavioural data, the analysis suggests a relationship between flow ratio and safety, between lane arrangement and safety, and is inconclusive about the relationship between approach distance and safety.

Several technical challenges were outlined, notably tracking error, quan-

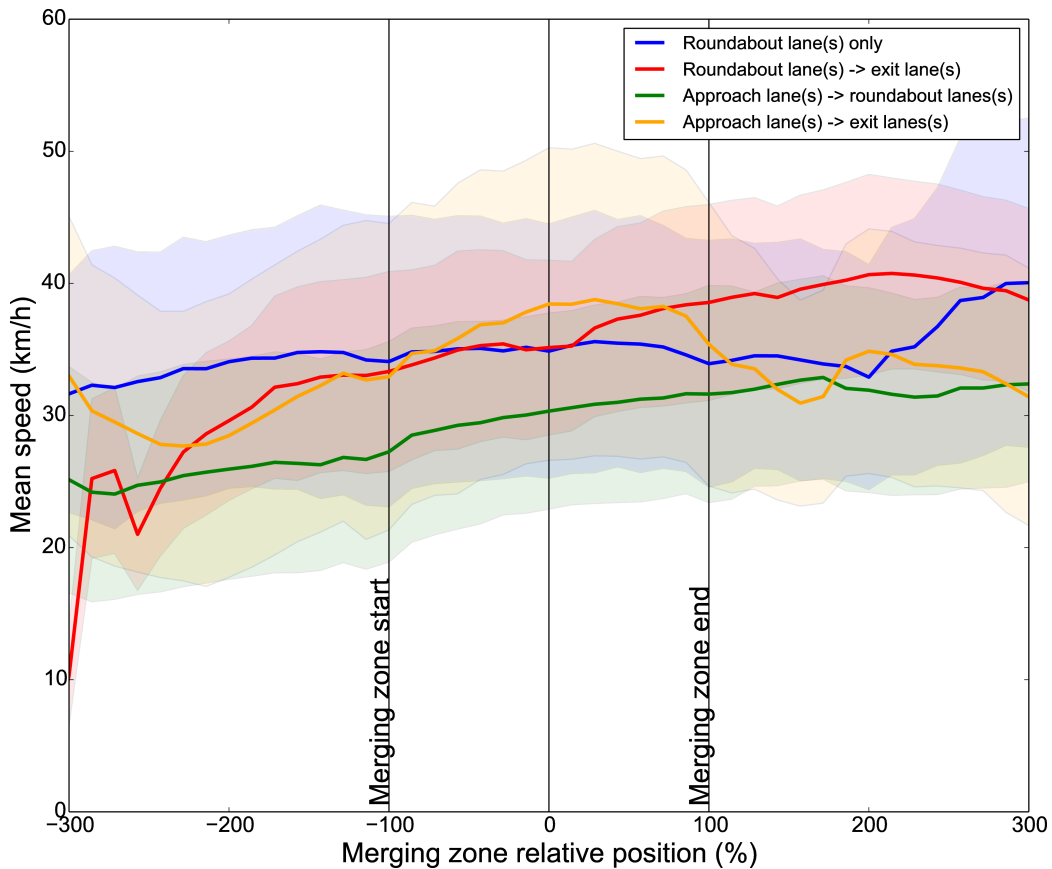


Figure 13: Mean and interval at mean \pm one standard deviation speed profiles relative to merging zone start and end according to movement type.

tified probability of collision from TTCs, and aggregation and sampling considerations, as they still require particular attention. It is expected that these issues will be addressed as processing and analysis tools become more accessible, more collaborators contribute solutions to the open-source software, and as techniques applied to transportation issues become more sophisticated.

The full results of the study over all 600 hours of video data will be the subject and focus of future papers. More advanced tracking, error detection,

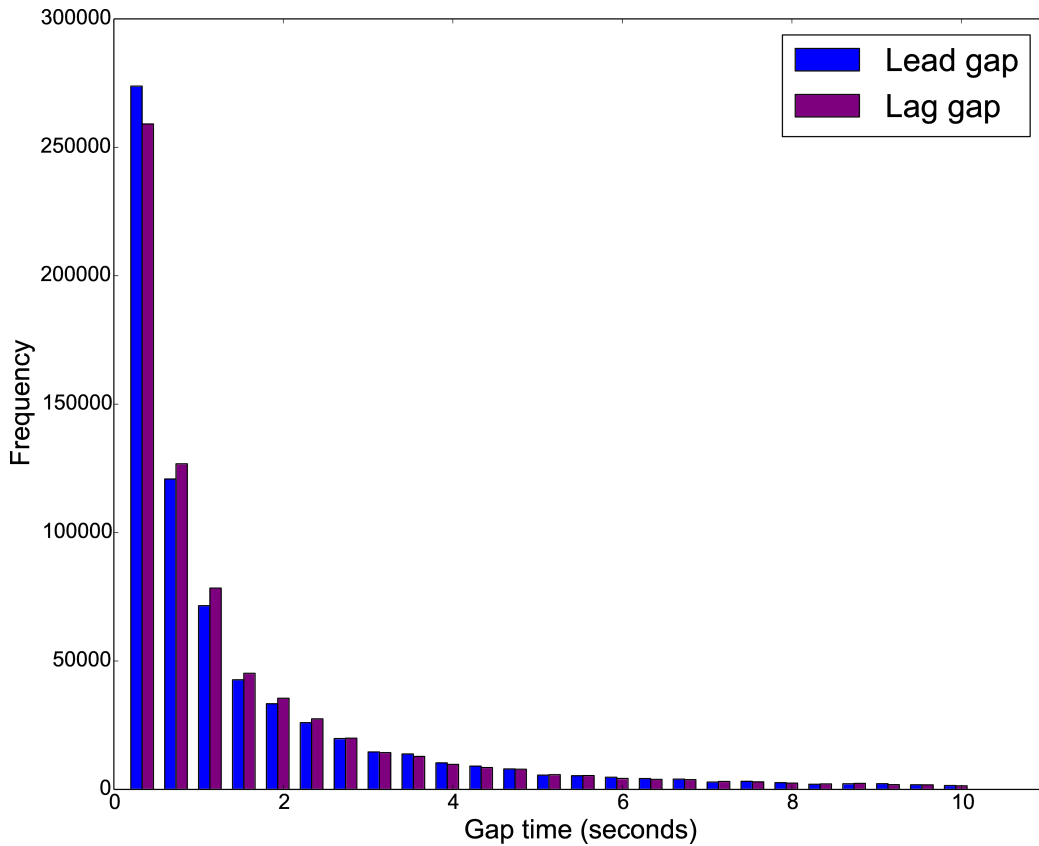


Figure 14: Accepted gap time at merging instant of approaching vehicles.

motion prediction models, and trajectory clustering will also be the subject of further research.

5. 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 (proposal number 2012-SO-163493),

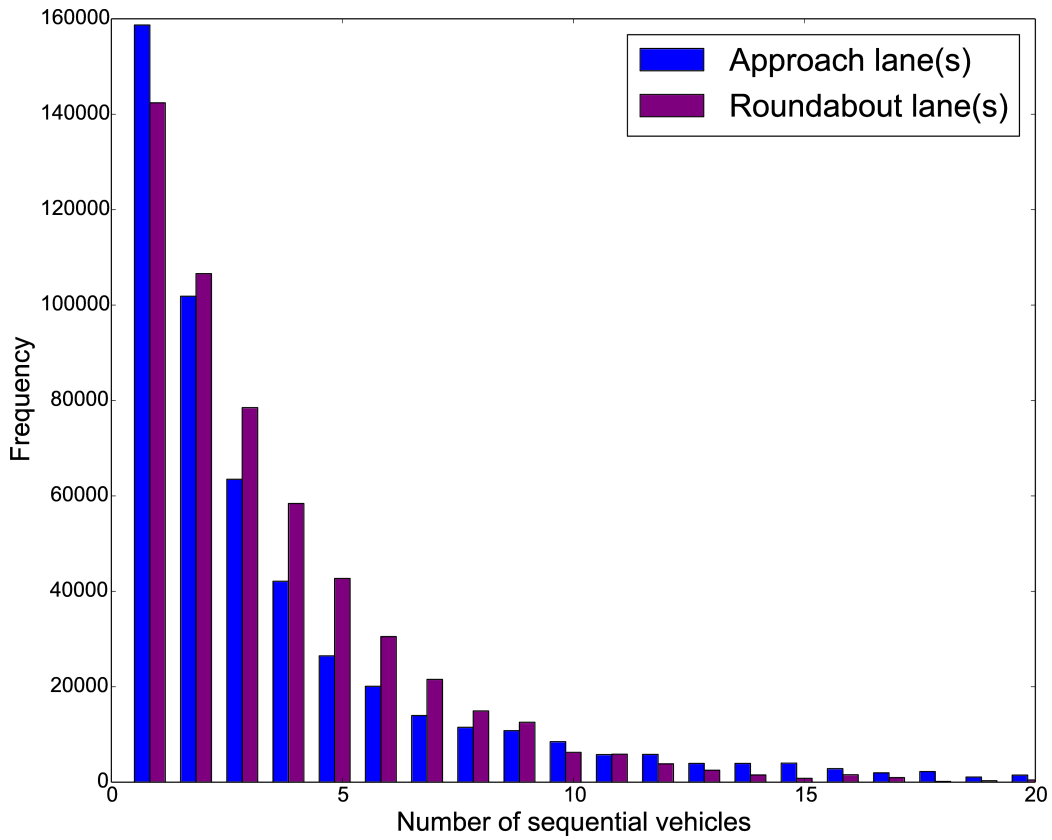


Figure 15: Platoon size (uninterrupted sequential flow) comparison between roundabout lanes and approach lanes.

as well as the varying 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

- [1] S. Perkins, J. Harris, Traffic conflicts characteristics: Accident potential at intersections, Highway Research Record 225 (1968) 35–43.
- [2] H.-C. Chin, S.-T. Quek, Measurement of traffic conflicts, Safety Science 26 (3) (1997) 169–185. doi:

10.1016/s0925-7535(97)00041-6.

URL [http://dx.doi.org/10.1016/s0925-7535\(97\)00041-6](http://dx.doi.org/10.1016/s0925-7535(97)00041-6)

- [3] N. Saunier, T. Sayed, K. Ismail, Large-Scale Automated Analysis of Vehicle Interactions and Collisions, *Transportation Research Record: Journal of the Transportation Research Board* 2147 (-1) (2010) 42–50. doi:10.3141/2147-06.
URL <http://dx.doi.org/10.3141/2147-06>
- [4] S. Jackson, L. F. Miranda-Moreno, P. St-Aubin, N. Saunier, Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioral Analysis, *Transportation Research Record: Journal of the Transportation Research Board* 2365 (-1) (2013) 90–98. doi:10.3141/2365-12.
URL <http://dx.doi.org/10.3141/2365-12>
- [5] F. Amundsen, C. Hydén, Proceedings of the first workshop on traffic conflicts, in: *Institute of Transport Economics, 1977, the famous unchanging TTC quote.*
- [6] A. Laureshyn, Å. Svensson, C. Hydén, Evaluation of traffic safety, based on micro-level behavioural data: Theoretical framework and first implementation, *Accident Analysis and Prevention* 42 (6) (2010) 1637–1646. doi:10.1016/j.aap.2010.03.021.
URL <http://dx.doi.org/10.1016/j.aap.2010.03.021>
- [7] P. St-Aubin, L. F. Miranda-Moreno, N. Saunier, Road User Collision Prediction Using Motion Patterns Applied to Surrogate Safety Analysis,, in: *Transportation Research Board Annual Meeting, Washington, D.C., 2014.*
- [8] N. Saunier, T. Sayed, C. Lim, Probabilistic Collision Prediction for Vision-Based Automated Road Safety Analysis, in: *The 10th International IEEE Conference on Intelligent Transportation Systems, IEEE, Seattle, 2007, pp. 872–878.* doi:10.1109/ITSC.2007.4357793.
- [9] N. Saunier, T. Sayed, A feature-based tracking algorithm for vehicles in intersections, in: *Canadian Conference on Computer and Robot Vision, IEEE, Québec, 2006.* doi:10.1109/CRV.2006.3.
- [10] J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, in: *5-th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1967, pp. 1:281–297.*
- [11] M. G. Mohamed, N. Saunier, Motion Prediction Methods for Surrogate Safety Analysis, *Transportation Research Record: Journal of the Transportation Research Board* 2386 (2013) 168–178. doi:10.3141/2386-19.
URL <http://dx.doi.org/10.3141/2386-19>
- [12] P. St-Aubin, L. Miranda-Moreno, N. Saunier, An automated surrogate safety analysis at protected highway ramps using cross-sectional and before–after video data, *Transportation Research Part C: Emerging*

- Technologies 36 (2013) 284–295. doi:10.1016/j.trc.2013.08.015.
URL <http://dx.doi.org/10.1016/j.trc.2013.08.015>
- [13] B. T. Morris, M. M. Trivedi, A Survey of Vision-Based Trajectory Learning and Analysis for Surveillance, Circuits and Systems for Video Technology, IEEE Transactions on 18 (8) (2008) 1114–1127. doi:10.1109/TCSVT.2008.927109.
- [14] N. Saunier, M. G. Mohamed, Clustering Surrogate Safety Indicators to Understand Collision Processes, in: Transportation Research Board Annual Meeting, 2014, 14-2380.
- [15] Å. Svensson, C. Hydén, Estimating the severity of safety related behaviour, Accident Analysis and Prevention 38 (2) (2006) 379–385. doi:10.1016/j.aap.2005.10.009.
URL <http://dx.doi.org/10.1016/j.aap.2005.10.009>
- [16] K. Ismail, T. Sayed, N. Saunier, Automated Analysis of Pedestrian-Vehicle Conflicts: Context For Before-and-after Studies, Transportation Research Record: Journal of the Transportation Research Board 2198 (2010) 52–64.
- [17] J. Autey, T. Sayed, M. H. Zaki, Safety evaluation of right-turn smart channels using automated traffic conflict analysis, Accident Analysis and Prevention 45 (2012) 120–130. doi:10.1016/j.aap.2011.11.015.
URL <http://dx.doi.org/10.1016/j.aap.2011.11.015>
- [18] L. Zheng, K. Ismail, X. Meng, Shifted Gamma-Generalized Pareto Distribution model to map the safety continuum and estimate crashes, Safety Science 64 (2014) 155–162. doi:10.1016/j.ssci.2013.12.003.
URL <http://dx.doi.org/10.1016/j.ssci.2013.12.003>
- [19] S. Brahmhatt, Practical OpenCVdoi:10.1007/978-1-4302-6080-6.
URL <http://dx.doi.org/10.1007/978-1-4302-6080-6>