Transferability Study of Video Tracking Optimization for Traffic Data Collection and Analysis

Philip Morse, Master Student
Department of Civil Engineering and Applied Mechanics, McGill University
Macdonald Engineering Building, 817 Sherbrooke Street West
Montréal, Québec, Canada H3A 0C3
Email: philip.morse@mail.mcgill.ca

Paul St-Aubin, PhD Candidate
Department of Civil, Geological and Mining Engineering
Polytechnique Montréal, C.P. 6079, succ. Centre-Ville
Montréal, Québec, Canada H3C 3A7
Email: paul.st-aubin@mail.mcgill.ca

Luis Miranda-Moreno, Associate Professor
Department of Civil Engineering and Applied Mechanics, McGill University
Room 268, Macdonald Engineering Building, 817 Sherbrooke Street West
Montréal, Québec, Canada H3A 0C3
Phone: (514) 398-6589
Email: luis.miranda-moreno@mcgill.ca

Nicolas Saunier, Associate Professor
Department of Civil, Geological and Mining Engineering
Polytechnique Montréal, C.P. 6079, succ. Centre-Ville
Montréal, Québec, Canada H3C 3A7
Phone: (514) 340-4711 x. 4962
Email: nicolas.saunier@polymtl.ca

5384 words + 4 figures + 5 tables = 7634 words

November 16, 2015

ABSTRACT

Despite the extensive studies on the performance of video sensors and computer vision algorithms, calibration of these systems is usually done by trial and error using small datasets and incomplete metrics such as simple detection rates. There is a widespread lack of systematic calibration of tracking parameters in the literature.

This study proposes an improvement in automatic traffic data collection through the optimization of tracking parameters using a genetic algorithm by comparing tracked road user trajectories to manually annotated ground truth data using the Multiple Object Tracking Accuracy (MOTA) as the fitness function. The optimization procedure is first performed on a given dataset and then validated by applying the resulting parameters on a separate dataset. A number of problematic tracking and visibility conditions are tested using five different camera views selected based on differences in weather conditions, camera resolution, camera angle, tracking distance, and camera site properties. The transferability of the optimized parameters is verified by evaluating the performance of the optimized parameters across these data samples.

Results indicate that there are significant improvements to be made in the parametrization. Winter weather conditions require a specialized and distinct set of parameters to reach an acceptable level of performance, while higher resolution cameras have a lower sensitivity to the optimization process and perform well with most sets of parameters. Regarding the impact on traffic variables, average spot speeds are found to be insensitive to MOTA while traffic counts are strongly correlated.

INTRODUCTION

The use of video data for automatic traffic data collection and analysis has been on an upward trend as more powerful computational tools, detection and tracking technology become available. Not only have video sensors been able for a long time to emulate inductive loops to collect basic traffic variables such as counts and speed as in the commercial system Autoscope (1), but they can also provide higher-level information regarding road user behaviour and interactions more and more accurately. Examples include pedestrian gait parameters (2), crowd dynamics (3) and surrogate safety analysis applied to motorized and non-motorized road users in various road facilities (4, 5, 6). Video sensors are relatively inexpensive and easy to install or already installed, for example by transportation agencies for traffic monitoring: large datasets can therefore be collected for large scale or long term traffic analysis. This so-called "big data" phenomenon offers opportunities to better understand transportation systems, presenting its own set of challenges for data analysis (7).

Despite the undeniable progress of the video sensors and computer vision algorithms in their varied transportation applications, there persists a distinct lack of large comparisons of the performance of video sensors in varied conditions defined for example by the complexity of the traffic scene (movements and mix of road users), the characteristics of cameras (8) and their installation (height, angle), the environmental conditions (e.g. the weather) (9), etc. This is particularly hampered by the poor characterization of the datasets used for performance evaluation and the limited availability of benchmarks and public video datasets for transportation applications (10). Tracking performance is often reported using ad hoc and incomplete metrics such as "detection rates" instead of detailed, standardized, and more suitable metrics such as the Multiple Object Tracking Accuracy (MOTA) and Precision (MOTP). This metric takes into account other factors such as the error in the object's position estimate, missed detections and false positives on a frame by frame basis (11). Finally, the computer vision algorithms are typically manually adjusted by trial and error using a small dataset covering few conditions affecting performance while the reported performance evaluated on the same dataset is thus over-estimated: comparing to other fields such as machine learning, it should be clear that the algorithms should be systematically optimized on a calibration dataset, while performance should be reported for a separate validation dataset (12).

While the performance of video sensors for more simple traffic data collection systems has been extensively studied, not all factors have been systematically analyzed and issues with parameter optimization and lack of separate calibration and validation datasets is widespread. Besides, the relationship of tracking performance with performance of traffic parameters has never been fully investigated.

The objectives of this paper are threefold:

1. to improve the performance of existing automated detection and tracking methods for video data in terms of the accuracy of tracking, through the optimization of tracking parameters using a genetic algorithm comparing the tracker output with manually annotated trajectories;

2. to study the relationship between tracking accuracy, its optimization, and different kinds of traffic data such as counts and speeds;

3. to explore the transferability of parameters for separate datasets with the same properties

1   (consecutive video samples) and across different properties, by reporting how optimiz-
2   ing tracking for one condition impacts tracking performance for the other conditions.

3   The method is applied to a set of traffic videos extracted from a large surrogate safety study of
4   roundabout merging zones (7), covering factors such as the distance of road users to the camera,
5   the types of cameras, the camera resolution and weather conditions. As a follow up on (12), this
6   new paper investigates more factors and how tracking performance is related to the accuracy of
7   traffic parameters.
8   This paper is organized as follows: in the next section a brief overview of the current
9   state of computer vision and calibration in traffic applications is provided; then the methodology
10   is provided in detail including the ground truth inventory, measures of performance and calibra-
11   tion procedure; and finally the last two sections discuss the results of the tracking optimization
12   procedure and conclusions regarding ideal tracking conditions and associated parameter sets.

13   **LITERATURE REVIEW**
14   **Computer Vision in Traffic Applications**
15   Computer vision is used extensively in traffic applications as an instrument of data collection and
16   monitoring. Cameras and computer vision are slowly being implemented on-board motorized
17   vehicles as part of the sensor suite necessary for vehicle automation, including advanced driver
18   assistance systems (e.g. pedestrian-vehicle collision avoidance system (13), vehicle overtaking
19   (14)) and optical camera communications systems (15). For traffic engineers, the two primary ap-
20   plications of computer vision using stationary cameras include vehicle presence detection systems
21   (sometimes referred to as virtual loops) and motion tracking. Presence detection has widespread
22   commercial application due to its relatively high degree of reliability which is on par with em-
23   bedded sensor technology such as inductive loops; its primary application is in providing traffic
24   counts, queue lengths, and basic presence detection (16) for a range of traffic engineering tasks
25   ranging from data collection to traffic light control and optimization.
26   Motion tracking is a more complex application which aims to extract the road users' tra-
27   jectories continuously with great precision, i.e. their position for every video frame, within the
28   camera field of view, from which velocity, acceleration, and a number of other traffic behaviour
29   measures may be derived. Due to the increased complexity of tracking, it is generally considered
30   less reliable than presence detection systems. There are three main categories of tracking methods:

31   1. tracking by detection, which typically relies on background subtraction to detect fore-
32   ground objects and appearance-based object classification (17);

33   2. tracking using flow, also called feature-based tracking (18), first introduced in (19);

34   3. tracking with probability based on Bayesian tracking frameworks (20).

35   The NGSIM project was one of the first large-scale video data collection projects making
36   use of semi-automated vehicle tracking from freeway and urban arterial video data to obtain ve-
37   hicle trajectories for traffic model calibrations (21). Surrogate safety analysis also makes use of
38   trajectory data, for example with the early SAVEME project (22, 23), and now more recently with
39   extensive open source projects such as Traffic Intelligence (18, 24).

1 **Tracking Optimization and Sensor Calibration**
2 The work done to optimize parametrization of the various trackers is sparse and usually set manu-
3 ally by trial and error from experimental results. The instances of automated calibration in (25) and
4 (26) used Adaboost training strictly for shape detectors and (27) used evolutionary optimization for
5 the segmentation portion. One of the only cases of systematic improvement of the tracking method
6 as a whole through evolution algorithms was done recently at Polytechnique Montréal (12): the
7 current work shares similarities with this work such as the use of the same measure of tracking
8 accuracy for optimization, but this paper deals with motorized traffic instead of pedestrians and
9 investigates further the transferability of calibrated parameters not only for the same camera view,
10 but across different types of cameras, camera views, and visibility/weather conditions.
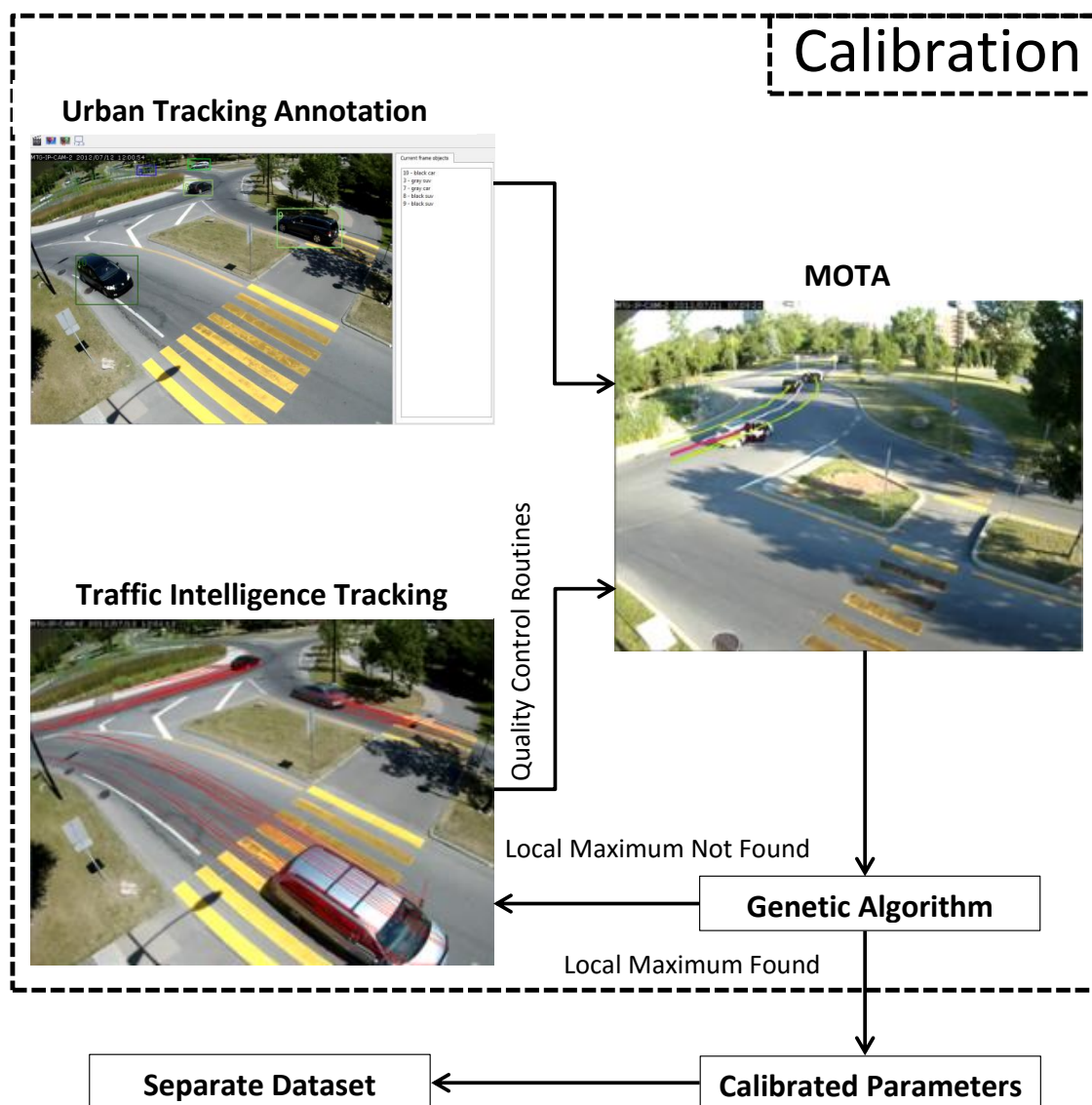

FIGURE 1 Overview of the optimization process

1 **METHODOLOGY**
2 The approach proposed in this paper consists in identifying different conditions that may have an
3 impact on tracking performance and traffic variables such as counts. For each condition, we need
4 two video samples or regions in the same video where the only or primary difference is the change
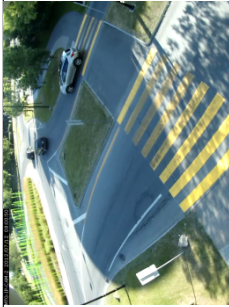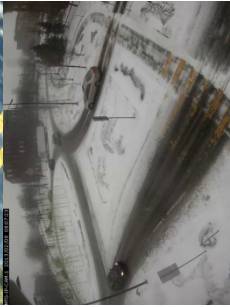5 in that condition. The four main steps are as follows:

6   1. Selection of sites and analysis zones: five different camera views are specifically se-
7   lected to allow for analysis based on the chosen conditions to be compared. Ten minutes
8   of video are manually annotated for an analysis zone in each camera view to be used as
9   a baseline for the analysis.

10   2. Optimization of tracking parameters over the whole annotated period (10 min): a subset
11   of the tracking parameters are optimized for each camera view using the chosen mea-
12   sure of performance. These results are used to evaluate traffic data and correlations of
13   parameters with the measure of performance.

14   3. Optimization of tracking parameters over the first five minute period: the tracking pa-
15   rameters optimized for the first five annotated minutes of each camera view are applied
16   to the whole 10 minute annotated video, as well as to sub-regions of the analysis zones
17   (two sub-regions, one close and one far from the camera), in order to evaluate over-
18   fitting.

19   4. The optimized tracking parameters from step 3 are applied to the full ten minutes of
20   each camera view to evaluate the transferability between sites, camera types and weather
21   conditions.

22   The overview of the methodology is presented in FIGURE 1.

23 **Ground Truth Inventory**
24 The ground truth data is obtained through manual annotation of the source video data using the
25 Urban Tracker annotation application (28). There are five video sequences selected from three
26 different roundabouts presented in Table 1 (a video sequence correspond to a camera views and
27 the terms are used interchangeably). S1S and S1W were recorded with similar field of views on the
28 first site to compare the weather condition. S1S and S2 show comparable views of two different
29 roundabouts (sites 1 and 2). S3V1 was recorded using the same camera as on S1 and S2, and can
30 be compared to S1S and S2 to evaluate the impact of the resolution. S3V1 and S3V2 allow to
31 compare the impact of the type of camera on the same site, with two different views. For each
32 camera view, an analysis zone covering a merging zone of the roundabout is defined inside the
33 zone where automated tracking is performed (site analysis mask) as can be seen in FIGURE 2. All
34 vehicles going through the analysis zone of each camera view were manually tracked for 10 min
35 (with bounding boxes drawn around each vehicle every 5-10 frames). Manual annotation is labour
36 intensive: the annotations require between half an hour to one hour of manual labour per minute
37 of video, depending on the frame rate and the traffic flow.

TABLE 1 Ground Truth Inventory: S1, S2 and S3 refer to the sites 1 to 3, S1S and S1W refer respectively to the videos recorded on the first site in summer and winter, and S3V1 and S3V2 refer respectively to the videos recorded simultaneously on the third site with two different cameras covering complementary zones of the roundabout (resolution is in pixels(pix) and video frame rate in frames per second (fps))

| Site	Time & Date	Number of Annotated Road Users	Camera Type	Conditions	Sample View
S1S	12:00pm, July 2012 (Thursday)	266	IP Camera 800x600 pix, 15 fps	Sunny, shadows	
S1W	8:00am, February 2013 (Friday)	209	IP Camera 800x600 pix, 15 fps	Low visibility, winter	
S2	7:00am, July 2012 (Wednesday)	64	IP Camera 800x600 pix, 15 fps	Sunny, some shadows	
S3V1	4:00pm, August 2013 (Friday)	80	IP Camera 1280x1024 pix, 15 fps	Sunny	
S3V2	4:00pm, August 2013 (Friday)	312	GoPro 1920x1080 pix, 15 fps, corrected for distortion	Sunny	
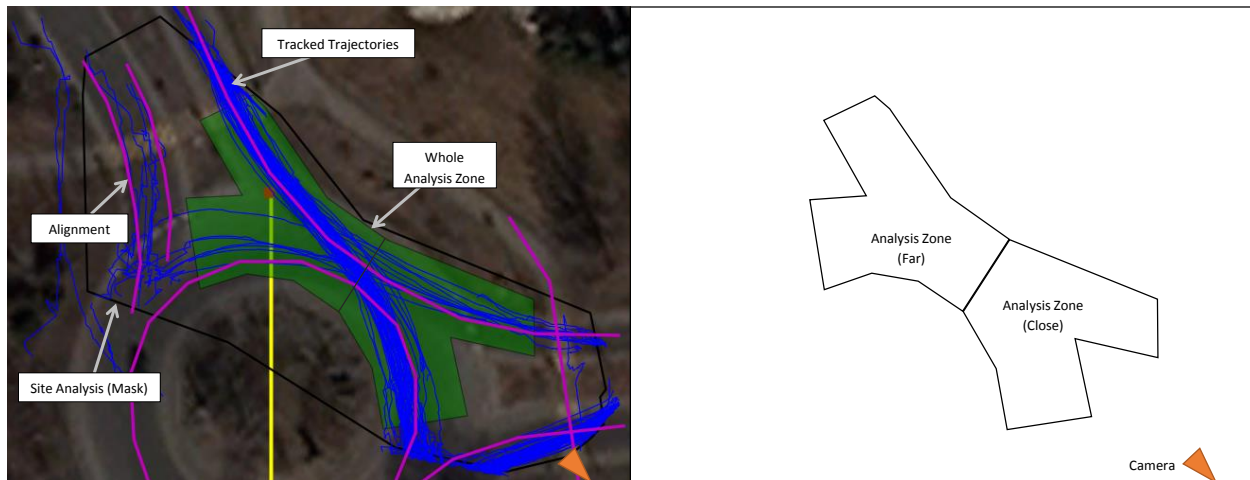
FIGURE 2 Example of an analysis zone and extracted trajectories for a given camera field of view

1 **Video Tracking**
2 The video analysis tool used in this work relies on feature-based tracking and is available in the
3 open source "Traffic Intelligence" project[1] (18). Feature-based tracking is composed of two main
4 steps:

5   1. distinct points such as corners are detected in the whole image and tracked frame after
6   frame until they are lost;

7   2. a road user will typically have several feature on it: the second step consists in grouping
8   the features corresponding to individual road users. Two feature trajectories are grouped
9   if they are close enough (within distance `mm-connection-distance`) and if the
10   different between their maximum and minimum distance is small enough (within dis-
11   tance `mm-segmentation-distance`). A group of features is saved, corresponding
12   to a road user trajectory, if it has at least on average `min-nfeatures-group` fea-
13   tures per frame.

14   The main parameters are listed in TABLE 2, including the grouping parameters mentioned
15 above. The resulting road user positions may be measured in image or world space, depending
16 on whether or not a homography transformation was computed before tracking to project image
17 positions to world positions, on the ground plane. A homography was computed for all camera
18 views used in this paper using a tool provided in Traffic Intelligence. In addition, because the
19 second type of camera used, the GoPro, is subject to strong radial distortion (so called "fish-eye"
20 effect), a processing step is added to correct the distortion (see the sample view of S3V2 in Table 1).
21 However, since ground truths are also built from undistorted video data, feature-based tracking
22 quality should in theory be identical regardless of camera distortion; instead distortion affects
23 primarily the quality of the world-space data after homography transformation, be they tracked
24 trajectories or manual annotations. Fortunately, this type of error is easily corrected by examination

---

[1]`https://bitbucket.org/Nicolas/trafficintelligence/`, accessed August 30, 2015

1   of the superposition of satellite imagery and does not warrant special optimization approaches.
2   Error tolerance for homography transformation is no more than 1 metre at a tracking distance of
3   up to 50 metres. On the other hand, since undistortion is applied to image space instead of to the
4   trajectories directly (for a number of technical reasons) some microscopic distortion effects might
5   occur to individual pixels, especially at the edges of image space (far data) which could have a
6   small impact on tracking quality.

7   **Quality Control Routines**
8   A small percentage of trajectories (typically less than 5 per 10-min of data) generated by the
9   tracker have several obvious issues and errors. For example: cases of obstruction by large vehicles
10 or lamp posts or ghost trajectories seemingly driving through each other. The frequency of these
11 errors is too infrequent to optimize algorithmically, but are severe enough to generate strong false
12 alarms. Fortunately, they are also easy to identify and correct. Most of these issues are corrected
13 or eliminated entirely using several quality control routines from the tools developed for the larger
14 project on the roundabout safety (7). These include the following functions:

15   • **Object integrity verification**: verify any corruption in the data structure.

16   • **Warm-up errors at scene edges**: vehicles entering image space are only partially tracked
17   until they come within full view, and therefore are lacking in number of tracked features
18   which causes issues with feature grouping.

19   • **Duplicate detection removal**: based on proximity and trajectory similarity, only the
20   most egregious examples of duplicate detections are removed. Tracking optimization
21   should correct most duplicate tracking issues.

22   • **Outlier point split**: when two distinct vehicles within the scene are grouped together,
23   they are tracked as a single vehicle which seems to "teleport" instantly across the scene.
24   It is split at the time of "teleportation".

25   • **Stub removal**: the minimum trajectory dwell time is 0.66 s.

26   • **Alignment filtering**: if alignment metadata (lane and sidewalk centerlines) exists, ve-
27   hicles that deviate significantly from any typical movements can be flagged for manual
28   review as either a severe traffic infraction or tracking error.

29 **Optimizing Tracking Accuracy**
30 The tracking parameters listed in TABLE 2 are optimized using a genetic algorithm that aims to
31 improve tracking accuracy, comparing the tracker output to the ground truth for a video sequence
32 (see overview in FIGURE 1. Each iteration of the genetic algorithm corresponds a set (popula-
33 tion) of individuals with each individual representing a complete set of tracking parameters θ: the
34 tracker and filtering routines are run on the video sequence for each set of tracking parameters θ.
35 The tracker output and the ground annotations are compared in the analysis zone and the genetic
36 algorithm will generate a new population of tracking parameters by favouring and combining the
37 best tracking parameters of the previous population.

TABLE 2 Tracking parameters considered for tracking accuracy optimization

Parameter	Range	Type	Description
Feature Tracking			
feature-quality	[0-0.4]	Float	Minimum quality of corners to track (unitless)
min-feature-distance-klt	[0-6]	Float	Minimum distance between features (in pixels)
window-size	[3-10]	Integer	Distance within which to search for feature in next frame (in pixels)
min-tracking-error	[0.01-0.3]	Float	Minimum error to reach to stop optical flow (unitless)
min-feature-time	[2-10]	Integer	Minimum time a feature must exist to be saved (in frames)
Feature Grouping			
mm-connection-distance	[1.5-3]	Float	Distance to connect features into objects (in world distance unit (m))
mm-segmentation-distance	[1-3]	Float	Segmentation distance (in world distance unit (m))
min-nfeatures-group	[2-4]	Float	Minimum number of features per frame to generate a road user

1   The metric of tracking performance, or fitness function of the genetic algorithm, is the
2 MOTA as described in (11). It is the most common metric for tracking accuracy, i.e. to evaluate
3 the whole trajectory and not just detections in each frame, used in computer vision. MOTA is
4 basically the ratio of the number of correct detections of each object over the number of frames in
5 which the object appears (in the ground truth):

$$MOTA = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t}$$

6   where m_t, fp_t and mme_t are respectively the number of misses, over detections (false
7 positives), and mismatches for frame t. These depend on matching the trajectories produced by
8 the tracker to the ground truth. In this work, a road user is considered to be tracked in a frame if
9 its centroid is within a given distance in world space from the ground truth bounding box centre.
10 Since there may be multiple matches, the Hungarian algorithm is used to associate uniquely the
11 ground truth and tracker output so that over detections (more than one trajectory for the same road
12 user) can be counted. The tracking results depend on these choices and a 5 m distance threshold
13 is used as it is approximately the length of a passenger vehicle. The complementary performance
14 measure of MOTP is reported in the results. It is the average distance between the ground truth
15 and road user trajectories (11). This is particularly important for traffic variables such as time and
16 distance headway, and safety analysis based on the proximity in time and space of interacting road
17 users as measured for example by the time to collision indicator (7).
18   Once the genetic algorithm finds a local maximum, the optimized or calibrated tracking
19 parameters can be applied to the other video sequences to determine the performance of the track-
20 ing parameters under different conditions. The relationship between the tracking parameters and

1 MOTA is evaluated using Spearman's Rank Correlation Coefficient ρ, calculated as:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

2   where d_i is the difference between the ranks x_i and y_i for each corresponding MOTA and
3 tracking parameter values in a sample of size n. The coefficient ρ varies between -1 and +1 when
4 the variables are perfectly correlated.
5   The road user trajectories obtained with the calibrated tracking parameters are analyzed to
6 generate traffic variables. The objective is to identify the relationship of different tracking perfor-
7 mance, measured by MOTA, with the traffic variables including traffic flow and spot speeds.

8 **Over-fitting and Transferability**
9 One of the risks in optimization is that the parameters may be very specific to the video sequence
10 used for optimization. The tracking performance on this video sequence may not be achieved when
11 applied to other video sequences and should not be considered as representing general performance
12 if applied to any other video sequence, even collected in the exact same conditions since traffic
13 will be different. To determine whether tracking parameters optimized for a specific sequence are
14 universally applicable, the performance of the results is examined for each camera view with the
15 following comparisons:

16   • The analysis area is split into a close section and a far section (see an example in FIG-
17   URE 2): optimization is done in the whole analysis zone and the MOTA is reported for
18   both zones independently.

19   • The ground truth is split into two 5-min sequences: optimization is done for the first
20   5 min and reported for each 5-min sequence.

21   • The same camera is used on two different sites at a similar height and angle (sequences
22   S1S and S2).

23   • The same camera view is used in both winter conditions and summer conditions (se-
24   quences S1S ans S1W).

25   • The same IP-camera is used at two different resolutions: 800x600 and 1280x1024 (se-
26   quences S1S ans S3V1).

27   • The same site is used for two different cameras, an IP and a GoPro camera (sequences
28   S3V1 and S3V2).

29   Transferability is thus verified by applying each set of optimized tracking parameters to
30 each of the other annotated sequences. The full ten minute annotated videos are used to calculate
31 the MOTA, which are then compared to the optimized tracking performance and reported also as
32 a percentage of the maximum MOTA. This is done to avoid potential bias from the site selection
33 and vehicle composition, which may lower the MOTA result compared to other cameras, despite
34 having similar relative tracking performance results.

1 **EXPERIMENTAL RESULTS**
2 **The Relationship of Tracking Accuracy with Traffic Data**
3 In the first phase of optimization, the genetic algorithm was run for the full length of each 10-min
4 video sequence. The population size is set to 20 individuals, using a minimum selection of 20 %
5 of the fitness function, a crossover rate of 60 % of individuals, and a mutation rate of 5 % (of indi-
6 viduals). These parameters were based off of typical values and tweaked to account for a relatively
7 low number of alleles (number of parameters being optimized) and the high cost of computing the
8 fitness function. All the sets of tracking parameters (individuals) and the corresponding trajecto-
9 ries generated by the tracker are saved over the whole optimization process: the traffic counts and
10 average speeds at the entrance and exit of each lane of the analysis zone are extracted and analyzed
11 with respect to tracking accuracy (see FIGURE 3).
12   As expected, there is a strong correlation between MOTA and the number of road users
13 tracked given the definition of MOTA: if tracking errors are uniformly spread over the analysis
14 zone, since MOTA represents the average percentage of correctly tracked road user instants, one
15 can expect that the resulting counts will be around the true number of road users multiplied by
16 MOTA. MOTA seems therefore to be a good indicator of total counting accuracy. On the other
17 hand, average spot speeds seem relatively insensitive to MOTA, except for lower values of MOTA
18 which tend to be associated with a larger range of average extracted spot speed, especially for S1W.
19 This can be related to the counts: lower MOTA values generate low counts and the spot speeds are
20 a small random sample of all road users with high variability. As the performance increases in
21 lanes that have a significant number of vehicles, the average spot speeds converge (one can still
22 observe a lot of variability independently of MOTA for lane 1 at the exit).
23   The two higher resolution camera views (S3V1 and S3V2) had very few results of poor
24 performance. This suggests a lower sensitivity to the tracking parameters chosen for optimization.
25 S1W presents a special case where only 49 % of generation individuals had any tracked objects.
26 These two cases are discussed in subsequent sections.

27 **Correlation of Tracking Parameters with Tracking Accuracy**
28 All the tested tracking parameters in the optimization history with their associated MOTA are used
29 to compute Spearman's rank correlation coefficient (see results by parameter and camera view in
30 FIGURE 4). The correlations therefore depend on these specific optimization runs, the number
31 of which is different for each camera view since the processing time for tracking took up to 12 h
32 for the 20 individuals evaluated at each generation of the genetic algorithm. The higher resolution
33 camera also had few data points for lower MOTA values as seen in FIGURE 3.
34   The tracking performance of the lower resolution IP camera was largely dependent on
35 `feature-quality`. In fact, in the case of S1W recorded in winter, a MOTA above 0.150 was
36 not found unless `feature-quality` was below 0.10 and the best performance came from a
37 `feature-quality` of approximately 0.01. This suggests that these video sequences recorded
38 in the winter should be calibrated separately from the other ones since its most important pa-
39 rameter is directly related to computation time (lower `feature-quality` will generate more
40 features that take more time to be grouped). The range of `feature-quality` should also be
41 adjusted for such a calibration. Another point of mention is that the GoPro camera had most
42 of the lowest correlation coefficients with respect to each parameter. The implication is that the
43 high resolution seems to be less sensitive to the choice of tracking parameters to achieve good
44 performance. The third observation is that certain parameters such as `min-tracking-error`
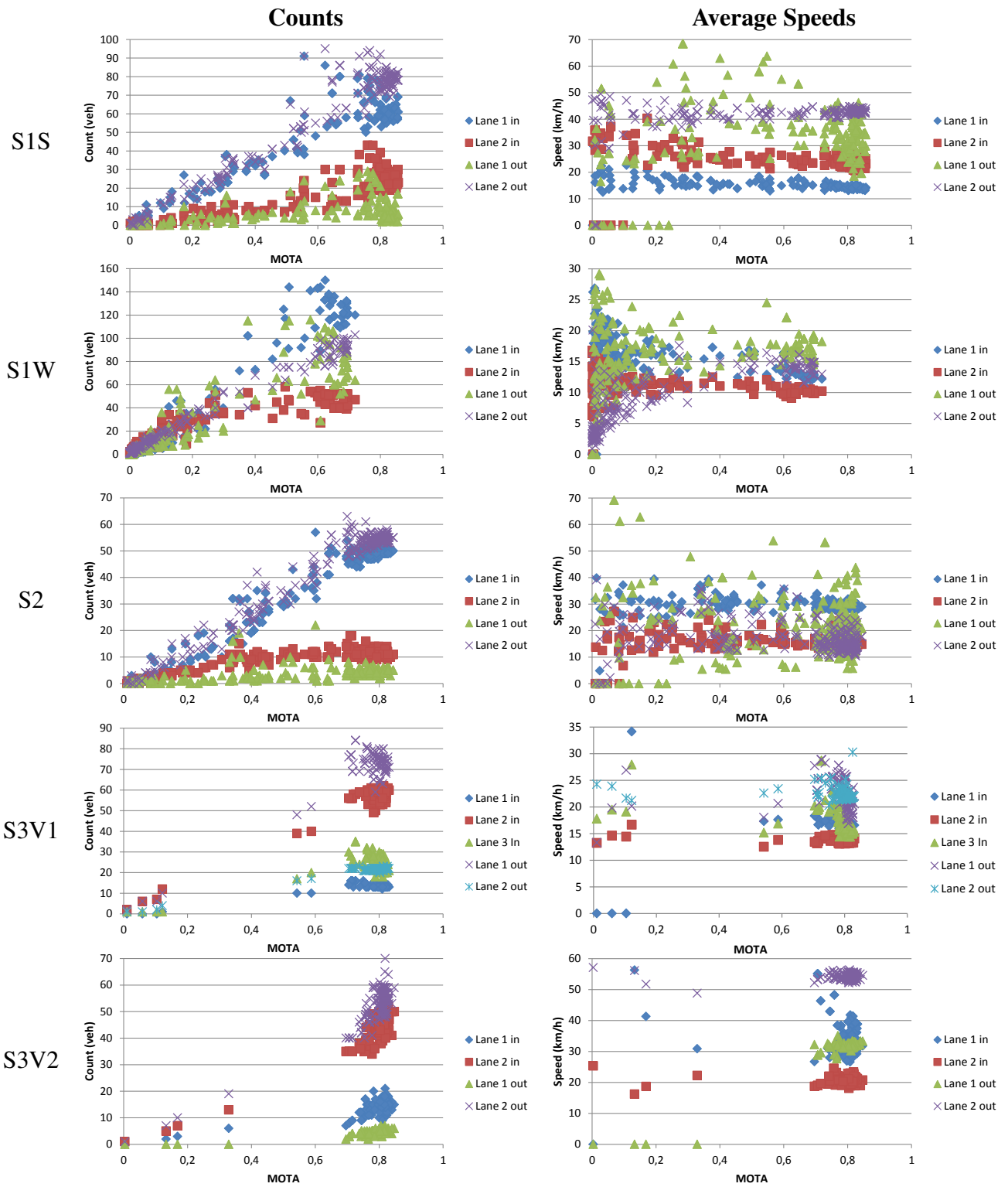
FIGURE 3 Counts (left column) and average speeds (right column) measured at the entrance and exits of the merging zone for each video sequence (corresponding to each row)
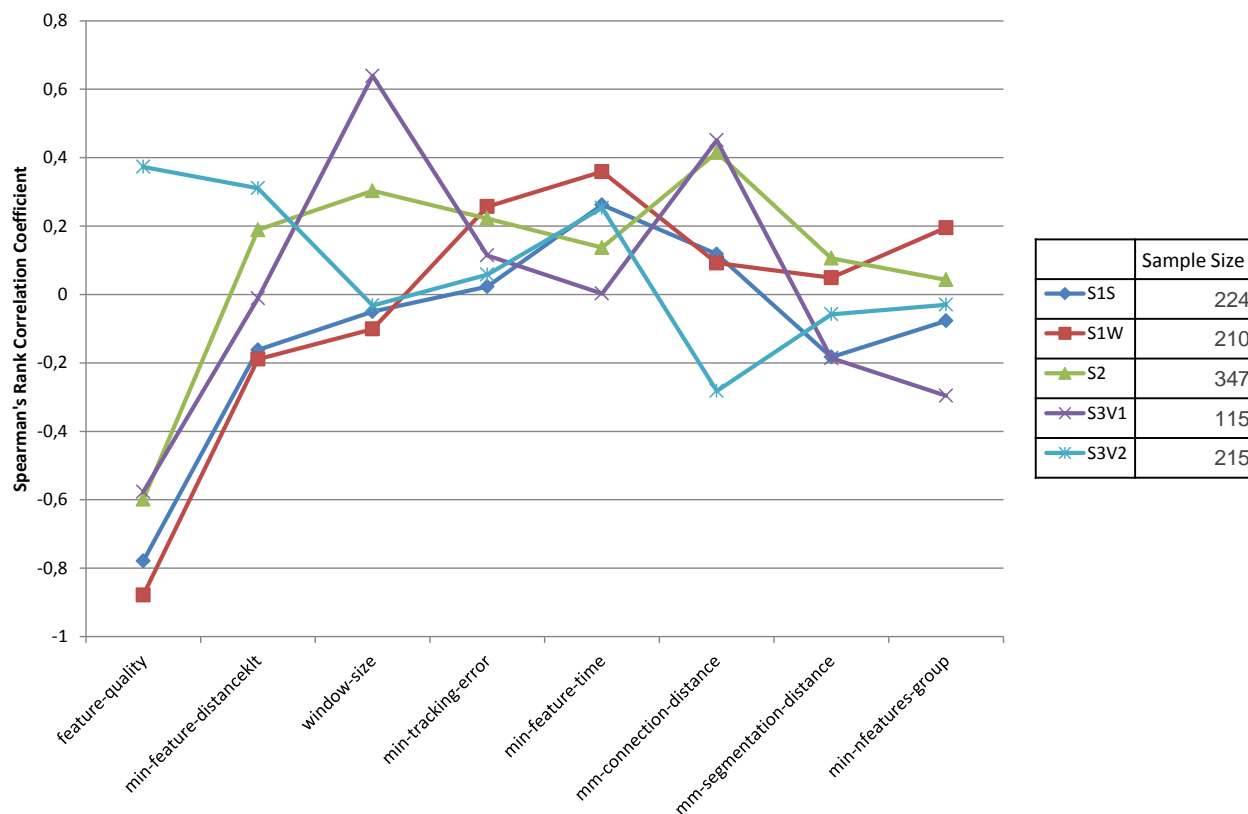
FIGURE 4 Spearman's rank correlation coefficient by tracking parameter and site computed over the calibration history

1 and `min-feature-time` could be eliminated from the optimization process due to the lack of
2 correlation for all camera views.

3 **Over-fitting Analysis**
4 The second phase of the validation process was to evaluate the performance of the optimized
5 parameters compared to the default parameters. The genetic algorithm was run for the first 5 min
6 of each video file for between 10 and 30 generations over the course of one to three days depending
7 on the site. TABLE 3 presents a summary of the optimal parameters found for each camera view.
8 Some parameters such as `min-feature-distance-klt`, `mm-connection-distance`
9 and `min- tracking-error` seem to have converged towards similar values on three of the
10 camera views. However, it is apparent that in most cases the optimal solution is unique.
11   The tracking performance obtained by the genetic algorithm is presented in TABLE 4 as
12 a comparison between the performance for the default tracking parameters and the parameters
13 optimized on the first 5 min of each camera view. Both the MOTA and MOTP values for every case
14 are improved, in some cases by a large margin. The results for winter (S1W) in particular express
15 the need for optimization as the MOTA did not reach over 0.05 using default parameters whereas a
16 potential for a MOTA over 0.70 was found. As expected, the camera views with a lower correlation
17 with the parameters, in particular S3V2, did not improve as much as the other sites. Furthermore,
18 a lower MOTA did not necessarily mean that the camera cannot be properly optimized. When

TABLE 3 Default tracking parameters and values optimized for each camera view

		Optimized for				
	Default	S1S	S1W	S2	S3V1	S3V2
window size	7	9	9	6	10	3
feature quality	0.1	0.0826	0.0096	0.0812	0.1167	0.2652
min-feature-distance-klt	5	3.0126	3.1374	3.5496	3.5424	2.8645
min-tracking-error	0.3	0.1856	0.1038	0.1833	0.1847	0.0284
min-feature-time	20	8	10	9	5	3
mm-connection-distance	3.75	2.7480	2.0005	2.6881	2.7306	1.8520
mm-segmentation-distance	1.5	2.4560	2.1804	1.8151	2.1762	1.2823
min-nfeatures-group	3	2.7231	3.4020	3.1675	2.5111	2.3042

1 comparing S1S and S2, which have similar physical properties and identical cameras, there is a
2 noteworthy difference (0.909 and 0.812 respectively). The source of the difference could be due to
3 the lower number of vehicles in S2 or a difference in traffic compositions (i.e. a higher volume of
4 trucks).

**TABLE 4 Tracking performance (MOTA and MOTP) for the different camera views for the
default tracking parameters and the parameters optimized on the first 5 min of each camera
view on the same sequence, as well as the last 5 min, the whole 10 min, the closer half and
further half of the analysis zone separately**

	S1S		S1W		S2		S3V1		S3V2	
	MOTA	MOTP	MOTA	MOTP	MOTA	MOTP	MOTA	MOTP	MOTA	MOTP
Default parameters										
First 5 min	0.746	1.642	0.043	2.666	0.647	1.251	0.754	1.149	0.820	1.101
Last 5 min	0.679	1.490	0.031	3.524	0.719	1.009	0.761	1.074	0.670	1.612
Full 10 min	0.719	1.581	0.041	3.001	0.703	1.100	0.760	1.114	0.750	1.340
Close	0.706	1.378	0.045	2.742	0.569	1.119	0.763	0.971	0.856	1.175
Far	0.632	1.819	0.029	4.594	0.700	1.092	0.700	1.230	0.634	1.519
Optimized parameters on first 5 min										
First 5 min	0.909	1.233	0.708	1.974	0.812	1.019	0.855	1.000	0.851	0.736
Last 5 min	0.884	1.233	0.693	1.885	0.717	0.853	0.772	0.842	0.691	0.585
Full 10 min	0.905	1.237	0.710	1.920	0.767	0.918	0.817	0.927	0.789	0.666
Close	0.871	1.174	0.674	1.979	0.617	0.856	0.784	0.857	0.875	0.713
Far	0.821	1.298	0.608	1.971	0.760	0.978	0.738	1.040	0.680	0.554

5   In the same table, the parameters optimized by the genetic algorithm were evaluated on
6 different conditions for each camera view:

7   • a validation sample made up by the last 5 min of each camera view

8   • the whole 10 min of annotated video

9   • the closer half of the analysis zone (whole 10 min) (see FIGURE 2)

10   • the further half of the analysis zone (whole 10 min)

1   This permits investigating over-fitting. It is found that, while the improvement were always
2   greater for the first 5 min on which the parameters were optimized, there was a systematic positive
3   upward trend in performance, both for MOTA and MOTP, on the validation sample (last 5 min)
4   and the other conditions. Yet improvement on the validation sample vary widely between S1S and
5   S1W and the other camera views where they are quite small. It can be seen that MOTA for the
6   far analysis zone is in general worse than for the closer zone as expected, except for S2. It should
7   be noted that the MOTA of the close and far zones do not always average to the MOTA for the
8   full analysis zone. This is explained by the small differences created at the borders of the analysis
9   zones as vehicles enter and leave and trajectories are split between the sub-zones.

10   **Transferability of Optimized Parameters Across Camera Views**
11   The third phase applies the parameters optimized on the first 5 min of each camera view on all an-
12   notated videos for the whole 10 min: the performance results are presented for MOTA in TABLE 5.
13   The MOTA was compared both as an absolute value as well as a percentage of the known highest
14   MOTA to investigate the relative gains of each camera view (the best MOTA for each camera view
15   is the one obtained with the parameters optimized for itself as expected). The winter conditions
16   (S1W) proved to be the most difficult ones, where tracking parameters optimized on a different
17   camera view always produce very few road user trajectories, leading to MOTA values far outside
18   the acceptable range. Inversely, using the tracking parameters optimized for S1W resulted in worse
19   tracking performance than obtained for the default parameters for two camera views. This suggests
20   the need for separate tracking parameters for different weather conditions. However, in the case of
21   good weather conditions, a result of around 90 % of the best known MOTA can be expected from
22   any set of optimized parameters.

TABLE 5 Tracking performance for default and optimized parameters for all camera views: each column corresponds to the camera view on which the parameters were optimized (first 5 min) while each row corresponds to the camera view (full 10 min) for which the performance is reported for the parameters

Site	Default	Parameters optimized for				
		S1S	S1W	S2	S3V1	S3V2
S1S	0.719	0.905	0.821	0.818	0.841	0.823
S1W	0.041	0.115	0.710	0.078	0.044	0.051
S2	0.703	0.740	0.623	0.767	0.746	0.718
S3V1	0.760	0.797	0.778	0.793	0.817	0.799
S3V2	0.750	0.705	0.737	0.776	0.700	0.789
Site	Default	Parameters optimized for				
		S1S	S1W	S2	S3V1	S3V2
S1S	79.50%	100.00%	90.77%	90.39%	93.01%	91.01%
S1W	5.79%	16.14%	100.00%	10.97%	6.26%	7.16%
S2	91.71%	96.55%	81.19%	100.00%	97.27%	93.69%
S3V1	92.94%	97.51%	95.21%	97.03%	100.00%	97.77%
S3V2	95.17%	89.41%	93.51%	98.43%	88.79%	100.00%

1 **CONCLUSION**
2 The first finding of this paper is the strong correlation between traffic counts and the measure of
3 tracking performance (MOTA), as well as a correlation as strong as -0.879 between MOTA and
4 certain tracking parameters (using Spearman's rank correlation). The second finding deals with
5 the question of over-fitting. While the results, as expected, are found to be optimal on the sequence
6 on which the calibration was done, there were improvements in all conditions, in separate parts
7 of the video and on both the analysis sub-zones close and far from the camera. The third finding
8 covers the transferability of tracking parameters optimized for a specific camera view on the others.
9 Parameters for summer sequences are demonstratively not applicable to winter conditions as results
10 for the winter sequence did not surpass a MOTA of 0.150 unless specifically optimized for, whereas
11 the high resolution camera is shown to be very tolerant to different parameters.
12   The genetic algorithm used with manually annotated video sequences shows that there is
13 room for noticeable improvements over the default tracking parameters. Considering the strong
14 impact that winter conditions had on the performance results, a logical next step would be the
15 evaluation of other meteorological conditions such as precipitation, low visibility (fog), nighttime
16 and high winds (affecting camera stability). High resolution cameras have fewer issues which the
17 choice of parameters and should be studied further to determine the effects of these conditions
18 compared to the effects on lower resolution cameras. This study relied on video data recorded only
19 at roundabouts: more sites, with different geometries and traffic conditions, and annotations are
20 needed to extend and further generalize this work.
21   Additional work is also needed on the relationship between tracking accuracy and other
22 types of traffic data that could not be extracted in sufficient quantities from ten minutes of video
23 to provide meaningful results: gap time, time-to-collision and road user interactions are examples
24 of such data relevant for safety. Different types of optimization algorithms, e.g. the evolutionary
25 algorithm used in (12), and performance metrics could also be evaluated based on both compu-
26 tation time and reliability of the solutions. The accuracy of traffic data (counts and speeds) was
27 not measured in this paper and its relationship with tracking accuracy should be investigated. In
28 particular, can better accuracy for traffic variables be obtained by optimizing it directly rather than
29 by optimizing tracking accuracy?
30   The real world application is to develop a reliable single or dynamic set of tracking param-
31 eters that could be calibrated and applied to a network of both fixed and mobile cameras to be used
32 under all conditions. Combined with automated trajectory analysis tools, the wide-scale deploy-
33 ment of automatically calibrated video analysis algorithms would provide researchers with very
34 large datasets of traffic data for a better understanding and management of transportation systems.

REFERENCES

[1] Michalopoulos, P. G., Vehicle detection video through image processing: the Autoscope system. *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 1, 1991, pp. 21–29.

[2] Saunier, N., A. El Husseini, K. Ismail, C. Morency, J.-M. Auberlet, and T. Sayed, Pedestrian Stride Frequency and Length Estimation in Outdoor Urban Environments using Video Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2264, 2011, pp. 138–147, presented at the 2011 Transportation Research Board Annual Meeting.

[3] Johansson, A., D. Helbing, H. Al-Abideen, and S. Al-Bosta, From crowd dynamics to crowd safety: a video-based analysis. *Advances in Complex Systems*, Vol. 11, No. 04, 2008, pp. 497–527.

[4] St-Aubin, P., N. Saunier, L. Miranda-Moreno, and K. Ismail, Use of Computer Vision Data for Detailed Driver Behavior Analysis and Trajectory Interpretation at Roundabouts. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2389, 2013, pp. 65–77.

[5] Sakshaug, L., A. Laureshyn, Å. Svensson, and C. Hydén, Cyclists in roundabouts—Different design solutions. *Accident Analysis & Prevention*, Vol. 42, No. 4, 2010, pp. 1338–1351.

[6] Autey, J., T. Sayed, and M. H. Zaki, Safety evaluation of right-turn smart channels using automated traffic conflict analysis. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 120–130.

[7] St-Aubin, P., N. Saunier, and L. F. Miranda-Moreno, Large-scale automated proactive road safety analysis using video data. *Transportation Research Part C: Emerging Technologies*, 2015, in Press.

[8] Wan, Y., Y. Huang, and B. Buckles, Camera calibration and vehicle tracking: Highway traffic video analytics. *Transportation Research Part C: Emerging Technologies*, Vol. 44, 2014, pp. 202–213.

[9] Fu, T., S. Zangenehpour, P. St-Aubin, L. Fu, and L. F. Miranda-Moreno, Using microscopic video data measures for driver behavior analysis during adverse winter weather: opportunities and challenges. *Journal of Modern Transportation*, Vol. 23, No. 2, 2015, pp. 81–92.

[10] Saunier, N., H. Ardö, J.-P. Jodoin, A. Laureshyn, M. Nilsson, A. Svensson, L. F. Miranda-Moreno, G.-A. Bilodeau, and K. Åström, Public Video Data Set for Road Transportation Applications. In *Transportation Research Board Annual Meeting Compendium of Papers*, 2014, 14-2379.

[11] Bernardin, K. and R. Stiefelhagen, Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics. *EURASIP Journal on Image and Video Processing*, Vol. 2008, 2008, pp. 1–10.

[12] Ettehadieh, D., B. Farooq, and N. Saunier, Systematic Parameter Optimization and Application of Automated Tracking in Pedestrian-Dominant Situations. In *Transportation Research Board Annual Meeting Compendium of Papers*, 2015, 15-2400.

[13] Llorca, D., M. Sotelo, I. Parra, J. Naranjo, M. Gavilan, and S. Alvarez, An Experimental Study on Pitch Compensation in Pedestrian-Protection Systems for Collision Avoidance and Mitigation. *IEEE Trans. Intell. Transport. Syst.*, Vol. 10, No. 3, 2009, pp. 469–474.

[14] Milanés, V., D. F. Llorca, J. Villagrá, J. Pérez, C. Fernández, I. Parra, C. González, and M. A. Sotelo, Intelligent automatic overtaking system using vision for vehicle detection. *Expert Systems with Applications*, Vol. 39, No. 3, 2012, pp. 3362–3373.

[15] Ifthekhar, M. S., N. Saha, and Y. M. Jang, Stereo-vision-based cooperative-vehicle positioning using OCC and neural networks. *Optics Communications*, Vol. 352, 2015, pp. 166–180.

[16] Hoose, N., Computer Vision as a Traffic Surveillance Tool. In *Control Computers, Communications in Transportation*, Elsevier BV, 1990, pp. 57–64.

[17] Zangenehpour, S., L. F. Miranda-Moreno, and N. Saunier, Automated classification based on video data at intersections with heavy pedestrian and bicycle traffic: Methodology and application. *Transportation Research Part C: Emerging Technologies*, Vol. 56, 2015, pp. 161–176.

[18] Saunier, N. and T. Sayed, A feature-based tracking algorithm for vehicles in intersections. In *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, Institute of Electrical & Electronics Engineers (IEEE), 2006.

[19] Huang, T., D. Koller, J. Malik, G. Ogasawara, B. Rao, S. J. Russell, and J. Weber, Automatic symbolic traffic scene analysis using belief networks. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, AAAI, 1994, Vol. 94, pp. 966–972.

[20] Boykov, Y. and D. P. Huttenlocher, A new bayesian framework for object recognition. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, IEEE, 1999, Vol. 2.

[21] Kim, Z., G. Gomes, R. Hranac, and A. Skabardonis, A Machine Vision System for Generating Vehicle Trajectories over Extended Freeway Segments. In *12th World Congress on Intelligent Transportation Systems*, San Francisco, CA, 2005.

[22] Ervin, R., C. MacAdam, J. Walker, S. Bogard, M. Hagan, A. Vayda, and E. Anderson, *System for Assessment of the Vehicle Motion Environment (SAVME)*. National Highway Traffic Safety Administration, 2000.

[23] Gordon, T., Z. Bareket, L. Kostyniuk, M. Barnes, M. Hagan, Z. Kim, D. Cody, A. Skabardonis, and A. Vayda, *Site-Based Video System Design and Development*. Strategic Highway Research Program (SHRP2), 2012.

1   [24] Jackson, S., L. F. Miranda-Moreno, P. St-Aubin, and N. Saunier, Flexible, Mobile Video
2   Camera System and Open Source Video Analysis Software for Road Safety and Behavioral
3   Analysis. *Transportation Research Record: Journal of the Transportation Research Board*,
4   Vol. 2365, No. -1, 2013, p. 90–98.

5   [25] Sidla, O., Y. Lypetskyy, N. Brandle, and S. Seer, Pedestrian Detection and Tracking for
6   Counting Applications in Crowded Situations. In *2006 IEEE International Conference on
7   Video and Signal Based Surveillance*, Institute of Electrical & Electronics Engineers (IEEE),
8   2006.

9   [26] Ali, I. and M. N. Dailey, Multiple Human Tracking in High-Density Crowds. In *Advanced
10   Concepts for Intelligent Vision Systems*, Springer Science + Business Media, 2009, pp. 540–
11   549.

12   [27] Pérez, O., M. Á. Patricio, J. García, and J. M. Molina, Improving the Segmentation Stage
13   of a Pedestrian Tracking Video-Based System by Means of Evolution Strategies. In *Lecture
14   Notes in Computer Science*, Springer Science + Business Media, 2006, pp. 438–449.

15   [28] Jodoin, J.-P., G.-A. Bilodeau, and N. Saunier, Urban Tracker: Multiple object tracking in
16   urban mixed traffic. In *IEEE Winter Conference on Applications of Computer Vision*, IEEE,
17   2014.