Automated Classification in Traffic Video at Intersections with Heavy Pedestrian and Bicycle Traffic

Sohail Zangenehpour, Ph.D. student (Corresponding author)
Department of Civil Engineering and Applied Mechanics, McGill University
Room 165, Macdonald Engineering Building, 817 Sherbrooke Street West
Montréal (Québec) Canada H3A 2K6
Email: sohail.zangenehpour@mail.mcgill.ca

Luis F. Miranda-Moreno, Ph.D., Assistant Professor
Department of Civil Engineering and Applied Mechanics, McGill University
Room 268, Macdonald Engineering Building, 817 Sherbrooke Street West
Montréal (Québec) Canada H3A 2K6
Tel: +1 (514) 398-6589
Fax: +1 (514) 398-7361
Email: luis.miranda-moreno@mcgill.ca

Nicolas Saunier, ing., Ph.D., Assistant professor
Department of civil, geological and mining engineering
École Polytechnique de Montréal, C.P. 6079, succ. Centre-Ville
Montréal (Québec) Canada H3C 3A7
Phone: +1 (514) 340-4711 ext. 4962
Email: nicolas.saunier@polymtl.ca

Word count

Text	5500
Tables (1 X 250)	250
Figures (6 X 250)	1500
References	1200
Total	*7250*

Date of submission: **August 1st, 2013**

ABSTRACT

Pedestrians and cyclists are vulnerable road users and despite their limited presence in traffic events, these two groups have the most collisions resulting in injuries and fatalities. Due to problems regarding data collection for pedestrians and cyclists, there is a shortcoming in the field of road safety with regards to the availability and quality of data for non-motorized modes. Also, due to the constant change of orientation and appearance of pedestrians and cyclists, detecting and tracking them is a hard task. This is one of the reasons why automated data collection methods have mainly been developed to detect and track motorized traffic. This paper presents a methodology based on Histogram of Oriented Gradients to extract features of an image box containing the tracked object and Support Vector Machine as a classifier, to classify moving objects in crowded traffic scenes. This method classifies moving objects into three main types of road users: pedestrians, cyclists, and motor vehicles. This is done by first tracking each moving object in the video, classifying its appearance in each frame and then computing the probability of belonging to each class based on its appearance and speed. Bayes' rule is used to fuse appearance and speed to predict the class for each object. Testing results show good performance, with an overall accuracy of more than 90 %.

INTRODUCTION

With the increase in computing power and capacity of sensors coupled with their decreasing price, the field of Intelligent Transportation System (ITS) has seen considerable improvements in automatic traffic monitoring systems. The aim is not only to collect macroscopic traffic data, e.g. flow, density and average speed at specific locations in the road network, but also detailed microscopic information about each road user (position and speed) continuously and over large areas of the network. A great amount of the workload of traffic monitoring will thus shift from human operators to these automated systems with improved performance and the possibility to perform new tasks such as road safety monitoring *(1)*.

Intersections are critical elements of the road network for safety given that a high concentration of conflicts, crashes and injuries occurs at these locations. With the promotion and increase of non-motorized transportation in North American cities, the safety of non-motorized users at intersections has gained a lot of attention. In cities like Montreal, 60 % of pedestrian and cyclist injuries occur at intersections *(2)*. Given the importance of this topic in research and practice, several recent studies have looked at different safety issues at intersections using traditional approaches based on historical crash data *(3)* and surrogate approaches such as conflict analysis *(4)*. Independent of the method for road safety diagnosis, obtaining macroscopic and microscopic traffic data is fundamental. In the traditional approach, exposure measures are often developed based on traffic counts of each user type (e.g., vehicular, pedestrian and bicycle volumes). In the surrogate approach, road user trajectories are necessary to compute typical measures such as Time To Collision (TTC), Post Encroachment Time (PET), and gap time *(5)*.

Road users can be detected and classified using a variety of sensors like inductive-loops, magnetic sensors, microwave and laser radars, infrared and ultrasonic sensors *(6)*. However, it seems that the most convenient way to obtain data, such as road user trajectories, over a certain area is through the use of video sensors. Video sensors have several advantages, in particular the ability to capture naturalistic movements of road users with a small risk of catching their attention, the relative ease of installation, the richness of extracted data and the relatively low cost *(7)*. Their weaknesses are caused by low light conditions, adverse weather, and occlusion in high traffic conditions.

Automated video analysis involves the use of computer vision techniques to overcome many of the shortcomings associated with manual field observations and manual video analysis *(8)*. Tracking and collecting observational data for cyclists and pedestrians is more difficult than for vehicles because of their non-rigidity, their more varied appearances and less organized movements. In addition, they often move in groups close to each other which make them even harder to detect and track.

There are two approaches to extract classified road user trajectories from video: either tracking all moving objects and then classifying them in several categories of road users, or detecting road users in the successive video frames and connecting the detections (i.e. tracking by detection *(9)*). Trackers with reasonable performance are available in the transportation field *(10)(11)*, including the open source project Traffic Intelligence (https://bitbucket.org/Nicolas/trafficintelligence/). However, classification algorithms of user trajectories are less popular and missing in current available software such as Traffic Intelligence. Classification is therefore performed after tracking, on the resulting tracked road user trajectories.

The objective of this paper is to develop and evaluate classifiers for at least three types of road users, in this case motor vehicles, cyclists and pedestrians, based on their speed and appearance in video. Five classifiers are designed to classify the tracked road users. The first classifier relies only on the speed of the tracked object to predict its type, while the second classifier uses only the appearance of each object in the video for classification. The third classifier combines speed with the object appearance through speed thresholds while the fourth classifier relies on Bayes' rule to fuse speed and appearance. Finally, the fifth classifier uses another probability-based combination of object speed and appearance.

The main contribution of this paper is the development of a method for the classification of different road users in crowded urban traffic scenes. Previous studies classified road users either in only two classes *(12)*, or in more classes in less complex environments such as highways where there are minor changes in the appearance of vehicles *(13)*. The classifiers designed in this paper classify tracked road users into three main road user types: pedestrian, cyclist and motor vehicle. The signalized intersection of Avenue des Pins and Rue Saint-Urbain in Montreal was selected during peak hours to test the method in a crowded traffic scene to compare the performance of the five classification algorithms. A final contribution is the release of the classifier code and the code used to produce the results presented in this paper.

The paper is organized as follows: first a review of previous work on the subject of tracking and classification of road users is provided. This is followed by a description of the developed system. The paper then presents and discusses the performance of the proposed classifiers, the results and finally the conclusions are drawn from the entire study.

BACKGROUND

The readers are referred to *(14)* for a general survey of object tracking. In *(15)*, the different approaches for the detection and tracking of road users are classified into:

1. Tracking by detection: in many cases, especially if the objects are well separated, this approach works well. Detection of objects is done using background modeling and subtraction with the current image *(16)* or deformable templates, i.e. a model of image appearance using color distribution, edge characteristics or texture *(17)*. Image classifiers can be trained on labeled data to detect road users *(18)(19)*.
2. Tracking using flow: when a deformable template specifying the appearance of an object is available, pixels in successive images can be matched. This approach is also called feature-based tracking and has been applied to traffic monitoring in *(10)*.
3. Tracking with probability: it is convenient to see tracking as a probabilistic inference problem in a Bayesian tracking framework. In simple cases, independent Kalman filters can be run successfully for each target *(20)*, but this approach will fail in scenes where the objects interact and occlude each other. This is called the data association problem and can be solved using particle filters and Markov chain Monte Carlo methods for sampling.

Although significant progress has been made in recent years, tracking performance is difficult to report and compare, especially when the systems are not publically available, and when benchmarks are rare and not systematically used.

Similarly to object detection and tracking, significant progress in object classification for images has been made over the recent years, but generic multi-class object classification is still a very challenging task. Most of the research boils down to the design and extraction of the best features or variables to describe the objects in the images. There are two main classes of description variables:

1. Variables describing the appearance of the object, i.e. the pixels. New features have been successfully developed, in particular being invariant to various image transformations like translation, rotation and scaling. Among them are the Histogram of Oriented Gradients features (HOG) *(19)*, Scale-Invariant Feature Transform features (SIFT) *(21)*, Speeded Up Robust Features (SURF) *(22)*, DAISY *(23)*, Local Binary Patterns (LBP) *(24)* and Fast Retina Keypoint (FREAK) *(25)*.
2. Variables describing the shape or contour of the object. A good overview can be found in *(26)*. The simplest are the area and aspect ratio of the bounding box of the object.

Once object instances are turned into numerical vectors, this becomes a more traditional classification problem that can be addressed using machine learning or other techniques to learn generative or discriminative models. A popular state of the art technique is Support Vector Machines (SVM) used for example in *(19)*. There is also a renewal of interest for nearest-neighbor techniques for object classification *(13)(27)*.

Road user classification is a useful addition to traffic monitoring systems and efforts have already been done in this area. An early simple system *(28)* classifies and then tracks vehicles and humans. The classification is done using a Mahalanobis-based distance and the correct classification ratio is respectively 86.8 % and 82.8 % for vehicles and humans.

Fitting a 3D model is another way to classify objects in traffic monitoring. Complex 3D models are used in *(29)* to classify vehicles into seven classes. The object description includes other visual features such as brightness and color histograms. A SVM classifier can also be used to differentiate between subclasses, such as between bicycles and motorcycles or between buses and trucks. A global detection rate as high as 92.5 % has been reported, however this value varies for different classes. In *(30)*, in simple highway settings, using feature-based tracking as well as the number of features making up the object's height, over 90 % of road users were correctly classified. The work presented in *(13)* extracts the standard description of blobs by simple morphological measurements and targets real-time traffic monitoring on highways. Its performance is not clear as it reports results for different confidence levels. Although the work presented in *(31)* is called unsupervised by its authors, using k-means, it implicitly relies on prior knowledge of the road users in the scene. The description variables are the velocity of the object area, the "compactness", defined as the ratio of the object area over the square of the object perimeter, the time derivative of the area and the angle between the motion direction and the direction of the major axis of the shape. It has to be mentioned that none of these studies focused on busy locations like at intersections with high levels of cyclist and pedestrian traffic.

The method to count and classify composite objects presented in *(12)* relies on various descriptors combined in a Naïve Bayes framework or simply concatenated as inputs to a SVM classifier. The reported classification accuracy is 92 % and the counting accuracy is 95 %. A follow up on *(7)* is presented in *(32)*. After tracking each moving object in video, the type is classified based on speed profile information, like maximum speed and stride frequency. In this work, a classification accuracy of 94.8 % and 88.6 % are reported respectively for binary classification of motorized vs. non-motorized road users and for the classification of three main types of road users.

Finally, an idea common to most of the research presented in this section is the use of multiple detections provided by a tracking system at each frame. Integrating the instantaneous classification, the system achieves more robust performance (e.g. see *(20)* for some quantitative results that illustrate this point).

METHODOLOGY

The classifiers have to be calibrated or trained before they can be applied to classify road users. These two steps are shown in Figure **1**. In this section, the main elements of the chosen classification method are described and then the five different classifiers are presented in detail.

Tracker

The proposed approach classifies the output of a generic feature-based moving object tracker *(10)*. This algorithm can be summarized in two steps:

1. Individual pixels are detected and tracked from frame to frame and recorded as feature trajectories using the Kanade Lucas Tomasi feature tracking algorithm *(33)*.
2. A moving object is composed of many features which must therefore be grouped. Feature trajectories are grouped based on consistent common motion.

The parameters of this algorithm are tuned through trial and error, leading to a trade-off between over-segmentation (one object being tracked as many) and over-grouping (many objects tracked as one). Readers are referred to *(10)* for more details.

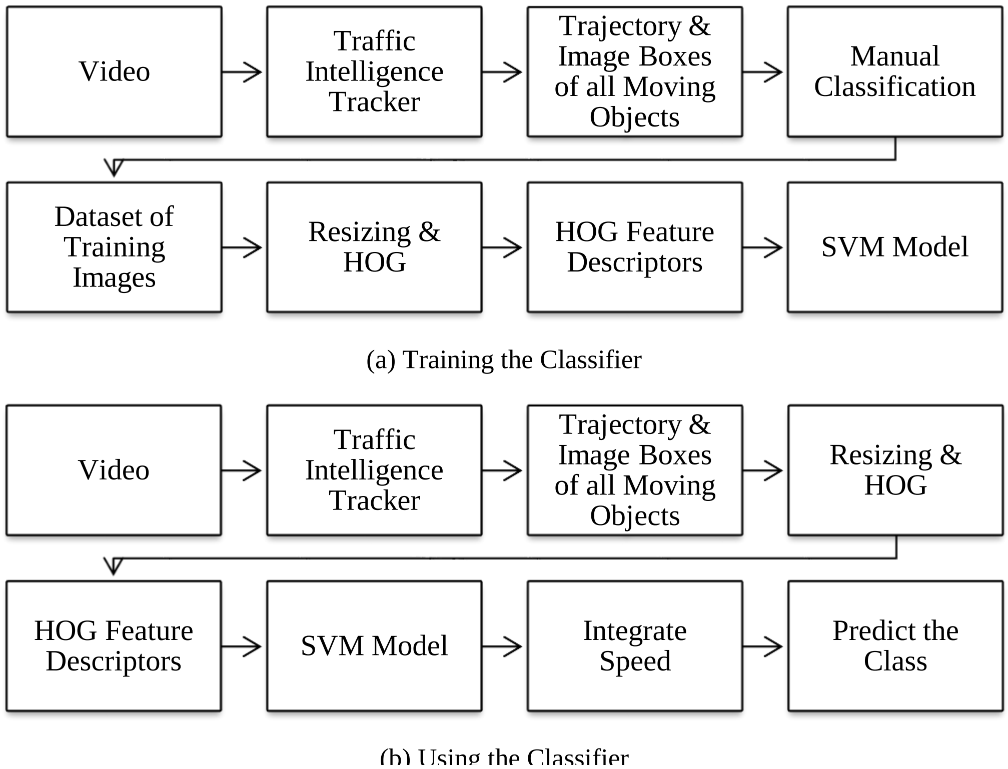
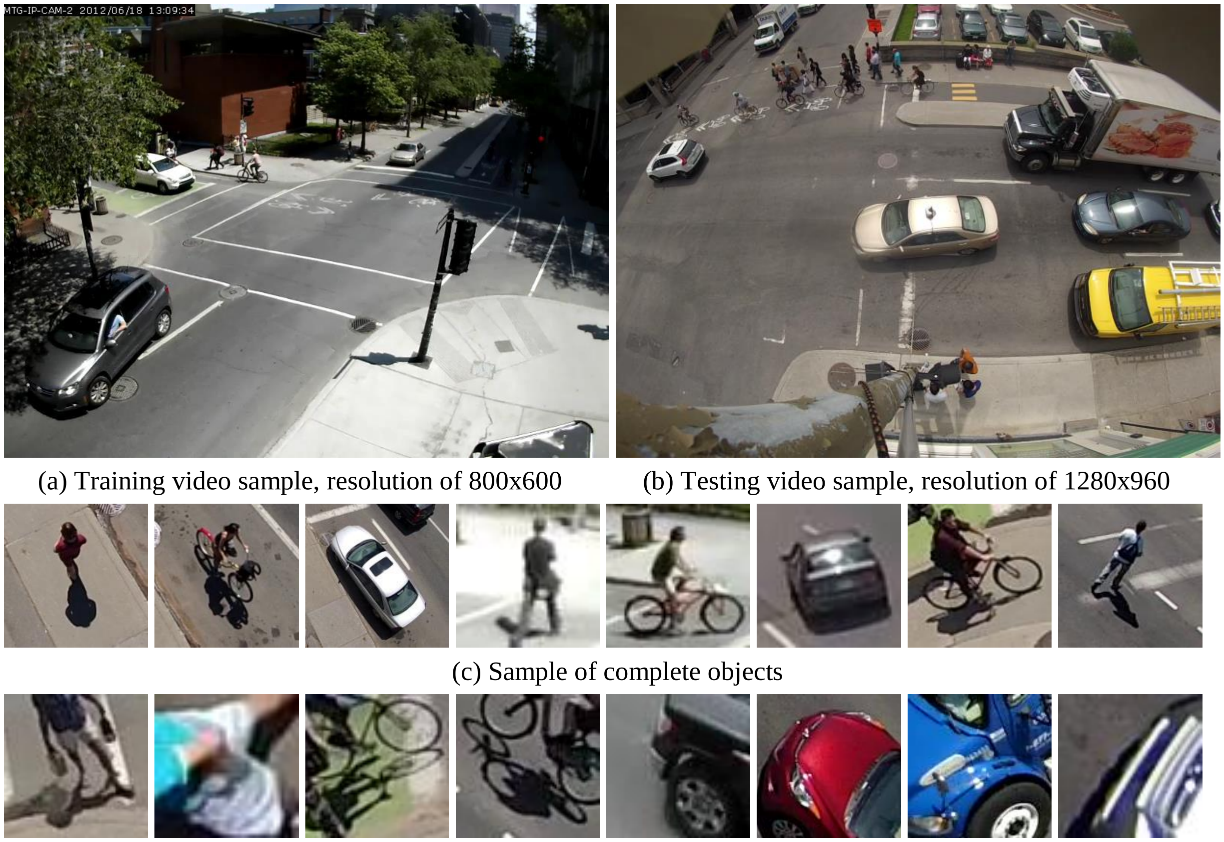


Figure 1. Steps involved in (a) training the classifier and (b) predicting the class of each object

Dataset and Modeling

A dataset containing images for each road user class, pedestrians, cyclists and vehicles, is used to train the appearance-based classifiers. Using the object trajectory provided by the tracker, the bounding boxes of the features on each moving object are automatically computed: the region of interest within the bounding box is saved and then manually classified into three groups: pedestrian, cyclist, and motor vehicle. It is worth mentioning that:

1. The videos used for extracting training data are different from the video used to test the algorithm performance.
2. For the training dataset, two different cameras with different resolution and view angle were used in locations different from where the testing videos were recorded. This implies that the algorithm does not have a high sensitivity to camera resolution or angle as well as to the site under study as can be seen in Figure 2a,b.
3. The tracker does not necessarily track the entire object. It is possible that parts of the pedestrian, cyclist or vehicle are not within the extracted image box. In this case, only part of a pedestrian body or a wheel or bumper of a vehicle is being tracked. Since this situation will occur also at prediction time, these object portions are added to training dataset as well (Figure 2c,d).



(a) Training video sample, resolution of 800x600 (b) Testing video sample, resolution of 1280x960

(c) Sample of complete objects

(d) Sample of objects which do not include the entire pedestrian/cyclist/vehicle

Figure 2. Sample of extracted road user images used for training and testing

Feature Descriptor

The first element to select in an appearance-based classifier is the description feature or descriptor best suited to discriminate between road user classes. Among the many image descriptors documented in the literature, Histogram of Oriented Gradients (HOG) is used as it has been applied with success to object classification, in particular pedestrian detection in static images *(19)* and vehicle detection *(34)*. HOG features concentrate on the contrast of silhouette contours against the background. It works by dividing each image into cells in which histograms of gradient directions or edge orientations are computed. The cells making up the image can have different illumination and contrast which can be corrected by normalization. This is achieved by grouping together adjacent cells into larger connected blocks and calculating a measure of intensity for these new blocks. The individual cells within the block can then be normalized based on the larger block. The HOG algorithm used in this work is an open source machine learning library for Python programming language (available at http://scikit-image.org/).

Feature Classification

The next step is to classify the chosen descriptors into the different road user classes to obtain the base appearance-based classifier. Supervised learning methods are used for classification tasks where the classes are known and labeled instances can be obtained *(35)*. In this work, labeled instances are the HOG features computed over an image sub-region, with their expected label. These labels correspond to the road user type (vehicle, pedestrian, and cyclist). A training algorithm builds a model of the labeled data that can then be applied to new, unlabeled input data, to predict their class. Artificial neural networks *(36)*, K-Nearest Neighbors (KNN) *(37)* and Support Vector Machine (SVM) *(38)* are well-known supervised classifiers.

Among the many methods and models developed in the field of machine learning, SVMs are one of the most commonly used classifiers which have good generalization *(38)* and one has been used as the classifier model in this paper.

SVM is by nature a binary classifier: for multi-class problems, several strategies exist in the literature, such as “one versus rest” where a classifier is trained for each class, or “one versus one” where a classifier is trained for each pair of classes. The SVM algorithm used in this work is the open source implementation LibSVM *(39)* available in the OpenCV library which uses the “one versus one” strategy: the final class is decided by majority vote of the underlying SVMs. This appearance-based classifier is called HOG-SVM.

Speed Information

Aside from the appearance of an object in the video, another criterion that can help predict the type of the object is its speed *(8)*. Speed can be aggregated over time and compared to a threshold to eliminate a possible object type. For example, it is nearly impossible for a pedestrian to walk at a speed of 15 km/h. Speed can also be combined with other information, such as appearance, using probability principles to increase the classification accuracy. In this study, alternative methods to combine criteria are used to design and test different classifiers.

To use speed as a criterion, one first needs to define a discriminative aggregated indicator of the instantaneous speed measurements made in each frame. The usual aggregation functions are: maximum, mean, median or percentiles of the speed measurements (e.g., 85^{th}). Since the speed given by the tracker may be noisy and the maximum and mean are sensitive to noise, the median is used. From this point forward, speed refers to the median of a road user’s instantaneous speeds.

Classifier Design

Based on the two criteria, the median of the speed measurements and the classification of the HOG-SVM in each frame, the following classifiers are derived:

Classifier I is the simplest one and relies on two speed thresholds to predict the type of each object. These two speed thresholds are extracted from Figure 4d, and are chosen as the limit values that define each speed interval over which the three types of road users are each most probable. Accordingly, this classifier assigns objects with speed between 0 and 6.5 km/h as a pedestrian; with speed between 6.5 km/h and 14.5 km/h as a cyclist and, with speeds greater than 14.5 km/h as a vehicle.

Classifier II only uses the appearance of each object through the video to predict its type (with HOG-SVM). A method is needed to decide based on the multiple predictions made for each frame in which the object is tracked. The proportion of frames in which the object is classified as each road user class can be considered as a probability $P(Class|Appearance)$ and the most likely (the category with the highest number of detections) is the predicted type for the object.

Classifier III combines both appearance-based and speed-based classifiers based on a simple algorithm illustrated in Figure 3 to switch between the following three possible situations:

1. The speed of the tracked object is lower than the speed threshold selected for pedestrians; the object can either be a pedestrian, a cyclist, or a vehicle. In this situation a HOG-SVM classifier trained for the three classes is used.
2. The speed of the tracked object is lower than the speed threshold selected for cyclists but is higher than the speed threshold selected for pedestrians. The object cannot be a pedestrian; it can either be a cyclist or a vehicle. In this situation a binary HOG-SVM classifier trained for the two classes, cyclist and vehicle, is used. It is expected that a binary classifier outperforms a multi-class classifier.
3. The speed of the tracked object is higher than the speed threshold selected for cyclists. In this situation the object can only be a vehicle and no classifier is needed.

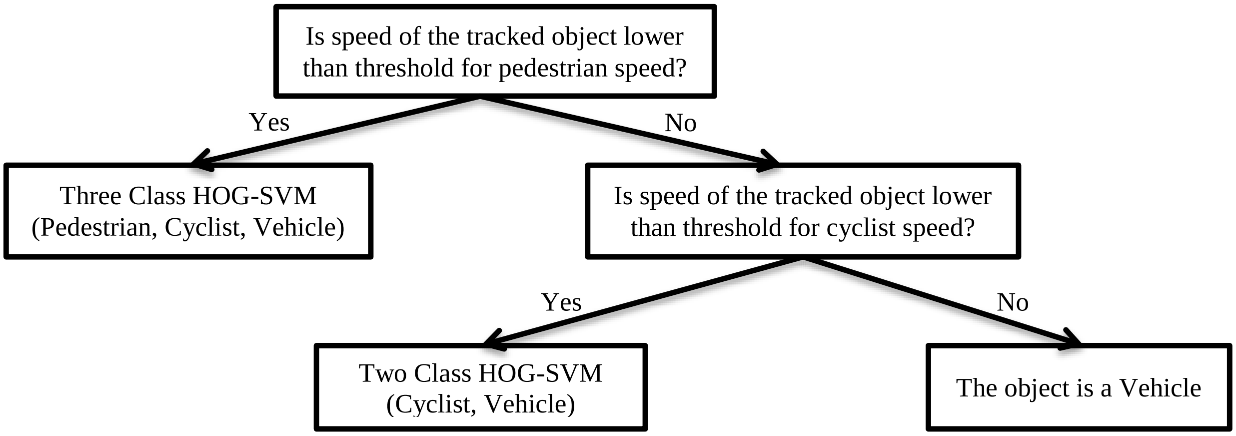


Figure 3. Flowchart showing the use of speed thresholds to switch between classifiers

Classifier IV combines the probability of each class given the speed and appearance information using the Bayes' rule and the typical (naïve) assumption of independence of these two pieces of information used for classification. To obtain this classifier, consider the typical Bayesian classifier given by the posterior distribution (likelihood × prior). This is obtained as:

$$P(Class \mid Speed, Appearance) = \frac{P(Class)}{P(Speed, Appearnce)} P(Speed, Appearnce \mid Class)$$

Then, by the assumption of independency among criteria:

$$P(Class \mid Speed, Appearance) = \frac{P(Class)}{P(Speed)P(Appearnce)} P(Speed|Class)P(Appearance|Class) \qquad (*)$$

Using conditional probability:

$$P(Appearance|Class)P(Class) = P(Class|Appearance)P(Appearance) \qquad (**)$$

By replacing $(**)$ into $(*)$:

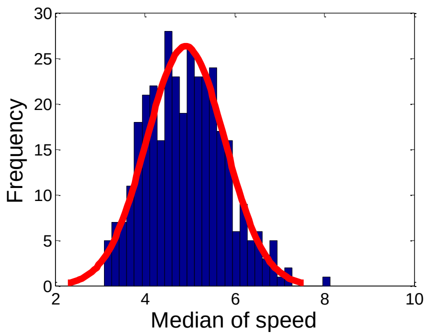
$$P(Class \mid Speed, Appearance) = \frac{P(Class|Appearance)}{P(Speed)}\ P\ (Speed|Class)$$

Finally given that $P(Speed)$ is independent of the classes, it can be said that:

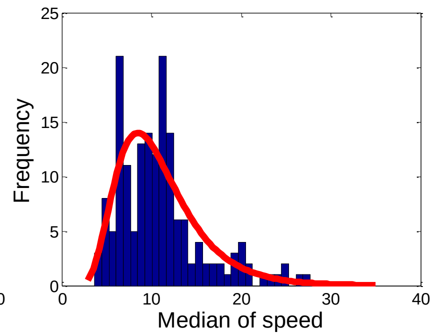
$$P(Class \mid Speed, Appearance) \propto P(Class|Appearance)\ P\ (Speed|Class)$$

$P\ (Speed|Class)$ is estimated through distributions fitted to the empirical speed distributions of the three road users classes, gathered through manual object classification in the sample video and shown in Figure 4a,b,c. The speed distributions of pedestrians and vehicles are fitted to normal distributions and the speed distribution of cyclists is fitted to a lognormal distribution. The parameters of these distributions are the following (see Figure 4d):

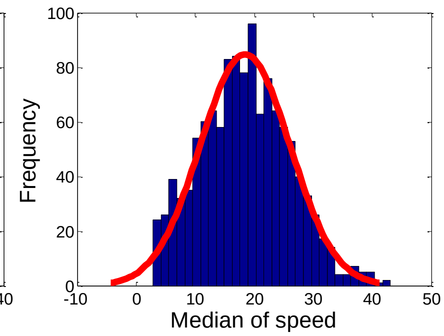
1. Pedestrian speed distribution: normal distribution with mean of $\overline{V_p}$=4.91 km/h and standard deviation of σ_p=0.88 km/h
2. Cyclist speed distribution: log-normal distribution with location parameter of $\overline{\mu_c}$=2.31 (mean of $\overline{V_c}$=11.00 km/h) and scale parameter of ς_c=0.42 (standard deviation of σ_c=4.83 km/h)
3. Vehicle speed distribution: normal distribution with mean of $\overline{V_v}$=18.45 km/h and standard deviation of σ_v=7.6 km/h



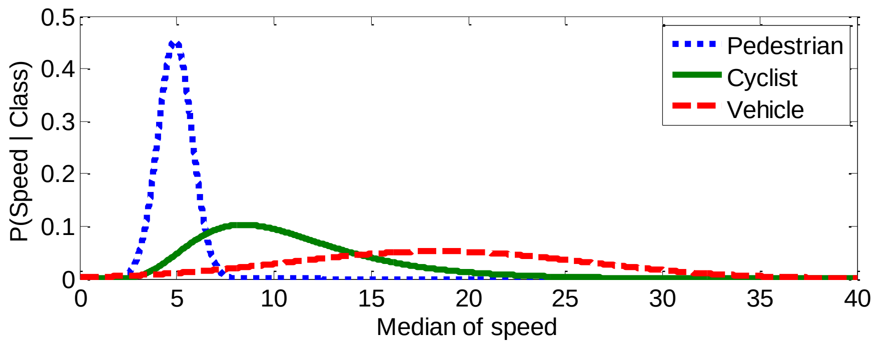
(a) Distribution of pedestrians' speed



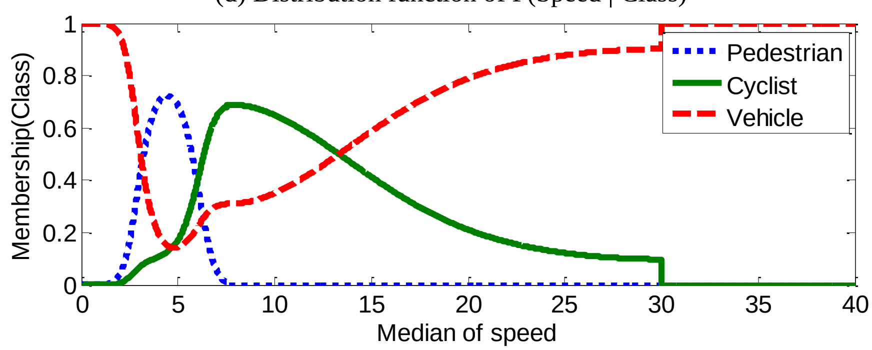
(b) Distribution of cyclists' speed



(c) Distribution of vehicles' speed



(d) Distribution function of P(Speed | Class)



(e) Membership function of object types

Figure 4. Speed distribution and membership function of each object type used for classifier's design

$P(Class|Appearance)$ is computed as for classifier II, over all HOG-SVM classification over the object existence. Finally, the class of the object is selected as the one with highest $P(Class|\ Speed, Appearance)$.

Classifier V is similar to Classifier IV; it uses probability functions to combine speed and appearance information. The distribution functions of the road users' speeds (Figure 4) are used to determine the membership functions for each object type (mean value and standard deviation for each user type are the same as the values used for classifier IV) (see Figure 4e). The sum of the membership functions for each speed is equal to one (because thresholds on pedestrian and cyclist speeds are taken into account, the membership function for vehicles is equal to one for speeds higher than the cyclist threshold) and can be calculated as:

$$Membership(Pedestrian) = \frac{\exp\left\{-\frac{(V_o - \bar{V}_p)^2}{2\sigma_p^2}\right\}}{\exp\left\{-\frac{(V_o - \bar{V}_p)^2}{2\sigma_p^2}\right\} + \exp\left\{-\frac{[\ln(V_o) - \bar{\mu}_c]^2}{2\varsigma_c^2}\right\} + \exp\left\{-\frac{(V_o - \bar{V}_v)^2}{2\sigma_v^2}\right\}}$$

$$Membership(Cyclist) = \frac{\exp\left\{-\frac{[\ln(V_o) - \overline{\mu}_c]^2}{2\varsigma_c^2}\right\}}{\exp\left\{-\frac{(V_o - \overline{V}_p)^2}{2\sigma_p^2}\right\} + \exp\left\{-\frac{[\ln(V_o) - \overline{\mu}_c]^2}{2\varsigma_c^2}\right\} + \exp\left\{-\frac{(V_o - \overline{V}_v)^2}{2\sigma_v^2}\right\}}$$

$$Membership(Vehicle) = \frac{\exp\left\{-\frac{(V_o - \overline{V}_v)^2}{2\sigma_v^2}\right\}}{\exp\left\{-\frac{(V_o - \overline{V}_p)^2}{2\sigma_p^2}\right\} + \exp\left\{-\frac{[\ln(V_o) - \overline{\mu}_c]^2}{2\varsigma_c^2}\right\} + \exp\left\{-\frac{(V_o - \overline{V}_v)^2}{2\sigma_v^2}\right\}}$$

Here V_o is the speed of the object being classified. Finally the class of the object is selected based on the highest value of $P(Class|Appearance) * Membership(Class)$. The implementations of the classifiers, as well as the training and testing functions, are available under an open source license on the paper webpage http://nicolas.saunier.confins.net/data/zangenehpour14trb.html.

RESULTS

Video data was collected at two intersections for labeled training data for HOG-SVM, and at the signalized intersection of Avenue des Pins and Rue Saint-Urbain in Montreal during peak hours for testing the classifiers' performance. Two different video cameras were used, with resolutions of 800x600 and 1280x960 and a frame rate of 15 fps (Figure 2a,b). One of the cameras has fisheye lens with an ultra-wide field of view.

The chosen parameters of the HOG feature descriptor are 9 even orientations, with 8x8 pixels for each cell, and each block made up of 2x2 cells. Before using HOG, bounding boxes are converted to grayscale images with a normalized size of 64x64 pixels. For the SVM model used for classification, the nonlinear kernel functions were Gaussian Radial Basis Function (RBF). Speed thresholds used in classifier I are 6.5 km/h for pedestrians and 14.5 km/h for cyclists. Speed thresholds for classifiers III, IV, and V are 7.5 km/h for pedestrians and 30 km/h for cyclists.

To test the accuracy of the designed classifiers, a different video from the training phase was used. This video is 232 minutes long and a total of 4756 objects were manually classified to create the ground truth. The predicted class (by each automated classifier) and the ground truth (observed, manually labelled) were then compared to compute the accuracy of each classifier.

For multi-class problems, it is crucial to report performance measures for each class and not only the global accuracy. The components of the confusion matrix c_{ij} are the number of objects of true class *i* predicted in class *j*. The performance measures are thus defined for class *k*:

$$Recall_k = \frac{c_{kk}}{\sum_j c_{kj}} \qquad Precision_k = \frac{c_{kk}}{\sum_i c_{ik}} \qquad Accuracy = \frac{\sum_k c_{kk}}{\sum_i \sum_j c_{ij}}$$

The results are shown in Table 1. Classifier IV and classifier V have the best performance among the tested classifiers. Classifier IV has the best recall rate for pedestrians and best precision for vehicles, while classifier V has the best recall rate for vehicles (and second best recall rate for cyclists after classifier III, with only 0.2 % difference) and best precision for pedestrians and cyclists. In the test video the majority of the traffic was motorized vehicles (around 68 %) with fewer pedestrians (around 22 %) and cyclists (around 10 %). In order to estimate the performance of the two best designed classifiers if the traffic had the same number of road users in each class, the performance for a balanced number of observations of each user type (100 observations for each type) is also shown in Table 1. This illustrates that the accuracy changes when the class distribution changes, and explains that the precision for cyclists is low in part because of relatively few cyclists in the video.

It is worth mentioning that several misclassifications occurred in cases where multiple objects were tracked as a single object (over-grouping problem of tracker) or when only a portion of an object was tracked (over-segmentation problem of tracker). Some samples of these situations are shown in Figure 5.

Table 1. Confusion matrices showing each classifier's performance

			Ground Truth					**Accuracy**
			Pedestrian	Bike	Vehicle	Total	**Precision**	
Predicted	**Classifier I**	Pedestrian	946	86	277	1309	72.3 %	**72.4 %**
		Bike	77	324	793	1194	27.1 %	
		Vehicle	0	78	2175	2253	96.5 %	
		Total	1023	488	3245	4756		
		Recall	92.5 %	66.4 %	67.0 %			
	Classifier II	Pedestrian	742	191	584	1517	48.9 %	**75.9 %**
		Bike	121	244	37	402	60.7 %	
		Vehicle	160	53	2624	2837	92.5 %	
		Total	1023	488	3245	4756		
		Recall	72.5 %	50.0 %	80.9 %			
	Classifier III	Pedestrian	726	43	64	833	87.2 %	**86.3 %**
		Bike	131	373	177	681	54.8 %	
		Vehicle	166	72	3004	3242	92.7 %	
		Total	1023	488	3245	4756		
		Recall	71.0 %	76.4 %	92.6 %			
	Classifier IV	Pedestrian	969	53	180	1202	80.6 %	**88.5 %**
		Bike	42	371	198	611	60.7 %	
		Vehicle	12	64	2867	2943	97.4 %	
		Total	1023	488	3245	4756		
		Recall	94.7 %	76.0 %	88.4 %			
	Classifier V	Pedestrian	889	38	82	1009	88.1 %	**90.3 %**
		Bike	58	372	130	560	66.4 %	
		Vehicle	76	78	3033	3187	95.2 %	
		Total	1023	488	3245	4756		
		Recall	86.9 %	76.2 %	93.5 %			
	Classifier IV (balanced observation)	Pedestrian	95	11	6	112	84.8 %	**86.3 %**
		Bike	4	76	6	86	88.4 %	
		Vehicle	1	13	88	102	86.3 %	
		Total	100	100	100	300		
		Recall	95.0 %	76.0 %	88.0 %			
	Classifier V (balanced observation)	Pedestrian	87	8	3	98	88.8 %	**85.3 %**
		Bike	6	76	4	86	88.4 %	
		Vehicle	7	16	93	116	80.2 %	
		Total	100	100	100	300		
		Recall	87.0 %	76.0 %	93.0 %			

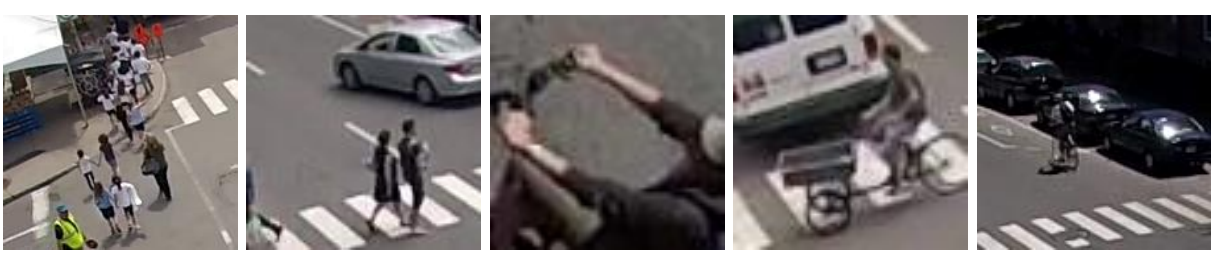
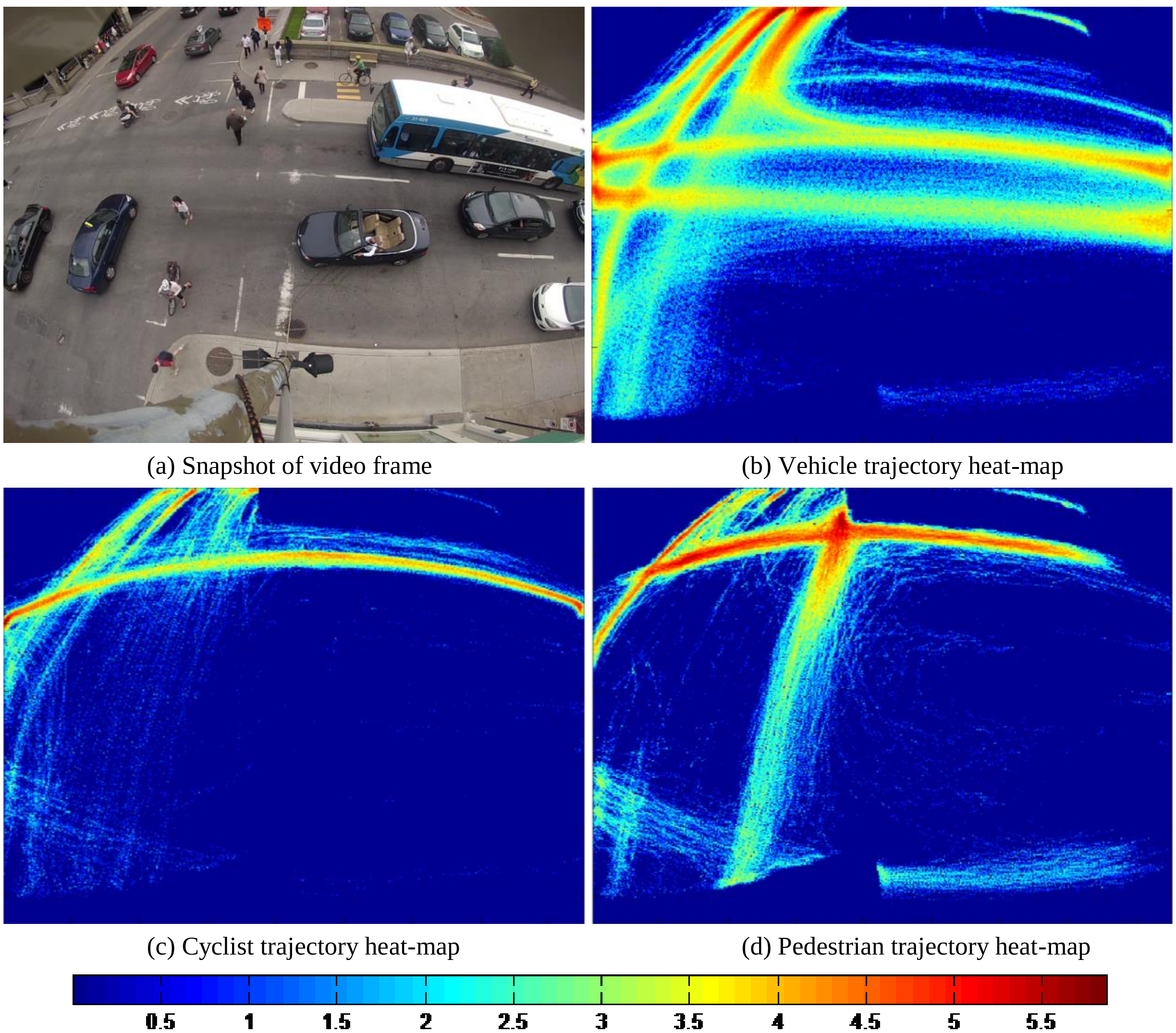


Figure 5. Example of situations hard to classify

The performance of the classifier relies on that of the tracker and therefore any error in the tracking may lead to errors in the classification process (or ambiguity in classification when road users of different types are not distinguished by the tracker). In most cases, even when the tracker only identified part of a pedestrian, cyclist or vehicle, the classifier was still able to classify the object correctly.

Another way to visualize the results of the proposed classifier is through heat-maps (frequency of trajectory or positions in discretized two-dimensional bins of the space) for the three road user classes (Figure 6).

The heat-maps show the good performance of classifier V since the trajectories of the different road user types are overall in the expected locations: pedestrians are on the sidewalks and crosswalks, cyclists are mostly in the cycle track, and vehicles are on the road and in the lanes. The other interesting information is the area where the classifier makes errors. For example a few cyclists in the cycle track have been classified as vehicles or there are some vehicles at the top of the camera view which are classified as pedestrians or cyclists.



(a) Snapshot of video frame

(b) Vehicle trajectory heat-map

(c) Cyclist trajectory heat-map

(d) Pedestrian trajectory heat-map

(e) Scale used for trajectorie heat-maps (log-scale)

Figure 6. Snapshot of video and position heat-maps for the three road user types (taken from Classifier V). The most and least used map locations are respectively red and blue (heat-map colours range from blue to red, passing through cyan, yellow, and orange) (the resolution of each heat-map cell is 3x3 pixel with respect to the camera resolution).

DISCUSSION

Since the tested classifiers have different precision and recall rates, the choice of the best classifier depends on the application and preference for missed detections or false alarms for one class or another. For example, if it is important to detect as many pedestrians as possible at the expense of other road users being classified as pedestrian, classifier IV is the best (recall rate of 94.7 % for pedestrians). On the other hand, if it is important that no other road user other than pedestrian is classified as pedestrian then classifier V is the best (precision of 88.1 % for pedestrians). Overall, classifiers IV (accuracy of 88.5 %) and V (accuracy of 90.3 %) have the best performance among the tested classifiers. There are several ways to improve the accuracy of the designed classifiers:

1. Using video data from different viewpoints to train the classifier. Using this approach, the classifier is generalized for different camera angles. One question is whether the performance will break down if the viewpoints become too different. The question of using more consistent viewpoints with the same angle is also raised as it may improve appearance-based classification by reducing the variability of object appearance.
2. As discussed previously, the classifier accuracy relies on the performance of the tracker algorithm, so a way to improve classification accuracy is to improve the tracking algorithm. These are some ideas that can help improve the tracking performance:
   i. Increase the camera angle to see objects separate from each other in crowded scenes as one of the major issues of the tracker is over-grouping in dense traffic.
   ii. Compensate the fisheye effect of the camera lens. A camera with a fisheye lens was used to cover as much of the intersection as possible. However, fisheye lenses produce strong visual distortion on the corners of the video frame (Figure 2b). This effect reduces the accuracy of the tracker to map the position of objects in real world coordinates and speed estimation. By correcting for the fisheye effect of the camera, the usage of position and speed of an object will be more reliable for classification.
3. In this paper HOG and SVM with a radial basis function were used as feature descriptor and classifier. Their parameters have been selected through trial and error and should be thoroughly tested. In addition, other feature descriptors and classifiers should be tested to see if better accuracy can be achieved.
4. Background subtraction is another possible way to increase the performance of the classifiers, especially to obtain more precise images of each object (more precisely around its contour).

CONCLUSION

The important value of microscopic data classified by user type is more and more recognized in the transportation literature in general and in traffic safety in particular.

This paper presented algorithms to design classifiers capable of classifying moving objects in crowded traffic video scenes (like intersections), into three main road user types: pedestrians, cyclists, and motor vehicles. Given the limitations of simple classification methods based on speed and appearance, this research combines these methods through several classifiers in order to improve the classification performance.

Among the five tested classifiers, the classifiers that combine the probability of both the speed and appearance of objects show systematically better performance. Accuracy for the best classifier (Classifier V) was more than 90 %. Due to the similarity in appearance between pedestrians and cyclists (a cyclist consists of a bicycle and a human who rides the bicycle) and of the large range of cyclist speed, cyclists are the most difficult road user to classify. False positive rates for the best classifier are 11.9 % for pedestrians, 33.6 % for cyclists, and 4.8 % for vehicles, while the rates for false negative are 13.1 %, 23.8 %, and 6.5 %, respectively.

An final contribution is to release the code used to train and test the different classifiers under an open source license to enable other researchers to reproduce the methods and improve upon them more easily.

Future work will explore changing the parameters of the appearance descriptor and classifier and incorporating additional information to improve the performance of road user classification, especially that of cyclists who form the main part of the error. Finally, these classification methods will enable the study of classified road user trajectories, in particular the influence of different factors on the safety of vulnerable users at intersections such as the influence of cycle tracks on conflicts between cyclists and right turning motor vehicles.

ACKNOWLEDGMENTS

This work is part of a research project on vulnerable road users' safety funded by the Québec Fund for Research on Nature and Technology (FRQNT), the Québec Ministry of Transportation (MTQ), and the Québec Fund for Health Research (FRQS) as part of their research program on road safety. The authors also would like to thank Taras Romancyshyn and Ting Fu for helping with video recording and Jillian Strauss for proofreading this paper. All remaining errors and the views expressed in this research are, however, solely ours.

REFERENCES

1. *Traffic Monitoring Guide*. U.S. Department of Transportation, Federal Highway Administration, Washington, DC, 2001.
2. Strauss, J., L. F. Miranda-Moreno, and P. Morency. Cyclist Activity and Injury Risk Analysis at Signalized Intersections: A Bayesian Modeling Approach. *Accident Analysis & Prevention*, Vol. 59, May 2013, pp. 9–17.
3. Miranda-Moreno, L. F., J. Strauss, and P. Morency. Disaggregate Exposure Measures and Injury Frequency Models of Cyclist Safety at Signalized Intersections. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2236, No. 1, Dec. 2011, pp. 74–82.
4. Ismail, K., T. Sayed, and N. Saunier. Automated pedestrian safety analysis using video data in the context of scramble phase intersections. *Annual Conference of the Transportation Association of Canada*, 2009.
5. Saunier, N., T. Sayed, and K. Ismail. Large-Scale Automated Analysis of Vehicle Interactions and Collisions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2147, Dec. 2010, pp. 42–50.
6. Klein, L. A., M. k. Mills, and D. R. P. Gibson. *Traffic Detector Handbook: Third Edition - Volume I*. 2006.
7. Saunier, N., A. El Husseini, K. Ismail, C. Morency, J.-M. Auberlet, and T. Sayed. Estimation of Frequency and Length of Pedestrian Stride in Urban Environments with Video Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2264, No. 1, Dec. 2011, pp. 138–147.
8. Ismail, K., T. Sayed, and N. Saunier. Automated collection of pedestrian data using computer vision techniques. *Transportation Research Board Annual Meeting*, 2009.
9. Breitenstein, M. D., F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online Multi-Person Tracking-by-Detection from a Single, Uncalibrated Camera. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 33, No. 9, Dec. 2010, pp. 1820–1833.
10. Saunier, N., and T. Sayed. A feature-based tracking algorithm for vehicles in intersections. *The 3rd Canadian Conference on Computer and Robot Vision (CRV'06)*, 2006, pp. 59–59.

11. Jackson, S., L. F. Miranda-Moreno, P. St-Aubin, and N. Saunier. A Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioural Analysis. *Transportation Research Board 92nd Annual Meeting*, 2013.

12. Somasundaram, G., R. Sivalingam, V. Morellas, and N. Papanikolopoulos. Classification and Counting of Composite Objects in Traffic Scenes Using Global and Local Image Analysis. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 14, No. 1, Mar. 2013, pp. 69–81.

13. Morris, B., and M. Trivedi. Learning, Modeling, and Classification of Vehicle Track Patterns from Live Video. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, No. 3, Sep. 2008, pp. 425–437.

14. Yilmaz, A., O. Javed, and M. Shah. Object tracking. *ACM Computing Surveys*, Vol. 38, No. 4, Dec. 2006, p. 13–es.

15. Forsyth, D. a., O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan. Computational Studies of Human Motion: Part 1, Tracking and Motion Synthesis. *Foundations and Trends® in Computer Graphics and Vision*, Vol. 1, No. 2/3, 2005, pp. 77–254.

16. Antonini, G., S. V. Martinez, M. Bierlaire, and J. P. Thiran. Behavioral Priors for Detection and Tracking of Pedestrians in Video Sequences. *International Journal of Computer Vision*, Vol. 69, No. 2, May 2006, pp. 159–180.

17. Gavrila, D. M., and S. Munder. Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle. *International Journal of Computer Vision*, Vol. 73, No. 1, Jul. 2006, pp. 41–59.

18. Wu, B., and R. Nevatia. Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. *International Journal of Computer Vision*, Vol. 75, No. 2, Jan. 2007, pp. 247–266.

19. Dalal, N., and B. Triggs. Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1, 2005, pp. 886–893.

20. Hsieh, J.-W., S.-H. Yu, Y.-S. Chen, and W.-F. Hu. Automatic Traffic Surveillance System for Vehicle Tracking and Classification. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 7, No. 2, Jun. 2006, pp. 175–187.

21. Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Vol. 60, No. 2, Nov. 2004, pp. 91–110.

22. Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, Vol. 110, No. 3, Jun. 2008, pp. 346–359.

23. Tola, E., V. Lepetit, and P. Fua. DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, May 2010, pp. 815–830.

24. Ojala, T., M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, Jul. 2002, pp. 971–987.

25. Alahi, a., R. Ortiz, and P. Vandergheynst. FREAK: Fast Retina Keypoint. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 510–517.

26. Bose, B. Classifying Tracked Moving Objects in Outdoor Urban Scenes. *EECS Department, MIT. Research Qualifying Examination report*, 2005.

27. Hasegawa, O., and T. Kanade. Type classification, color estimation, and specific target detection of moving targets on public streets. *Machine Vision and Applications*, Vol. 16, No. 2, Feb. 2005, pp. 116–121.

28. Lipton, A., H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real-time video. *Proceedings Fourth IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201)*, 1998, pp. 8–14.

29. Messelodi, S., C. M. Modena, and M. Zanin. A computer vision system for the detection and classification of vehicles at urban road intersections. *Pattern Analysis and Applications*, Vol. 8, No. 1-2, Jul. 2005, pp. 17–31.

30. Kanhere, N., and S. Birchfield. Real-Time Incremental Segmentation and Tracking of Vehicles at Low Camera Angles Using Stable Features. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, No. 1, Mar. 2008, pp. 148–160.

31. Zhang, Z., Y. Cai, K. Huang, and T. Tan. Real-Time Moving Object Classification with Automatic Scene Division. *2007 IEEE International Conference on Image Processing*, 2007, pp. V – 149–V – 152.

32. Zaki, M. H., and T. Sayed. A framework for automated road-users classification using movement trajectories. *Transportation Research Part C: Emerging Technologies*, Vol. 33, Aug. 2013, pp. 50–73.

33. Birchfield, S. KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker. http://www.ces.clemson.edu/~stb/klt/.

34. Kembhavi, A., D. Harwood, and L. S. Davis. Vehicle detection using partial least squares. *IEEE transactions on pattern analysis and machine intelligence*, Vol. 33, No. 6, Jun. 2011, pp. 1250–65.

35. Aha, D. W., D. Kibler, and M. K. Albert. Instance-based learning algorithms. *Machine Learning*, Vol. 6, No. 1, Jan. 1991, pp. 37–66.

36. White, H. Learning in Artificial Neural Networks: A Statistical Perspective. *Neural Computation*, Vol. 1, No. 4, Dec. 1989, pp. 425–464.

37. Hastie, T., and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 6, Jun. 1996, pp. 607–616.

38. Burges, C. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, Vol. 167, 1998, pp. 121–167.

39. Chang, C.-C., and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, Vol. 2, No. 3, Apr. 2011, pp. 1–27.