A Prototype System for Truck Signal Priority (TkSP) using Video Sensors

Nicolas Saunier, University of British Columbia

Tarek Sayed, University of British Columbia

Clark Lim, University of British Columbia

Paper prepared for presentation

at the “Innovations in Traffic Control Safety” Session

of the 2009 Annual Conference of the

Transportation Association of Canada

Vancouver, British Columbia

ABSTRACT

The efficient and safe movement of freight is one of the important goals of urban transportation systems and vital to not only the local economy, but nationally as well. Given the importance of the freight transportation system, opportunities to utilize state of the art Intelligent Transportation Systems (ITS) technologies are increasing to improve the operation of the existing infrastructure to promote the efficient and safe transportation of freight. Among these technologies, a Truck Signal Priority (TkSP) strategy gives priority to a traffic signal approach when trucks are detected. By using the rich data available through the use of video sensors, the use such a strategy can improve the efficiency and safety of freight movement by reducing truck travel time and the number of truck stops at the intersection.

This paper presents a prototype TkSP system using video sensors to detect, identify and track trucks. A classification module takes the road users' trajectories as input, and classifies them as either truck or non-truck. Using a mixture of Gaussians to model the background, the appearance and shape of road users in each frame is extracted. Using labeled shape data, a classifier is trained. A road user will be classified as a truck if its shape is classified as such in enough frames. The system is tested on real world data from the Next Generation SIMulation project (NGSIM). The truck recall reaches 78% to 95%, with a false alarm rate below the 0.5% value (used to test different scenarios in traffic simulation not presented in this document). This shows that the performance required for effective advanced truck signal priority is reached or within reach of automated video-based sensors.

1.Introduction

The efficient and safe movement of freight is one of the important goals of urban transportation systems. Provision of efficient infrastructure for freight movement provides the basis for regional and national economies, directly sustains hundreds of thousands of jobs, and distributes the necessities of life to every resident and business every day. Greater Vancouver faces a special challenge given its nature as a gateway city to the Pacific Rim. It is a point of significant intermodal transportation movements, air, water, rail and long-haul trucking. Trade moves through Greater Vancouver, both to and from the rest of Canada, the U.S. and Asia.

Given the importance of the freight transportation system and the limited urban space available for the expansion of the transportation network, it is becoming important to take advantage of state of the art Intelligent Transportation System (ITS) technologies to improve the operation of the existing infrastructure to promote the efficient and safe transportation of freight.

Truck Signal Priority (TkSP) is one of the methods that can be used to improve the efficiency and safety of freight movement without major capital investment. A TkSP strategy gives priority to a traffic signal approach when trucks are detected. By implementing a TkSP strategy, the truck travel time can be decreased and consequently the cost of goods movement reduced. In addition, there are safety benefits from reducing the number of stops of trucks approaching the intersection at the end of the green phase, which may reduce red light running. Reducing the number of stops for trucks should also have a positive effect on emissions and noise. Finally, it may be used to encourage trucks to use specific routes.

The goal of this project is to develop a prototype truck detection and tracking system using video sensors. Video sensors should be able to detect, identify, and track heavy trucks traveling within a corridor. Real time data collected by video sensors provide a rich source of information that can be readily available for decision on signal control. However, the information from such sensors can be marred with errors, one of the major challenges in the use of these sensor technologies. It seems that for the application of truck signal priority, false alarms are the more detrimental of the two types of detection errors, as they may trigger a signal priority in the absence of any truck. Hence particular attention was paid to minimize the false alarms rate.

The difficulty of road user tracking in video data depends on the environment. The system presented in this paper targets typically environments such as major roads and highways which may be qualified as more controlled and present easier behavior patterns (most road user movements are straight with occasional lane changes) than for example urban intersections. Other common problems for tracking are global illumination variations, shadow handling, multiple object tracking and classification.

This work relies on an existing system developed at the University of British Columbia for traffic data collection and especially for automated road safety analysis. It has already been used for traffic conflict detection [SS07], and to demonstrate a new probabilistic framework for automated road safety analysis [SS08]. The readers are referred to [SS06] for more details.

This paper presents the classification module that takes the road users' trajectories as input, and classifies them as either truck or non-truck. The next section presents the literature review on object detection in images. Section 3. describes the proposed approach for truck

classification. Section 4. will demonstrate the performance of the system on a few video sequences. Section 5. concludes and offers some perspectives on future work.

2. Literature Review on Object Classification in Images

Significant progress in object classification in images has been made over the recent years, but generic multi-class object classification is still a very challenging task. There are many types of road users, the inter-class differences in appearance can be subtle (e.g. there are nine classes of trucks in the classification of the Federal Highway Administration) while the intra-class variance in appearance can be high (e.g. with various painting colors and sizes). Large datasets have been made available recently for competitions and prove to be very complex for the most advanced techniques (See the PASCAL Visual Object Classes datasets and the corresponding competitions[1]).

Most of the research boils down to the design and extraction of the best features or variables to describe the objects in images. There are two main classes of description variables:

•Variables describing the appearance of the object, i.e. the pixels. New features have been successfully developed, in particular to be invariant to various image transformations like translation, rotation and scaling. Among them are the Histogram of Gradients features (HoG) [DT05], Scale-invariant feature transform features (SIFT) [Low04] and other edge-based features [MG05], and more plausible biological models based on Gabor filters. Other approaches use simple filters and classifiers to learn automatically the best image descriptors for the task [VJ01] [WMM+08].

•Variables describing the shape, or contour, of the object. A good overview can be found in [Bos05]. 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 techniques to learn generative or discriminative models. The readers are referred to [DH00] for a comprehensive introduction to the field of machine learning. A famous application to real-time face recognition is Viola and Jones's work using boosted cascades of classifiers [VJ01]. Another very popular state of the art technique is Support Vector Machines (SVM) used for example in [DT05]. There is also a renewal of interest for nearest-neighbor techniques for object classification [BSI08] [HK05] [MT08].

Road user classification is a useful addition to a traffic monitoring system, and work has already been done in this area. An early simple system [LFP98] tracks and classifies vehicles and human beings. Tracking relies on temporal differencing and template correlation. There are only two description variables: the object area and the "dispersedness", defined as the ratio of the perimeter over the area. The classification is done using a Mahalanobis-based distance, and the correct classification ratio is respectively 86.8% and 82.8% for vehicles and human beings.

The most common classification approach in traffic monitoring is to use the estimated object dimensions, possibly by fitting a (3D-) model. The system described in [GMPP02] tracks

1 http://pascallin.ecs.soton.ac.uk/challenges/VOC/. In the 2008 challenge results, average precision of the car and bus classes are respectively lower than 60% and 52%.

correctly 90% of the total number of vehicles, among which 70% are correctly classified using vehicle dimensions. Very accurate methods to estimate vehicle dimension are presented in [PLY07]. More complex 3D models are used in [MMZ05] to classify vehicles into eight classes. For real-time requirements they are pre-projected off-line in the image space for direct comparison. The object description includes other visual features such as brightness and color histograms. In addition to dimension-based classification, a support vector machine classifier is used to further differentiate sub-classes like bicycles and motorbikes, buses and trucks. It reports a global correct detection rate of 92.5%, though it varies considerably for each class. In [KB08], more than 90% of road users are correctly classified in a simpler highway setting. This good performance is achieved using feature-based tracking and the number of features associated to the object height.

Yet, as noted in [HK02], it is difficult to establish a geometric model for each class. Following [LFP98], other approaches use more description variables. The system described in [HK02] attempts to classify road users as well as estimate their color. It uses mainly shape-based variables, projects the object description to a space with fewer dimension by Linear Discriminant Analysis (LDA) and finally classifies the objects using a weighted nearest neighbor technique. It approaches 91% of correct answer rate. The recent work presented in [MT08] is inspired by the previous system. It extracts standard description of blobs by simple morphological measurements and targets real-time traffic monitoring in highways. Its performance is not clear as it reports results for different confidence levels. Another recent work is described in [HYCH06], relying on two description variables, the size, normalized by the lane width previously recovered from the scene, and the “linearity” measuring the dispersion of boundary pixels around a straight line. Vehicle classes are represented by a few templates. A road user is classified by computing its average distance to the templates of a class. It reports accuracy above 90% for four classes. Although the work presented in [ZCHT07] is called unsupervised by its authors, using k-means, it implicitly relies on prior knowledge of the road users present 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.

Finally, an idea common to most of the research presented in this section is to use the multiple detections provided by a tracking system at each frame. Integrating the instantaneous classifications, the systems achieve more robust performance (e.g. see [HYCH06] for some quantitative results that illustrate this point).

3. Truck Classification

A classification module was developed for this project. It takes the road users' trajectories as input, and classifies them as either truck or non-truck (See Figure 1). In addition to camera calibration, off-line data comprises also a set of instances of truck and non-truck images, manually labeled off-line. These instances are used to build automatically a truck classifier, instead of manually specifying simple classification rules. The features used by the classifier rely on a model of the static elements of the scene, or background, to identify the shape and appearance of the road users. This will be explained in details in the next section.

The classification module was developed to be simple and extensible, in order to be able to

adapt to various challenges and situations. The vehicles trajectories provided by the feature-based method are only trajectories. The features tracked on a vehicle rarely cover the whole object. In order to differentiate road users, it is necessary to extract as many pixels of the object as possible, in order to analyze its shape and appearance.

For this purpose, background subtraction was used. The background of the scene, i.e. the image of the static elements, is modeled using the well-known mixture of Gaussians [SG99]. Each pixel is independently modeled as a mixture of Gaussians and can be updated continuously, which is essential for a system that would operate continuously under changing conditions. For each frame, each pixel is matched to the Gaussians, among which the ones with the most importance and least variance constitute the background. An alternative more simple approach uses a simple background image obtained from the mixture of Gaussians model, and subtracts the current frame with it, using a suitable threshold. The connected components in the difference image constitute the foreground objects, or road users in movement (See Figure 2). The result is typically noisy, which is addressed by applying morphological operations.

Road users' trajectories can be matched in each frame with the foreground objects to extract the description variables that will be used for classification. There are various types of variables, either related to the shape or the appearance of the foreground objects. As a starting point, the system was developed to use a moment-based representation of the object shape, also called contour [Bos05] (chapter 8 of [BK08]). The (*p*+*q*)th-order spatial moment of a foreground object is given by $m_{p,q} = \iint f(x,y)\, x^p\, y^q\, dxdy$, for *p,q = 0,1,2…*, where *f(x,y) = 1* for points inside the foreground object and zero elsewhere (e.g. respectively the white and black pixels in the right image of Figure 2). In images, the integrals are replaced by discrete sums. Particular low order moments are the size in pixels (zeroth moment) and the position of the centroid (first-order moment). The set of all spatial moments provides a complete description of the shape.

Spatial moments are closely related to features such as dispersedness or compactness that have been used successfully in other works. To make these description variables more robust, they must be made invariant with respect to 2D transformations such as translation, rotation, reflection, and scaling. The central moments are translation-invariant:
$m_{p,q} = \iint f(x,y)(x-\bar{x})^p (y-\bar{y})^q\, dxdy$, where $\bar{x} = m_{1,0}/m_{0,0}$ and $\bar{y} = m_{0,1}/m_{0,0}$. The normalized moments are scale-invariant $\eta_{p,q} = \mu_{p,q}/\mu_{0,0}^{\gamma}$, where $\gamma = (p+q+2)/2$. To achieve rotation- and reflection-invariance, the Hu moments $\{h_i | i = 1...7\}$ can be finally computed as follows

$$h_1 = \eta_{2,0} + \eta_{0,2}$$

$$h_2 = (\eta_{2,0} - \eta_{0,2})^2 + 4\,\eta_{1,1}^2$$

$$h_3 = (\eta_{3,0} - 3\,\eta_{1,2})^2 + (3\,\eta_{2,1} - \eta_{0,3})^2$$

$$h_4 = (\eta_{3,0} + \eta_{1,2})^2 + (\eta_{2,1} + \eta_{0,3})^2$$

$$h_5 = (\eta_{3,0} - 3\,\eta_{1,2})(\eta_{3,0} + \eta_{1,2})((\eta_{3,0} - 3\,\eta_{1,2})^2 - 3(\eta_{2,1} + \eta_{0,3})) + (3\,\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})(3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{1,2} + \eta_{0,3}))$$

$$h_6 = (\eta_{2,0} - \eta_{0,2})((\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2) + 4\eta_{1,1}(\eta_{3,0} + \eta_{1,2})(\eta_{2,1} + \eta_{0,3})$$

$$h_7 = (3\eta_{2,1} - \eta_{0,3})(\eta_{2,1} + \eta_{0,3})(3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2) - (\eta_{3,0} - 3\eta_{1,2})(\eta_{2,1} + \eta_{0,3})(3(\eta_{3,0} + \eta_{1,2})^2 - (\eta_{2,1} + \eta_{0,3})^2)$$

Using a subset of these variables describes effectively the object shape. A classifier can now be built to distinguish trucks from other road users. Instead of the tedious, error-prone and sub-optimal manual setting of rules, it is preferred to use machine learning techniques. The field of machine learning is concerned with building models of data from a given training dataset that can be used to predict results on other examples from the same distribution as the training dataset. This field is well researched and vast, and readers are referred to [DH00] for a complete and detailed overview. Among the different models and algorithms available in the literature, discriminative models were chosen. Decision trees were chosen in particular since they are simple, fast, and can be used as base classifiers in ensemble methods [Bre01]. Ensemble methods like Bagging and Boosting [Die00] are newer techniques that aggregate multiple randomized simple, or weak, classifiers to boost performance. Ensemble methods feature among the best performing machine learning systems, including for computer vision applications [VJ01].

In an off-line phase, a classifier is trained on manually labeled data, with examples of trucks and other road users. Once trained, a classifier returns a prediction for each instantaneous foreground shape. A road user will therefore have many classifications as it moves through the scene, up to one for each frame in which it appears. These classifications are integrated temporally to improve robustness; a road user will be classified as a truck if the classifier predicted that the shape was the shape of a truck in more than *nDetections* frames.

4. Experimental Results

The system was developed using the open source computer vision library OpenCV [BK08]. Another program was developed to manually track and classify road users to provide training and testing data for the truck classifier (See Figure 3).

The data used to test the system comes from the datasets collected by the Federal Highway Administration for the Next Generation SIMulation project (NGSIM), made available online[2]. The US highway 101 dataset was chosen for the experiments; the data, i.e. the video data and the road users' trajectories, was collected in a long corridor covered by eight cameras (See Figure 4). Unfortunately, the calibration data provided by the maintainers of the project proved impossible to use. This seems to have little impact on the performance as the videos were captured from very high above the ground.

The dataset consists of three periods of 15 minutes in the morning: period 1 from 7:50am to 8:05am, period 2 from 8:05am to 8:20 and period 3 from 8:20 to 8:35. A small set of road users of the subset camera 8 / period 1 was manually classified. Tests were made on other parts of the dataset, for the same camera and different periods, as well as for the data covered by the other cameras, including sections with significant perspective variations and different road user's appearance. In the video data collected from cameras 5 to 8, the road users are moving

2 http://www.ngsim.fhwa.dot.gov

closer to the cameras; in the video collected by the rest of the cameras, 1 to 4, the road users are moving further away from the cameras.

The random forests classifier was chosen for its high accuracy [Bre01] and availability in OpenCV. It is compared to its base classifier, a simple decision tree. For each road user detected and tracked by the system, the classification module indicates whether it is a truck or not. This is therefore a binary classification problem where the classes are {non-truck, truck}. The number of classes is *C=2*. The number of instances is *N*. The results are based on the confusion matrix; the element $m_{i,j}$ of the confusion matrix is the number of instances of class i that are classified in class j by the system. In the case of binary classification, where truck is the "positive" class to be detected, $m_{truck,truck}$ is the number of true positive, $m_{non\text{-}truck,truck}$ is the number of false positives or false alarms, and $m_{truck,non\text{-}truck}$ is the number of false negatives or missed detections. The following performance measures are computed from the confusion matrix: the total accuracy, the precision and recall for each class. The formulas are

$$Accuracy = \frac{\sum_{i=1}^{i=C} m_{i,i}}{N}$$

$$Precision_c = \frac{m_{c,c}}{\sum_{i=1}^{i=C} m_{i,c}} \; for\ class\ c$$

$$Recall_c = \frac{m_{c,c}}{\sum_{i=1}^{i=C} m_{c,i}} \; for\ class\ c$$

(2)

Cross validation is a standard technique to estimate the performance of a classifier, more robust than simply splitting the dataset in a training and a testing dataset. K-fold cross-validation consists in splitting the dataset in *K* subsets and in testing on each subset a classifier trained on the *K-1* other subsets. As each instance is classified once, the confusion matrix can be filled for all instances and the performance indicators computed. When the number of folds *K* is the number of instances, it is called leave-one-out cross-validation. The results of leave-one-out cross-validation for the training dataset, for instantaneous frame detections, are presented in Table 1. Two sets of description variables for object shapes are tested: the full set of all moments, and a smaller subset consisting only of the Hu moments, called respectively "full" and "small" in the following tables and figures (both sets are supplemented with the ratio of the number of object pixels divided by the pixel size of its bounding box). As expected, random forests appear to perform better than decision trees, for most performance measures. The conclusions are not so clear for the sets of description variables. The full set provides better performance for random forests for all performance measures. Regarding the results for the decision trees, the small set provides better total accuracy, but a much degraded Recall for trucks (i.e. many non-trucks are incorrectly classified as trucks).

The next results are obtained by training a classifier on labeled instances from the subset camera 8 / period 1, and by testing it on the other subsets covered by the same camera 8, period 2 and 3, as well as the subsets camera 4 / all periods (to investigate the effect of a very different perspective). The sizes of the sets are reported in Table 2. In the classification process, some special processing was added to deal with special conditions. In congested situations,

foreground objects tend to merge and the individual object shapes cannot be obtained through background subtraction. To avoid numerous false alarms in such situations, slow road users are therefore classified as non-truck. In any case, signal priority cannot have much impact in such traffic conditions.

Since there is a detection threshold *nDetections* to set, there is a tradeoff between recall and precision, between detecting as many trucks as possible while keeping the false alarms as low as possible. This tradeoff can be captured by computing the receiver operating characteristic (ROC). The ROC curve is the plot of $Recall_{truck}$, also called true positive rate, versus $1\text{-}Recall_{non\text{-}truck}$, also called false alarm rate, for various settings of the classifier, obtained for example in our case by changing the parameter nDetections. The best possible prediction method would yield a point in the upper left corner of the ROC space, at coordinate (0,1), representing 100% recall for the truck class, i.e. no missed trucks, and 100% recall for the non-truck class, i.e. no missed non-trucks that would constitute false alarms. A completely random guess, like flipping coins, would give a point along a diagonal line from the left bottom to the top right corners, also called line of no-discrimination.

The ROC curves for the subsets camera 4 and 8, for all periods, are displayed in Figure 5. The operation mode should be chosen depending on the application. In this work, a false alarm rate of 0.5% was used in the experiments carried out to test various truck signal priority strategies, described in the next sections. The maximum $Recall_{truck}$ such that the false alarm rates remains below 0.5% is reported in Table 3. From these results, it is clear that the full set of description variables yields better results, for both classifiers. The surprising part is that the best performing classifier changes with the camera, i.e. the camera angle. The reason may be that the simple decision tree generalizes better to unseen data. Note that the performance on camera 8 / period 1 comes from the fact that part of it is used to train the classifier, and is not indicative of general accuracy, as can be seen in the other results (it is still interesting as only a small part of the subset is labeled for training). The good news for the application is that the recall reaches a relatively high recall for trucks, from 78% to 95%, with a false alarm rate below the 0.5% value used for the system simulation.

5. Conclusion and Future Work

This paper has presented the development of a prototype system for the detection and tracking of trucks using video sensors, designed for a larger truck signal priority system. The prototype video-based truck detection system is tested on real world data from the NGSIM project. The truck detection rate reaches a relatively high recall for trucks, from 78% to 95%, with a false alarm rate below the 0.5% value (used to test different scenarios in a traffic simulation not presented in this report). This shows that the performance required for effective truck signal priority is reached or within reach of automated video-based sensors.

The system can still be further improved. The classification module should be generalized to classify all the types of road users, using more varied features, especially based on the appearance, which will make the system more robust to particular conditions. Road users' size or volume is obtained from the features, but should be improved as the features may not cover completely the road users. Such improvements will enable the processing of more challenging data.

Another area is the development of a multi-camera system to cover larger areas, for example to track trucks along a corridor for signal priority and other services, or to better estimate the road users' height if the field of views of different cameras overlap (along the third dimension orthogonal to the ground considered as a plane).

Acknowledgements

The research presented in this paper was conducted at the Bureau of ITS and Freight Security of the Faculty of Applied Science (BITSAFS-Engineering) at the University of British Columbia. The authors would like to acknowledge agency funding partners of BITSAFS-Engineering who made this research possible: Transport Canada, B.C. Ministry of Transportation and Investment, and TransLink.

References

[AH77] F. Amundsen and C. Hydén, editors. Proceedings of the first workshop on traffic conflicts, Oslo, Norway, 1977. Institute of Transport Economics.

[AMBT06] G. Antonini, 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, 69(2):159–180, 2006.

[Arc04] J. Archer. Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-simulation Modelling. Academic thesis, Royal Institute of Technology, Stockholm, Sweden, December 2004.

[ASC78] B. L. Allen, B. T. Shin, and D. J. Cooper. Analysis of traffic conflicts and collision. Transportation Research Record, 667:67–74, 1978.

[BK08] G. Bradski and A. Kaehler. Learning OpenCV: Computer Vision with the OpenCV Library. O'Reilly Media, Inc., 2008.

[Bos05] B. Bose. Classifying tracked moving objects in outdoor urban scenes. Research qualifying examination report, EECS Department, MIT, January 2005.

[Bre01] L. Breiman. Random forests. Machine Learning, 45(1):5–32, October 2001.

[BSI08] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pages 1–8, Anchorage, AK, 2008. IEEE.

[DH00] R. O. Duda and P. E. Hart. Pattern Classification. Wiley-Interscience, 2000.

[Die00] T. G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editors, First International Workshop on Multiple Classifier Systems, Lecture Notes in Computer Science, pages 1–15. Springer Verlag, 2000.

[DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, pages 886–893, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005.

[GM07] D. M. Gavrila and S. Munder. Multi-cue pedestrian detection and tracking from a moving vehicle. International Journal of Computer Vision, 73(1):41–59, 2007.

[HK05] O. Hasegawa and T. Kanade. Type classification, color estimation, and specific target detection of moving targets on public streets. Machine Vision and Applications, 16(2):116–121, February 2005.

[HYCH06] J.-W. Hsieh, 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, 7(2):175–187, June 2006.

[KB08] N.K. Kanhere and S.T. Birchfield. Real-time incremental segmentation and tracking of vehicles at low camera angles using stable features. IEEE Transactions on Intelligent Transportation Systems, 9(1):148–160, March 2008.

[LFP98] A. Lipton, H. Fujiyoshi, and R. Patil. Moving target classification and tracking from real-time video. In Proc. of the Workshop on Application of Computer Vision, pages 8–14. IEEE, October 1998.

[Low04] D. G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, 2004.

[LSS05] B. Leibe, E. Seemann, and B. Schiele. Pedestrian detection in crowded scenes. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 878–885, June 2005.

[MG05] X. Ma and W.E.L. Grimson. Edge-based rich representation for vehicle classification. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), volume 2, pages 1185–1192, October 2005.

[MMZ05] S. Messelodi, C. M. Modena, and M. Zanin. A computer vision system for the detection and classification of vehicles at urban road intersections. Pattern Analysis & Applications, 8(1-2):17–31, September 2005.

[MT08] B.T. Morris and M.M. Trivedi. Learning, modeling, and classification of vehicle track patterns from live video. IEEE Transactions on Intelligent Transportation Systems, 9(3):425–437, September 2008.

[PLY07] C.C.C. Pang, W.W.L. Lam, and N.H.C. Yung. A method for vehicle count in the presence of multiple-vehicle occlusions in traffic images. IEEE Transactions on Intelligent Transportation Systems, 8(3):441–459, September 2007.

[SG99] C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pages 246–252, 1999. background modeling algorithm.

[SG00] C. Stauffer and E. Grimson. Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Recognition and Machine Intelligence, 22(8):747–757, August 2000.

[SS06] N. Saunier and T. Sayed. A feature-based tracking algorithm for vehicles in intersections. In Third Canadian Conference on Computer and Robot Vision, Québec, June 2006. IEEE.

[SS07] N. Saunier and T. Sayed. Automated Road Safety Analysis Using Video Data. Transportation Research Record, 2019:57–64, 2007.

[SS08] N. Saunier and T. Sayed. A Probabilistic Framework for the Automated Analysis of the Exposure to Road Collision. In Transportation Research Board Annual Meeting Compendium of Papers, Washington, D.C., January 2008. 08-2916. Accepted for publication in Transportation Research Record.

[VJ01] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple

features. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pages 511–518, Los Alamitos, CA, USA, 2001.

[YJS06] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. ACM Computing Surveys, 38(4):13, 2006.

[ZCHT07] Z. Zhang, Y. Cai, K. Huang, and T. Tan. Real-time moving object classification with automatic scene division. In IEEE International Conference on Image Processing, volume 5, pages 149–152, October 2007.

Tables

			Non-truck Class		Truck Class	
Classifier	Variables	Accuracy	Recall	Precision	Recall	Precision
Decision Tree	Full	72.06	73.32	85.62	68.8	50.52
	Small	74.87	85.82	80.42	47.17	56.82
Random Forest	Full	82.02 +/-0.39	89.20 +/- 0.41	86.19 +/- 0.34	63.86 +/- 1.04	70.06 +/- 0.83
	Small	74.99 +/-0.51	87.19 +/- 0.53	79.78 +/- 0.34	44.13 +/- 1.10	57.68 +/- 1.21

Table 1 Performance results (in percentage) for a decision tree and a random forest classifier, for the full or small set of description variables, on the training dataset (748 shape instances). The leave-one-out cross-validation is run 100 times for the random forests classifier since its training algorithm is randomized, the average and standard deviation are provided for all measures.

Camera	Period	Number of trucks	Number of non-truck road users
4	1 (7:50 to 8:05)	55	2115
	2 (8:05 to 8:20)	46	1972
	3 (7:20 to 8:35)	43	1874
8	1 (7:50 to 8:05)	57	2115
	2 (8:05 to 8:20)	42	1972
	3 (7:20 to 8:35)	39	1874

Table 2 Number of truck and non-truck instances in the sets used for testing.

Camera	Period	Classifier	Variables	*nDetections*	$Recall_{truck}$
4	1 (7:50 to 8:05)	Decision Tree	Full	54	78.18
			Small	122	3.64
		Random Forest	Full	26	74.55
			Small	76	3.64
	2 (8:05 to 8:20)	Decision Tree	Full	33	91.30
			Small	126	23.91
		Random Forest	Full	17	80.43
			Small	83	23.91
	3 (7:20 to 8:35)	Decision Tree	Full	26	95.35
			Small	130	25.58
		Random Forest	Full	12	88.37
			Small	86	18.60
8	1 (7:50 to 8:05)	Decision Tree	Full	28	98.25
			Small	104	36.84
		Random Forest	Full	13	100
			Small	58	64.91
	2 (8:05 to 8:20)	Decision Tree	Full	44	78.57
			Small	117	23.81
		Random Forest	Full	19	78.57
			Small	73	40.48
	3 (7:20 to 8:35)	Decision Tree	Full	29	92.31
			Small	111	38.46
		Random Forest	Full	12	94.87
			Small	89	10.26

Table 3 Maximum $Recall_{truck}$ (in percentage) and corresponding parameter value nDetections (in frames), while keeping the false alarm rate below 0.5%, for the classifiers decision tree (DT) and random forest (RF), with full and small variable sets, for data subsets camera 4 and 8, periods 1 to 3.

Figures

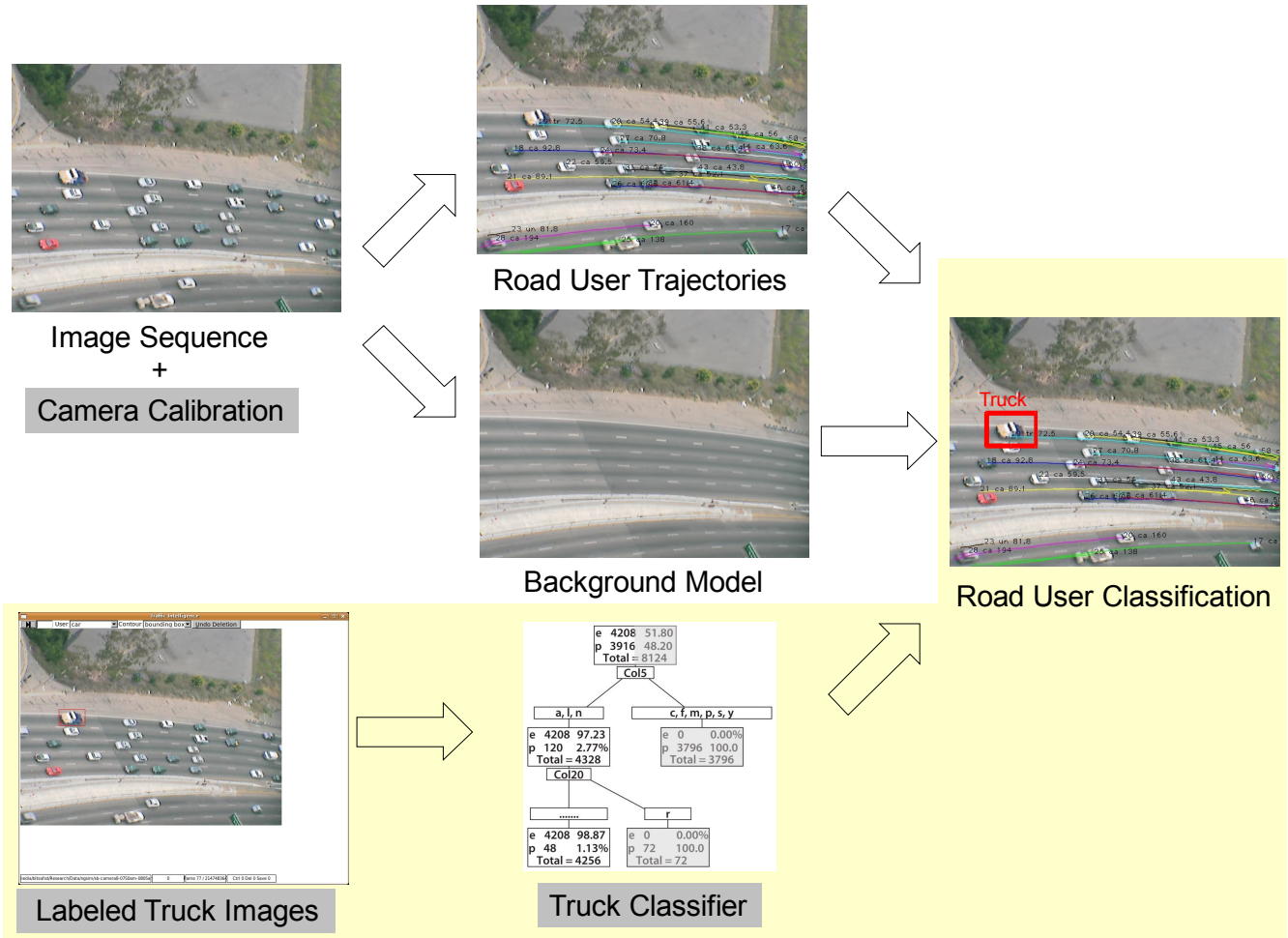

Figure 1 Overview of the system. Grayed boxes represent system components that are computed off-line, the classification module is composed of the elements on a yellow background.

Figure 2 Example of background subtraction. From left to right, the current frame, the background image and the resulting foreground image are displayed.

Figure 3 Interface of the program for road user manual tracking and labeling. This data is then used for training and testing (as ground truth).

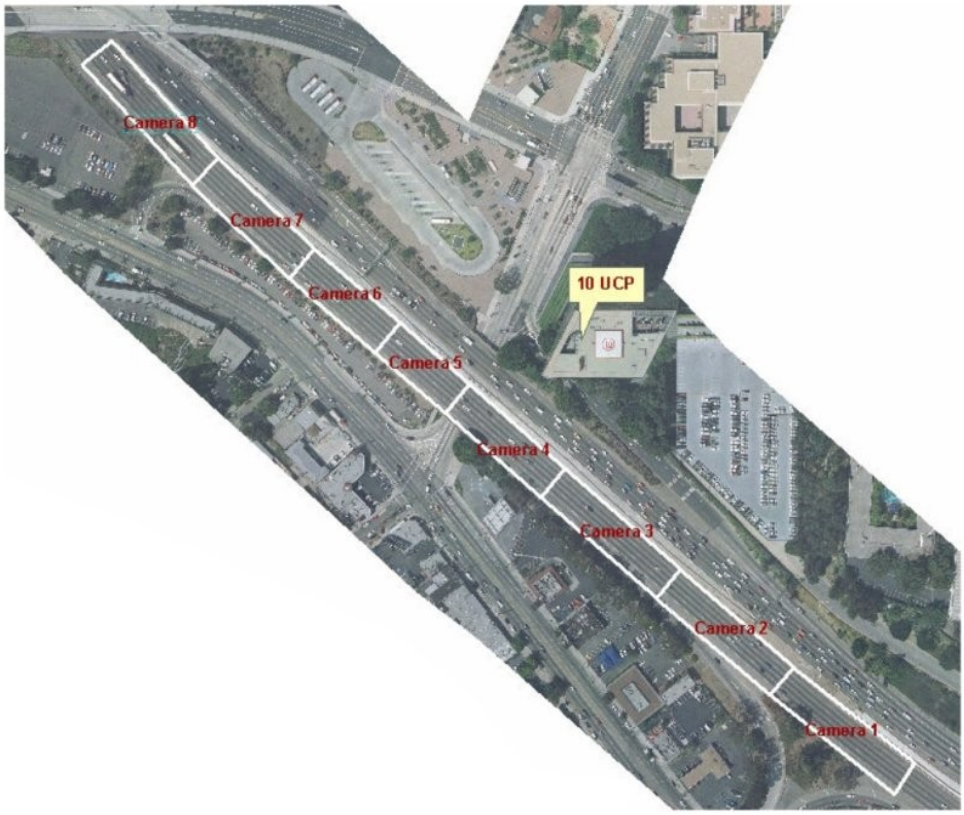

Figure 4 US highway 101 NGSIM dataset and the camera coverage. “10 UCP” points at the camera location.

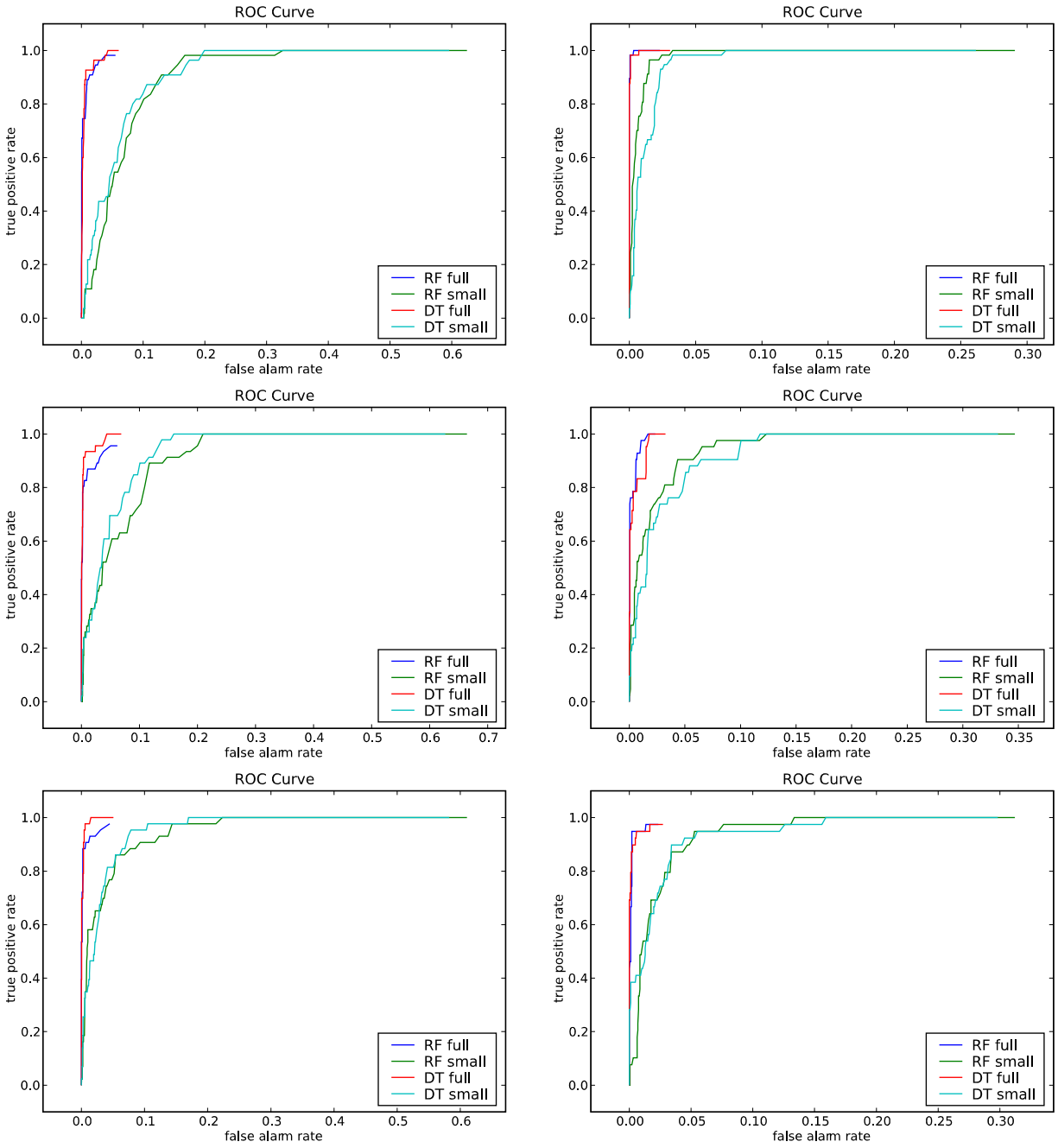

Figure 5 ROC Curve for the various subsets, for the two classifiers random forests (RF) and decision trees (DT), and the two variables sets (full and small), trained on labeled instances of the subset camera 8 / period 1. The first column corresponds to camera 4, the second to camera 8; the rows correspond respectively from top to bottom to period 1 from 7:50am to 8:05am, period 2 from 8:05am to 8:20 and period 3 from 8:20 to 8:35.