Fully Automated Real-Time Vehicles Detection and Tracking with Lanes Analysis

Jakub Sochor*
Supervised by: Adam Herout†

Faculty of Information Technology
Brno University of Technology
Brno / Czech Republic

Abstract

This paper presents a fully automated system for traffic surveillance which is able to count passing cars, determine their direction, and the lane which they are taking. The system works without any manual input whatsoever and it is able to automatically calibrate the camera by detecting vanishing points in the video sequence. The proposed system is able to work in real time and therefore it is ready for deployment in real traffic surveillance applications. The system uses motion detection and tracking with the Kalman filter. The lane detection is based on clustering of trajectories of vehicles. The main contribution is a set of filters which a track has to pass in order to be treated as a vehicle and the full automation of the system.

Keywords: motion detection, tracking, vehicles, traffic surveillance camera, direction detection, lanes detection, real-time

Figure 1: Example of video scene processed by the proposed traffic analysis system. Information about passing cars and their directions are displayed in output video.

1 Introduction

This paper presents a fully automated system for traffic analysis. These types of analysis systems have a wide spectrum of usage. For example, it is possible to monitor the traffic or try to predict characteristics of the future traffic flow. The presented system is able to count passing cars, determine their direction and lane which they are taking. The goal is to run the system without any manual calibration or input whatsoever. The full automatism of the system is required if the system should be usable with already mounted uncalibrated cameras which are spread over highways. Therefore, the camera is automatically calibrated prior to running the traffic surveillance system. Real time processing is another requirement which needs to be satisfied for usage in real traffic surveillance applications.

Some methods for calibration of the camera require user input [29, 3] and therefore they can not be used in fully automated systems. Approaches for the calibration are usually focused on detection of vanishing point of the direction parallel to moving vehicles [6, 10, 23, 25]. There are several ways how to detect the vanishing point. Detected lines [25, 6] or lanes [25, 10] can be used for obtaining this vanishing point. On the other hand, Schoepflin and Dailey [23] use motion of vehicles and assume that they have straight parallel trajectories. Kanhere et al. [16] detect vehicles by a boosted detector and observe their movement, and Zhang et al. [30] analyze edges present on the vehicles.

A popular approach to detection and tracking of vehicles is to use some form of background subtraction and Kalman filter [15] to track the vehicles [12, 21, 14, 28, 1, 4, 7, 20, 17, 22]. Other approaches are based mainly on detection of corner features, their tracking and grouping [2, 13, 5]. Also, Cheng and Hsu [4] use pairing of headlights for the detection of vehicles at night.

Two main approaches are used for the detection of lanes. The first one is based on detection of the lane dividing lines [13, 18]. The other approach is based on motion of vehicles and their trajectories. Tseng et al. [28] use a virtual line perpendicular to vehicles' motion and compute intersections of the line with trajectories of vehicles. Hsieh et al. [12] use a two-dimensional histogram of accumulated centers of vehicles and Melo et al. [20] approximate the trajectories with low-degree polynomials

*xsocho06@stud.fit.vutbr.cz
†herout@fit.vutbr.cz

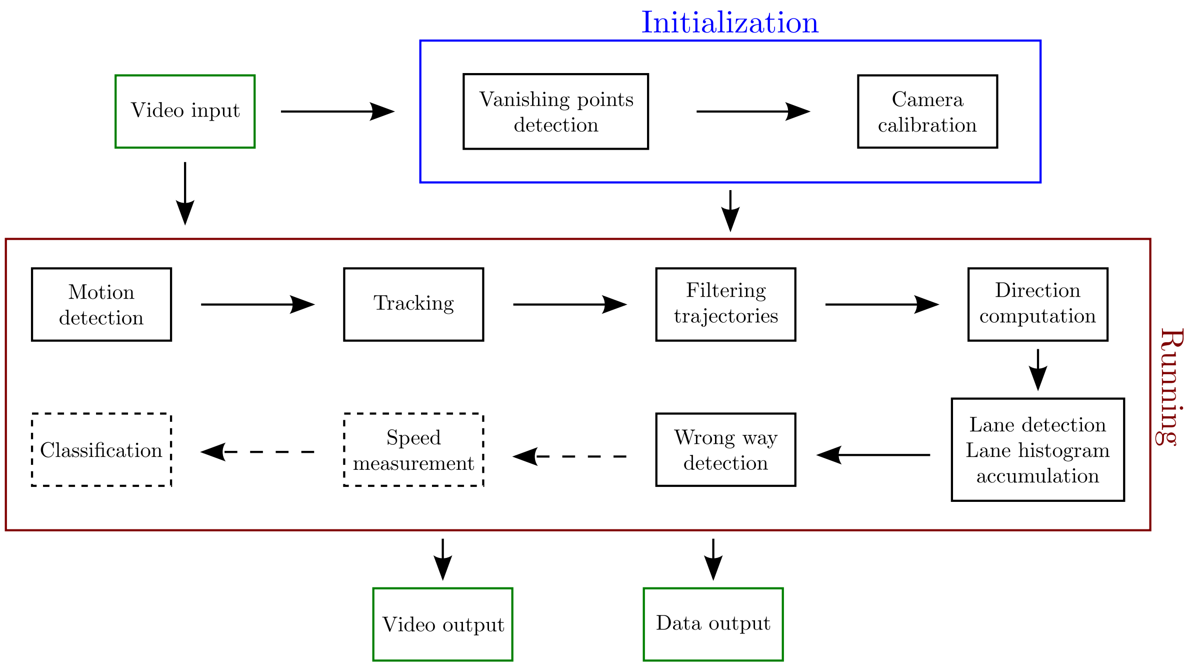


Figure 2: Pipeline of processing of the input video stream. Parts of the pipeline which will be implemented in the future, namely Classification and Speed measurement, are shown in dashed boxes.

and cluster these approximations.

The system proposed in this paper uses detection of vehicles by background subtraction [26, 31] and Kalman filter [15] for tracking. Prior to running the algorithm, the camera is calibrated by the detected vanishing points and the vanishing point of direction parallel to the motion of vehicles is used for higher accuracy of tracking. The detection of lanes is based on trajectories of vehicles and their approximation by a line.

2 Proposed Method for Traffic Surveillance

This section of the paper presents methods used in the system for detection and tracking of cars. The direction and lane detection is also discussed in detail. The overall processing pipeline is shown in Figure 2.

The main goal of the system is to create statistics of traffic on a road which is monitored by a camera. These statistics include the number of passed cars, their direction and lane.

2.1 Initialization

It is required to initialize the system prior to processing a video stream. The main purpose of the initialization is to find vanishing points of the scene and use the vanishing points to calibrate the camera. This is performed in a fully automated way and no user input is used. Arrows directed to the vanishing points are used for visualisation of the vanishing points. An example of the visualisation of the vanishing points is in Figure 3.

The vanishing point of the direction parallel to the vehicle movement is denoted as the first vanishing point. The second vanishing point has perpendicular direction to the movement of vehicles and the third vanishing point is perpendicular to the ground plane. However, only the first vanishing point is required for the tasks described in this paper; therefore, only detection of this vanishing point will be described. The detection of the other vanishing points is described in a paper written by Dubská et al. [8], currently submitted to IEEE Transactions on Intelligent Transportation Systems.

First Vanishing Point Detection

Corner feature tracking is used for the detection of the first vanishing point. Hence, Good Features to Track [24] are detected in the video stream and KLT tracker [27] is used for the tracking of the corner features. Detected motion of the tracked features is extended into a line which is defined by image points (x_t, y_t) and (x_{t+1}, y_{t+1}) which are positions of the feature in frame t and $t+1$.

All these lines are accumulated into the *diamond space* [9] until the initialization is terminated. The initialization is terminated when the global maximum of the diamond space is bigger then a predefined threshold and therefore a sufficient number of lines was accumulated. Afterwards, the coordinates of the global maximum in the diamond space are transformed into coordinates of the vanishing point in the image plane.

The diamond space is based on Cascaded Hough trans-

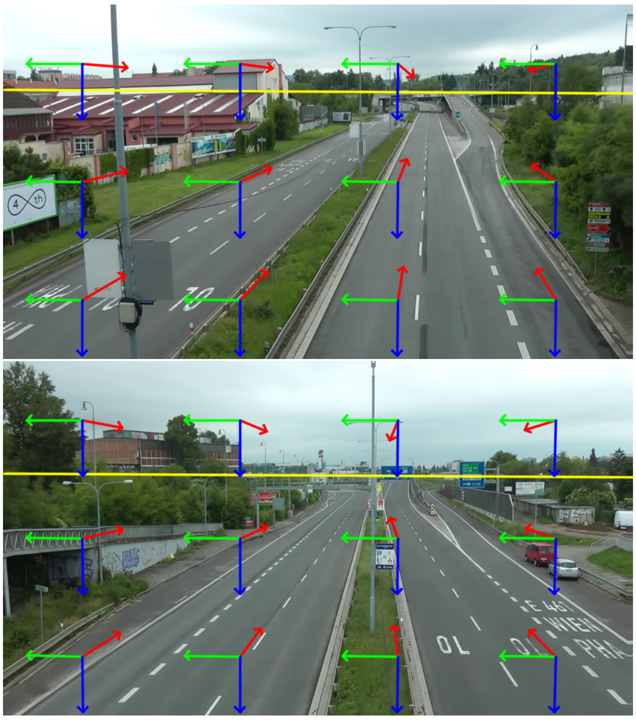


Figure 3: Detected vanishing points. Red arrows are pointing to the first vanishing point, green to the second one, and the third vanishing point is defined by the blue arrows. Yellow horizon line connects the first and second vanishing point.

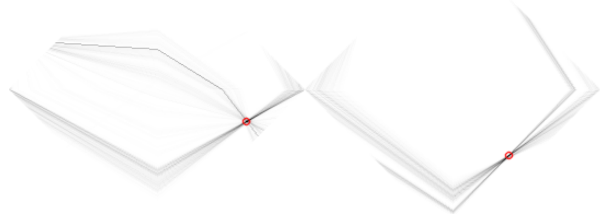


Figure 4: Examples of diamond spaces for detection of the first vanishing point with located global maximum

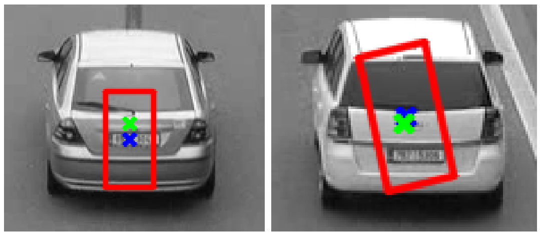


Figure 5: Examples of matching rectangles (red) for predicted object location (blue). The actual center of the detected connected component is drawn by green color. The figure shows that the longer side of the rectangle is directed to the first vanishing point.

form and parallel coordinates. Each line which is accumulated into the diamond space has to be transformed into coordinates in this space. The transformation divides the line into three line segments which are accumulated into the diamond space. Examples of the diamond space are in Figure 4.

It should be noted that the system uses a video downsampled to a framerate close to 10 FPS, so that the movement of corner features is detectable and measurable.

2.2 Vehicle Detection and Tracking

The vehicle detection is based on motion detection in the video scene. Mixture of Gaussians background subtraction [26, 31] is used for the motion detection. Also, shadow removal [11] is used for higher accuracy of the motion detection. Noise in the detected motion is removed by morphological opening followed by morphological closing. Detected connected components are considered to be a potential vehicle. The motion detection approach was selected mainly for its speed.

Kalman filter [15] is used for prediction of the new position of a car and for associating cars in consequent frames. The state variable of the Kalman filter $(x, y, v_x, v_y)^T$ contains the current position of the car and its velocity in image coordinates.

Several conditions are used for matching an object in the consequent frame to its predicted location. The first condition states that the matched object must have similar colors. This condition is enforced by correlating histograms of objects in HSV color space. The second and last condition is that the center of matched object must be inside of so called *matching rectangle*. The predicted location of a car is the center of this matching rectangle and the longer side of the rectangle is directed towards the first vanishing point, as it is shown in Figure 5, and the matching rectangle has size 30×15 pixels. This condition is built on the assumption that the vehicle is going either in the direction towards the vanishing point or from the vanishing point, and therefore it is expected that in this direction can be higher displacement from the predicted location. Lastly, the closest connected component which meets the conditions presented above is found for each object and its predicted location in the consequent frame.

When a match is not found in several consequent frames, the tracked object is removed from the pool of tracked objects. Several filters are used for determining if the object should be accounted in the statistics of passed cars. The trajectory of the object is approximated by a line using least squares approximation. After that, the distance of the first vanishing point from the line is measured. Let us denote this distance as d_{vp}. Also, the ratio r, Eq. (1), between passed distance and maximal possible distance which an object can pass in the given trajectory is measured, Figure 6 shows the positions of P_e, P_s, L_e and L_s. The object is accounted in the statistics as a vehicle if the

acc variable is equal to 1, Equation (2), where t_{vp} and t_r are predefined thresholds.

$$r = \frac{||P_e - P_s||}{||L_e - L_s||} \quad (1)$$

$$acc = \begin{cases} 1 & d_{vp} \leq t_{vp} \text{ and } r \geq t_r \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

2.3 Direction Estimation and Lane Detection

For a new vehicle which is about to be added to the statistics, the direction of the vehicle and its lane is calculated. Rule (3), which compares the relative positions of the first vanishing point and the last and first position of the vehicle, is used for computing the direction.

$$dir = \begin{cases} \text{To VP} & ||VP_1 - P_e|| < ||VP_1 - P_s|| \\ \text{From VP} & \text{otherwise} \end{cases} \quad (3)$$

The detection of lanes is based on clustering of the trajectories of cars. Therefore, the trajectory is also approximated by a line with least squares approximation, see green line in Figure 6. Each cluster of the lines corresponds to a lane in the monitored traffic surveillance scene and the clustering is performed in a one-dimensional space, where the values of the trajectory lines are their angles with axis x in the image. The clusters itself are searched as local maxima in the histogram of the angles. Hence, the clusters have to be a local maximum in the histogram in a predefined surroundings and also the maximum has to have at least a predefined amount of accumulated lines. The closest lane is assigned to a new passing vehicle as the lane which the vehicle is using. The closest lane computation is also based on the angles of the trajectory line and the lane with axis x.

This clustering is always performed after every 200 trajectory lines are accumulated and a unique identification number is assigned to each cluster. Let us denote the set of clusters as $C_N = \{(c_1, a_1), \ldots, (c_n, a_n)\}$ where N is the number of accumulated lines, and pair (c_i, a_i) denotes one cluster, where c_i is its identification number and a_i the angle corresponding to the found local maximum. Correspondences for clusters C_N and C_{N-200} are searched in order to obtain the temporal consistency of detected lanes in the scene. The clusters' identification numbers would change after every 200 accumulated lines if the correspondences were not found; and therefore, it would be impossible to create long-term statistics for cars passing in the detected lanes.

The identification number of the found correspondence is assigned to a cluster if the correspondence is found. A new unique identification number is assigned to the cluster otherwise. The correspondence for a cluster $(c_i, a_i) \in C_N$ is a cluster $(c_j, a_j) \in C_{N-200}$ for which (4) and (5) hold. The distance function is computed according to Equation (6) which compensates that the angles 0 and 2π have distance from each other 0.

$$a_j = C_{N-200}\left(\arg\min_c |dist(C_{N-200}(c), a_i)|\right) \quad (4)$$

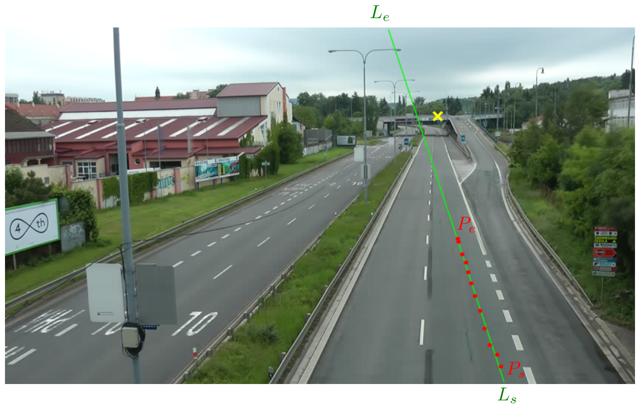


Figure 6: Measured distances for a passed object. The distance between approximated line (green) and the first vanishing point (yellow) is measured. Also, the distance between the first and last (P_s, P_e) point of the track of a vehicle is measured. The maximal distance which is possible to pass with a given trajectory is also measured (distance of L_s and L_e).

$$dist(a_j, a_i) \leq t_d \quad (5)$$

$$dist(x, y) = \min(2\pi - |x - y|, |x - y|) \quad (6)$$

The dominant direction is also computed for each cluster c of the trajectory lines. The dominant direction dir_c is computed according to (7), where l_{VP} is the amount of the trajectories in the cluster which have direction towards the first vanishing point and l is the number of all trajectories in the cluster. Reasonable value for threshold t_{dom} is 0.1.

$$dir_c = \begin{cases} \text{To VP} & \frac{l_{VP}}{l} \geq 1 - t_{dom} \\ \text{From VP} & \frac{l_{VP}}{l} \leq t_{dom} \\ \text{Mixed} & \text{otherwise} \end{cases} \quad (7)$$

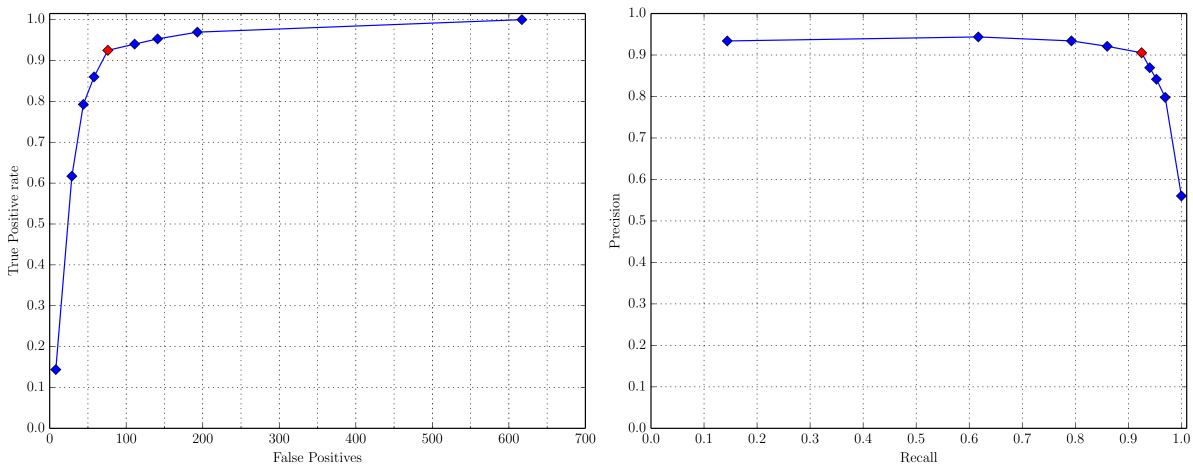
When the dominant direction for a lane is known, it is possible to detect vehicles which are traveling in wrong way. The detection is based on the detected direction dir of the vehicle and the dominant direction dir_c of the lane which the vehicle belongs to. The wrong way variable ww is determined by (8).

$$ww = \begin{cases} \text{True} & dir = \text{To VP} \wedge dir_c = \text{From VP} \\ \text{True} & dir = \text{From VP} \wedge dir_c = \text{To VP} \\ \text{False} & \text{otherwise} \end{cases} \quad (8)$$

3 Results

This section presents the achieved results and methods of evaluation of the algorithms, which were presented above. The speed of video processing is also discussed.

The presented traffic analysis system was evaluated on several video streams. The processed video streams have resolution 854×480 and the video camera was located several meters above the road. The angle of the video camera varies as Figure 8 shows.



(a) ROC curve (b) Precision-Recall curve

Figure 7: ROC and Precision-Recall curves for detection and tracking of vehicles in video. Configuration providing the best results has F-Measure equal to 0.915 and is marked by red color.

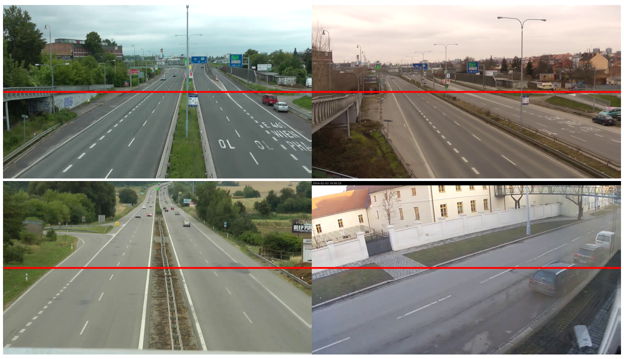


Figure 8: Examples of videos for detection and tracking evaluation. Virtual line which was used for manual ground truth annotation is drawn by red color.

3.1 Detection and Tracking

A manually annotated dataset was created for the evaluation of accuracy of the detection and tracking of vehicles. Imaginary line, see Figure 8, which is crossing the center of image and dividing frames into two equal parts was displayed and for each car, the location and time of crossing the line was annotated. Almost 30 minutes of video was annotated in this way resulting in almost 800 vehicles in the dataset.

The comparison with the ground truth annotation was performed in the following way. For each vehicle which was detected by the traffic analysis system, the trajectory is approximated by a line and the intersection of the approximation with the imaginary line is computed. A match with the ground truth is a vehicle which has trajectory with close intersection to the ground truth intersection and projected time of passing this intersection does not differ too much. If there are more vehicles which satisfy this condition, the vehicle with the smallest time and intersection difference is selected as the match with the ground truth. This way of evaluation was selected because the system targets mainly on overall statistics of passed vehicles.

Nine various configurations which have different maximal distance to the first vanishing point and minimal passed distance of a vehicle were created and evaluated. The ROC and Precision-Recall curves are in Figure 7. Configuration providing the best results has F-Measure [19] equal to 0.915 (Precision is 0.905 and Recall 0.925). The False Negative cases are caused mainly by vehicle occlusions. The occlusions are caused either by a shadow which connects vehicles into one connected component or by a situation when a vehicle partially covers some other vehicle. The False Positives are caused primarily by the motion detection incorrectly dividing a vehicle into two objects and both these objects are tracked and treated as vehicles.

3.2 Direction Estimation and Lane Detection

Several video sequences with a sufficient number of cars were obtained and stability of detected lanes was evaluated for these videos. The results of the evaluation are in Figure 9 and as the graphs show, the detected lanes are almost totally stable and do not change with passing cars. It should be noted that the detected lanes are recomputed always after next 200 cars were observed. Also the directions of the lanes were correctly detected as shown in Figures 9 and 10.

3.3 Evaluation of Speed

Processing speed of the system was also evaluated and the results are in Table 1. The measured framerates include also reading and decoding the video. The system was evaluated on a machine with Intel Dual-Core i5 1.8 GHz and

resolution	traffic intensity	FPS
854×480	high	57.97
	low	82.43
1920×1080	high	28.59
	low	47.88

Table 1: Processing speed evaluation. Approximately 110 minutes of video were used for the evaluation. The videos were divided into groups with respect to the traffic intensity. It should be also noted that the system uses video stream downsampled to ~ 10 FPS, so that the movement is detectable and measurable.

8GB DDR3 RAM. As the table shows, the system can be used for real-time analysis of Full-HD traffic surveillance video. The framerates are higher in videos with lower traffic intensity. The video sequences with higher traffic intensity contain more motion and vehicles which need to be tracked; therefore, more computational resources are used.

4 Conclusions

This paper presents a system for fully automated traffic analysis from a single uncalibrated camera. The camera is automatically calibrated, vehicles are detected, tracked and their direction is computed. Also, the lanes are detected and therefore cars travelling in the wrong way can be detected. The system works in real time and in a fully automated way and therefore it can be used for online traffic analysis with any camera which is monitoring a highway or a street. The system is ready for deployment and it is currently used for online traffic analysis.

The system is able to work under bad lightning and weather conditions. However, for example at night or during rainy weather, the accuracy of detection and tracking decreases slightly because of light reflections from the road. On the other hand, the initialization process can be performed at night without any problem, it will just take longer time because there is a lower amount of vehicles on streets at night.

The main contribution and advantage of the proposed traffic analysis system is that the system works without any manual input whatsoever and the set of conditions which a trajectory of a moving object in video is considered to be a vehicle. Future development of the system will focus mainly on complex crossroads and shadow elimination. Also, elimination of pedestrians from statistics should be addressed.

References

[1] C. Aydos, B. Hengst, and W. Uther. Kalman filter process models for urban vehicle tracking. In *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, pages 1–8, Oct 2009.

[2] D. Beymer, P. McLauchlan, B. Coifman, and J. Malik. A real-time computer vision system for measuring traffic parameters. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 495–501, 1997.

[3] F.W. Cathey and D.J. Dailey. A novel technique to dynamically measure vehicle speed using uncalibrated roadway cameras. In *Intelligent Vehicles Symposium*, pages 777–782, 2005.

[4] Hsu-Yung Cheng and Shih-Han Hsu. Intelligent highway traffic surveillance with self-diagnosis abilities. *Intelligent Transportation Systems, IEEE Transactions on*, 12(4):1462–1472, Dec 2011.

[5] Benjamin Coifman, David Beymer, Philip McLauchlan, and Jitendra Malik. A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C: Emerging Technologies*, 6(4):271 – 288, 1998.

[6] Rong Dong, Bo Li, and Qi-mei Chen. An automatic calibration method for PTZ camera in expressway monitoring system. In *World Congress on Computer Science and Information Engineering*, pages 636–640, 2009.

[7] Yuren Du and Feng Yuan. Real-time vehicle tracking by kalman filtering and gabor decomposition. In *Information Science and Engineering (ICISE), 2009 1st International Conference on*, pages 1386–1390, Dec 2009.

[8] M. Dubská, A. Herout, R. Juránek, and J. Sochor. Fully automatic roadside camera calibration for traffic surveillance. Submitted to: *IEEE Transactions on ITS*, 2014.

[9] Markéta Dubská and Adam Herout. Real projective plane mapping for detection of orthogonal vanishing points. In *British Machine Vision Conference, BMVC*, 2013.

[10] George S. K. Fung, Nelson H. C. Yung, and Grantham K. H. Pang. Camera calibration from road lane markings. *Optical Engineering*, 42(10):2967–2977, 2003.

[11] T. Horprasert, D. Harwood, and L. S. Davis. A statistical approach for real-time robust background subtraction and shadow detection. In *Proc. IEEE ICCV*, volume 99, pages 1–19, 1999.

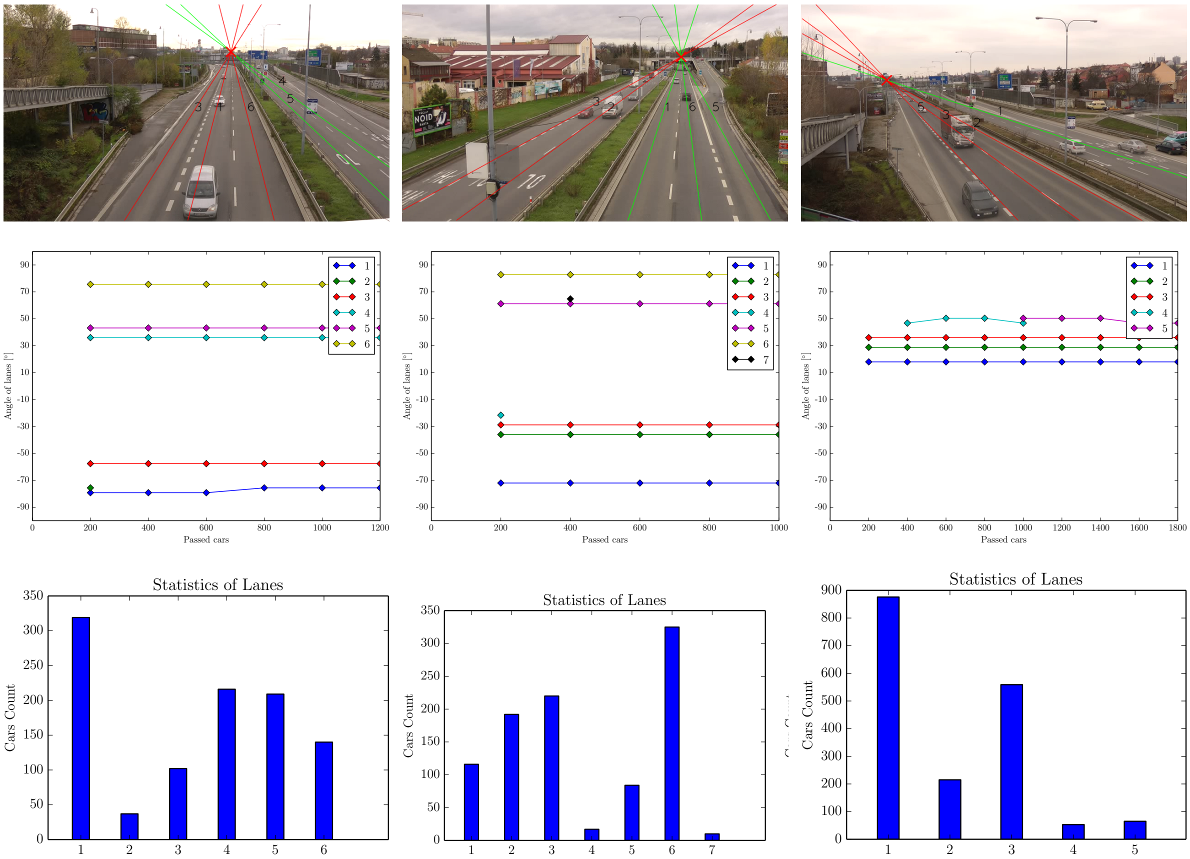


Figure 9: Stability of lanes detection for long video sequences. The top line of images presents the detected lanes. Only lanes which were valid for the last frame of video are drawn. The middle images show changes in detected lanes over time as new cars were observed in video. Finally, the bottom line shows the statistics of observed cars in the detected lanes.

[12] Jun-Wei Hsieh, Shih-Hao Yu, Yung-Sheng Chen, and Wen-Fong Hu. Automatic traffic surveillance system for vehicle tracking and classification. *Intelligent Transportation Systems, IEEE Transactions on*, 7(2):175–187, 2006.

[13] Lili Huang. Real-time multi-vehicle detection and sub-feature based tracking for traffic surveillance systems. In *Informatics in Control, Automation and Robotics (CAR), 2010 2nd International Asia Conference on*, volume 2, pages 324–328, March 2010.

[14] Young-Kee Jung and Yo-Sung Ho. Traffic parameter extraction using video-based vehicle tracking. In *Intelligent Transportation Systems, 1999. Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on*, pages 764–769, 1999.

[15] R. E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering*, (82 (Series D)):35–45, 1960.

[16] Neeraj K Kanhere, Stanley T Birchfield, and Wayne A Sarasua. Automatic camera calibration using pattern detection for vision-based speed sensing. *Journal of the Transportation Research Board*, 2086(1):30–39, 2008.

[17] Dieter Koller, Joseph Weber, and Jitendra Malik. Robust multiple car tracking with occlusion reasoning. pages 189–196. Springer-Verlag, 1993.

[18] A. H S Lai and N. H C Yung. Lane detection by orientation and length discrimination. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 30(4):539–548, Aug 2000.

[19] Christopher D MANNING, Prabhakar RAGHAVAN, and Hinrich SCHÜTZE. *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[20] J. Melo, A. Naftel, A. Bernardino, and J. Santos-Victor. Detection and classification of highway lanes using vehicle motion trajectories. *Intelligent Transportation Systems, IEEE Transactions on*, 7(2):188–200, June 2006.

[21] B. Morris and M. Trivedi. Robust classification and tracking of vehicles in traffic video streams. In *In-*

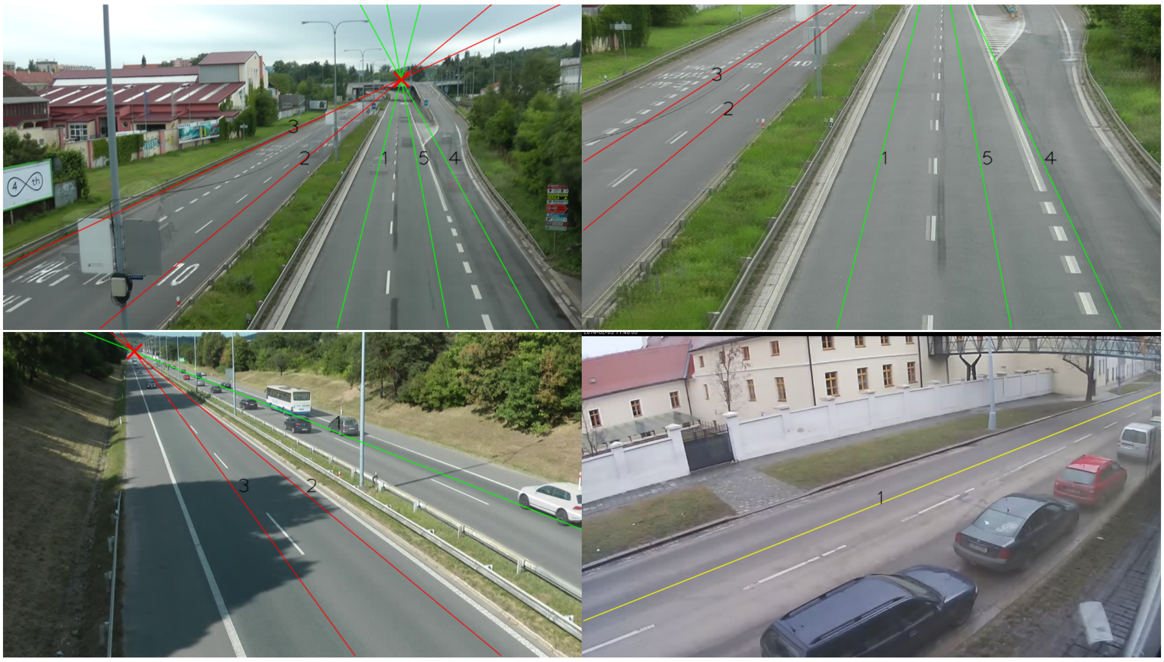


Figure 10: Example of detected lanes and dominant direction of cars in the lanes. Green color means that the majority of cars is heading towards the first vanishing point and red the opposite. Yellow color means that there is no dominant direction for the given lane. Example of this situation is shown in bottom right image. It should be noted, that the centers of cars, which are used for lanes detection, are not in the middle of the lanes because of the angle of view.

telligent Transportation Systems Conference, 2006. ITSC '06. IEEE, pages 1078–1083, 2006.

[22] Roya Rad and Mansour Jamzad. Real time classification and tracking of multiple vehicles in highways. *Pattern Recognition Letters*, 26(10):1597 – 1607, 2005.

[23] T.N. Schoepflin and D.J. Dailey. Dynamic camera calibration of roadside traffic management cameras for vehicle speed estimation. *IEEE Transactions on Intelligent Transportation Systems*, 4(2):90–98, 2003.

[24] J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 593–600, Jun 1994.

[25] Kai-Tai Song and Jen-Chao Tai. Dynamic calibration of Pan–Tilt–Zoom cameras for traffic monitoring. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(5):1091–1103, 2006.

[26] C. Stauffer and W. E. L. Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition*, volume 2, pages 246–252, 1999.

[27] Carlo Tomasi and Takeo Kanade. *Detection and tracking of point features*. School of Computer Science, CMU, 1991.

[28] B.L. Tseng, Ching-Yung Lin, and J.R. Smith. Real-time video surveillance for traffic monitoring using virtual line analysis. In *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, volume 2, pages 541–544 vol.2, 2002.

[29] Kunfeng Wang, Hua Huang, Yuantao Li, and Fei-Yue Wang. Research on lane-marking line based camera calibration. In *International Conference on Vehicular Electronics and Safety, ICVES*, 2007.

[30] Zhaoxiang Zhang, Tieniu Tan, Kaiqi Huang, and Yunhong Wang. Practical camera calibration from moving objects for traffic scene surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(3):518–533, 2013.

[31] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31 Vol.2, 2004.