

Vehicle Detection and Tracking for Traffic Monitoring

Gian Luca Foresti and Lauro Snidaro

Department of Mathematics and Computer Science,
University of Udine, Via delle Scienze 206, 33100 Udine, Italy
{foresti, snidaro}@dimi.uniud.it

Abstract. This paper addresses some of the indications of the European Union for road safety by proposing a real-time traffic monitoring system for vehicle detection and tracking in bad illuminated scenarios. Several urban and extra-urban roads during the night or tunnels are characterized by low illumination, light spots, shadows, light reflections, etc. The main objectives of the proposed system are: (a) to monitor the traffic flow, (b) to estimate the vehicle's speed or determine the state of the traffic, (c) to detect anomalous situations, e.g. rising alarms in case of road accidents or stopped cars. Experimental results on real image sequences demonstrate the effectiveness of the proposed system.

1 Introduction

Currently, more than 40.000 persons are killed every year on EU roads and less than 1000 in the other modes of transport. The short term strategic objective of the Community is to halve the number of fatalities by 2010. The medium term objective for MISS is to cut by around 50% the number of persons killed or severely injured by 2010 [9].

The Monitor Integrated Safety System (MISS) project, funded by the European Sixth Framework Program, aims to develop an innovative platform to dynamically sense and predict natural and infrastructure conditions, so to improve safety and efficiency of transport operations in a multi-environmental scenario. This project wants to increase citizens and operators safety by enabling a just in time intelligent computation of an open dynamic road surveillance network and streamlining alerting tasks under the daily duty provided by clerical staff.

Improved safety can be achieved by assisting the human operator with a surveillance system providing enhanced sensory information about the environment [1]. In addition, the systems that control the flow of traffic can likewise be enhanced by providing sensory information regarding the current conditions in the environment.

Image processing and computer vision techniques are extensively used in traffic monitoring systems in order to increase safety [2-9]. For instance, the possibility to extract complex and high-level traffic information such as congestion or accidents allows to efficiently plan a path through the road network, to quickly bring assistance where needed or to deviate the traffic flow. In order to extract this type of information it is first necessary to separate moving objects from the scene. In this way, vehicles can be counted, and their trajectories and speeds estimated. Moreover, by using these parameters, it is possible to make a traffic classification between safe, congested or dangerous situations.

One of the major difficulties of automatic monitoring traffic scenes is the variety of light conditions of outdoor scenes. D'Agostino [4] discussed the potentials of a commercial machine vision system for traffic monitoring and control. The basic requirements are low cost and robust performance, which have not been fully met till now. Moreover, problems of shadows and nighttime operation have not been solved. For example, the system cannot classify vehicles during nighttime. Vehicle speed was estimated by setting up two inspection zones. The system needs expensive special image processing hardware.

A background updating method for road traffic scenes is discussed in [5]. This system extracts vehicles using a spatio differentiation method to remove the adverse effects of vehicle shadows. The background updating method is based on the change ratio between the road surface brightness of the current input image and that of the old background image. In addition, the system must determine if vehicle(s) appearing in two successive frames are the same. The problem of estimating vehicle speed was not addressed.

In other systems, vehicle shapes are modeled using complex models [6], which cannot be processed in real-time with low-cost hardware. Yuan et al. [7] extract a vehicle and estimate its length, width, height and the number of units of the vehicle from a single perspective image captured by a camera placed at the roadside. The ultimate goal of their system is to classify vehicles into many categories, therefore reducing the gap between the requirement and the availability. However, their approach encounters the general problems in image segmentation, and the methods to identify the roof, side and front of a vehicle are quite ad hoc. Zhu et al. [8] presented a new approach to automatic traffic monitoring using 2D spatio-temporal images. A TV camera is placed above the highway to monitor the traffic through two slice windows for each traffic lane. One slice window is a detection line perpendicular to the lane and the other is a tracking line along the lane. The problems of vehicle counting, speed estimation and vehicle classification have been addressed. In particular, the vehicle classification is performed by using 3D measurements (length, width and height). The system, which has been called VISATRAM (vision system for automatic traffic monitoring), has been tested with real road images under various light conditions, including shadows in daytime and lights at night.

This paper presents a visual-based traffic monitoring system for analyzing vehicles' activities in bad illuminated roads or tunnels, where not only low light conditions, but also light spots, shadows and reflections characterize the observed environments.

2 System Architecture

The general system architecture is shown in Fig.1. Video signals are pre-processed to achieve 2D motion detection of mobile vehicles. To this end, a layered background subtraction approach [5] is used. A change detection module (CD) [5] is applied to compare each frame $I(x,y)$ of the input image sequence with a background image $BCK(x,y)$ which represents the monitored scene without moving objects.

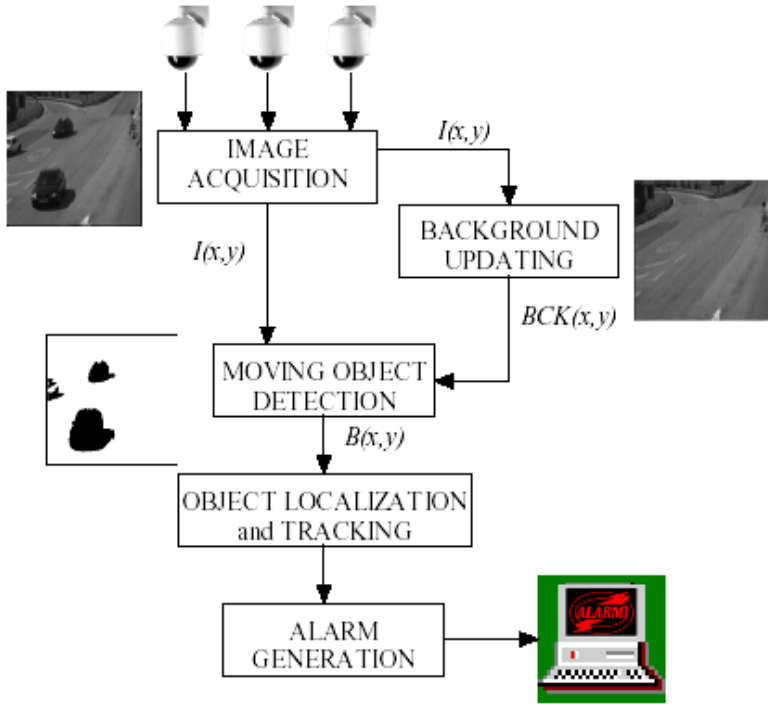


Fig. 1. General system architecture

This procedure generates a binary image $B(x,y)$ where each pixel can assume two possible states: a static or a moving point. Groups of connected pixels (*blobs*) belonging to the class of moving points represent possible vehicles in the scene. A background updating procedure based on the Kalman Filter is applied to estimate significant illumination changes on the background scene $BCK(x,y)$ [5].

The object localization module transforms 2D blob positions into 3D object positions referred to a top view map of the scene. An off-line camera calibration procedure is applied to compute for each camera the transformation matrix (obtained during the initial setup of the sensor). In order to solve the ill-posed problem of 2D into 3D transformation, the ground plane hypothesis is applied.

The object tracking module associates a potential vehicle detected at frame $t-1$ to one of detected objects identified at frame t . This operation is very complex if performed in environments like tunnels or urban roads at night characterized by bad illumination conditions (e.g., low illumination, light spots, shadows, light reflections, etc.). In these conditions, classical object detection methods fail.

Finally, the obtained data are processed to count the number of vehicles moving in the scene, to estimate their speed and to determine the state of traffic. A man-machine interface is used to display in the most appropriate way both statistical information and alarms to a remote operator.

3 Traffic Monitoring

The main innovative part of the system is represented by the vehicle tracking module. It is based on different object features: (a) the 2D image position $M(x,y)$ of the blob's center of mass computed at the time t and estimated at time $t-I$, and (b) the ratio $R = \frac{h}{w}$ between the height and the width of the minimum bounding rectangle (MBR) of the detected blob.

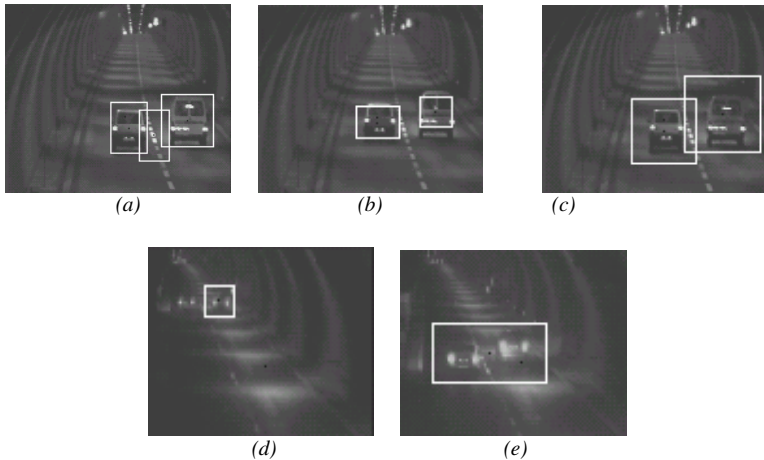


Fig. 2. Some problems of vehicle tracking in a badly illuminated tunnel: (a) false MBRs are detected, (b,c) the size of the MBRs can vary between successive frames, (d) true MBRs that disappear, (e) wrong (fused) MBRs can be detected

When a blob has been detected by the object detection module, the above features are computed in each frame of the sequence, and a matching function is applied to find true correspondences. Two constraints have been introduced in order to reduce the number of possible matches: (i) maximum displacement of the object over two consecutive frames (this constraint allows to define a confidence circle in which to search for candidates for association), and (b) maximum variation of the ratio R between two consecutive frames (this allows to consider only objects with a similar size). If the illumination conditions of the scene are good, the tracking problem reduces to match each detected object in the current frame with a candidate object of the previous frame.

Unfortunately, several urban roads and some tunnels are badly illuminated. Figure 2 shows the most common problems that can occur during the tracking operation in these conditions : (a) false MBRs can be detected (Fig. 3a), (b-c) the size of the MBRs can vary significantly between consecutive frames due to light spots (Figs. 3b and 3c), (d) true MBRs can disappear, (e) wrong MBRs can appear as consequence of erroneous blob detection due to partial occlusions, spot light, shadows, etc. In order to

allow the tracking operation also in complex situations, a Kalman filter is applied to estimate the position and the size of the MBR of each detected blob in the entire image sequence. Estimated MBRs can be compared with detected ones in order to eliminate false blobs, to take into account missed blobs due to occlusions, and to correct MBR' size and/or position.

3.1 Tracking Model

The following quantities of interest (QOIs) have been selected: (a) (x_l, y_l) - coordinates of the top-right corner of the MBR, (b) (x_2, y_2) - coordinates of the bottom-left corner of the MBR, (c) (x_c, y_c) - coordinates of the center of mass of the MBR, (c) (vx_l, vy_l) , (vx_2, vy_2) and (vx_c, vy_c) - translation of the selected points along the x and y axes, respectively, (d) ρ - scale factor due to the perspective projection introduced by the camera. The state equations (position and speed) of the KF become:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{v}_k + \boldsymbol{\eta}_k \tag{1a}$$

$$\mathbf{v}_{k+1} = \mathbf{C}\mathbf{v}_k + \boldsymbol{\omega}_k \tag{1b}$$

where $\boldsymbol{\eta}_k$ and $\boldsymbol{\omega}_k$ represent the process errors, the matrix \mathbf{A} is the identity \mathbf{I} . The matrix $\mathbf{C} = \alpha\mathbf{A}$, where α represents the (constant) deceleration factor, and

$$B = \begin{pmatrix} \tau & 0 & \tau(x_1 - x_2) & 0 & 0 & 0 & 0 \\ 0 & \tau & \tau(y_1 - y_2) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \tau & 0 & \tau(x_2 - x_c) & 0 \\ 0 & 0 & 0 & 0 & \tau & \tau(y_2 - y_c) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \tau \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2}$$

where τ is the time interval between two consecutive frames. The measurement equation of the KF becomes:

$$\mathbf{z}_{k+1} = \mathbf{H}\mathbf{v}_k + \hat{\mathbf{x}}_k + \gamma_k \tag{3}$$

where γ_k is the measurement error of the position vector and \mathbf{H} is the observation matrix, i.e.,

$$\mathbf{H} = \begin{pmatrix} \tau & 0 & \tau(x_{1,k}^- - x_{c,k}^-) \\ 0 & \tau & \tau(y_{1,k}^- - y_{c,k}^-) \end{pmatrix} \tag{4}$$

We assume that $\boldsymbol{\eta}_k$, $\boldsymbol{\omega}_k$ and γ_k are independent random variables with zero means and covariance matrices, \mathbf{Q}_1 , \mathbf{Q}_2 and \mathbf{R} . Finally, the equation of the Kalman filter for the position vector becomes:

(a) Prediction phase

$$\hat{\mathbf{x}}_{k+1}^- = \mathbf{A}\hat{\mathbf{x}}_k + \mathbf{B}\hat{\mathbf{v}}_k \tag{5a}$$

$$\mathbf{P}_{k+1}^- = \mathbf{P}_k + \mathbf{Q} \tag{5b}$$

where \mathbf{P}_k is the covariance matrix of the error estimation.

(b) Updating phase

$$\mathbf{K}_k = \mathbf{P}_k^{-1} \mathbf{H}^T (\mathbf{H} \mathbf{P}_k^{-1} \mathbf{H}^T + \mathbf{R})^{-1} \tag{6a}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_k^-) \tag{6b}$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^- \tag{6c}$$

The KF requires an initialization step. In particular, the initial values (at the time instant $t=0$) of the state vectors, \mathbf{x}_0 and \mathbf{v}_0 , and of the covariance matrix \mathbf{P}_0 must be defined.

4 Experimental Results

Black/white cameras with near infrared response have been employed for experiments. Image grabbing was performed at 256x256 pixels resolution. The cameras have been placed to monitor the entire area of a bad illuminated road tunnel. Figure 3 shows an image sequence containing multiple moving vehicles.

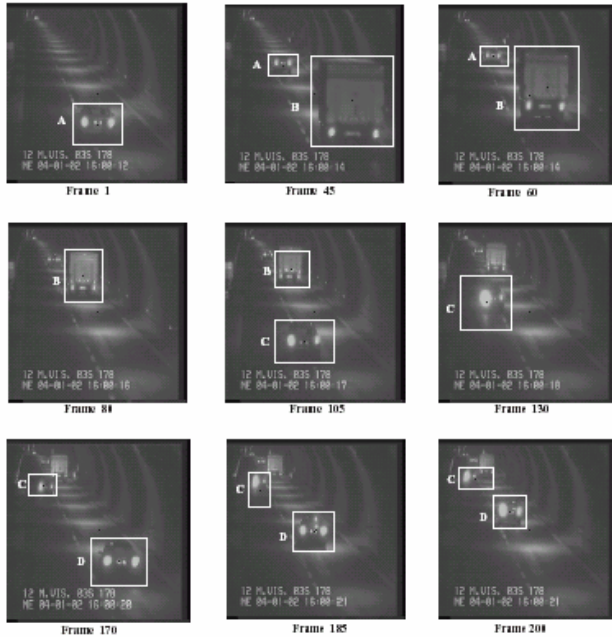


Fig. 3. Some frames of a long image sequence containing multiple moving vehicles

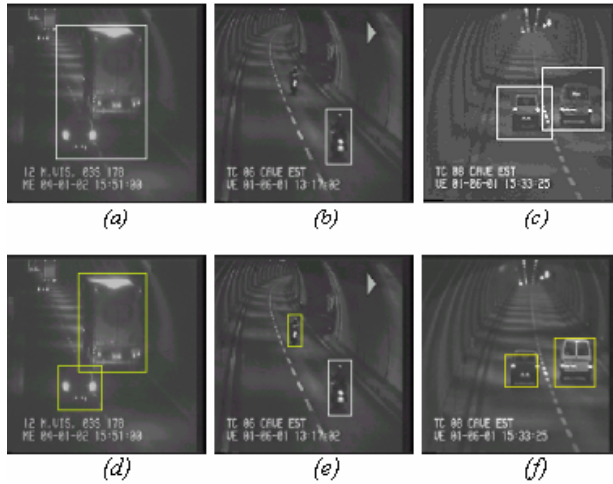


Fig. 4. Results of the proposed system in different situations

The images in Figure 3 belong to a 4 hours sequence (composed of about 4×10^5 frames) characterized by low illumination, light spots and light reflection produced by the tunnel walls and by the rear lights of the vehicles. Figure 4 shows the results of the proposed system in three different complex situations: (1) fused MBRs due to blobs generated by closed vehicles (e.g., a vehicle overtakes another vehicle) (Fig. 4a), (2) missed MBRs due to occlusions or wrong blob detection (Fig. 4b), (3) wrong MBRs due to enlarged blobs generated by shadows or by light spots (Fig. 4c). The output of the system is shown in Fig. 4d, 4e and 4f, respectively.

It is worth noting that a percentage of about 9% of False Alarms (FAs) and a percentage of about 6% of Missed Alarms (MAs) have been obtained by the system on the 4 hours test sequence in Figure 4.

5 Conclusions

Within the European MISS project aimed at improving road safety, a system for real-time traffic monitoring in bad illumination conditions has been presented here. The system takes into account illumination changes and it performs vehicle tracking by means of an extended Kalman filter with a 2D state vector whose components represent each vehicle's position and speed. An affine model taking into account translations and scale changes is used to represent accurately vehicles' motion.

Acknowledgements

This work was supported by the European FP6 Project MISS "Monitor Integrated Safety System" (TST4-CT-2004-516235). Authors wish to thank Mr. Christian Nadalutti for his implementing support.

References

- [1] G. L. Foresti, C.S. Regazzoni, and P.K. Varshney, *Multisensor Surveillance Systems: The Fusion Perspective*, Kluwer Academic Publishers, 2003.
- [2] A.E. Grace, D. Pycock, H.T. Tillotson, and M.S. Snaith, "Active shape from stereo for highway inspection", *Machine Vision and Applications*, vol. 12, no. 1, pp.7-15, 2000.
- [3] M. Aoky, "Imaging and analysis of traffic scene", *In.1 Conf. on Image Processing*, Kobe, Japan, October 1999, pp. 28-32.
- [4] S. D'Agostino, "Commercial machine vision system for traffic monitoring and control", *SPIE*, Vol. 1615, pp.180-186,1991.
- [5] G.L.Foresti, "Real-time detection of multiple moving objects in complex image sequences", *Int. Journal of Imaging Systems and Technology*, Vol. 10, 1999, pp. 305-317.
- [6] D. Bullock and S. Mantri, "Multimedia data model for video detection research", *Journal of Transportation Engineering*, vol. 121, no. 5, pp. 385-390, 1995.
- [7] X. Yuan, Y-J. Lu and S. Sarraf, "Computer vision system for automatic vehicle classification", *Journal of Transportation Engineering*, vol. 120, no. 6, pp. 861-876, 1994.
- [8] Z. Zhu, G. Xu, B. Yang, D. Shi, and X. Lin, "VISATRAM: a real-time vision system for automatic traffic monitoring", *Image and Vision Computing*, vol. 18, no. 10, pp.781-794, 2000.
- [9] Online White Paper: "European transport policy for 2010: time to decide": http://europa.eu.int/comm/energy_transport/en/lb_en.html, 2001.