Vehicle Tracking Using a Multi-scale Bayesian Algorithm for a Perspective Image of a Road

Przemysław Mazurek and Krzysztof Okarma

Higher School of Technology and Economics in Szczecin
Faculty of Motor Transport,
Klonowica 14, 71-244 Szczecin, Poland
{okarma,mazurek}@wste.szczecin.pl

Abstract. Tracking the vehicles in the short distance from the camera using the perspective view of the camera installed above the road requires considering the effect of perspective. Vehicles that are close to the camera are large and easily distinguishable on individual video frames. Moreover, the separation between adjacent vehicles is also high. However, tracking capabilities are deteriorated for longer distances of the vehicle from the camera. The article presents a solution for tracking the vehicles moving away from the camera, based on the image analysis at different scales in order to increase the range of the tracking system. The proposed algorithm utilizes a block matching technique using the correlation coefficients but the size of matched blocks varies for different video frames. Matching can be implemented in several variations depending on the choice of the reference blocks only in the previous frame or averaging of several frames. A particularly useful technique, used in the paper, is the spatio-temporal recursive Track-Before-Detect algorithm, especially for distant objects represented by a small number of pixels.

Keywords: Track-Before-Detect, Multi-scale video tracking, Intelligent Transport Systems.

1 Introduction

Tracking algorithms based on the video systems usually utilise a classical approach assuming the sequential process of detection of the vehicle, tracking and the assignment to a given trajectory. For this reason such systems belong to the class of multi-target tracking [1,2]. Utilising some limitations related to the vehicles' motion trajectory (movements are possible only on a specified plane determined during the calibration of the system), the location and velocity of the vehicles can be determined using only the data acquired from a single camera. Working quality of such system is dependent on a number of factors such as camera location relatively to the road and mutual obscuration of the vehicles associated with it. A typical working location of the cameras is the installation on a pylon or a building on a side of a road. The cameras are often directed perpendicularly to the axis of the road, especially if they are mounted directly

J. Mikulski (Ed.): TST 2011, CCIS 239, pp. 346–353, 2011.

over the road (e.g. with the use of some overpasses), usually at an acute angle to the road, because their aim is to observe traffic in their near neighbourhood, in particular the license plate numbers. the application of the cameras for the estimation of motion parameters from large distances is troublesome, because of necessary complicated image processing operations, limiting its applicability. The usage of such cameras together with some modern tracking algorithms allows the extension of the working range of the cameras and more importantly some existing systems may be used for this purpose. Unfortunately, the quality of acquired video frames decreases significantly, particularly the resolution of the tracked vehicles and their distinguishability against the background.

An alternative solution is the usage of some Track-Before-Detect (TBD) algorithms, assuming that the object is present on the consecutive video frames [3,4]. Performing the tracking the probability values are accumulated in each step, which are then used for the assignment of the object to the motion trajectory (for high values). A disadvantage of such approach is high computational cost of the TBD algorithm related to the necessary simultaneous tracking of many trajectories, even for an empty road without any vehicles.

The application of the TBD algorithms allows the proper tracking in the presence of strong noise [6]. Nevertheless, in such conditions an appropriate model of the possible trajectories and a large number of observations are necessary. In the non-recurrent TBD algorithms the trajectories may be constructed without any limitations. In most cases the recurrent TBD algorithm can be applied using the data acquired directly from the device, i.e. directly mapping the coordinates from the image into the input data.

2 Track-Before-Detect Algorithms

There are many existing Track-Before-Detect algorithms and one of the most interesting ones is the likelihood TBD [8], which is usually implemented as the likelihood ratio TBD die to some computational reasons.It is initialised by the probability value of the object being in a given state $p(t_0, s)$ divided by the probability of not being in this state $p(t_0, \phi)$. The likelihood values can vary within a broad range and, in contrast to the probability, their integral values equal to zero during the processing are not necessary. The algorithm can be described as:

Likelihood ratio initialisation:

$$\Lambda(t_0, s) = \frac{p(t_0, s)}{p(t_0, \phi)} \qquad for\ s \in S \tag{1}$$

For $k \geq 1$ *and* $s \in S$
Motion Update:

$$\Lambda^{-}(t_k, s) = \int_S q_k(s|s_{k-1}) \Lambda(t_{k-1}, s_{k-1}) ds_{k-1} \tag{2}$$

Information Update:

$$\Lambda(t_k, s) = L_k(y_k, s)\Lambda^-(t_k, s) \tag{3}$$

EndFor

where: t – time moment, s – state, k – step number, Λ – likelihood ratio, $q_k(s|s_{k-1})$ - state transition, q – Markov matrix, L_k – measurement likelihood.

Measurement likelihood can contain the processing characteristic of the input system [8] and is used for the input of the data. The result is expressed as the Λ value (from the prediction) or the current Λ^- value.

A similar method is the Spatio-Temporal TBD algorithm, where the information update formula mixes input data and predicted positions. The motion update formula utilises Markov matrix for the dispersion of probabilities (or likelihoods) between the current and future time steps. The new informations from measurements improve sharpness of the state space (probabilities or likelihoods). Depending on the value of the weight coefficient α the next prediction step is based mainly on the incoming data (α close to 0) or previous prediction (α close to 1). The basic algorithm can be described as [8]:

Initialisation:

$$P(k = 0, s) = 0 \tag{4}$$

For $k \geq 1$
Motion Update:

$$P^-(k, s) = \int_S q_k(s_k|s_{k-1})P(k-1, s_{k-1})ds_{k-1} \tag{5}$$

Information Update:

$$P(k, s) = \alpha P^-(k, s) + (1-\alpha)X(k, s) \tag{6}$$

EndFor

where: s – particular space, k – number of iteration, X – input data, $q_k(s_k|s_{k-1})$ - state transition (Markov matrix), P^- – predicted TBD output, P – TBD output, α – weighting (smoothing) coefficient $\alpha \in (0; 1)$.

The tracking algorithm can be used directly, but the crucial aspect of the practical implementation is the limitation of the amount of the input data. One of the possible approaches is the usage of the background estimation for this purpose.

The background estimation can be applied in many variants, but the most typical ones are the utilisation of the exponential smoothing filters, long nonrecurrent moving average (MA) filters (implemented using the pipeline techniques), median filters and median filters with downsampling. The most efficient from the numerical point of view is the exponential filter but its disadvantage is

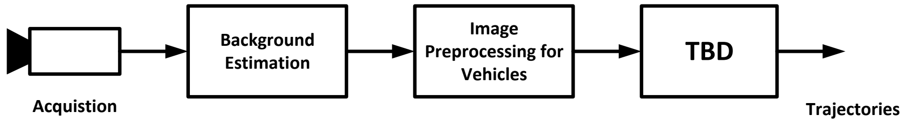

Fig. 1. The idea of the tracking system with the limited amount of computations and improved detection of the vehicles for the TBD

high inertia related to varying light conditions. The estimation of the background allows to determine roughly the size of the vehicle and its location utilised further for a faster initialisation of the TBD algorithm and increasing the separation between the vehicles and the background. The multiple objects tracking using this method is also possible but in this paper only the case of a single vehicle tracking is analysed.

3 Image Preprocessing for Vehicle's Tracking

The application of the correlation coefficient allows the utilisation of the numerical value without the necessity of thresholding, which introduces a noise into the tracking system. Such noise is the effect of the binarisation of the input data (1 stands for the presence of the tracked object in a given location and 0 for no object in this position). The images acquired by the video camera are greyscale or colour ones and the thresholding operation can also be used for them, causing the simplified construction of the system but lowering its efficiency, especially reducing the working range.

An alternative technology is the usage of the systems working with raw data, acquired directly from the measuring sensor (from the camera in this case) or the processed data obtained as the result of some algorithms which do not perform any binarisation. A typical approach is the usage of the two-dimensional correlation interpreted as the degree of matching of two images. Such operations used for the comparison of images (some other ones are Mean Squared Error, Mean Absolute Error as well as some modern full-reference image quality assessment methods [7]) are usually applied for the extraction of some relevant features of the image which cannot be determined without the knowledge of two images. Such operation allows the utilisation of the informations related to the tracked object from some other measurements for the increase of the object's searching effectiveness the based on some other ones.

For the images acquired by the video camera, the representation of the closely located vehicle on the image plane can occupy relatively large number of pixels, while some distant vehicles' representations are much smaller. Tracking objects, which move away from the camera, requires the comparison of the initial image of the object with all assumed possible locations for each possible scale (size) of the object.

The comparison operation using the assumed criteria (e.g. correlation) is used for two purposes. The first aim is to determine the location and the second one is related to the decrease of the object's size in a new space of the input data for the tracking algorithm. Bayesian tracking algorithms using the raw data work efficiently for the pixel-size data, where the vehicle is represented by a single point. During the tracking process the locations are blurred because of the predictor's working properties, so the appropriate data preprocessing is necessary. Direct input of the data, where objects are represented by several pixels is not desired, since the input data contain the informations corresponding to some details of the objects (related to its different parts e.g. some windshields or wheels). In such case the values of the signal can be both positive or negative in relation to the background for different parts of the same object (vehicle), so the total signal can be blurred by the predictor. The resulting signal in the tracking algorithm will be weakened and the tracking accuracy will decrease. Utilising the correlations of an object with the image for an appropriate scale, a significant limitation of the area of interest with maximum correlation values can be usually obtained.

Applying the multi-scale analysis and comparison the two general approaches can be used: scaling can be performed on the smaller image representing the vehicle or on the larger image of a road. In this paper the first approach has been used due to smaller distortions introduced by such operation.

The two-dimensional correlation is defined as the following operation using the two rectangular images (A and B)

$$r = \frac{\sum\limits_{m,n} \left(A_{m,n} - \overline{A}\right)\left(B_{m,n} - \overline{B}\right)}{\sqrt{\sum\limits_{m,n} \left(A_{m,n} - \overline{A}\right)^2 \sum\limits_{m,n} \left(B_{m,n} - \overline{B}\right)^2}}, \tag{7}$$

and can be implemented also for non-rectangular images. It is important for the vehicles' tracking, since depending on the camera's location and orientation the tracked vehicle can be observed from above, side view, front-side or back-side etc. at different angles. In this case a binary mask can be used, which defines the areas of interest and prevents the comparisons with the surrounding of the object. In such case the additional condition for all the summations in the Eq. 7 is introduced: $m, n \in M$, where M denotes the binary mask. The example of such limitation is presented in Fig. 2, where the mask obtained from the background estimation (white pixels) is used for the limitation of the potential area of interest and the correlation values with some other fragments of the scene are filtered out.

Scaling operation performed on the image corresponds also with the same scaling on the assigned binary mask and the number of allowed scales depends on the desired precision as well as the working range (usually several scales are used). In the paper 14 scales have been used experimentally starting from 1 (initial original image) with the step -0.05. The set of images in such scales as well as the appropriate masks are created for each detected image in the scene.

In order to utilise the correlation between the basic data and the consecutive video frames it is necessary to consider the influence of changes of the distance on

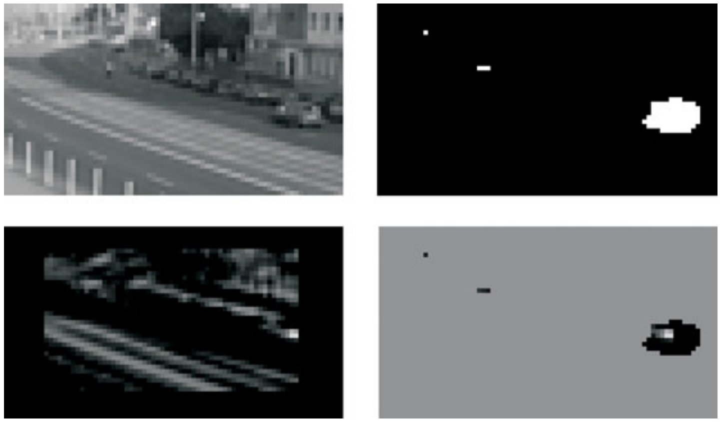

Fig. 2. Original image, mask obtained from the background estimation, correlation of the vehicle's model for the whole image and the same correlation for the masked fragment of the image

the image size. Since the correlation is calculated for the images having different resolutions (size), there may be the situation when the highest correlation values are obtained for the smallest size of the model (a vehicle represented by a few pixels only), while the model having the proper scale leads to lower correlation values with all the fragments of the video frame. Such phenomenon is caused by the presence of image distortions, especially in the nearest neighbourhood of the tracked vehicle, which can be a result of even small changes of lighting conditions. In such situation a set of highly correlated values are obtained for the smallest scales instead of a single one for the proper scale of the model.

This effect is especially strong for the large number of available scales, so in order to reduce this problem the solution based on the utilisation of two masks

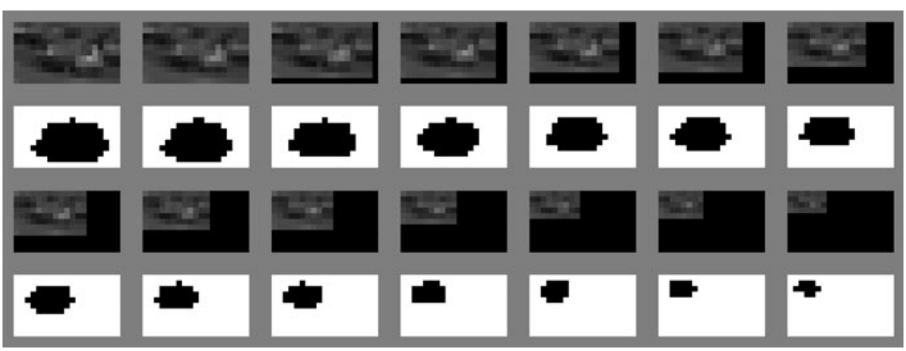

Fig. 3. Different scales of the initial image of the vehicle and the corresponding binary masks for an exemplary vehicle

is proposed. Both, the mask of the vehicle from the current base of the tracked objects (at different scales) and the current mask, originate from the background estimation algorithm. The mask taken from the base can be treated as a good estimate because of its acquisition at a small scale, but the current mask is only a rough approximation, especially for small objects. This mask is usually noisy, regardless of the possible usage of the floating numbers in the background estimation algorithm instead of typical 8-bit integers. Apart from the sensor noise and light changes, small objects cause the blurring effects as well as the local changes of brightness. This effect corresponds to the discrete character of the image (grid) and some limitations related to the sampling process. The maximum excitation of a single pixel may be equivalent to the 50% of excitation of two neighbouring pixels (causing the blurring and change of brightness) or even 4 pixels excited by 25% of the signal, depending on the position of the object relatively to the image grid. Using the simple thresholding of the differences between the current image and the estimate od the background in the presence of noise such blurred differences are often below the threshold level, so the proper mask cannot be determined. Typically, the size of the mask decreases before it disappears, and the tracking algorithms based on the mask cannot determine the proper motion trajectory because of such noise.

The above discussed reasons cause the necessity of using some other methods for object's searching with a small mask. The main problem is related to the high number of false matches, so the combination of mask fitting and multi-scale matching is necessary and the TBD algorithm can be used for the further improvement. The size of the mask from the base is compared to the current image mask in the location of the vehicle. For high differences the scale (mask size) is rejected from further processing due to too small or too large size of the mask. The results of the tracking (correlation range $\langle -1; 1\rangle$ has been limited to $\langle 0; 1\rangle$ for better visualisation) obtained using the proposed approach are shown in Fig. 4, where the highest correlation values are achieved for different scales decreasing according to the increase of the distance of the tracked vehicle from the camera.

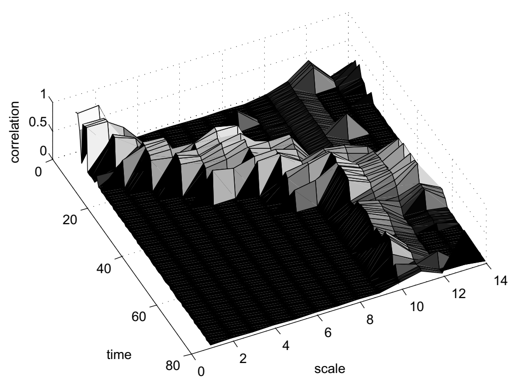

Fig. 4. Results obtained by the proposed approach

4 Conclusions

The algorithm presented is the paper can be efficiently used in the Intelligent Transportation Systems, which utilise vehicles' tracking based on the Track-Before-Detect algorithm. The main advantage of the proposed approach is the improvement of tracking distant objects represented by a small number of pixels observed from a perspective view. Due to changes of their size on the consecutive video frames the discussed multi-scale analysis is particularly helpful for proper tracking of objects moving away or towards the camera. It can also be considered as potential extension for the systems dedicated only for tracking objects observed from the side view.

Acknowledgements. This work is supported by the Polish Ministry of Science and Higher Education (Grant No. N509 399136 "Estimation of the vehicles' motion trajectories using the Bayesian analysis and digital image processing algorithms").

References

1. Bar–Shalom, Y.: Multitarget–Multisensor Tracking: Applications and Advances, vol. II. Artech House, Norwood (1992)
2. Blackman, S., Poupoli, R.: Modern Tracking Systems. Artech House, Norwood (1999)
3. Blackman, S.: Multiple–Target Tracking with Radar Applications. Artech House, Norwood (1986)
4. Boers, Y., Ehlers, F., Koch, W., Luginbuhl, T., Stone, L.D., Streit, R.L. (eds.): Track Before Detect Algorithm, EURASIP Journal on Advances in Signal Processing, Hindawi (2008)
5. Klein, L.A.: Sensor Technologies and Data Requirements for ITS. Artech House ITS library, Norwood (2001)
6. Mazurek, P.: Suppression of Impulse Noise in Track-Before-Detect Algorithms. Computer Applications in Electrical Engineering 8, 201–211 (2010)
7. Okarma, K.: Combined Full-Reference Image Quality Metric Linearly Correlated with Subjective Assessment. In: Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2010. LNCS (LNAI), vol. 6113, pp. 539–546. Springer, Heidelberg (2010)
8. Stone, L.D., Barlow, C.A., Corwin, T.L.: Bayesian Multiple Target Tracking. Artech House, Norwood (1999)
9. Traffic Detector Handbook: Third Edition, Publication No. FHWA-HRT-06-10, Federal Highway Administration, vol. I, US Department of Transportation (2006)