Background Suppression for Video Vehicle Tracking Systems with Moving Cameras Using Camera Motion Estimation

Przemysław Mazurek and Krzysztof Okarma

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

Abstract. Camera oscillations and slight movements are typical in the video based parts of the Intelligent Transportation Systems, especially in the cases when the cameras are mounted on the high pylons or pillars, similarly as some street lamps. The influence of strong wind and some vibrations caused by some heavy vehicles may result in some shifts of the images captured as the consecutive video frames. In such situations some typical background estimation and removal algorithms based on the comparison of corresponding pixels of each video frame may lead to significant errors. The influence of such camera motions increases seriously for high focal length corresponding to tracking of distant objects. In order to minimize the influence of such movements the background suppression algorithm using the camera motion estimation is proposed in the paper increasing the stability of the estimated background which is further used in the vehicle tracking algorithm.

Keywords: background suppression, vehicle tracking, motion estimation.

1 Introduction

One of the main purposes of the Intelligent Transportation Systems (ITS) is allowing the improvement of many traffic parameters, such as reduction of the number of traffic jams, shortening the travel time, reduction of toxic emissions by vehicles as well as the reduction of the number of disturbing events (e.g. traffic accidents). Such systems consist of three major subsystems: measuring (acquisition of the traffic data), controlling (containing the traffic control algorithms) and executive (directly affecting the road users). A crucial part is the measuring subsystem containing many road sensors [1]. The most typical ones are the induction loops, WIM (Weight in Motion) sensors or radar sensors (Doppler based) allowing the measurements at a specified fragment of a road regardless of weather conditions. Such measurements can also be conducted at more locations but their cost increases significantly. Furthermore, not all parameters can be revealed using such sensors, so an interesting supplement for them can be video

J. Mikulski (Ed.): TST 2012, CCIS 329, pp. 372–379, 2012.

Fig. 1. The illustration of the image shifts introduced by the camera motion

based measurements. Using the cameras located near the roads many motion and traffic parameters can be measured simultaneously but the working range of video systems and their measurement capabilities strongly depend on weather conditions.

The cameras used in the ITS solutions are typically mounted on dedicated pylons, high pillars, gates or, in some cases, at the buildings' walls. Installations of cameras on the street lamps are rather not used because of their elasticity causing noticeable wavings on the wind affecting the stability of acquired images. In such case the camera's vibrations may significantly influence its working area and range. Such changes are of great importance especially for the cameras with narrow angle lenses, since the acquisition of the video data corresponding with distant objects may be troublesome because of the wind's influence.

Each change of the visible area influences the results of tracking [2], also using the high dynamic range approach [3], especially on the background estimation results, since the commonly used algorithms for such purposes can be applied assuming the constant location of the camera without any motions. Each motion of the camera can be interpreted as the motion of the objects causing the intensity changes of some pixels so further detection is more difficult. Additionally, such motion influences the tracked vehicles' location on the consecutive images acquired by the camera, introducing the changes of objects' velocities both in horizontal and vertical direction. Considering the issues addressed above, a great importance of the image stabilisation leading to the elimination of the wind's influence in the ITS solutions can be noticed. An exemplary illustration of the possible results of the camera motion on the stability of the acquired images is shown in Fig. 1 where the translations can be easily noticed at the image boundaries (only two vehicles are in motion). Fig. 2 contains the results of the background estimation [4,5] for two methods discussed in detail in some earlier papers [6,7,8]: exponential smoothing (with smoothing parameter $\alpha = 0.95$) and median algorithm together with the differences between the "ground truth" background and its estimates. The images have been generated as synthetic using the captured sequence of images for the static camera in order to determine the reference background image. As illustrated both algorithms converge to the reference background image, especially for low density traffic, but the character of the absolute estimation errors during the first phase is different for both methods (global for the exponential smoothing and local for the median algorithm).

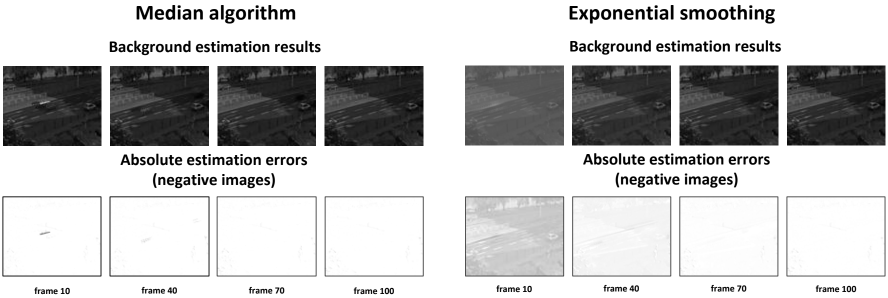

Fig. 2. The exemplary results obtained for two different background estimation algorithms (median algorithm and exponential smoothing)

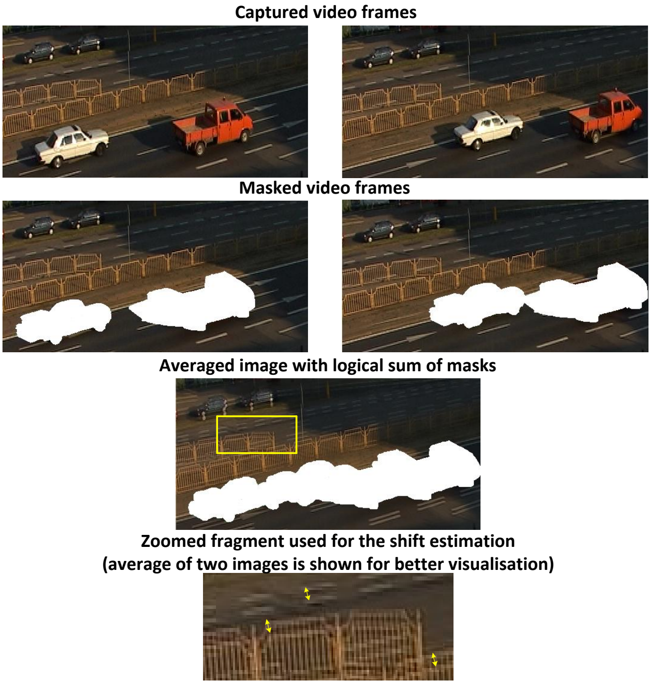

Fig. 3. The idea of masking applied for the image shift estimation

2 Methods for Estimation of the Image Shifts

Estimation of the shifts between a pair of images requires the knowledge about the area where the vehicles can move. It is essential for the elimination of the influence of the vehicles' motion, visible as changes of their locations in consecutive video frames, on the results of the shift estimation. It is necessary to determine a mask covering the representations of some physically moving objects on the image plane. This can be conducted "by hand" with necessary human assistance or automatically using a dedicated algorithm e.g. using shape classification for each video frame [9] in order to determine the possible fragments of images representing the vehicles. Without such masking the impact of the vehicles' shifts relative to the whole image can lead to significantly disrupted results of camera shift estimation. The idea of such approach is illustrated in Fig. 3.

Image matching algorithms may utilise the camera motion model. The application of some estimators, such as Benedict-Bordner, Kalman or Bayes filters [10] allows the reduction of the amount of calculations due to the prediction of the possible camera positions. Since the scene is roughly static and the motion range is limited, such solutions can be replaced by some simplified algorithms based on the image matching for a specified possible motion range.

Image stabilisation methods are based on two approaches:

- image matching according to the characteristic points,
- image matching according to the whole image plane.

The first solution is characterised by the reduced amount of calculations. It is typically applied for the image matching supervised by humans. Its main advantage is probably the robustness to the presence of some strong distortions in the image. The second approach is more sensitive to the preence of noise but the results are obtained using more informations from the larger area of the image.

One of the most typical matching criterions can be expressed as:

$$E(k_x, k_y) = \sqrt{\sum_{x,y} \left[I(x - k_x, y - k_y) - R(x, y)\right]^2} = \sqrt{\sum_{x,y} d(x, y, k_x, k_y)^2} \quad (1)$$

where E is the matching function, k_x, k_y are the horizontal and vertical shifts, I denotes the input image, R is the reference image and x, y are the pixels' coordinates. Nevertheless, this criterion takes into account only the translations without rotations and scaling (which is typically not used). The solution is:

$$E(k_x^{est}, k_y^{est}) = \arg\min_{k_x, k_y} E(k_x, k_y) \quad (2)$$

The translation causes the change of the image area used for the comparison, so the obtained result should be corrected considering the number of pixels used for the comparison:

$$E^*(k_x^{est}, k_y^{est}) = \frac{1}{N} E(k_x^{est}, k_y^{est}) \quad (3)$$

where N is the number of pixels used in the comparisons considering the thresholding and binary masks applied for both images.

The comparison of images is conducted with some errors being the result of noise and light fluctuations. For those reasons a modified criterion using the threshold T can be applied as:

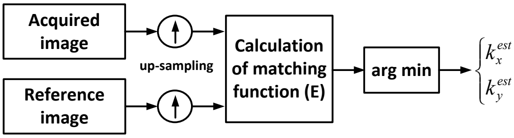

Fig. 4. The idea of image matching with up-sampling and nearest neighbour interpolation

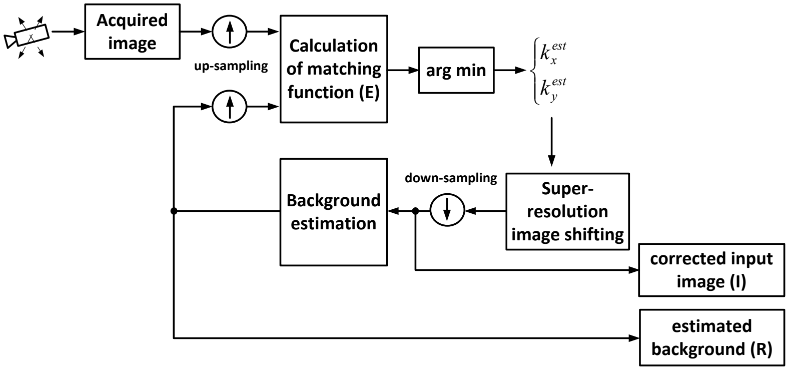

Fig. 5. The idea of the background estimation with sub-pixel image matching for camera motion estimation

$$E^{*}(k_x^{est}, k_y^{est}) = \frac{1}{M}\sqrt{\sum_{x,y} d(x,y,k_x,k_y)^2 \cdot [d(x,y,k_x,k_y) > T]} \quad (4)$$

where M is the number of pixels fulfilling the thresholding condition (for which the difference between the compared pixels exceeds the threshold T) and d is defined as shown in equation (1). The purpose of this operation is the decrease of the influence of noise on the matching results.

The image matching operation is conducted typically with the assumption of integer values of the translations k_x, k_y corresponding to the coordinates of the image pixels. Nevertheless, for the images acquired from the real cameras can be characterised by sub-pixel translations. In such case the image matching can be conducted using up-sampling with nearest-neighbour interpolation as illustrated in Fig. 4. The idea of the backgorund estimation using the camera motion estimation is shown in Fig. 5.

Since the images acquired from the camera during its vibrations may be considered as the observations from slightly different locations, a negative wind phenomenon can be utilised for increasing the resolution of the acquired images. For this purpose the super-resolution algorithms [11,12] can be successfully

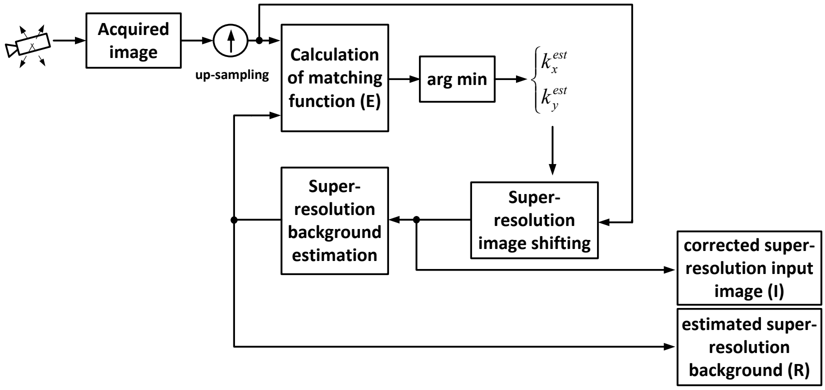

Fig. 6. The idea of the super-resolution background estimation based on the camera motion estimation

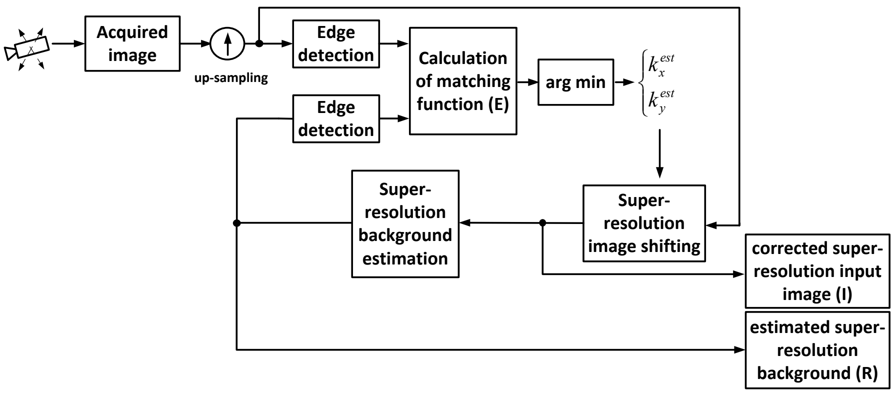

Fig. 7. The idea of the background estimation based on the camera motion estimation and edge detection (without masking)

applied leading to increase of the image resolution based on the series of shifted images. The idea is illustrated in Fig. 6.

Another solution may be based on the image matching based on some characteristics points on the image plane. Nevertheless, some lighting changes between two images can influence on the result of matching, so the method should be independent on such changing lighting conditions. For this purpose some edge detection algorithms may be applied, such as e.g. Prewitt, Sobel, Scharr or Canny filters, allowing detection of image fragments with rapid intensity changes, equivalent to the results of high-pass spatial image filtering, assuming no presence on noise. The idea of such system is illustrated in Fig. 7. Nevertheless, for simplicity, the masking operation is not included in figures.

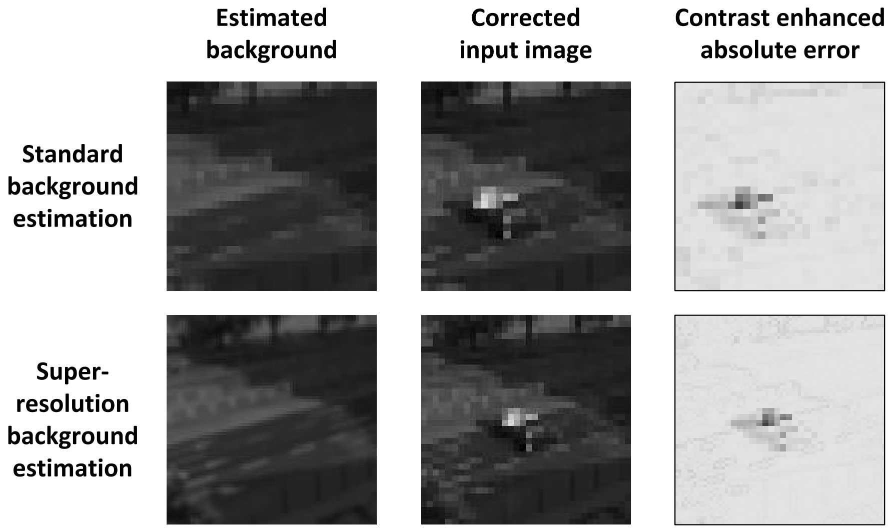

Fig. 8. Illustration of the exemplary results obtained for two different background estimation methods with the super-resolution camera motion detection

3 Results and Conclusions

In order to verify the properties of the discussed methods and their proposed modifications some experiments has been conducted using the implemented algorithms for the background estimation. Analysing the obtained results the advantages of using the methods utilising the super-resolution technology for the image shifting can be easily noticed. The greatest benefits can be observed for the super-resolution background estimation where the details of the obtained image are quite smooth.

Comparing the absolute errors presented on the right side of Fig. 8 for the frame with two vehicles visible on the road, an interesting property of the super-resolution background estimation algorithm can be observed. The largest errors resulting from the presence of vehicles are similar but the errors related to the background area are different for both methods. For the standard method some modified pixels can be observed, which occupy relatively big area of the image. Nevertheless, the corresponding fragments of the image affected by errors are much smaller so their influence on the overall result of the background estimation is also reduced. In this case only the sub-pixels are changed and for the standard method the whole pixels are changed as the smallest possible parts of the image.

The application of the super-resolution background suppression algorithm based on the super-resolution camera motion estimation can be applied as an effective pre-processing step of the video based vehicle tracking in the ITS solutions. A promising direction of further research seems to be the verification of this algorithm in the sophisticated tracking systems for small objects e.g. using Track-Before-Detect approach [13], which is especially useful for tracking distant objects in the presence of a strong noise.

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. Klein, L.: Sensor Technologies and Data Requirements for ITS. Artech House ITS library, Norwood (2001)
2. Mazurek, P., Okarma, K.: Application of Bayesian a Priori Distributions for Vehicles' Video Tracking Systems. In: Mikulski, J. (ed.) TST 2010. CCIS, vol. 104, pp. 347–355. Springer, Heidelberg (2010)
3. Okarma, K., Mazurek, P.: Vehicle Tracking Using the High Dynamic Range Technology. In: Mikulski, J. (ed.) TST 2011. CCIS, vol. 239, pp. 172–179. Springer, Heidelberg (2011)
4. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 4, pp. 3099–3104 (October 2004)
5. Reddy, V., Sanderson, C., Lovell, B.C.: A low-complexity algorithm for static background estimation from cluttered image sequences in surveillance contexts. EURASIP Journal on Image and Video Processing, Article ID 164956, 14 (2011)
6. Okarma, K., Mazurek, P.: Background estimation algorithm for optical car tracking applications. Machinebuilding and Electrical Engineering (7-8), 7–10 (2006)
7. Okarma, K., Mazurek, P.: Nonlinear background estimation methods for video vehicle tracking systems. Archives of Transport Systems Telematics 4(4), 42–48 (2011)
8. Okarma, K., Mazurek, P.: A modified hybrid method of nonlinear background estimation for vision based vehicle tracking systems. Logistyka (3), 1747–1752 (2012) (in Polish)
9. Okarma, K., Mazurek, P.: Application of Shape Analysis Techniques for the Classification of Vehicles. In: Mikulski, J. (ed.) TST 2010. CCIS, vol. 104, pp. 218–225. Springer, Heidelberg (2010)
10. Blackman, S., Popoli, R.: Design and Analysis of Modern Tracking Systems. Artech House (1999)
11. Irani, M., Peleg, S.: Super resolution from image sequences. In: Proceedings of the International Conference on Pattern Recognition (ICPR), Atlantic City, New Jersey, vol. 2, pp. 115–120 (June 1990)
12. Irani, M., Peleg, S.: Improving resolution by image registration. Graphical Models and Image Processing 53, 231–239 (1991)
13. Mazurek, P., Okarma, K.: Vehicle Tracking Using a Multi-scale Bayesian Algorithm for a Perspective Image of a Road. In: Mikulski, J. (ed.) TST 2011. CCIS, vol. 239, pp. 346–353. Springer, Heidelberg (2011)