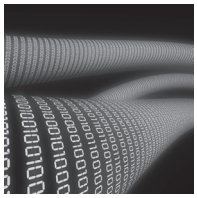

Archives of
Transport System
Telematics

Volume 4

Issue 4

November 2011

# Nonlinear background estimation methods for video vehicle tracking systems

**K. OKARMA[a], P. MAZUREK[a]**

[a]Faculty of Motor Transport, Higher School of Technology and Economics in Szczecin, Klonowica 14, 71-244 Szczecin, Poland,

EMAIL: okarma@wste.szczecin.pl

## ABSTRACT

One of the major advantages of the video cameras' usage for tracking of vehicles is to reduce the costs of Intelligent Transport Systems. However, this requires the development of software techniques allowing an automatic extraction of the vehicle or group of vehicles from the current video frame, which is possible by using the background estimation methods, assuming a fixed camera installed over or at the side of the road. Background estimation based on the linear image filtering algorithms can be performed by averaging a certain number of video frames. However, this technique is relatively slow, which complicates its use, especially in variable lighting conditions. The paper presents an alternative background estimation technique, utilised for its further replacement, based on the nonlinear image filtering algorithms.

**KEYWORDS: background estimation, video tracking, Intelligent Transport Systems**

## 1. Introduction

Video based vehicle tracking systems [1] are based on two types of cameras sensitive on the visible light or the infra-red ones. Regardless of its type one of the basic operations used for the reduction of the amount of processed data, as well as their transmission in distributed traffic monitoring systems [2,3], is related to the estimation of background and its elimination from each video frame captured by the camera.

The most typical approach to background elimination is based on more or less complicated motion detection algorithms. In the simplest case (called also the naïve approach) the neighbouring frames are compared with the use of the threshold and all the corresponding pixels which have the same colour are classified as representing the background. The main disadvantage of such approach in practical applications is its sensitivity to noise and changes of lighting conditions. In such cases, typical for the outdoor acquisition of the video signals e.g. for traffic monitoring purposes, the threshold should be adaptively changed or some more advanced algorithms can be applied.

A reliable estimation of the background objects should be not only weather-proof but also insensitive to some other disruptions e.g. related to some rapid local colour changes. The most typical reasons may be the directional light reflections related to the CCD thermal noise, influence of street and car lights, the presence of water on a road, leaves moving on the wind etc. [4]. Such rapid change of the background may also be caused e.g. by a vehicle starting from a parking previously classified as a non-moving element of the background (changes in the background geometry).

The influence of some other long term disturbances, especially those having rather global character, is usually easier to predict e.g. changes of light conditions caused

by street lamps, slowly moving clouds, sun, shadows etc. Another relevant element which should be considered is the influence of camera oscillations as well as the warm air motion caused by high temperature of the asphalt.

# 2. Background estimation algorithms

The basic method of background estimation (working as the differential detection) assuming the previous frame as the background works well only in the constant light conditions without any moving objects on the scene except the tracked vehicle. It is very fast and similar to some simple motion detection algorithms and some video compression algorithms which do not utilise any motion vectors. Some additional limitations are related to the object's speed and the camera's frame rate as well as the threshold. Since the differences of corresponding pixels' colours between two neighbouring frames can be either positive or negative the dynamic range of the resulting image increases, or the absolute value can be used.

Another approach is based on the averaging of the specified number ($t$) of frames [5] and can be expressed as:

$$B_t^{AVG}(u,v) = \frac{1}{t}\sum_{k=1}^{t} I_k(u,v) \quad (1)$$

where $u$ and $v$ denote the pixel's coordinates.

This method is slow and memory consuming so it can be modified towards the moving (running) average (MA) or the exponential smoothing filter [6]. The MA filter can be described as:

$$B_t^{MA}(u,v) = \frac{1}{N}\sum_{k=0}^{N-1} I_{t-k}(u,v) \quad (2)$$

or in the recurrent form as:

$$B_t^{MA}(u,v) = B_{t-1}^{MA}(u,v) + \frac{1}{N}\left(I_t(u,v) - I_{t-N}(u,v)\right) \quad (3)$$

where $B$ stands for the estimated background and $I$ is the input image.

In some systems the weighted average of the each pixel's recent history is used, where the most recent frames have higher weighting coefficients. Another modification can be based on the additional selectivity so pixels which have been classified as the foreground can be ignored in the background model in order to prevent the corruption of the background by the pixels logically not belonging to the background scene [7].

One of the most relevant limitations of the classical linear methods of background estimation is troublesome choice of threshold. It is typically based on a single value, not dealing with some multiple modal background distributions.

Another interesting idea is based on Gaussian average with fitting the Gaussian distribution over the histogram with running average update. For the multimodal background distributions the Mixture of Gaussians approach can be used, but there are also some problems with initialisation and update over time. Since, some of Gaussian distributions model the foreground and some others correspond to background, there is a need to divide them into such groups [7].

# 3. Experimental evaluation of algorithms

## 3.1. Initialisation of the algorithms

Background estimation can be applied with the use of the exponential smoothing filter IIR (Infinite Impulse Response) of the first order, characterised by inherent stability, expressed as:

$$B_t^{EXP}(u,v) = \alpha \cdot B_{t-1}^{EXP}(u,v) + (1-\alpha)\cdot I_t(u,v) \quad (4)$$

The initialisation can be done using two approaches:

$$B_{t=0}^{EXP}(u,v) = 0 \quad (5)$$

or

$$B_{t=0}^{EXP}(u,v) = \frac{1}{2}\cdot Range \quad (6)$$

where *Range* denotes the dynamic range of the image depending on its type (0-1 for the normalised images represented by the floating point numbers or 0-255 for 8-bit unsigned integer notation typical e.g. for 24-bit RGB images).

According to the formula (5) the background estimate is initialised by the black pixels, so the convergence can be achieved after the time necessary for obtaining the luminance level of the brightest pixel of the background. Such time can be calculated using the step response of the filter. The modified initialisation (6) can be used for the acceleration of the convergence due to the choice of the middle level of luminance as a starting point for the algorithm.

The chosen value of the parameter a should be large (close to 1), since the input image usually has the range 0-255 and the estimation update with the component $(1-\alpha)\cdot I_t(u,v)$ should be large enough to suppress the noise (preferably represented as a floating-point number).

The results of the background estimation using two different initialisation schemes are illustrated in Fig. 2 for five chosen frames (no. 1, 1000, 2500, 4000 and 5000).

Images on the left side illustrate the current frames, while the middle and the right columns illustrate the results of background estimation using the initialization by the luminance equal to 0 and 128.

## 3.2. Median-based estimation

Considering some disadvantages of the linear filters, mainly their sensitivity to impulse noise, some nonlinear algorithms may be used instead of them. Such filters, mainly the median ones, are robust for rapid local changes of luminance values, which are typical for moving objects over the static background [8].

The basic median algorithm can be described as:

$$B_t^{MED}(u,v) = MEDIAN\left(\begin{Bmatrix} I_t(u,v), I_{t-1}(u,v), \dots \\ \dots, I_{t-(N-1)}(u,v) \end{Bmatrix}\right) \quad (7)$$

where the pixels with the same coordinates $(u,v)$ from $N$ neighbouring frames are sorted and the middle element of the sorted vector value is chosen as the result. For the even number of elements ($N$) in the sorted vector (frames) the result is the average of the two middle values, so such filter can be treated as partially averaging filter. In order to increase the processing speed and reduce the influence of noise, the median filter with temporal downsampling can be used, where some frames are not used. In such case the impact of the vehicles moving on the scene is significantly reduced, since they occupy different areas of the image in the frames used for the analysis. Such filter is described as:

$$B_t^{DM}(u,v) = MEDIAN\left(\begin{Bmatrix} I_t(u,v), I_{t-M}(u,v), \dots \\ I_{t-1,M}(u,v), \dots, I_{t-(N-1)M}(u,v) \end{Bmatrix}\right) (8)$$

where $M$ is the number of omitted frames.

Comparing the results of the background estimation using median filtering the advantages of using the temporal downsampling can be easily noticed. Illustration of such differences are shown in Fig. 3, where the original frames are shown in the left column, the results obtained for "standard" median filter in the middle, and the effects of using the median filter with temporal downsampling (with $N$=11 and $M$=5) in the right column.

Obtained results can be verified by a human operator

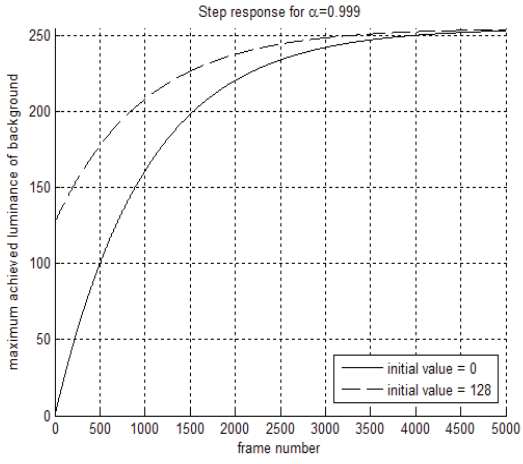


**Fig.1. Comparison of the step responses for the convergence testing of two initialisation schemes.**

**Fig.2. Comparison of the obtained results for two initialisation schemes.**

**Fig.3. Comparison of the obtained results for selected frames using two versions of median filters.**

using subjective evaluation or utilising some automatic image quality assessment methods. Nevertheless, some of the metrics are well correlated with human perception of distortions and similarity of images but are not reliable for the error estimation purposes.

### 3.3. Automatic verification using image quality assessment methods

Two typical approaches to image quality assessment are subjective evaluation and using objective measures. Subjective evaluation requires performing some tests based on filling the questionnaires by the observers what allows calculation of the Mean Opinion Score (MOS) and some further statistical analysis. For this reason its application to image or video processing applications is seriously limited because of the necessity of using time-consuming evaluation by observers.

Much more desired method for computer applications is objective evaluation based on preferably single scalar value related to the overall quality of the image. Such automatic measure can be used e.g. as the optimisation criterion in many digital image and video processing applications. A good example can be lossy compression where it is often relevant to decide whether e.g. 1% better compression ratio introduces artifacts causing serious reduction of the quality.

Some classical image quality measures [9] such as Mean Square Error (MSE) and some similar ones e.g. Peak Signal-to-Noise Ratio (PSNR) are poorly correlated with Human Visual System so recently some new metrics have been proposed. Nevertheless, some traditional measures based on the analysis of single pixels without their neighbourhood are still in use, especially for the detection of changes between two images, especially in the applications where the human perception is not critical.

All such methods belong to the group of full-reference methods, which require the knowledge of the original image without any distortions. Such approach is typical for the optimisation of many image processing algorithms, where the knowledge of the original image is assumed. In this paper the "ideal" background image is also assumed as known, since the image of the road without any moving vehicles or long-term average can be used for this purpose. Nevertheless, in practical applications, especially for a high density city traffic the acquisition of such "empty" background frame is often impossible.

Application of "blind" image quality assessment methods [10], where the original image is not necessary, is quite complicated task and is not analysed of this paper. Such no-reference methods are rather specialised and insensitive to many types of distortions, so their main application area is limited e.g. to the estimation of the amount of noise, quality prediction of the JPEG compressed images [11] or blurred ones [12,13].

In this paper two full-reference metrics have been used for the verification of the background estimation algorithms. The first classical method is the Peak Signal-to--Noise Ratio (PSNR) defined as:

$$PSNR = 10 \cdot \log_{10} \sum_{u,v} \left( \frac{k^2}{[B(u,v) - Q(u,v)]^2} \right) \qquad (9)$$

assuming that $Q$ is the reference background image, $B$ is the current estimation and $k$ denotes the dynamic range (255 for the 8-bit image or 1 for the normalised one).

Due to poor correlation of classical metrics with the Human Visual System (HVS) some new image quality measures have been proposed in recent years. The first one [14] is the Universal Image Quality Index (UIQI), further extended [15] into Structural Similarity (SSIM). This metric is probably the most popular modern approach to automatic image quality assessment. The local SSIM index for the fragment of the image (typically 11×11 pixels) can be calculated as:

$$SSIM = \frac{(2 \cdot \bar{x}\bar{y} + C_1) \cdot (2 \cdot \sigma_{xy} + C_2)}{(\bar{x}^2 + \bar{y}^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \qquad (10)$$

where $C_1$ and $C_2$ are small constants preventing the division by zero chosen such that they do not introduce significant changes of obtained results (recommended values are ). Symbols $\bar{x}$ and $\bar{y}$ denote the mean values and $\sigma^2$ stands for the variances ($\sigma_{xy}$ is the covariance) within the current window (x and y are the original and distorted image samples respectively). This measure allows creating a quality map of the image using sliding window approach and the overall scalar quality index for greyscale images is obtained as the average value of the local indexes using the Gaussian weighting (windowing) function. The size and type of the weighting function can be changed [16,17], influencing the properties of the metric, but these changes are not significant for the tests conducted in this work.

The PSNR and SSIM metrics discussed above have been used for the comparison of the obtained estimates with the reference background image. The results are presented in Figs. 4 and 5 respectively.

Analysing the results presented in Figs. 4 and 5 the advantages of the median approach can be noticed in the first time period because of it fast convergence to a good estimate of the background. Unfortunately, there are some negative peaks present in the plot, caused mainly by the moving large vehicles, where the length of the sorted vector within the median filtering procedure is too small. In the long--time period the background estimation obtained by the exponential smoothing filter is better, so the combination of both methods could be used. The median estimation with temporal downsampling should be used for the initial part

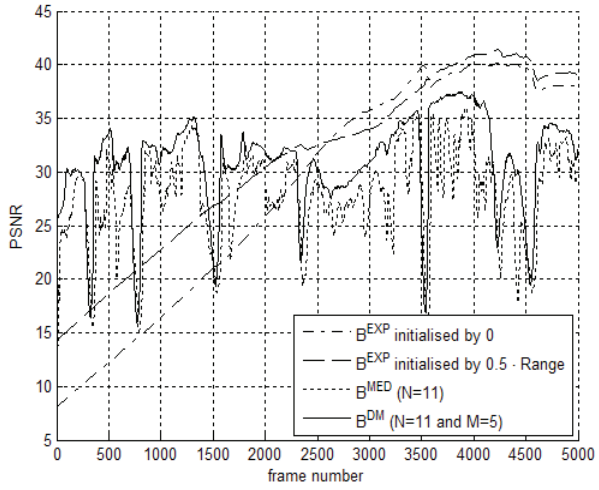


**Fig.4. Peak Signal-to-Noise Ratio (PSNR) values for consecutive frames of the background estimation using four various filters.**

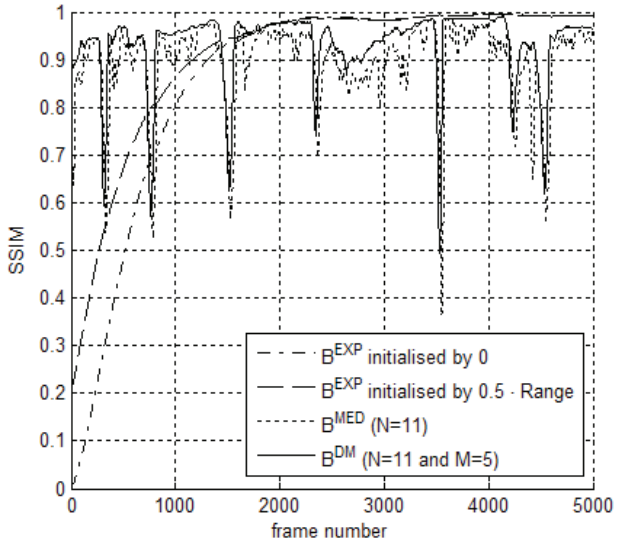


**Fig.5. Structural Similarity (SSIM) index values for consecutive frames of the background estimation using four various filters.**

and then the switch to the exponential filter should be done. The only problem in practical application is the appropriate choice of the switching moment without the knowledge of the reference background image.

The comparison of the results obtained for two different lengths of the sorted vector (11 and 31 frames) are illustrated in Figs. 6 and 7.

Since the median-based approach leads to faster convergence, it can be used for the initialisation of the exponential smoothing filter, which is more accurate due to using more frames and floating-point representation of data, similarly as in the High Dynamic Range (HDR) imaging. The idea of proposed hybrid background estimator is illustrated in Fig. 8. It can also be described as the following formula:

$$B_t^H(u,v) = \begin{cases} MEDIAN\left(\left\{I_t(u,v), I_{t-M}(u,v), \ldots \right.\right. \\ \left.\left. I_{t-i\cdot M}(u,v), \ldots, I_{t-(N-1)M}(u,v)\right\}\right) & : \quad t < T \\ \alpha \cdot B_{t-1}^H(u,v) + (1-\alpha)\cdot I_t(u,v) & : \quad t \geq T \end{cases} \tag{11}$$

The comparison of obtained results by means of the image quality assessment metrics over time is presented in Figs. 9 and 10.

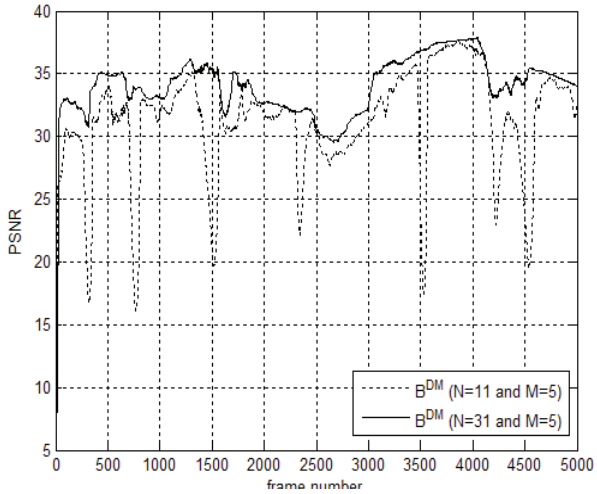


**Fig.6.Peak Signal-to-Noise Ratio (PSNR) values obtained for two median filters with temporal downsampling and different number of used frames (N).**

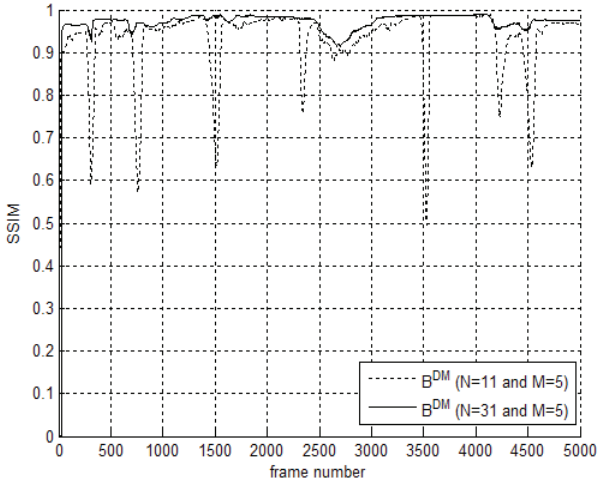


**Fig.7. Structural Similarity (SSIM) values obtained for two median filters with temporal downsampling and different number of used frames (N).**

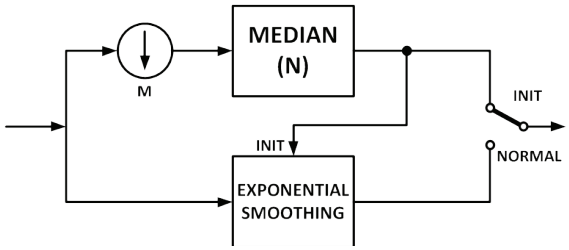


**Fig.8. The idea of the hybrid filter as the exponential smoothing initialised by median filter with temporal downsampling.**

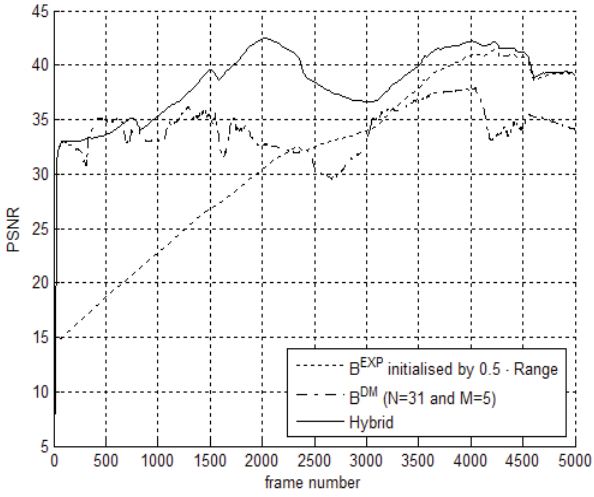


**Fig.9. Comparison of the Peak Signal-to-Noise Ratio (PSNR) values obtained for various approaches and the proposed hybrid filter.**

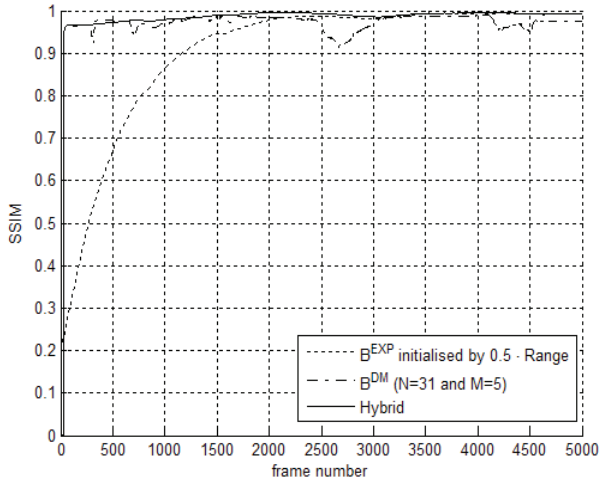


**Fig.10.Comparison of the Structural Similarity (SSIM) values obtained for various approaches and the proposed hybrid filter.**

# 4. Conclusion

Background estimation and subtraction algorithms are still an active research area [18,19] especially in the applications related to the video surveillance systems. The analysis presented in the paper illustrates the disadvantages of some typical methods, so one of the most interesting alternatives is their combination, allowing better initialisation using the nonlinear median-based filtering with temporal downsampling preventing from the influence of noise.

Omitting 5 frames using the temporal downsampling approach with the frame rate 25 frames per second, the time period corresponding to the boundary frames is 2.2 s and 6.2 s respectively, what is a reasonable choice for the city ITS solutions and has been used in this paper.

The best results can be obtained using the exponential smoothing filter initialised by the median filter with temporal downsampling.

**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").

# Bibliography

[1] BLACKMAN S., POPOLI R., Design and Analysis of Modern Tracking Systems, Artech House, 1999.

[2] KLEIN L.A., Sensor Technologies and Data Requirements for ITS. Artech House ITS library, Norwood, Massachusetts 2001.

[3] KLEIN L.A., MILLS M.K., GIBSON D.R.P., Traffic Detector Handbook: Third Edition - Volume I, FHWA-HRT-06-108, FHWA, 2006.

[4] OKARMA K., MAZUREK P., Background Estimation Algorithm for Optical Car Tracking Applications. Machinebuilding and Electrical Engineering no. 7–8, p. 7-10, 2006.

[5] LO B.P.L., VELASTIN S.A., Automatic Congestion Detection System For Underground Platforms. Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing, pp. 158–161, 2000

[6] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, 2003-2010.

[7] PICCARDI M., Background subtraction techniques: a review. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, The

Hague, Netherlands, pp. 3099–3104, October 2004.

[8] CUCCHIARA R., GRANA C., PICCARDI M., PRATI A., Detecting Moving Objects, Ghosts and Shadows in Video Streams. IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 25, no. 10, pp. 1337–1342, 2003.

[9] ESKICIOGLU A., Quality Measurement for Monochrome Compressed Images in the Past 25 Years. Proceedings of the International Conference on Acoustics Speech & Signal Processing, pp. 1907–1910, Istanbul, Turkey, 2000.

[10] LI X., Blind Image Quality Assessment. Proceedings of the IEEE International Conference on Image Processing, pp. 449–452, 2002.

[11] WANG Z., SHEIKH H., BOVIK A., No-reference Perceptual Quality Assessment of JPEG Compressed Images. Proceedings of the IEEE International Conference on Image Processing, pp. 477–480, 2002

[12] MARZILIANO P., DUFAUX F., WINKLER S., EBRAHIMI T., A No-Reference Perceptual Blur Metric. Proceedings of the IEEE International Conference on Image Processing, pp. 57–60, 2002.

[13] ONG E.-P., LIN LU W., YANG Z., YAO S., PAN F., JIANG L., MOSCHETTI F., A No-reference Quality Metric for Measuring Image Blur. Proceedings of the 7th International Symposium on Signal Processing and Its Applications, pp. 469–472, 2003.

[14] WANG Z., BOVIK A., A Universal Image Quality Index. IEEE Signal Processing Letters vol. 9 no. 3, pp. 81–84, 2002.

[15] WANG Z., BOVIK A., SHEIKH H., SIMONCELLI E., Image Quality Assessment: From Error Measurement to Structural Similarity. IEEE Trans. Image Processing vol. 13 no. 4, pp. 600–612, 2004.

[16] OKARMA K., Two-dimensional Windowing in the Structural Similarity Index for the Colour Image Quality Assessment. Lecture Notes in Computer Science vol. 5702, pp. 501–508, Springer-Verlag, 2009.

[17] OKARMA K., Influence of the 2-D Sliding Windows on the Correlation of the Digital Image Quality Assessment Results Using the Structural Similarity Approach with the Subjective Evaluation. Electrical Review (Przegląd Elektrotechniczny), vol. 86 no. 7, pp. 109–111, 2010.

[18] 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 pages, 2011.

[19] MADDALENA L., PETROSINO A., A Self-organizing Approach to Background Subtraction for Visual Surveillance Applications. IEEE Transactions on Image Processing, vol. 17, no. 7, pp. 1168–1177, 2008.