# Histogram of Oriented Gradients with Cell Average Brigthess for Human Detection

**Type:**

Research paper

**Abstract:**

The modification of the descriptor in human detector using Histogram of Oriented Gradients (HOG) and Support Vector Machine is presented. The proposed modification requires inserting the values of average cell brightness resulting in the increase of the length of the descriptor from 3780 to 3908 values, but it is easy to compute and instantly gives »25% of miss rate improvement at 10-4 False Positives Per Window (FPPW). The modification has been tested on two versions of HOG-based descriptors: the classic Dalal-Triggs and the modified one, where, instead of spatial Gaussian masks for blocks, an additional central cell has been used. The proposed modification is suitable for hardware implementations of HOG-based detectors, enabling increase of the detection accuracy or resignation from the use of some hardware-unfriendly operations, such as the spatial Gaussian mask. The results of the tests of the influence on the brightness changes of test images are also presented. The descriptor may be used in sensor networks equipped with hardware acceleration of image processing to detect humans in the images.

**Authors:**

Dr Marek Wojcikowski, Gdansk University of Wójcikowski — Gdańsk University of Technology, Department ETI

**Keywords:**

digital image processing, object detection, human detection

Editorial System

**Manuscript body**

Download source file (7.1 MB)

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

*www.metrology.pg.gda.pl*

# HISTOGRAM OF ORIENTED GRADIENTS WITH CELL AVERAGE BRIGHTNESS FOR HUMAN DETECTION

**Marek Wójcikowski**

*Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdansk, Poland*
*(✉ wujek@ue.eti.pg.gda.pl, +48 58 347 1974)*

**Abstract**

The modification of the descriptor in human detector using Histogram of Oriented Gradients (HOG) and Support Vector Machine is presented. The proposed modification requires inserting the values of average cell brightness resulting in the increase of the length of the descriptor from 3780 to 3908 values, but it is easy to compute and instantly gives ≈25% of miss rate improvement at $10^{-4}$ False Positives Per Window (FPPW). The modification has been tested on two versions of HOG-based descriptors: the classic Dalal-Triggs and the modified one, where, instead of spatial Gaussian masks for blocks, an additional central cell has been used. The proposed modification is suitable for hardware implementations of HOG-based detectors, enabling increase of the detection accuracy or resignation from the use of some hardware-unfriendly operations, such as the spatial Gaussian mask. The results of the tests of the influence on the brightness changes of test images are also presented. The descriptor may be used in sensor networks equipped with hardware acceleration of image processing to detect humans in the images.

Keywords: digital image processing, object detection, human detection.

*© 2015 Polish Academy of Sciences. All rights reserved*

## 1. Introduction

Detection of persons in images is an important and challenging task needed for applications such as driving assistance, autonomous driving or video surveillance, where the pedestrian detection must be both robust and in real-time. There are two main methods of person detection: single-scanning window and part-based detector. Scanning window methods are based on various feature descriptors, such as Histogram of Oriented Gradients (HOG) [1], Haar wavelet [2], Edge Orientation Histogram (EOH) [3][4] or Local Binary Pattern (LBP) [3][4]. The descriptors are classified using machine learning techniques, such as Support Vector Machine (SVM) [5][6] or boosting classifier. SVM is a well-known method of classification with a solid mathematical background, where the learning phase is reasonably quick, but the classification stage requires a significant number of multiplications and additions. The SVM has been successfully applied to many different problems [7][8]. The boosting classifier consists of a cascade of "weak" classifiers, where early stages of the cascade reject most negative data. Owing to this, only a limited number of samples traverse the full cascade, thus this method requires a very long time at the learning stage and it is quick at the classification stage. Part-based methods [9] mainly use a deformable model, which improves the detection performance. They generally work better at partial occlusions and for pedestrians in various poses. Feature descriptors are used for detection of the parts of the model. Part-based detectors require increased computational costs and it is more difficult to implement a real-time robust application, while many successful real-time applications using sliding window have been reported.



Editorial System

Among various feature descriptors, the HOG descriptor outperforms most other techniques and it is widely used for pedestrian detection. The HOG descriptor used together with the SVM classifier is one of the best-known human detection methods. The HOG has been introduced in [10] and [1], many modifications of the HOG descriptor may be found in the literature, intended for the improvement of the detection quality or the speed. Combining the HOG descriptor with boosting-based methods gives faster classification speeds than SVM-based methods [11][12]. In [13], the HOG descriptor combined with Haar-like features and boosted cascade classifier has been proposed for a better detection accuracy and efficiency. In [14], the HOG descriptor is combined with multi-scale curvelet features for full body detection. The simultaneous use of HOG and LBP gives long descriptors but it also gives very good detection results, as described in [15] and [16]. Zeng et al. [17] are using a two-stage cascade of rejectors: HOG classifiers and LBP classifiers, to reduce the problem of processing long descriptors.

To decrease the processing time and achieve real-time operation, many algorithm modifications and acceleration techniques have been invented. The widely-used technique is the integral image method [18], which provides the possibility of obtaining area-based descriptor values in constant time. Hardware implementations, such as feature extraction accelerator VLSI [19], can be used in portable, on-board vehicle systems or sensor networks; hardware accelerators can also increase the efficiency of computer-based approaches. However, not all algorithms can be efficiently implemented in hardware, therefore the researchers are searching for pipeline-friendly methods, which can be embedded into vision chips [20].

In this paper, a modification of classic HOG descriptor is proposed. The main contribution of this paper is the introduction of additional, easily calculated values to the descriptor, which instantly improves the miss rate parameter of window detector. To provide the clarity of the presented evaluations, all the results are compared with the well-known and well-described classic method presented in [1] and are tested using INRIA dataset [21].

## 2. The Descriptor

The calculation of classic HOG descriptor begins with dividing the image under the detection window into a dense grid of rectangular cells, for each cell a separate orientation of gradients is calculated. The histogram consists of evenly spaced orientation bins accumulating the weighted votes of gradient magnitude of each pixel belonging to the cell. In [1], 8 x 8 pixel cells and 9 bins for the orientation range 0-180 degrees have been used. Additionally, the cells are grouped into blocks and for each block the normalisation of all cell histograms is performed. The blocks are overlapping, so the same cell can be differently normalised in several blocks. The descriptor is calculated using all the overlapping blocks from the image detection window. From the detection window of size 64 x 128 pixels and for block of 2 x 2 cells, shifted by 8 pixels, ones obtains 3780 features per detection window.

The basic version of HOG descriptor would not give such good results, unless some additional techniques are used, as proposed by Dalal and Triggs [1]:

- For colour images, separate gradients for each colour are calculated, but only the gradient with the largest norm is used.
- A Gaussian mask is used on each pixel of the block to downweight the pixels near the edges.
- Each vote of gradient magnitude is bilinearly interpolated into neighbouring bins and in the same manner is also divided into neighbouring cells. This procedure is called a tri-linear interpolation.



Editorial System

In this paper, a modification of classic HOG descriptor is proposed, where an additional value $I_i$ is included to each cell $i$ in the descriptor, as shown in Fig. 1. $I_i$ is the average brightness of the cell $i$, calculated using an average of $R$, $G$ and $B$ pixel components. Using the infinity norm of $R$, $G$ and $B$, instead of an average for calculating a cell's average brightness, gives slightly worse results. Calculation of $I_i$ can be done easily during the calculation of the histograms of gradients; the main disadvantage is the increased length of the descriptor from 3780 to 3908 features (for the detection window of size 64 x 128 pixels and for block of 2 x 2 cells, shifted by 8 pixels), which can cause an increased processing time. It is also possible to use $I'_i = I_i - I_{avg}$ instead of $I_i$, where $I_{avg}$ is the average brightness of the window, but the test results are similar to using $I_i$.

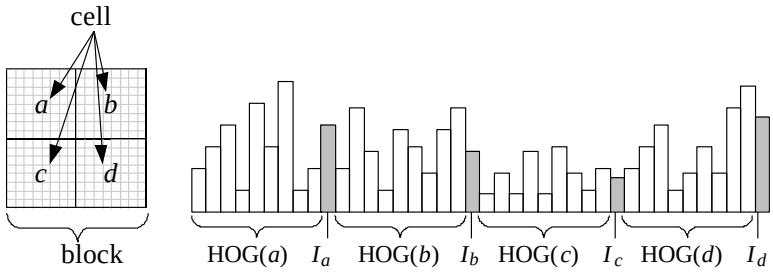



Fig. 1 The proposed structure of the descriptor with average brightness value of each cell. $I_i$ , represents an average intensity of the pixels belonging to the cell $i$={$a$,$b$,$c$,$d$}

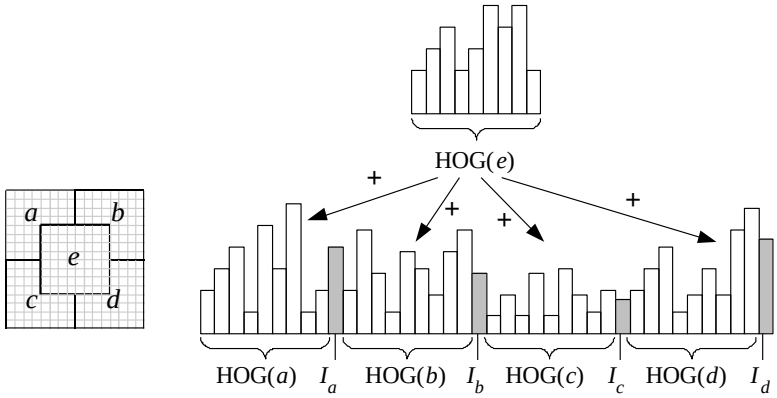



Fig. 2 The structure of the descriptor using central cell approach from [16], instead of Gaussian spatial mask used in [1], with cells' average brightness values included in the descriptor

In hardware implementation, the most challenging operations needed for calculating the HOG are: Gaussian mask and tri-linear interpolation, since they do not fit well in a pipelined style of hardware operation and integral image approach. In [16] an additional HOG of the cell centred in the original block has been used to replace the spatial pixel weighting, which in fact improved the overall detection quality. The proposed modification of adding brightness-based value has also been tested with the central cell approach from [16] instead of Gaussian spatial mask, which enables easier hardware implementation.

## 3. Learning and testing the classifiers

The proposed modification of the HOG descriptor has been used with linear SVM for the classification of the analysed images. Linear SVM is based on solving the optimisation problem [5][6]:



Editorial System

$$\min_{\mathbf{w},b,\xi} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{L}\xi_i \tag{1}$$

subject to:

$$y_i(\mathbf{x}_i \cdot \mathbf{w} + b) - 1 + \xi_i \geq 0 \qquad \xi_i \geq 0, \quad i = 1,..., L \tag{2}$$

where $\{\mathbf{x}_i, y_i\}$ is training data set of size $L$ and $\mathbf{x} \in \Re^D$ is an input point of $D$ attributes with the corresponding label $y_i = -1$ or $+1$. $\xi_i$ is a positive slack variable, which relaxes the constraints and allows some points to be misclassified, when the problem is not fully linearly separable. Variables $\mathbf{w}$ and $b$ define the optimal orientation of the hyperplane, separating the points belonging to two different classes with the soft error margin controlled by the parameter $C$. The minimisation problem (1)(2) is solved using iterative quadratic problem solver, where the termination criteria for the algorithm may be: a maximum number of iterations and/or a tolerance error $\varepsilon$.

The HOG descriptor has been combined with linear SVM to obtain a classifier. In this paper the value of the parameter controlling the error margin in SVM has been set to $C = 0.01$, which enables an easy comparison with the results presented in [1], where the same value for $C$ has been applied. For the experiments, the SVM implementation from *OpenCV* library version 2.4.11.0 has been used, with the termination criterion $\varepsilon = 0.001$.

The SVM has been trained using INRIA data set (2416 positive examples and 24360 negative examples, including their mirrored versions) in the same way as in [1], i.e. the re-training phase has been complete using hard training examples detected after the first training. The examples of the training images are shown in Fig. 3.

Fig 3 Few examples of positive images containing pedestrians (top row) and negative, nonpedestrian images (bottom row) from INRIA database [21].

For the testing, INRIA test images have been used with 1132 positive examples and 453 nonpedestrian negative images, where each negative image has been extensively searched with 8 pixel shift of the test window and 1.2x scale down factor of the image. The testing procedure was the same as in [1]. Usually the Receiver Operating Characteristics (ROC) curves are used to quantify the performance of the detectors, based on the classification return values, which are the signed distances to the margin in 2-class SVM classifier. The shape of ROC curves does not allow easy distinguishing of small probabilities, so in further considerations Recall-Precision (RP) and Detection Error Tradeoff (DET) curves will be used,



Editorial System

which contain the same information as ROC curves. The RP curve plots precision versus recall on a log-log scale. Precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

where

*TP* – the number of windows where there was a person and a person has been detected;

*FP* – the number of windows where there was no person, but the detector indicated the presence of a person;

*TN* – the number of windows where there was no person and the detector did not detect any person;

*FN* – the number of windows where there was a person but the detector did not detect any person.

The RP curves have been presented in Fig. 4 for three classifiers: Dalal-Triggs classifier from [1] (HOG) and the classifiers from Fig. 1 and Fig. 2.

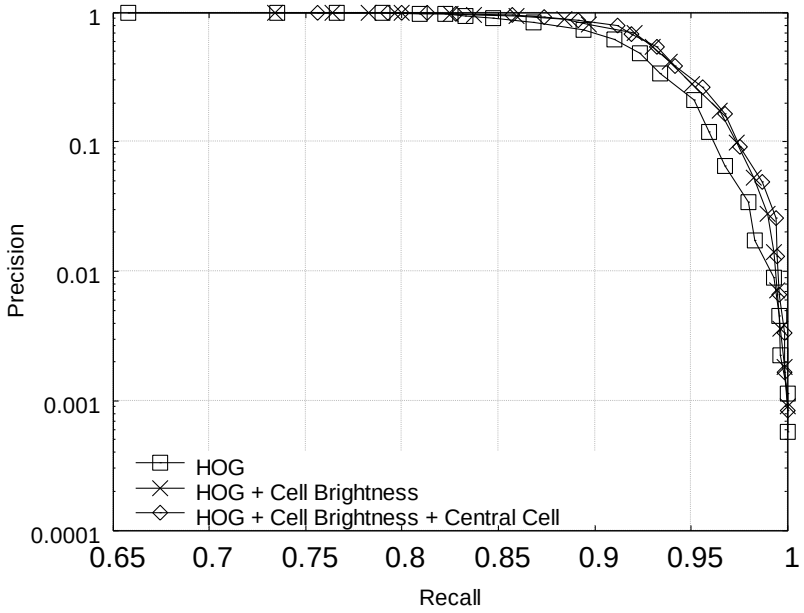



Fig. 4 Recall-Precision curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

In Fig. 5 the DET curves have been shown, which use the measures: miss rate and False Positives Per Window (FPPW), defined as follows:

$$MissRate = 1 - Recall = \frac{FN}{TP + FN} \tag{5}$$

$$FPPW = \frac{FP}{FP + TN} \tag{6}$$



Editorial System

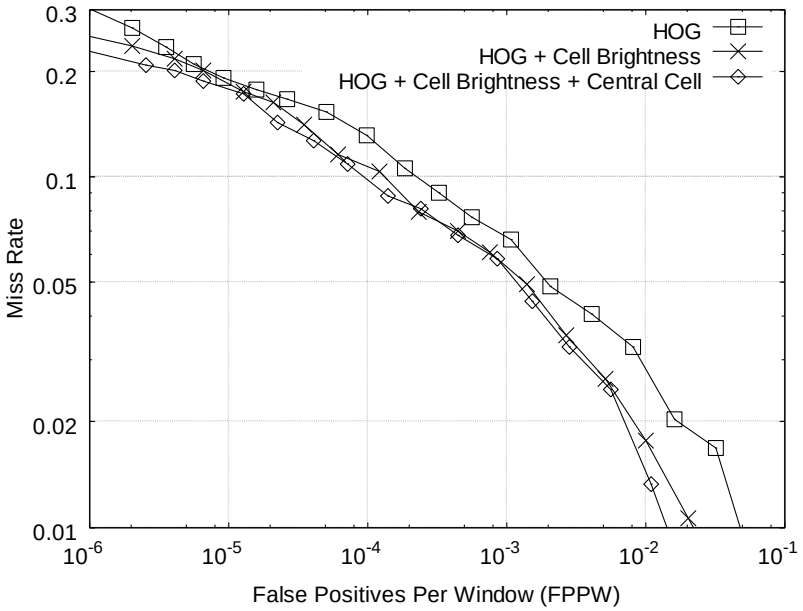



Fig. 5 Detection Error Tradeoff curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

The test results presented in Fig. 5 show ≈25% improvement of the miss rate at $10^{-4}$ FPPW for the proposed descriptor, which is equivalent to three times better FPPW at the same miss rate. It must be noted that part of the miss rate improvement has been achieved by the use of central cell approach.

To seek an optimal value of the parameter $C$, the $k$-fold cross-validation with INRIA train dataset and $k = 9$ has been applied, where $C$ has been raised from the value $C_0 = 2 \cdot 10^{-4}$ up to $C_j = 20$ with step $S = 1.2$ in iterations $j = 0, 1, ..., n$, where each next value of $C_j$ has been calculated as:

$$C_{j+1} = S \cdot C_j \tag{7}$$

As the result of this search, the new value of $C = 0.03125$ has been obtained, for which the test set error is minimal. However, the performance of the classifiers with the new value of $C$ was similar to the previous results (see Fig. 6), therefore the previous value $C = 0.01$ has been used in all further calculations presented this paper.

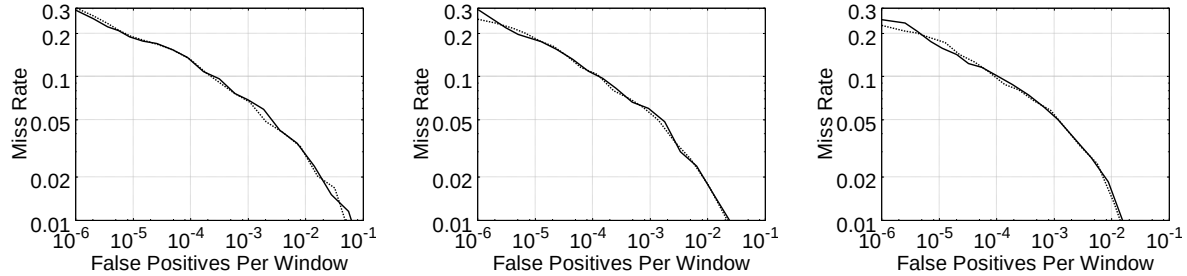



Fig. 6 Detection Error Tradeoff curves for the classifiers learned with the SVM's parameter value $C = 0.01$ (dashed line) and the value $C = 0.03125$ obtained from $k$-fold cross-validation and sweeping (solid line) using: (a) - the descriptor from Dalal-Triggs method [1]; (b) - the descriptor from Fig. 1, (c) - the descriptor from Fig. 2.



Editorial System

Using brightness values directly in the descriptor might suggest that the descriptor has lost its brightness-invariance. To test the behaviour of the descriptor, a test set has been prepared, containing the transformed INRIA test images, where the intensities of the positive test images have been randomly changed according to the equation:

$$\mathbf{I}_{new} = \alpha\mathbf{I} + \beta \tag{8}$$

where:

$\mathbf{I}$ and $\mathbf{I}_{new}$ – the intensities of the image before and after transformation, respectively;

$\alpha$, $\beta$ – the coefficients, randomly changed for each image, $\alpha$ has been changed in the range from 0.5 to 3 and $\beta$ from –50 to 100.

The results of the test using the brightness-transformed positive test images are presented in Fig. 7.

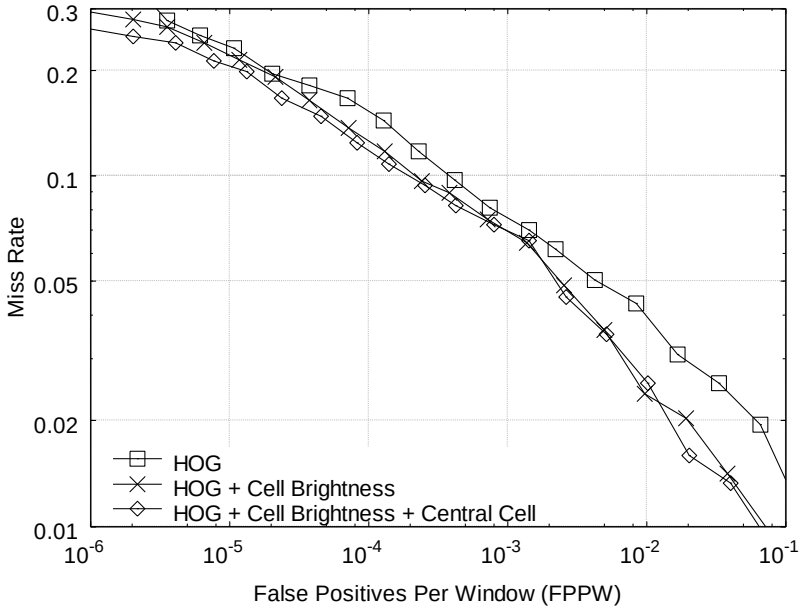



Fig. 7 Test results showing the influence of brightness changes according to Eqn. 8 for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1] with randomly changed brightness of the positive test images

In another test, instead of uniformly changing the brightness of the whole window, the brightness has been randomly changed at random regions of the image. First, each tested window has been divided into two parts by a diagonal line at variable position and angle, then the brightness of a randomly chosen part has been changed according to Eqn. (8) with the coefficients changed randomly for each window. The random brightness change has been also applied to the randomly selected rectangular areas of the window (with the random number of rectangles from 1 to 4). The examples of images after those transformations have been shown in Fig. 8. The DET curves presenting the performance of the detectors for the manipulated images have been shown in Fig. 9.

The results from both tests with the use of images with randomly changed brightness show that the DET curves in Fig. 7 and Fig. 9 are always above the DET curve for classic HOG, proving that the proposed modification gives better results than the classic HOG descriptor.



Editorial System

Fig. 8 Examples of the positive test images from INRIA database with randomly changed brightness at random areas.

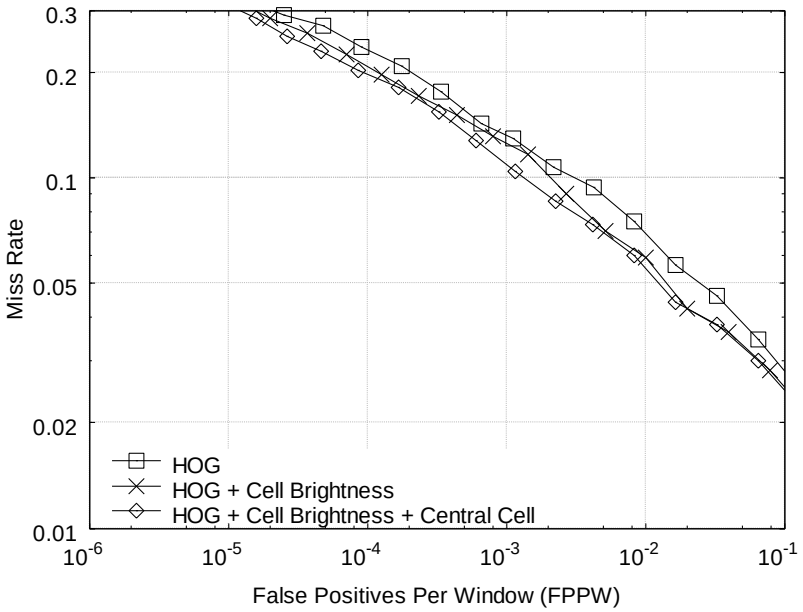



Fig. 9 Test results showing the influence of random brightness change at random rectangular regions for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1], where the positive test images have randomly changed brightness of randomly selected regions of the image.

The calculation times of the proposed solution are presented in Table 1. Due to the inclusion of the brightness data into the descriptor from Fig. 1, the classification time of a single detection window (measured as pure classification time, not including the reading of the image from disk and writing the results) has increased by approx. 1% comparing to the classic HOG descriptor. The descriptor from Fig. 2 is ≈2% slower, but it has the potential to be more effective in hardware, pipelined implementation.

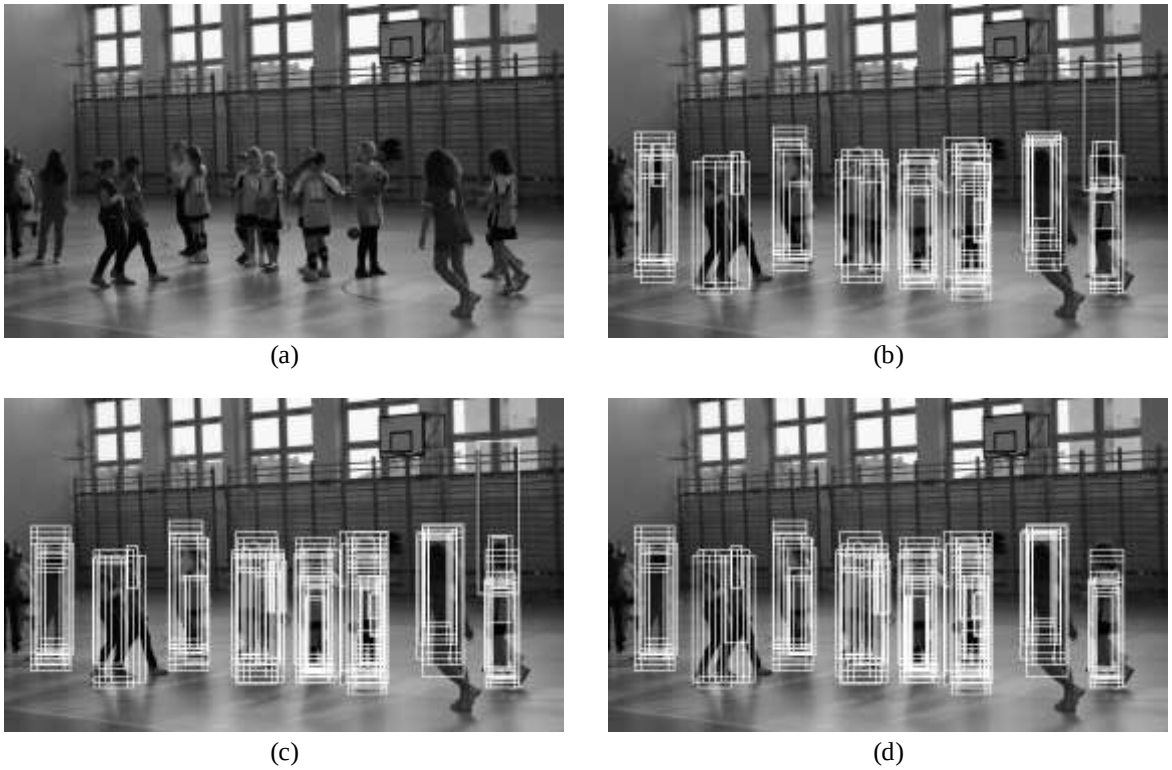


Editorial System

Table 1 The comparison of calculation times of SVM implementations. The descriptors have been written in C++ using the SVM libraries *OpenCV* version 2.4.11.0. The evaluations have been made using PC computer with Intel i7 3.2 GHz and 64-bit Windows 7 operating system. Pixel intensity values have been represented as 8-bit unsigned numbers. For the magnitude of the gradient, 64-bit floats have been used; the angle has been calculated in degrees and saved as 32-bit integer. The histograms have been represented as STL vectors of 64-bit floating-point numbers.

| | Units | Dalal-Triggs HOG descriptor [1] | HOG + Cell Brightness descriptor from Fig. 1 | HOG + Cell Brightness + Central Cell descriptor from Fig. 2 |
|---|---|---|---|---|
| Descriptor length (the number of values) | - | 3780 | 3908 | 3908 |
| Average classification time of single test window | [ms] | 4.562 | 4.605 | 4.650 |
| Total learning time (26776 images), including reading the image files from disk, calculation of the descriptor and saving the results to disk | [s] | 820 | 778 | 708 |
| Total classification time (10192 images), including reading the image files from disk, calculating the descriptor and saving the results to disk | [s] | 95.7 | 96.0 | 96.3 |

The operation of the descriptor has been shown in Fig. 10, where a dense scan of an image containing standing persons has been made. As can be seen, the detector using the proposed descriptor gives more *TP* detections for almost each person on the image. At the same time, a few *FP* detections are skipped – in Fig. 10b the first person to the left and the first persons to the right have some false detections, which are correctly not detected in Fig. 10d.



(a) (b) (c) (d)

Fig. 10 An example of dense scan detection results: (a) the input image, (b) - the detection results from Dalal-Triggs's descriptor; (c) – the detection results from the descriptor from Fig. 1, (d) - the detection results from the descriptor from Fig. 2. The overlapping detections have not been merged (i.e. using non-maximum suppression) to show all the detections. The image of size 1474 x 828 pixels has been scanned with the moving window with the step of 3 pixels and rescaled down with the factor 1.1.



Editorial System

## 4. Conclusions

The main contribution of the author is the idea of using additional information in the HOG descriptor based on average pixel intensity. This simple modification slightly increases the length of the descriptor but it results in a significant improvement of the miss rate of the window detector. The proposed idea has been compared with the well-known HOG descriptor described in [1], also the modification based on the idea of using central cell instead of spatial Gauss mask [16] has been used. The test results show that this modification gives valuable hints to SVM classifier, resulting in the miss rate improvement by ≈25% at $10^{-4}$ FPPW over the original version of the HOG method, at the expense of up to ≈2% increase of calculation time. Due to the shallow shape of DET curves, such an improvement in the miss rate is equivalent to 3 times better FPPW at the same miss rate. It has also been shown that the proposed modification, despite the fact that it uses the pixel intensities in the descriptor in addition to gradients, it still provides the improvement for the images of randomly changed brightness. This shows that the brightness values present in the proposed descriptor contain additional information that help SVM to discriminate between positive and negative samples and may be considered as another procedure to improve the miss rate of the detectors. It is highly probable that adding brightness-based values can improve many other descriptors based on the HOG method.

## References

[1] Dalal, N., Triggs, B. (2005). Histograms of oriented gradients for human detection. *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, 886–893.

[2] Viola, P., Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, 511–518.

[3] Ma, Y., Chen, X., Jin, L., Chen, G. (2011). *A monocular human detection system based on EOH and oriented LBP features. Proc. 7th Int. Conf. Adv. Visual Comput.*, I, 551–562.

[4] Ma, Y., Deng, L., Chen, X., Guo, N. (2013). Integrating Orientation Cue With EOH-OLBP-Based Multilevel Features for Human Detection. *IEEE Trans. Circuits Syst. Video Technol.*, 23(10), 1755-1766.

[5] Boser, B. E., Guyon, I., Vapnik V. (1992). A training algorithm for optimal margin classifiers. *Proc. Fifth Annual Workshop on Computational Learning Theory*, *ACM Press*, 144-152.

[6] Cortes, C., Vapnik, V. (1995). Support-vector network. *Machine Learning*, 20:273-297.

[7] Cui, J., Wang, Y., (2010). A Novel Approach of Analog Fault Classification Using a Support Vector Machines Classifier. *Metrology and Measurement Systems*, 17(4), 561-581.

[8] Wójtowicz, B., Dobrowolski, A., Tomczykiewicz, K., (2015). Fall Detector Using Discrete Wavelet Decomposition And SVM Classifier. *Metrology and Measurement Systems*, 22(2), 303-314.

[9] Zhang, H., Bai, X., Zhou, J., Cheng, J., Zhao H. (2013). Object Detection via Structural Feature Selection and Shape Model. *IEEE Trans. Image Process.*, 22(12), 4984-4995.

[10] Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *Int. Journal of Comput. Vision*, 60(2), 91-110.

[11] Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S. (2006). Fast human detection using a cascade of histograms of oriented gradients. *Proc. IEEE Int. Conf. Comput. Vision Pattern Recognit.*, 2, 491-1498.

[12] Zeng, H.-C., Huang, S.-H., Lai, S.-H. (2008). Real-time video surveillance based on combining foreground extraction and human detection. *Proc. 14th Int. Multimedia Modeling Conf., MMM 2008, Kyoto, Japan*, 70-79.

[13] Chen, Y.-T., Chen, C.-S. (2008). Fast human detection using a novel boosted cascading structure with meta stages. *IEEE Trans. Image Process.*, 17(8), 1452-1464.



Editorial System

[14] Cheng, H.-Y., Zeng, Y.-J., Lee C.-C., Hsu S.-H. (2013). Segmentation of Pedestrians with Confidence Level Computation, *Journal of Signal Processing Systems*, 72(2), 87-97.

[15] Wang, X., Han, T. X., Yan, S. (2009). An HOG-LBP human detector with partial occlusion handling. *Proc. IEEE Int. Conf. on Comput. Vision, ICCV 2009*, Kyoto, 32-39.

[16] Geismann, P., Knoll, A. (2010). Speeding Up HOG and LBP Features for Pedestrian Detection by Multiresolution Techniques. *Proc. 6th Int. Symposium Advances in Visual Computing, ISVC 2010, Las Vegas, NV, USA*, 243-252.

[17] Zeng, C., Ma, H, Ming, A. (2010). Fast human detection using mi-sVM and a cascade of HOG-LBP features. *Proc. 17th IEEE Int. Conf. on Image Processing (ICIP)*, 3845-3848,

[18] Crow, F. (1984). Summed-area tables for texture mapping. *Proc. of SIGGRAPH*, 18(3), 207-212.

[19] Takagi, K., Tanaka, K., Izumi, S., Kawaguchi H., Yoshimoto M. (2014). A Real-time Scalable Object Detection System using Low-power HOG Accelerator VLSI. *Journal of Signal Processing Systems*, doi 10.1007/s11265-014-0870-7.

[20] Jendernalik, W., Blakiewicz, G., Handkiewicz, A., Melosik, M. (2013). Analogue CMOS ASICs in Image Processing Systems. *Metrology and Measurement Systems*, 20(4), 613-622.

[21] Everingham, M., Zisserman, A., Williams, C. K .I., Van Gool, L. (2006). The PASCAL Visual Object Classes Challenge 2006 (VOC 2006) Results. *Technical Report, Univ. of Oxford*.

[22] Chang, C.-C., Lin, C.-J. (2011). LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3) 27:1-27:27.



Editorial System

**Figure 1**

Download source file (40.02 kB)

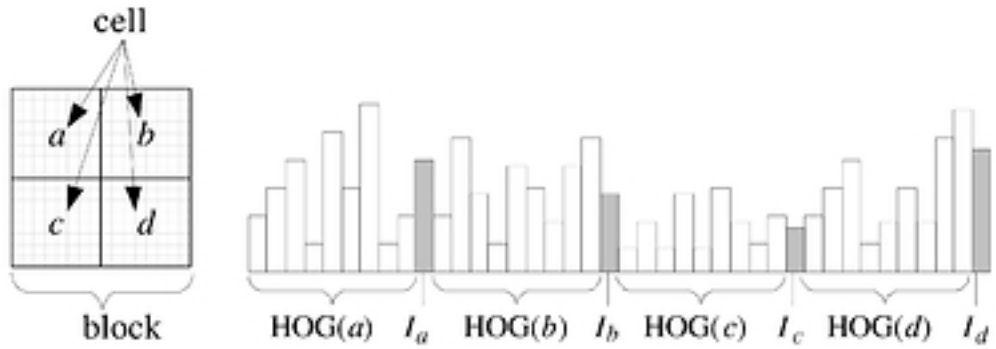



Fig. 1 The proposed structure of the descriptor with average brightness value of each cell. Ii , represents an average intensity of the pixels belonging to the cell i={a,b,c,d}

Editorial System

**Figure 2**

Download source file (47.78 kB)

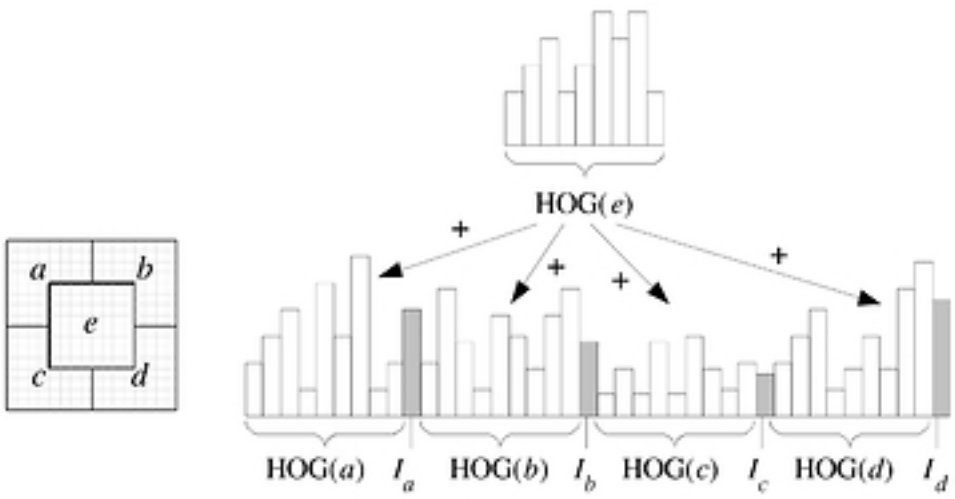



Fig. 2 The structure of the descriptor using central cell approach from [16], instead of Gaussian spatial mask used in [1], with cells' average brightness values included in the descriptor

Editorial System

**Figure 3**

Download source file (328.57 kB)

Fig 3 Few examples of positive images containing pedestrians (top row) and negative, nonpedestrian images (bottom row) from INRIA database [21].

Editorial System

**Figure 4**

Download source file (16.81 kB)

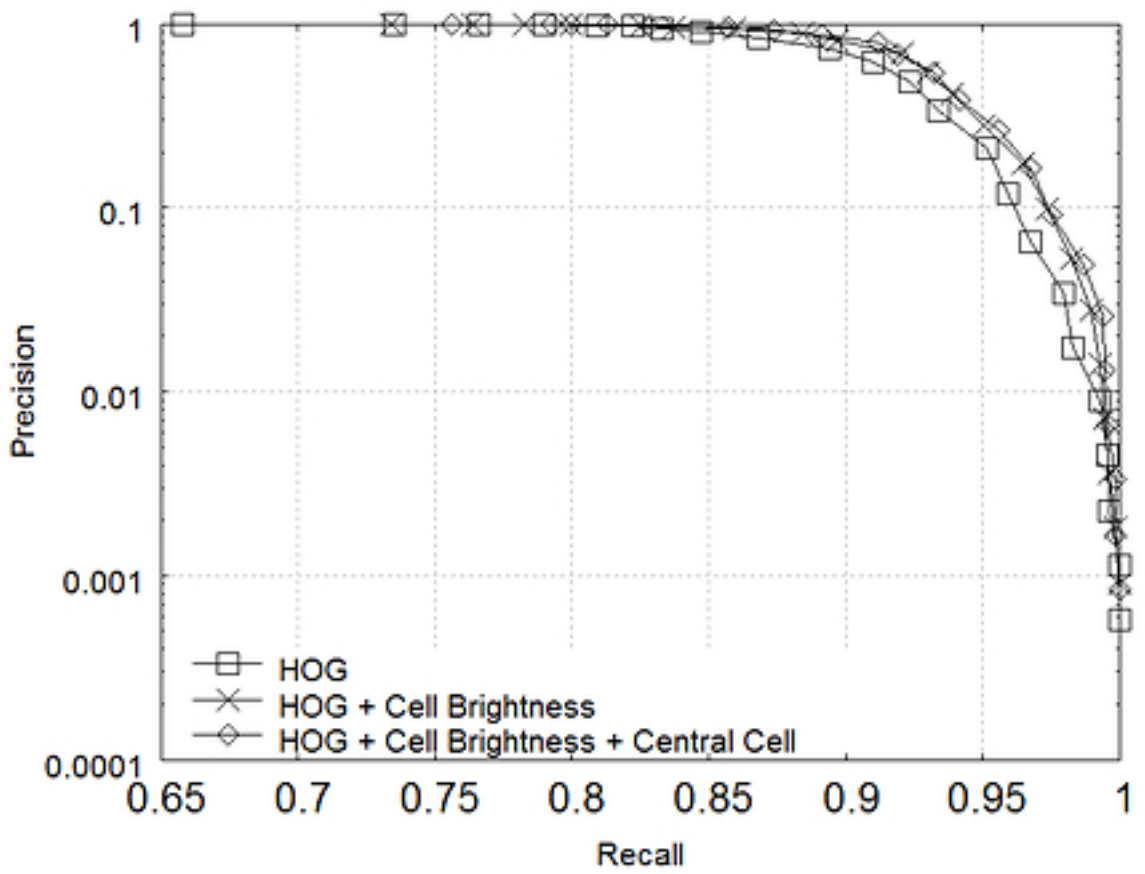



Fig. 4 Recall-Precision curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

Editorial System

**Figure 5**

Download source file (169.06 kB)

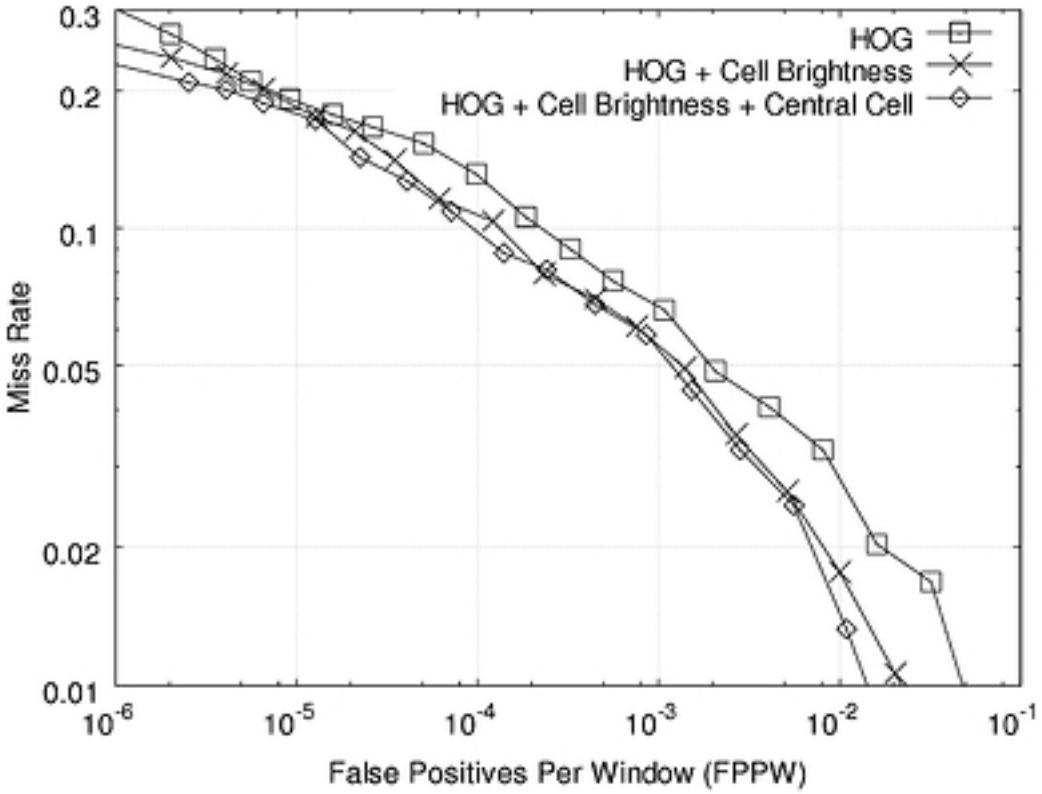



Fig. 5 Detection Error Tradeoff curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

Editorial System

**Figure 6**

Download source file (91.46 kB)

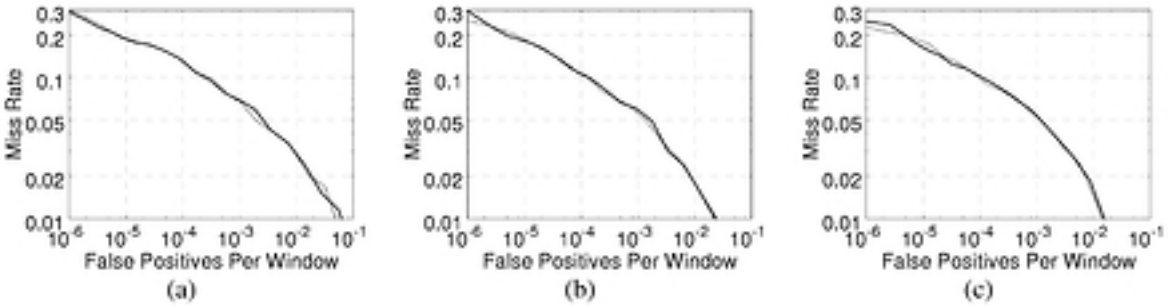



Fig. 6 Detection Error Tradeoff curves for the classifiers learned with the SVM's parameter value C = 0.01 (dashed line) and the value C = 0.03125 obtained from k-fold cross-validation and sweeping (solid line) using: (a) - the descriptor from Dalal-Triggs method [1]; (b) - the descriptor from Fig. 1, (c) - the descriptor from Fig. 2.

Editorial System

**Figure 7**

Download source file (173.94 kB)

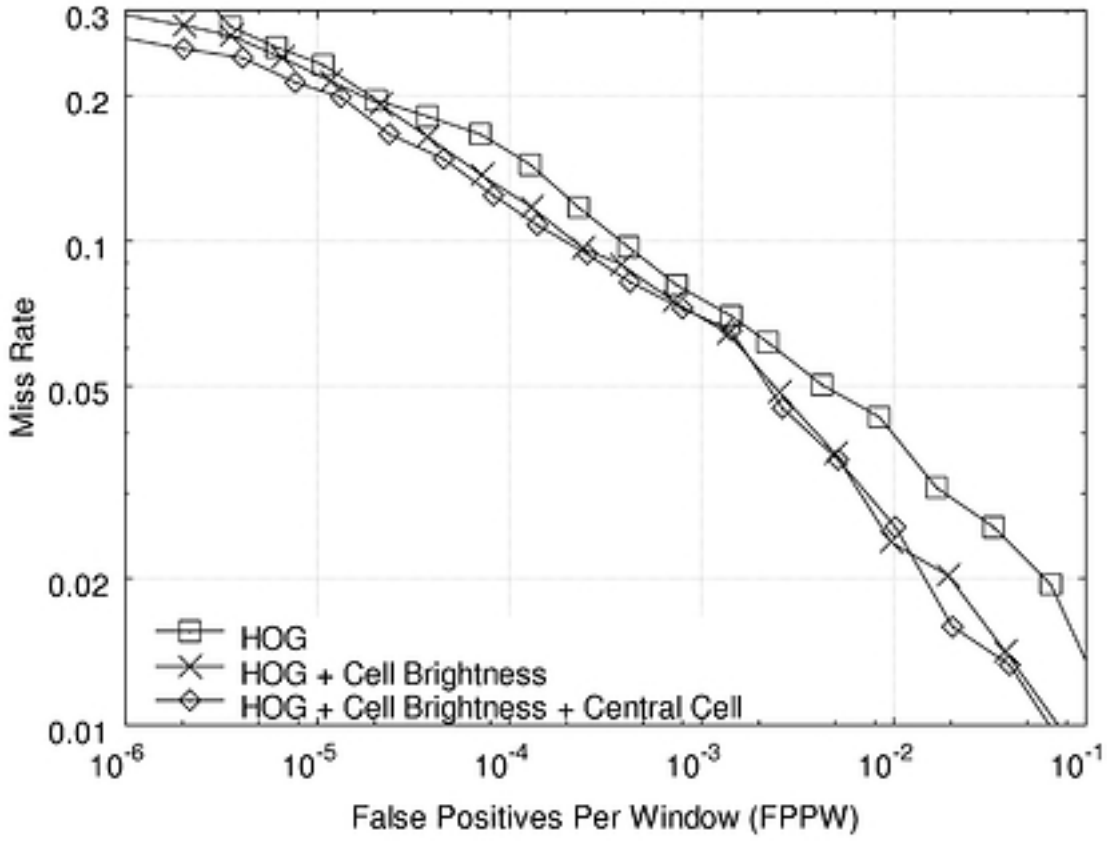



Fig. 7 Test results showing the influence of brightness changes according to Eqn. 8 for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1] with randomly changed brightness of the positive test images

Editorial System

**Figure 8**

Download source file (110.36 kB)

Fig. 8 Examples of the positive test images from INRIA database with randomly changed brightness at random areas.

Editorial System

**Figure 9**

Download source file (154.85 kB)

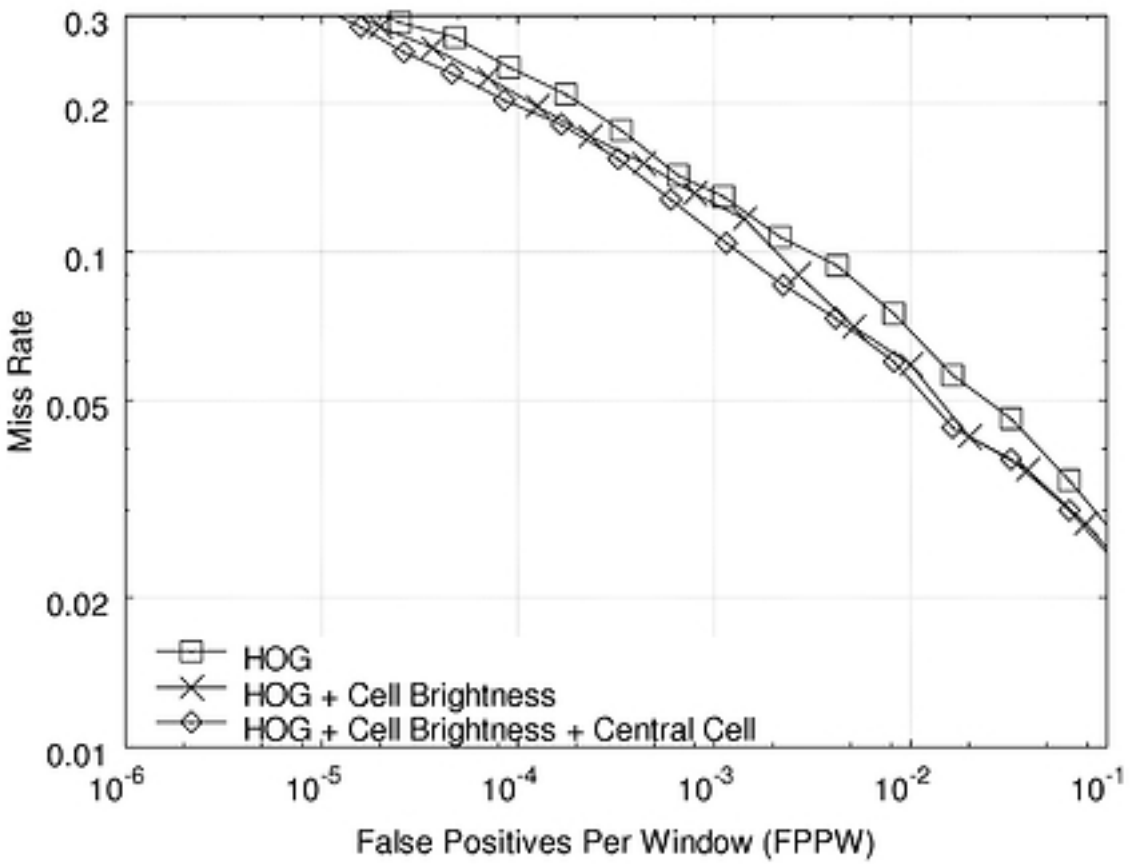



Fig. 9 Test results showing the influence of random brightness change at random rectangular regions for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1], where the positive test images have randomly changed brightness of randomly selected regions of the image.

Editorial System

**Figure 10**

Download source file (2.65 MB)

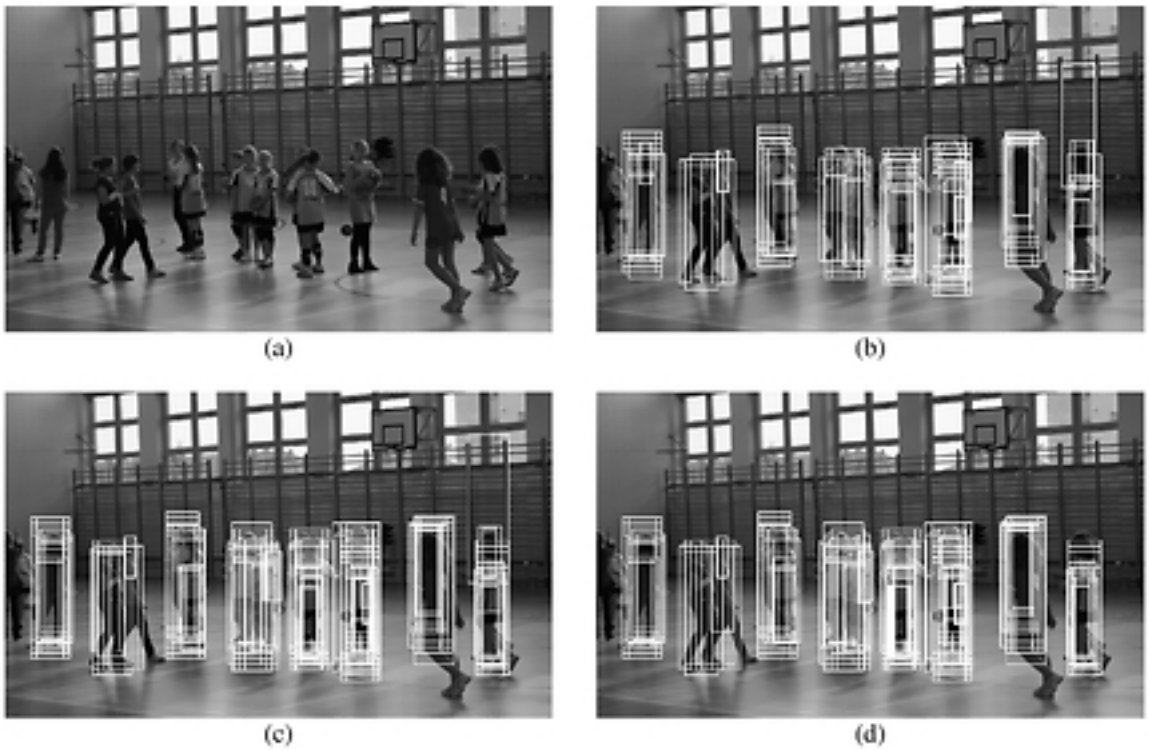



Fig. 10 An example of dense scan detection results: (a) the input image, (b) - the detection results from Dalal-Triggs's descriptor; (c) – the detection results from the descriptor from Fig. 1, (d) - the detection results from the descriptor from Fig. 2. The overlapping detections have not been merged (i.e. using non-maximum suppression) to show all the detections. The image of size 1474 x 828 pixels has been scanned with the moving window with the step of 3 pixels and rescaled down with the factor 1.1.

Editorial System

**Figure 11**

Download source file (110.54 kB)

| | Units | Dalal-Triggs HOG descriptor [1] | HOG + Cell Brightness descriptor from Fig. 1 | HOG + Cell Brightness + Central Cell descriptor from Fig. 2 |
|---|---|---|---|---|
| Descriptor length (the number of values) | - | 3780 | 3908 | 3908 |
| Average classification time of single test window | [ms] | 4.562 | 4.605 | 4.650 |
| Total learning time (26776 images), including reading the image files from disk, calculation of the descriptor and saving the results to disk | [s] | 820 | 778 | 708 |
| Total classification time (10192 images), including reading the image files from disk, calculating the descriptor and saving the results to disk | [s] | 95.7 | 96.0 | 96.3 |

Table 1 The comparison of calculation times of SVM implementations. The descriptors have been written in C++ using the SVM libraries OpenCV version 2.4.11.0. The evaluations have been made using PC computer with Intel i7 3.2 GHz and 64-bit Windows 7 operating system. Pixel intensity values have been represented as 8-bit unsigned numbers. For the magnitude of the gradient, 64-bit floats have been used; the angle has been calculated in degrees and saved as 32-bit integer. The histograms have been represented as STL vectors of 64-bit floating-point numbers.

Editorial System

## Index

### Manuscript body

Manuscript body 1 - Download source file (7.1 MB)

### Figures

Figure 1 - Download source file (40.02 kB)
Fig. 1 The proposed structure of the descriptor with average brightness value of each cell. Ii , represents an average intensity of the pixels belonging to the cell i={a,b,c,d}

Figure 2 - Download source file (47.78 kB)
Fig. 2 The structure of the descriptor using central cell approach from [16], instead of Gaussian spatial mask used in [1], with cells' average brightness values included in the descriptor

Figure 3 - Download source file (328.57 kB)
Fig 3 Few examples of positive images containing pedestrians (top row) and negative, nonpedestrian images (bottom row) from INRIA database [21].

Figure 4 - Download source file (16.81 kB)
Fig. 4 Recall-Precision curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

Figure 5 - Download source file (169.06 kB)
Fig. 5 Detection Error Tradeoff curves for the classifier using the descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs method [1] are also given (HOG). All tests have been performed using INRIA Person dataset with re-training on hard examples, negative test images have been scanned using 1.2x rescale factor and 8-pixel window shift (the same procedure as described in [1])

Figure 6 - Download source file (91.46 kB)
Fig. 6 Detection Error Tradeoff curves for the classifiers learned with the SVM's parameter value C = 0.01 (dashed line) and the value C = 0.03125 obtained from k-fold cross-validation and sweeping (solid line) using: (a) - the descriptor from Dalal-Triggs method [1]; (b) - the descriptor from Fig. 1, (c) - the descriptor from Fig. 2.

Figure 7 - Download source file (173.94 kB)
Fig. 7 Test results showing the influence of brightness changes according to Eqn. 8 for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1] with randomly changed brightness of the positive test images

Figure 8 - Download source file (110.36 kB)
Fig. 8 Examples of the positive test images from INRIA database with randomly changed brightness at random areas.

Figure 9 - Download source file (154.85 kB)
Fig. 9 Test results showing the influence of random brightness change at random rectangular regions for the proposed descriptor from Fig. 1 (HOG + Cell Brightness) and from Fig. 2 (HOG + Cell Brightness + Central Cell). For comparison, the results from Dalal-Triggs descriptor [1] are also given (HOG). Trained and tested on INRIA Person dataset using the same procedure as described in [1], where the positive test images have randomly changed brightness of randomly selected regions of the image.

Figure 10 - Download source file (2.65 MB)
Fig. 10 An example of dense scan detection results: (a) the input image, (b) - the detection results from Dalal-Triggs's descriptor; (c) – the detection results from the descriptor from Fig. 1, (d) - the

MOST WIEDZY Downloaded from mostwiedzy.pl

detection results from the descriptor from Fig. 2. The overlapping detections have not been merged (i.e. using non-maximum suppression) to show all the detections. The image of size 1474 x 828 pixels has been scanned with the moving window with the step of 3 pixels and rescaled down with the factor 1.1.

Figure 11 - Download source file (110.54 kB)

Table 1 The comparison of calculation times of SVM implementations. The descriptors have been written in C++ using the SVM libraries OpenCV version 2.4.11.0. The evaluations have been made using PC computer with Intel i7 3.2 GHz and 64-bit Windows 7 operating system. Pixel intensity values have been represented as 8-bit unsigned numbers. For the magnitude of the gradient, 64-bit floats have been used; the angle has been calculated in degrees and saved as 32-bit integer. The histograms have been represented as STL vectors of 64-bit floating-point numbers.