# Pedestrian Detection in Low Resolution Night Vision Images

Paweł Pawłowski, Karol Piniarski, Adam Dąbrowski
Poznan University of Technology
Department of Computing, Division of Signal Processing and Electronic Systems
Poznań, Poland
{pawel.pawlowski, adam.dabrowski}@put.poznan.pl, karol.piniarski@doctorate.put.poznan.pl,

*Abstract*— **This paper presents a test of pedestrian detection in low resolution night vision infrared images. An image feature extractor based on histograms of oriented gradients followed by a Support Vector Machine (SVM) classifier are evaluated, optimized and used. Tests performed on three different night vision infrared datasets show that the classification quality of the proposed method is very high even in very low resolutions of images. In practice, large frame size for analysis not always improves the classification effectiveness, but always requires more time for processing.**

***Keywords: night vision, video processing, object detection, classifier, pedestrian, low resolution, support vector machine, histogram of oriented gradients***

## I. INTRODUCTION

Digital vision system are very popular and commonly used in today's life. They support medical imaging, production processes, safety in urban areas and automotive applications. Some of them, like closed-circuit television (CCTV) support human work, some, like machine vision in robotics and production work autonomously. Many of applications, especially those connected with a security, are related to a human detection [1, 4, 9, 10, 11, 14]. Among others we can mention CCTV, search and rescue services and driver assistance [13, 14].

Due to extremely high variations in illumination of visible light in different weather conditions and too low lighting at night some systems make use of infrared (IR) light and hence are often called night-vision or thermo-vision systems. They can be classified twofold: as active or passive systems. The active systems are equipped with infrared illuminators and capture the light reflected from the objects. They are typically used in CCTV with the same image sensor for visible and infrared light. Therefore they work with infrared light close to the visible range (near infrared or NIR for short). The main advantages of the NIR systems are relatively low cost and natural image similar to black & white pictures, among disadvantages are limited working distance and sensitivity. On the other hand, the passive systems typically capture a long-wavelength infrared (LWIR) radiation (some sources classify it as FIR – far-infrared), which is naturally emitted by objects and enough far from the visible light to avoid interferences. The passive image sensors are much more sensitive, so offer a long distance detection, but of a rather high cost and lower resolution. Because of the specific physical characteristics, they weakly represent the textures and produce images with low contrast [14]. Despite the disadvantages, the passive systems, especially due to the high imaging range, are in the area of research related to people detection.

Authors of [5] proposed using histograms of oriented gradients (HOG) for human detection. They studied the HOG for changes in window size, window overlap, normalization, number of orientation bins and more and presented detection error tradeoff (DET) of them [5]. Similar methods for pedestrian detection, i.e. HOG and a support vector machine classifier are presented in [1, 14, 15]. In the papers [1, 5, 15] authors propose similar setup for the HOG method. Gradient-based methods such as HOG, normalize visual signatures but in the case of the human detection, it may lead to complex solutions when aiming to work in low resolution [19]. A solution of the problem with the low resolution images is given in [16], where additionally a method for people detection from a moving camera is presented. In the thesis [11] many feature extractors and classifiers were tested with a plenty of methods and comparing figures. The authors of [1, 6, 10] built their own infrared video datasets, so many of the other papers used them as the bench tests [3, 4, 15, 16]. Also in this paper we used these three video datasets – we describe them in details in Sect. III.

In this paper we test the object classification procedures and optimize them to detect pedestrians in both NIR and LWIR images with low and very low resolutions. The NIR and LWIR images differ a lot and thus the video processing algorithms should be optimized separately for each of the type.

The tested object classifier is a very important part of the night vision video processing system and is used for detection of pedestrians e.g. in automotive applications [14]. The whole processing scheme is presented in Fig. 1. For image segmentation – a first part of ROI (region of interest) generation – a modified adaptive dual-threshold was proposed [14]. For selection of candidates we applied connected component labeling (CCL) [18], for feature extraction the histogram of oriented gradients (HOG) [5], and finally the support vector machine (SVM) for training of the classifier [2, 17].

Beside of the improvement of the classifier for high detection quality, we try to improve a speed of the data processing.

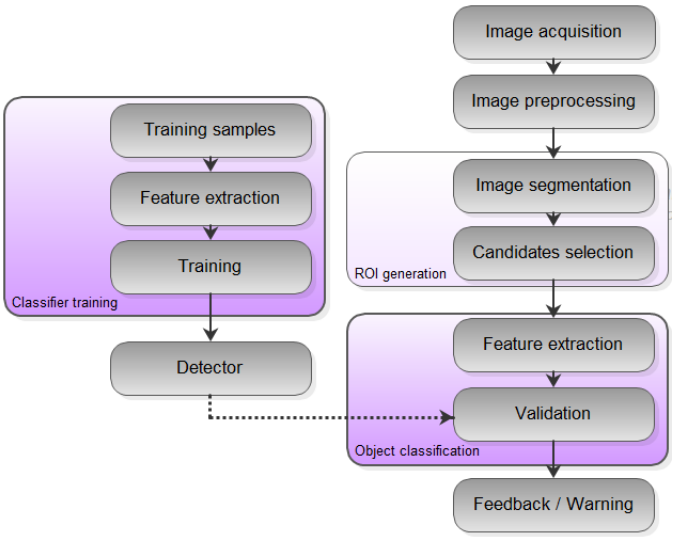


Figure 1. Video processing for detection of pedestrians

The paper is organized as follows: after an introduction we present the classifier used for the pedestrian detection. Then we describe the video datasets and present results of experiments with detailed comments. Finally we formulate conclusions.

## II. OBJECT CLASSIFICATION

The object classification is the next step after the ROI generation (cf. Fig. 1). It consists of two stages: feature extraction and validation. These stages are crucial and strongly affects quality of the pedestrian recognition [14].

### A. *Feature extraction*

In the feature extraction stage it is possible to bring out the most valuable features and reduce the amount of data that describes the object. As in many similar papers [1, 5, 14, 15], for the feature extraction we used HOG [5]. This method calculates gradients and forms histograms of the gradients orientation. To improve reliability of the HOG the local normalization is used. Finally the ROI is represented by a locally normalized feature vector constructed from the histograms of orientation. In HOG method descriptors involve many parameters concerning the cells, blocks, and histogram [15]. A cell size sets the number of pixels horizontally and diagonally that are contained in a cell. With a block the three parameters are associated: size – the number of cells contained in a block, spacing – the number of cells overlapped by block, and the normalization scheme. In the histogram we should fix at least the number of bins. For details see also [5, 14].

### B. *Classifier*

The second stage of the object classification that finally validates the object is a classifier. One of the most common classifiers, especially in pedestrian detection applications is the SVM [2, 14, 17]. In this paper the linear SVM classifier has been selected. As we proved in [14], on the one hand, the linear SVM classifier is slightly worse from the kernel type SVM, but on the other hand, it is easier to control and requires less parameters to tune. Additionally, it reaches high efficiency in combination with HOG algorithm, and offers very good quality of detection [14].

Further we tested the influence of the image resolution and parameters of the HOG (especially spacing parameter), because these elements were not enough developed in the other, related to the pedestrian detection papers. The rest of parameters, as well optimized, we took from [5] and our previous work [14]. See there for the details.

## III. NIGHT VISION PEDESTRIAN DATASETS

In order to test the classification stage the authors used three night vision video datasets.

### A. *NTPD*

Dataset NTPD (Night-time Pedestrian Dataset) [1] consists of images of pedestrians stored by the NIR active system in the resolution 64×128 (see Fig. 2) and is divided into two sub-bases: training and testing. Numbers of the training and test samples are presented in Tab. I. Please note, that to improve the quality of classifier testing we extend the number of test negative samples. In a real case of pedestrian detection at night, e.g. in automotive applications such asymmetric ratio (more negative than positive samples) is typical.

Figure 2. NTPD dataset pedestrian (positive) samples

TABLE I. TRAINING AND TEST SAMPLES IN EXTENDED NTPD DATASET

| | **No. of training samples** | **No. of test samples** |
|---|---|---|
| Positive samples | 1998 | 2370 |
| Negative samples | 8730 | 12600 (*) |

Comments:
(*) Extended number of test negative samples in compare to [1].

### B. *LSI FIR Pedestrian Dataset*

Dataset LSI FIR (Laboratorio de Sistemas Inteligentes / Intelligent System Lab Far Infrared Pedestrian Dataset) [10, 11] consists of FIR images collected from a vehicle driven in outdoors urban scenarios. The dataset is divided in two subsets: classification dataset with positive and randomly sampled negative images rescaled to 32×64 pixels, and detection dataset, i.e. original (164×129 pixels) positive and negative images with annotations. Numbers of the training and test samples in LSI FIR dataset are presented in Tab. II. In this work we used the classification subset only.

TABLE II. TRAINING AND TEST SAMPLES IN LSI FIR DATASET

| | No. of training samples | No. of test samples |
|---|---|---|
| Positive samples | 10208 | 5944 |
| Negative samples | 43390 | 22050 |

### C. OSU Thermal Pedestrian Dataset

OSU (Ohio State University) Thermal Pedestrian Dataset consists of 10 video sequences of size 320×240 pixels taken at the university campus walkway intersection and street and were captured over several days (morning and afternoon) on various weather conditions using a Raytheon 300D passive thermal sensor [6]. Example image is shown in Fig. 3.

Figure 3. OSU dataset example image [6]

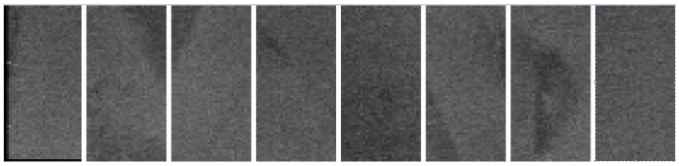

Figure 4. Negative samples produced from OSU dataset

Because the selected images are non-uniformly sampled the object motion information is not available. With this dataset several authors built their own training and testing subsets [3, 4], but they have small number of samples and are rather not standardized. Therefore we decided to prepare our own dataset upon the original dataset [6]. Because the original resolution is 360×240 only, pedestrians are relatively small, we decided to extract samples with the resolution equal to 32×64 pixels. From the first 5 records we extracted pedestrians. The extracted parts together with their mirror images formed positive training samples. From the other 5 sequences we created training samples, analogously. One frame with no pedestrian were cut with a window sized 32×64 with 8 pixels spacing. With the use of additional rotation and vertical or horizontal mirror images we produced 3864 negative samples of the background (see Fig. 4). Half of them were used for training and half for testing. The numbers for modified OSU dataset is presented in Tab. III.

TABLE III. TRAINING AND TEST SAMPLES IN OSU DATASET

| | No. of training samples | No. of test samples |
|---|---|---|
| Positive samples | 1004 | 964 |
| Negative samples | 1932 | 1932 |

## IV. EXPERIMENTS

The above introduced datasets were used to test the HOG feature extractor and the linear SVM classifier. All of images used for training and testing were prepared to the feature extraction by scaling to the size of 64×128 pixels for NTPD dataset and 32×64 pixels for remaining two datasets, i.e. LSI FIR and OSU Thermal Pedestrian Dataset.

Then the images were scaled down in several steps: 56×112, 48×96, 40×80, 32×64, 24×48, 16×32 (c.f. Fig. 5). In all those resolutions the detection effectiveness and the calculation time were computed (sets numbered as 1 to 14), except on the LSI FIR and OSU datasets where the image resolution scaling started at 32×64 pixels (sets 9 to 14), because they have the lower source resolution.

Block size for all sets was established to 16×16 pixels, cell size was 8×8 pixels, and number of bins was equal 9 (all these values were taken as optimal from [5, 14]). For the size of 64×128 pixels, the following values of HOG feature extractor were obtained: 15 blocks horizontally, 7 blocks vertically with 4 cells for each block, and 9 bins in the histogram. Those give $15 \cdot 7 \cdot 4 \cdot 9 = 3780$ features for each image.

Figure 5. Three positive samples in resolutions 64×128, 56×112, 48×96, 40×80, 32×64, 24×48, 16×32. Oryginal images are from NTPD database

Calculated detection efficiency presents point on DET curve where false alarm probability equals miss probability. It was computed with 180 test samples (90 positives and 90 negatives). Obtained mean calculation time takes into account the extraction of HOG features and prediction with corresponding classifier for one test sample.

The video processing was implemented in C# language with EmguCV v. 2.4.10 environment [12] (for classifier training without setting of probability of class membership) and LIBSVM – a library for SVMs [21] (for classifier training with setting of probability of class membership). The computations were performed on the following hardware: CPU Intel Core 2 Duo 2.4 GHz, GPU GeForce 9600M GT, 6 GB of RAM (with no GPGPU usage).

Configuration sets, classification efficiency, calculation time, and DET curves for three mentioned datasets are presented in Tabs. IV, V and VI, Figs. 6 – 12. Figures 6, 9, 11

present DET curves for half block spacing (8×8). Half block spacing means that consecutive blocks for analysis overlap in half size of the block size (16×16 pixels). Figure 6 compares DET results for the half and full block spacing on NTPD dataset. Figures 8, 10, 12 present effectiveness of the classifier and their execution time in various image resolutions on NTPD, LSI FIR, and OSU datasets respectively.

### *A. NTPD*

TABLE IV. CONFIGURATION SETS, CLASSIFICATION EFFICIENCY AND TIME ON A NTPD DATASET

| Set | Frame size [px] | Spacing [px] | No. of features | Detection efficiency (*) [%] | Calculation time (**) [ms] |
|---|---|---|---|---|---|
| 1 | 64x128 | 8x8 | 3780 | 98.94 | 0.76 |
| 1A | 64x128 | 16x16 | 945 | 98.82 | 0.61 |
| 2 | 56x120 | 8x8 | 3024 | 98.78 | 0.60 |
| 3 | 56x112 | 8x8 | 2808 | 98.61 | 0.55 |
| 4 | 56x104 | 8x8 | 2592 | 98.56 | 0.55 |
| 5 | 48x96 | 8x8 | 1980 | 98.57 | 0.43 |
| 5A | 48x96 | 16x16 | 495 | 97.68 | 0.36 |
| 6 | 40x88 | 8x8 | 1440 | 98.74 | 0.34 |
| 7 | 40x80 | 8x8 | 1296 | 98.91 | 0.31 |
| 8 | 40x72 | 8x8 | 1152 | 98.78 | 0.28 |
| 9 | 32x64 | 8x8 | 756 | 98.34 | 0.22 |
| 9A | 32x64 | 16x16 | 189 | 96.41 | 0.19 |
| 10 | 24x56 | 8x8 | 432 | 97.77 | 0.16 |
| 11 | 24x48 | 8x8 | 360 | 97.65 | 0.18 |
| 12 | 24x40 | 8x8 | 288 | 97.25 | 0.14 |
| 13 | 16x32 | 8x8 | 108 | 95.02 | 0.09 |
| 14 | 16x32 | 16x16 | 27 | 84.26 | 0.08 |

Comments (for Tab. IV, V, and VI):

Block size for all sets is set to 16×16 px, cell size is 8×8 px, and number of bins equals 9.
(*) Detection efficiency presents point on DET curve where false alarm probability equals miss probability. It was computed with 180 test samples (90 positives and 90 negatives).
(**) Presented mean calculation time takes into account the extraction of HOG features and prediction with corresponding classifier for one test sample.

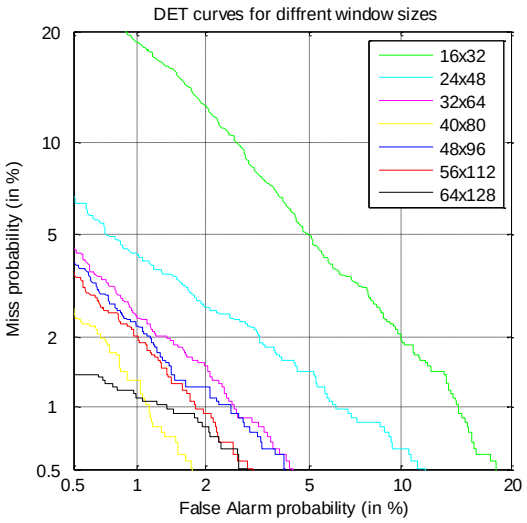


Figure 6. Performance of classifier in various window sizes and half block spacing on NTPD dataset

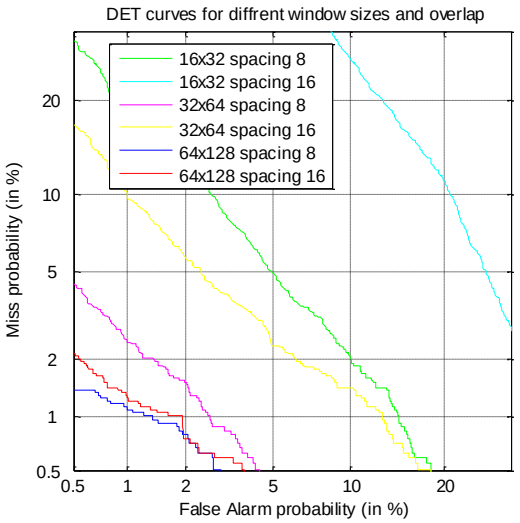


Figure 7. Performance of classifier in various window sizes and various block spacing on NTPD dataset

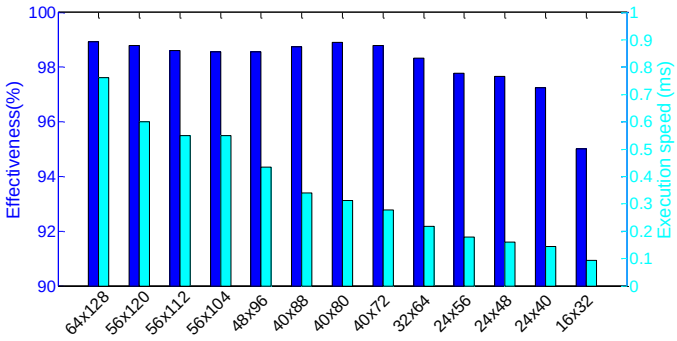


Figure 8. Effectiveness and execution time in various image resolutions on NTPD dataset

### *B. LSI FIR Pedestrian Dataset*

TABLE V. CONFIGURATION SETS, CLASSIFICATION EFFICIENCY AND TIME ON A LSI FIR DATASET

| Set | Frame size [px] | Spacing [px] | No. of features | Detection efficiency(*) [%] | Calculation time(**) [ms] |
|---|---|---|---|---|---|
| 9 | 32x64 | 8x8 | 756 | 98.74 | 0.22 |
| 9A | 32x64 | 16x16 | 189 | 97.98 | 0.19 |
| 10 | 24x56 | 8x8 | 432 | 99.01 | 0.19 |
| 11 | 24x48 | 8x8 | 360 | 98.72 | 0.17 |
| 12 | 24x40 | 8x8 | 288 | 98.31 | 0.13 |
| 13 | 16x32 | 8x8 | 108 | 96.58 | 0.10 |
| 14 | 16x32 | 16x16 | 27 | 90.96 | 0.07 |

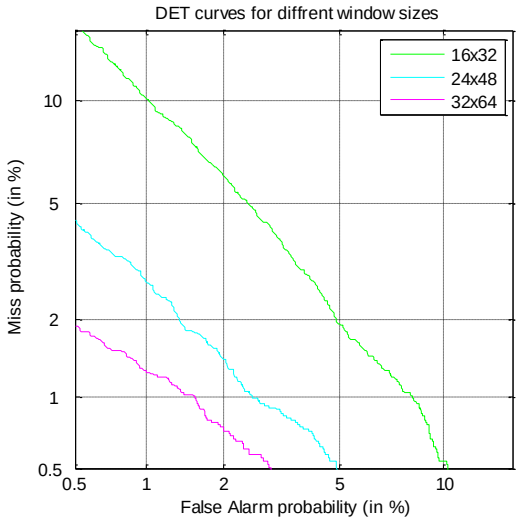


Figure 9. Performance of classifier in various window sizes and half block spacing on LSIFIR dataset

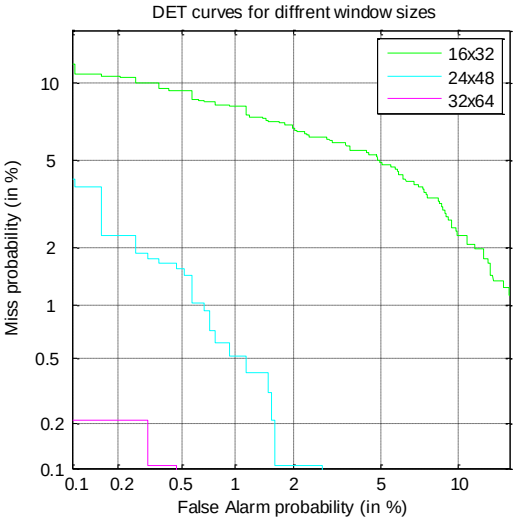


Figure 11. Performance of classifier in various window sizes and half block spacing on OSU dataset

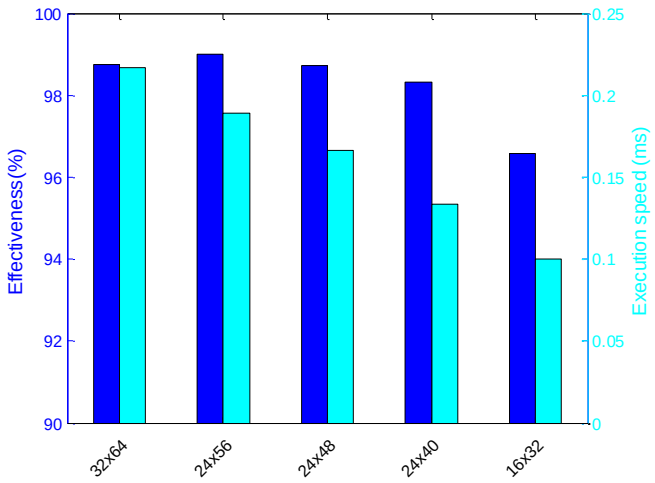


Figure 10. Effectiveness and execution time in various image resolutions on LSIFIR dataset

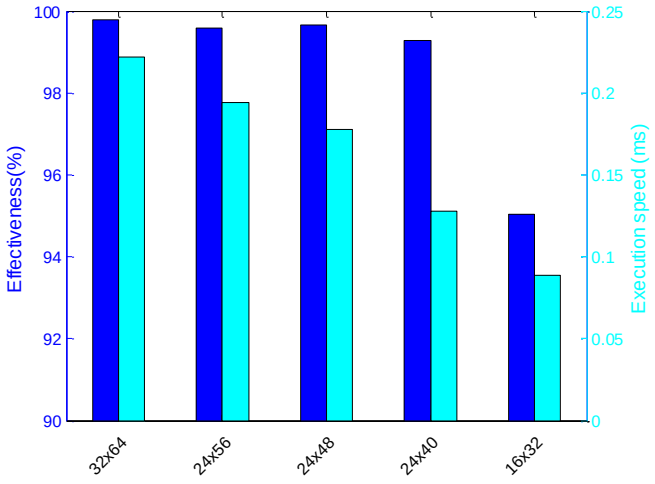


Figure 12. Effectiveness and execution time in various image resolutions on OSU dataset

## C. OSU Thermal Pedestrian Dataset

TABLE VI. CONFIGURATION SETS, CLASSIFICATION EFFICIENCY AND TIME ON A OSU DATASET

| Set | Frame size [px] | Spacing [px] | No. of features | Detection efficiency(*) [%] | Calculation time(**) [ms] |
|---|---|---|---|---|---|
| 9 | 32x64 | 8x8 | 756 | 99.79 | 0.22 |
| 9A | 32x64 | 16x16 | 189 | 98.85 | 0.19 |
| 10 | 24x56 | 8x8 | 432 | 99.58 | 0.19 |
| 11 | 24x48 | 8x8 | 360 | 99.65 | 0.18 |
| 12 | 24x40 | 8x8 | 288 | 99.27 | 0.13 |
| 13 | 16x32 | 8x8 | 108 | 95.03 | 0.09 |
| 14 | 16x32 | 16x16 | 27 | 94.17 | 0.08 |

## D. Discussion

In experiments we used the linear SVM classifiers as they are only slightly worse than the kernel ones in classification but are easier to control and produce stable results.

Obtained detection efficiency for all resolutions (even very low) with half block spacing for HOG are very good (above 95%).

With the same resolution, but different HOG spacing, in all cases half block spacing produce better detection efficiency than full block spacing (c.f. Fig. 7).

When we compare computation time (Figs. 8, 10, 12), we see that for smaller images the computing time shrink. Also full block spacing reduces computation speed (c.f. Tabs. IV, V, VI), but with a loss of detection efficiency, so the full block spacing is rather not recommended.

With the consideration presented in [17] we can prove that the upper limit of the classifier error is related to the dimension of features vector and to the number of training samples (therefore we tried to maximize the training set of samples). These relations are also visible in the conducted experiments (c.f. Tab. IV and Fig. 6). In Fig. 8 we see that the classification effectiveness does not diminish significantly, even if the image resolution decrease. In practice, large window size not always improve the classification effectiveness, but always require more time for processing.

The used datasets are not balanced, i.e. they have much more negative samples. This relation is typical in a real cases, but may lead to problems with the classifier training. If the classifier does not present the probability of class membership (like in EmguCV) and it is trained to achieve the training error as low as possible it could lead to lower false alarm ratio. It is related to more negative slack variables, which affect the objective function. To balance the classifier training we can weigh the samples in both classes:

$$\text{, where} \qquad (1)$$

and  is the upper bound for Lagrange multipliers, as the penalty parameter, which determines importance of the misclassification [17]. The optimum of weights may be found using some gradient methods.

## V. CONCLUSIONS

This paper presents tests of pedestrian detection in low resolution night vision infrared images. Tests performed on three different night vision infrared datasets show that the classification quality of the proposed method is very high in a wide range of low and very low resolutions of images. This fact can be used to reduce the cost of the night vision system by use of a low resolution image sensor without loss of detection quality.

The experiments also proven that large window size used in analysis not always improves the classification effectiveness, but always requires more time for processing.

In future it is planned to further improve the speed of processing with the use of the DSP hardware aided blocks and/or GPGPU technique [7].

## REFERENCES


[1] M.Bertozzi, et al., "A Pedestrian Detector Using Histograms of Oriented Gradients and a Support Vector Machine Classifier", Proc. of the IEEE Intelligent Transportation Systems Conf., Seattle, WA, USA, Sept. 30 – Oct. 3, 2007, pp. 143–148.

[2] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, June 1998, vol. 2, Issue 2, pp 121–167.

[3] C. Dai, Y. Zheng, X. Li, "Layered Representation for Pedestrian Detection and Tracking in Infrared Imagery", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 20–26 June 2005, vol. 3, pp. 13.

[4] C. Dai, Y. Zheng, X. Li, "Pedestrian detection and tracking in infrared imagery using shape and appearance," Computer Vision and Image Understanding, Elsevier, 2007, vol. 106, pp. 288–299.

[5] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, USA, 20–26 June 2005, vol. 1, pp. 886–893.

[6] J. Davis and M. Keck, "A two-stage approach to person detection in thermal imagery," Proc. of Workshop on Applications of Computer Vision, Breckenridge, Colorado, USA, 5–7 Jan. 2005, vol. 1, pp. 364–369.

[7] A. Dąbrowski, P. Pawłowski, M. Stankiewicz, F. Misiorek, "Fast and accurate digital signal processing realized with GPGPU technology," *Electrical Review*, *R. 88*, *No 6/2012*, pp. 47 – 50.

[8] Y. Freund and R. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", J. of Computer and System Science, No. 55(1), 1997, pp. 119–139.

[9] T. Marciniak, D. Jackowski, P. Pawłowski, A. Dąbrowski, "Real-time people tracking using DM6437 EVM," Proc. of IEEE Conf. Signal Processing Algorithms, Architectures, Arrangements and Applications (SPA), 24–26 Sept. 2009, Poznań, Poland, pp. 116–120.

[10] D. Olmeda, et al., "Pedestrian Classification and Detection in Far Infrared Images," Integrated Computer-Aided Engineering 2013, vol 20, pp. 347–360.

[11] D. Olmeda, "Pedestrian Detection in Far Infrared Images," Doctoral thesis, Universidad Carlos III de Madrid, Spain 2013.

[12] Open source computer vision – OpenCV, (on-line) http://opencv.org, 2015

[13] P. Pawłowski, D. Prószyński, A. Dąbrowski, "Real-time procedures for automatic recognition of road signs," Elektronika – konstrukcje, technologie, zastosowania, Sigma NOT, 3/2009, pp. 57–61

[14] K. Piniarski, P. Pawłowski, A. Dąbrowski, "Pedestrian Detection by Video Processing in Automotive Night Vision System", Proc. of IEEE Conf. Signal Processing Algorithms, Architectures, Arrangements and Applications (SPA), 22–24 Sept. 2014, Poznan, Poland, pp. 104–109.

[15] F. Suard, A. Rakotomamonjy, A. Bensrhair, A. Broggi, "Pedestrian Detection using Infrared images and Histograms of Oriented Gradients", Intelligent Vehicles Symposium, Tokyo, Japan, June 13–15, 2006, pp. 206–212.

[16] M. Teutsch, T. Mueller, M. Huber, J. Beyerer, "Low Resolution Person Detection with a Moving Thermal Infrared Camera by Hot Spot Classification", Proc. of IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), Columbus, Ohio, USA, 23–28 June 2014, pp. 209–216.

[17] V. N. Vapnik, "The nature of statistical learning theory", Springer-Verlag New York, Inc., New York, NY, 1995.

[18] R. Walczyk, A. Armitage, D. Binnie, "Comparative Study on Connected Component Labeling Algorithms for Embedded Video Processing Systems," IPCV'10. CSREA Press, 2010

[19] T. Watanabe, S. Ito, K. Yokoi. "Co-occurrence Histograms of Oriented Gradients for Pedestrian Detection", Springer LNCS, 2009, vol. 5414, pp. 37–47.

[20] Ch.-Ch. Chang, Ch. – L. Lin, "LIBSVM: A library for support vector machines," ACM Trans. on Intelligent Systems and Technology, 2011, vol. 2, issue 3, p. 27.



The research was supported with the DS 2015 means.