# Evaluation of Human Silhouette Detection Methods for a Non-cooperative Biometric System

Damian Kacperski, Michal Wlodarczyk, Kamil Grabowski
Department of Microelectornics and Computer Science
Lodz University of Technology
ul. Wolczanska 221/223
90-924 Lodz, Poland
email: {mwlodarczyk, dkacperski, kgrabowski}@dmcs.pl

João Neves
Department of Computer Science
University of Beira Interior
Rua Marquês de Ávila e Bolama
6201-001 Covilhã, Portugal
email: jcneves@penhas.di.ubi.pt

***Abstract*—Biometric authentication systems are becoming increasingly popular and widespread as the demand for automated people identification increases. Although, several research works focused their efforts on these type of solutions, none of the commonly available systems provide a non-cooperative approach to object identification. For this reason, they are not suitable for use in some specific situations, such as people entering the stadium. In this paper, we present an evaluation of different algorithms suitable for person detection in such environment. We focus on investigating their performance and effectiveness under unconstrained conditions, such as different lighting.**

***Index Terms*—biometrics, person detection, non-cooperative identification, Viola - Jones, Histogram of Gradients, background subtraction**

## I. INTRODUCTION

Nowadays, the most technologically advanced practical biometric system, called Iris-On-The-Move, is based on iris pattern analysis. This solution is able to identify at most 30 people per minute, assuming that each person does not exceed the speed of 1 m/s and is not further than 3 meters from the vision system. Due to these restrictions, we decided to develop a solution that would allow to perform a human identification in a fully non-cooperative way. We intend to use two wide- and two narrow-field of view cameras. The purpose of WFOV cameras is to observe the entire scene in order to find potential objects that we want to identify and locate them in three dimensional space. After determining that the found object is a human being, the system is going track it. At the time when the system decide that the distance and pose of tracked person is sufficient to perform recognition, the NFOV cameras will be directed to a specific point of scene to capture high quality image in order to perform recognition. This is somewhat similar to Wheelers image acquisition system setup presented in [1]. The identification process will consist of the fusion of face, ear, periocular and iris biometrics. The proposed structure of the vision system will allow us to track several people at the same time. On the other hand, because of using multiple biometrics features the images will not have to be schematic, the identification will be possible for various poses and distances. All of these factors should let us to achieve both high performance of the system (understood as total number of people passing through the gate at the minute) and high identification rate. Furthermore, the identification process will not require any cooperation from users. People who have already been recognized will be only tracked by wide field of view cameras, while narrow field of view cameras will focus on not identified objects. Fig. 1 presents an example use case of our system. We would like to achieve the performance of at least 30 people per minute. We plan to obtain such results using specialized multicore DSPs and / or FPGA technology [14]. Such system can be installed in airports, stadiums and other public buildings in which fast people authentication is required in order to improve the quality of provided services.

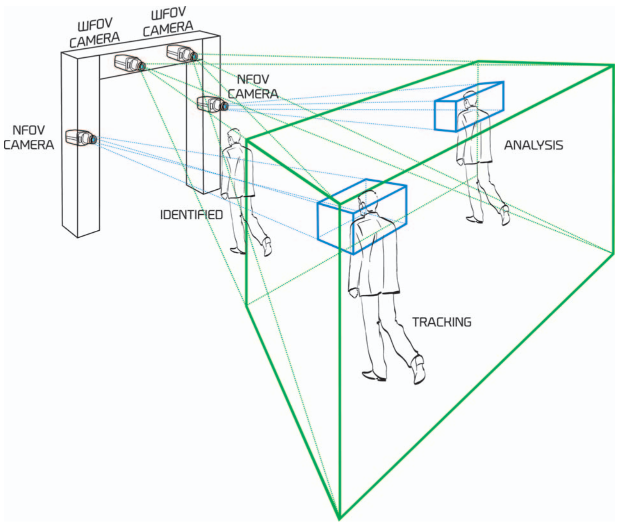


Fig. 1. The COMPACT system overview

## II. DETECTION METHODS

The first step of image analysis in our system will involve detection of object that we intend to identify. Achieving high performance in real time determines the need of using efficient and reliable algorithms. In related work of other authors there are many methods that could be applied to our solution. Those that are best suited for our needs are briefly described in this Section and tested in Section IV.

## A. Viola-Jones Method

In 2001, Paul Viola and Michael Jones proposed a method for fast detection in real time using Haar wavelets [5]. The input image is analyzed by using a detection window that is moved through each fragment of it. Its characteristic features such as local dimming or brightening are investigated. For each dark and bright region the average value of pixel is analyzed which allows to determine the Haar wavelets (classifiers). When the difference of its value exceeds a specific limit, which is typically the limit of the noise of adjacent pixels, it is possible to define the presence of Haars feature. The use of such classifiers provides efficient and simple way to compare complex image features with reference model. Moreover, these characteristics can easily be scaled in order to allow the recognition of objects of different sizes. For the purpose of detection different kinds of Haar classifiers are used, depending on implementation and reference data set. In original implementation only four Haar classifiers are used, but for detecting a person additional elements were introduced later [13]. Fig. 2 shows the operation sequence of Viola-Jones method.

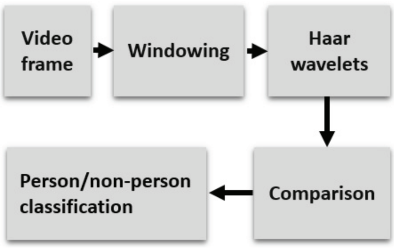


Fig. 2. The operation sequence of Viola-Jones algorithm

In order to identify an object in the input image the cascades of classifiers are compared with a reference model, which is created in an advanced learning process, mainly based on an iterative algorithm AdaBoost (AdaptiveBoost) [12]. Each of the classifier used in cascade is built on a set of weak classifiers. They use single image parameters that have a binary classification efficiency similar to the probability of a random distribution. In order to develop a good classifier it is necessary to use a large number (hundreds, thousands) of learning images both containing and not containing the modeling object. Although this process requires a lot of work and time, it directly affects the effectiveness of the identification.

## B. Histogram of Oriented Gradients method

Histogram of Oriented Gradients (HOG) method was introduced by Navneet Dalal and Bill Triggs [2]. HOG descriptors describe the shape and are used to find a particular object in the image. The basis of HOG descriptors is the assumption that the appearance of the object can be described by the distribution of the edges at a certain angle. The edges presented in the image are determined by calculating the direction derivative along the appropriate vectors. The distinguishing feature of this solution from others, such as [3], is the use of normalization of contrast in so-called blocks, which consist of cells. According to the other authors [4], it allows to obtain greater resistance to external influences, such as changing lighting and shadows.

The first step in calculating HOG descriptors is to perform pre-processing of the input image, which relies on pixel intensity normalization and gamma correction. Both operations are used to remove excess contrast from the image, which may occur due to distortions introduced by the vision system. Then, pixels are grouped into equally spaced fragments, known as cells. The following step is to calculate the gradient for each cell. It is counted in two directions, vertically and horizontally, through the use of two horizontal filter [-1, 0, 1] and vertical $[-1, 0, 1]^T$. Having this done, the histogram is created for each cell to show the distribution of these gradients. In the final step, the cells are grouped into larger spatial units, denoted as blocks. Then the contrast normalization is performed for each block. A study carried out by Dalal [4] shows that the best detection results are obtained using circular blocks. As a result of these operations HOG descriptor is saved as a vector consisting of a histogram calculated for all blocks. Depending on the implementation, the blocks may overlap meaning that selected cells can occur more than once in the final results. In the original implementation the classification is performed using the Support Vector Machines (SVM). The sequence of object detection using HOG descriptors is presented in Fig. 3.

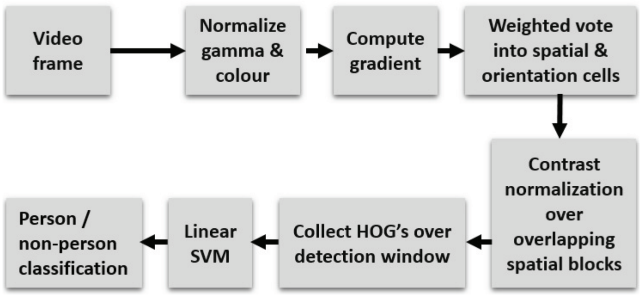


Fig. 3. The operation sequence of HOG algorithm

## C. Background Subtraction Methods

The detection methods described as background subtraction are based on dynamically finding the changes that occur on the images. They conduct the division of the input image into two parts consisting of background and foreground. Fig. 4 shows an operation sequence for these methods. The basic requirement for the use of background subtraction methods is use of static cameras that can register the image without introducing undesirable shifts. Another negative impact on the successful detection can be any change in the background environment such as the wind moving leaves on trees or changing lighting conditions. In order to minimize these effect sophisticated algorithms were developed such as Mixture of Gaussians (MOG).

MOG was introduced by C. Stauffer and W. Grimson [7]. The authors, instead of modeling the value of each pixel as one particular distribution, model values of each pixel as mixture of Gaussians. Relying on the persistence and variance

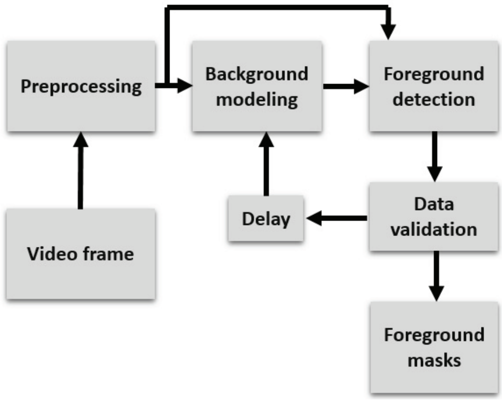


Fig. 4. The operation sequence of background subtraction methods

of each Gaussian, the background model of the input image is determined. By using this approach the algorithm becomes somewhat resistant to changes in lighting and repetitive movements of the scene. The values of pixels that do not match the distribution of background are grouped in blocks and identified as foreground objects. Such groups of pixels are then identified and tracked in subsequent frames. This method was then improved by P. KadewTraKuPong and R. Bowden in [9], who improved the detection by introducing strategies for shadow removal. After that Z. Zivkkovic presented the improvements in efficiency and effectiveness of the algorithm in [8].

## III. Detection Methods Comparison

In order to access object detection methods performance in unconstrained scenarios, we tested three detection methods described in previous section. The analysis of related works, such as [10] or [11], let us assume that using Viola-Jones Haar cascades or HOG descriptors the detection rate should be at least 80%. Because the computational requirements of both of these methods are rather high, we conclude that a good complement to these algorithms may be MOG for background subtraction that can significantly narrow the search area to get better performance results. As all the available tests presented in other papers refer mainly to general scenarios and do not match our systems use cases, we decided to test each of the algorithm to find the best approach for our needs. The obtained results and experiment scenarios are described in the next section.

## IV. Experiment Results

To perform objective and reliable experiments we developed a C++ application, which allows us to study the effectiveness and performance of each detection method. Each algorithm implementation comes from OpenCV. Viola-Jones method is used in improved version [13], just as MOG [9]. HOG is based on original implementation [2]. We conducted series of tests that allow us to draw specific conclusions in context of our system. The input images sequence was gathered with the use of high resolution camera Stingray F-504B. This camera captures images in gray scale, but it does not any negative impact on our analysis. The camera was mounted at a height of 3 meters, so as to simulate WFOV camera target location in our system. Its zoom was set in such a manner that objects can be detected from a distance of about 8-10 meters. All tests were carried out on a PC computer with Intel i5 processor and 4 GB of RAM memory.

The first test case allowed us to compare the detection effectiveness for both Viola-Jones and HOG algorithms in good lighting conditions. Test data set consisted of 161 images, one person visible on each frame. We checked and compared both Positive Detection Rate (PDR) and False Detection Rate (FDR). The results are presented in Table I. Fig. 5 shows two example frames, on which the test was carried out.

TABLE I
HOG and Viola-Jones results for good lighting conditions

| Algorithm | Results | |
|---|---|---|
| | FDR[%] | PDR[%] |
| Viola-Jones | 0.87 | 71.43 |
| HOG | 9.74 | 95.65 |

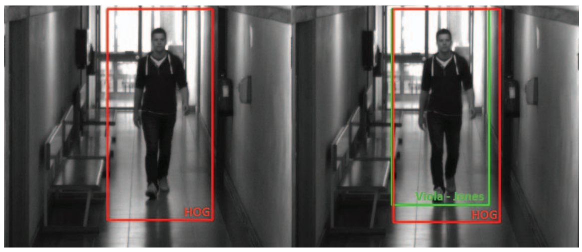


Fig. 5. HOG and Viola-Jones comparison in good lighting conditions

The second test intended to check the impact of lighting conditions on detection results from HOG and Viola-Jones. Test data set consisted of 106 input images with one person visible on each frame. We conducted analogous simulation as for the first test, however, the amount of light was very small. Because we assume our system to be installed in different places, we wanted to examine how big the impact of lighting on the detection effectiveness is. Obtained results are shown in table II. Fig. 6 shows two sample frames from the testing data set.

TABLE II
HOG and Viola - Jones results for bad lighting conditions

| Algorithm | Results | |
|---|---|---|
| | FDR[%] | PDR[%] |
| Viola-Jones | 15.38 | 12.26 |
| HOG | 59.22 | 97.17 |

The third test was to compare the execution time of both HOG and Viola-Jones methods. Test data set was the same as for the first experiment. We checked how MOG (background subtraction) can improve their performance by limiting search area for descriptors and how it affects the detection effectiveness. The execution time for each method was measured for every frame. Table III shows the obtained results. Fig. 7 shows two sample frames from the test data set.

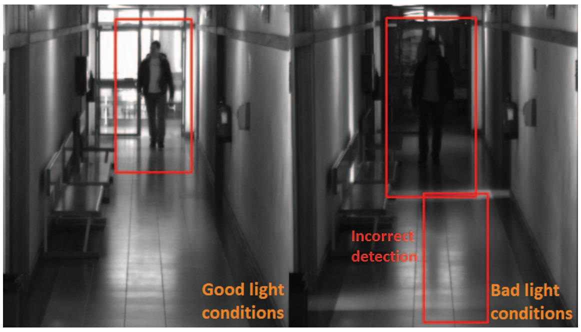


Fig. 6. HOG and Viola-Jones comparison in different lighting conditions

TABLE III
HOG AND VIOLA - JONES RESULTS WITH MOG

| Algorithm | Computation time [ms] | | | Results | |
|---|---|---|---|---|---|
| | min | max | avg | FDR[%] | PDR[%] |
| Viola-Jones | 21 | 69 | 31.04 | 0.87 | 71.43 |
| Viola-Jones + MOG | 15 | 60 | 27.05 | 0.94 | 65.84 |
| HOG | 457 | 933 | 512.89 | 9.74 | 95.65 |
| HOG + MOG | 33 | 130 | 130.47 | 1.29 | 96.27 |

Fig. 7. HOG and Viola-Jones with MOG

## V. CONCLUSIONS

The performed tests show how many problems and issues must be analysed in order to provide effective and efficient object detection in non-cooperative scenarios. After analyzing related works and the most popular algorithms we decided to implement and test Viola-Jones, HOG and MOG for background subtraction in context of our system.

The obtained results are very encouraging for further development and allow us to draw a number of conclusions in context of our system. Firstly, it turns out that Viola - Jones algorithm is not suitable for our purposes. It provides relatively sufficient results in good lighting conditions, maintaining high performance, however, the results obtained in bad lighting conditions are highly unsatisfactory, as shown in Table II. The decrease in its effectiveness is unacceptable for preliminary object detection in our system. Nevertheless, we still consider to use this method for segmentation of features, such as face, ear or periocular area. HOG algorithm turns out to perform much better for people detection in our data set. The impact of variable lighting conditions on the results is relatively insignificant and can be neglected. These results are presented in Table I and Table II. The main drawback of HOG is its computation time. It improves significantly where the search region for the descriptor was narrowed by using MOG for background subtraction. It is nearly five times smaller, as shown in Table III. Furthermore, the introduction of this additional processing step also allows us to reduce the false detection rate, because it turns out that parts of the images, where wrongly detected objects were placed, are now not analyzed any more. The MOG algorithm provides satisfactory results, so we state that this solution is sufficient for our purpose.

Summarizing, the conducted tests yields unambiguous and valuable findings. We conclude that the combination of HOG and background subtraction methods performed with the use of MOG algorithm is the optimal solution for our system to perform effective and efficient person detection.

## VI. FUTURE PLANS

In future work, we intend to optimize both HOG and MOG algorithms to get even better positive detection rates and higher performance results. The next stage of our research will also be associated with further processing of the detected objects in scene in order to locate another features such as face, ear or periocular area.

## ACKNOWLEDGMENT

This presented research was funded by Polish National Centre for Research and Development in the frame of the project LIDER/027/591/L-4/12/NCBR/2013, entitled: ”Non-COoperative bioMetric system for Positive AuthentiCaTion” (COMPACT)

## REFERENCES

[1] Wheeler, F. W., Perera, A. G. A., Abramovich, G., Bing Yu and Tu, P. H., *Stand-off Iris Recognition System*, Biometrics: Theory, Applications and Systems, 2nd IEEE International Conference on, 2008
[2] Dalal N., Triggs B., *Histograms of oriented gradients for human detection*, Proceedings of IEEE Converence Computer Vision and Pattern Recognition, 2005
[3] W.T. Freeman, M. Roth., *Orientation histograms for hand gesture recognition*, Intl. Workshop on Automatic Face and Gesture Recognition, IEEE Computer Society, 1995
[4] Dalal N., *Finding People in Images and Videos, Phd Thesis*, Institute National Polytechnique de Grenoble, 2006
[5] Viola P., Jones M., *Rapid object detection using a boosted cascade of simple features*, IEEE Computer Society Conference, pp. 511 - 518, 2001
[6] Piccardi M., *Background subtraction techniques: a review*, In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, pp. 3099 - 3104, 2004
[7] Stauffer C., Grimson W., *Adaptive background mixture models for real-time tracking*, CVPR, vol. 02, 1999
[8] Zivkovic Z., *Improved adaptive Gaussian mixture model for background subtraction*, Proceedings of the 17th International Conference on Pattern Recognition, 2004
[9] KadewTraKuPong P., Bowden R., *An improved adaptive background mixture model for real-time tracking with shadow detection*, Video-Based Surveillance Systems, 2002
[10] Brehar R., Nedevschi S., *A comparative study of pedestrian detection methods using classical Haar and HoG features versus bag of words model computed from Haar and HoG features*, Intelligent Computer Communication and Processing (ICCP), 2011

[11] Schiele B., Andriluka M., Majer N., Roth S., Wojek C., *Visual People Detection Different Models, Comparison and Discussion*, Proceedings of the IEEE ICRA Workshop on People Detection and Tracking, 2009
[12] Grabowski K., Sankowski W., *Human tracking in non-cooperative scenarios*, Chapter in Springer book, 2014
[13] Lienhart R., Maydt J., *An Extended Set of Haar-like Features for Rapid Object Detection*, IEEE ICIP 2002, Vol. 1, pp. 900-903, 2002
[14] Grabowski, K., Napieralski, A., *Hardware architecture optimized for iris recognition*, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 21 No. 9, pp. 1293-1303, 2011
[15] *The OpenCV Reference Manual Release 2.4.8.0*, 2013