Research on Human Silhouette Detection Methods for a Non-cooperative Biometric System

Damian Kacperski, Michał Włodarczyk, and Kamil Grabowski

***Abstract*—Together with the development of effective and efficient people identification algorithms, biometric authentication systems become increasingly popular and widespread, leading to a significant growth in the number of institutions interested in implementing and using such systems. 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. Therefore, we decided to go up against these limitations and develop biometric identification system for less constrained scenarios. In this paper, we present an evaluation of different algorithms suitable for human silhouette detection in such environment. We focus on investigating their effectiveness and performance under unconstrained conditions, such as different lighting.**

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

I. Introduction

CURRENTLY *Iris-On-The-Move* is the most known practical biometric system for less-constrained biometric identification. This system 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. The authors decided to develop a solution that would allow to perform a human identification in a less cooperative way. For the purposes of that system authors plan to use two wide- and two narrow-field of view cameras. The purpose of WFOV (Wide-Field-of-View) 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 (Narrow-Field-of-View) 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 Wheeler's 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 significant cooperation from users. People who have already been recognized will be only tracked by WFOV cameras, while NFOV will focus on not identified objects. Fig. 1 presents an example use case of the presented system. We would like to achieve the performance of at least 30 people per minute. We plan to obtain such results using powerful PC with multi-core processor and high amount of RAM. In the future we intend to use specialized multi-core DSPs or FPGA technology [14]. Such system can be deployed on airports, stadiums and other public buildings in which fast people authentication is required in order to improve the quality of provided services.

D. Kacperski, M. Włodarczyk and K. Grabowski are with the Department of Microelectornics and Computer Science, Lodz University of Technology, ul. Wolczanska 221/223, 90-924 Lodz, Poland, (e-mails: {dkacperski, mwlodarczyk, kgrabowski}@dmcs.pl)

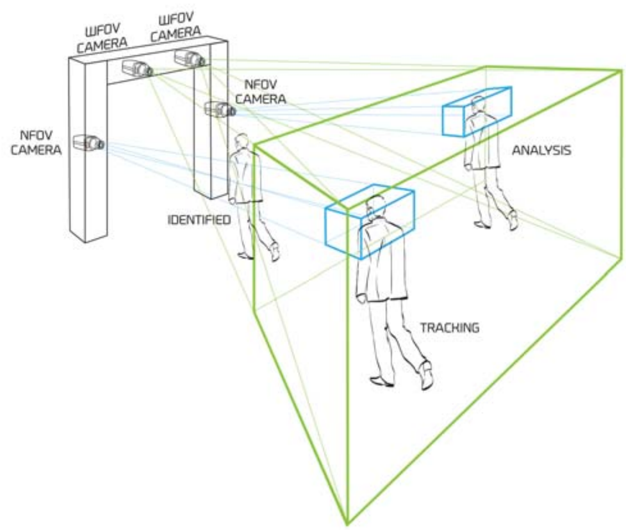

Fig. 1. The COMPACT system overview

II. Detection Methods

The first step of image analysis in the presented 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. 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 HaarâĂŹs 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.

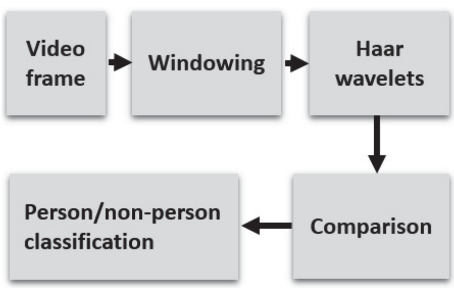

Fig. 2. The operation sequence of Viola-Jones algorithm

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.

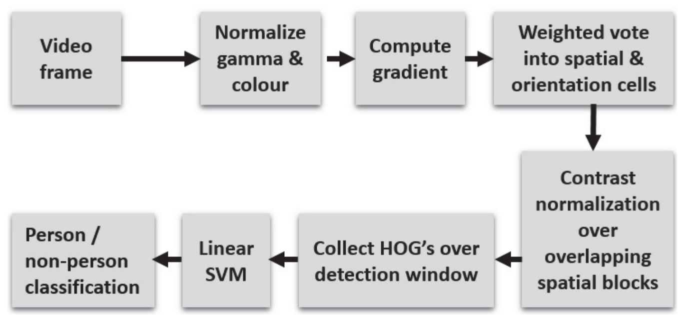

Fig. 3. The operation sequence of HOG algorithm

C. Background Subtraction Methods

The detection methods described as *background subtraction* are based on dynamically finding the changes that occur on the images. They perform 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

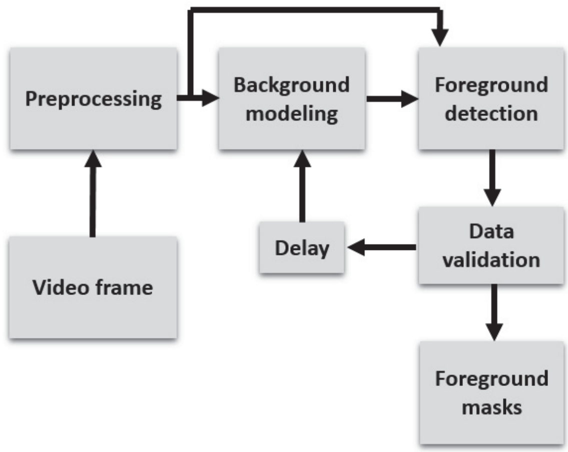

Fig. 4. The operation sequence of background subtraction methods

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

The analysis of related works, such as [10] or [11], let authors 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 decided to examine how decreasing image size affects both performance and computation time. Moreover, we checked if these algorithms may be complemented by MOG for background subtraction that can significantly narrow the search area to get better performance results. The next sections describes experiment scenarios and obtained results.

IV. Experimental Results

To perform reliable and objective results the authors developed a C++ application, which allowed to study each detection method effectiveness and it's computation performance. Each algorithm implementation comes with OpenCV library, which was used for first experiments. Viola-Jones method is used in improved version [13], just as MOG [9]. HOG is based on original implementation [2]. All experiments were carried out using own database, which at that time consisted of 300 images. Pictures were taken in two separate sessions. First session was carried out with bad lighting conditions, whereas the second with typical indoor illumination system turned on. The deterioration of the lighting conditions was simulated by amending exposure time of the camera from 200 milliseconds to 150 milliseconds. This means that the amount of light which was registered by the camera was reduced 1.33 times. Some of the samples contain more than one person. Database was gathered with the use of high resolution camera Teledyne Dalsa TS-C2500, which was mounted at the height of approx. 2.5 meters, on a specially prepared gate that is the main part of the COMPACT system. We used lenses set up to approx 20 mm of focal length, which allowed to observe the object distant from 2 up to 10 meters. The gate was located on the corridor at the Department of Microelectronics and Computer Science (DMCS) of the Technical University of Lodz (TUL). The image processing algorithms were run using a PC computer with Intel i5 processor and 16 GB of RAM memory.

The first test case allowed to compare the detection effectiveness for both HOG and Viola-Jones algorithms in good lighting conditions. The authors checked and compared both False Detection Rate (FDR) and Positive Detection Rate (PDR). The results are shown in Table I. Fig. 5 presents two example frames.

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

Algorithm	Results	
	FDR [%]	PDR [%]
Viola-Jones	5.88	62.09
HOG	5.23	96.08

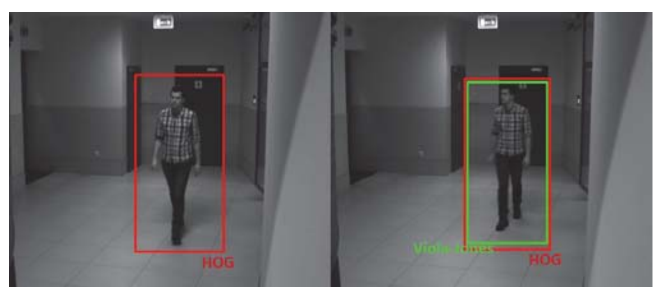

Fig. 5. HOG and Viola-Jones comparison in good lighting conditions

The second test intended to check the impact of lighting conditions on detection results from HOG lighting conditions on detection results from HOG and Viola-Jones. We conducted analogous simulation as for the first test, however, the amount of light was very weak. 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 presents 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	6.47	55.88
HOG	7.65	92.35

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

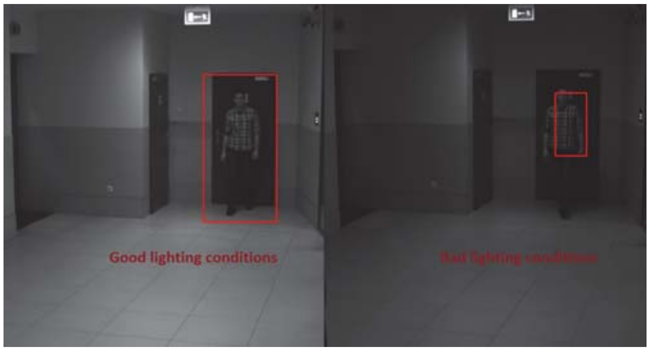

Fig. 6. HOG and Viola-Jones comparison in different lighting conditions

subtraction) can improve their performance by limiting search area for descriptors and how it affects the detection effectiveness. We also examined it for decreasing image size. 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.

TABLE III
HOG AND VIOLA-JONES PERFORMANCE RESULTS

Algorithm	Image Scale	Computation time [ms]			Results	
		min	max	avg	FDR [%]	PDR [%]
Viola-Jones	100%	60	95	69	5.23	62.09
	50%	60	119	84	4.58	62.75
	25%	67	195	111	5.88	59.48
Viola-Jones + MOG	100%	58	296	139	1.31	61.44
HOG	100%	1911	2247	2048	5.23	96.08
	50%	396	469	410	2.61	96.73
	25%	65	82	74	1.96	67.97
HOG + MOG	100%	730	1052	862	2.61	95.42

Fig. 7. HOG and Viola-Jones with MOG

V. CONCLUSIONS

The performed tests show how many problems and issues must be analyzed in order to provide effective and efficient human silhouette detection in non-cooperative scenarios. After analyzing related works and the most popular algorithms we decided to examine both effectiveness and computational performance of HOG and Viola-Jones in context of our system and real time processing.

The obtained results allow us to draw a number of conclusions. Firstly, we notice that HOG gives significantly better detection results than Viola-Jones for both good and bad lighting conditions. At the same time, the false detection rates are similar for both algorithms. Moreover, it turns out that HOG is also more resistant to variable lighting conditions. 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 or the image size was decreased. It is nearly 2 to 5 times smaller, as shown in Table III. What is also important, in most cases, the positive detection rate did not change considerably. Only the result for the image size reduced four times differs, because for further distances the human silhouette was too small. Furthermore, the introduction of this additional 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. When using Viola-Jones, we does not notice any computational performance improvements when using MOG or decreasing image size.

Summarizing, the conducted tests yields unambiguous and valuable findings. We conclude that the effectiveness of Viola-Jones algorithm is unacceptable for preliminary human silhouette detection in our system. We intend to use HOG with one of the methods reducing its computation time.

VI. FUTURE PLANS

Future work will focus on motion detection and object tracking. It will be also 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 âĂŞ 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

Damian Kacperski is a Ph.D. student and teaching assistant at Lodz University of Technology. He received B.Sc. degree at Electronics in 2012 and M.Sc. degree at Computer Science in 2013. His research is focused on real-time, high performance vision systems and tracking algorithms in non-cooperative scenarios.

Michał Włodarczyk received the B.Sc. degree at Electronics in 2012 and M.Sc. degree in 2013 at Computer Science from the Lodz University of Technology, Poland. He currently serves as a PhD Student and teaching assistant at the Department of Microelectronics and Computer Science. His main reserach intrests are focused on a computer vision, pattern recognition and non-cooperative biometric recognition algorithms.

Kamil Grabowski received the Ph.D. degree in electronics from the Lodz University of Technology in 2010. His research interests are mainly focused on the field-programmable gate array and digital signal processor hardware systems for advanced image processing with an emphasis on the biometric applications, namely the study of iris-based authentication. Currently, he serves as Assistant Professor at the Department of Microelectronics and Computer Science of the Technical University of Lodz Poland, and is with the ”DMCS Biometric Lab.” research group. He is an author and coauthor of over 20 publications in International Journals and Conferences.