F O U N D A T I O N S O F C O M P U T I N G A N D D E C I S I O N S C I E N C E S
Vol. 42 (2017) No. 2
ISSN 0867-6356
DE GRUYTER OPEN DOI: 10.1515/fcds-2017-0006 e-ISSN 2300-3405

EVALUATION OF FACE DETECTION ALGORITHMS FOR THE BANK CLIENT IDENTITY VERIFICATION

Maciej SZCZODRAK*, Andrzej CZYŻEWSKI*

Abstract. Results of investigation of face detection algorithms efficiency in the banking client visual verification system are presented. The video recordings were made in real conditions met in three bank operating outlets employing a miniature industrial USB camera. The aim of the experiments was to check the practical usability of the face detection method in the biometric bank client verification system. The main assumption was to provide a simplified as much as possible user interaction with the application. Applied algorithms for face detection are described and achieved results of face detection in the real bank environment conditions are presented. Practical limitations of the application based on encountered problems are discussed.

Keywords: video analysis, face detection, banking

1. Introduction

Face image is an important biometric feature, which is easy to acquire and it does not require any special or physical interaction between the subject and the acquisition device. The most popular application of face image processing in the computer vision systems is person verification or identification, however a more sophisticated analysis based on face detection can be made [1]. Face scanning biometric technology is highly versatile and is utilized in the wide spectrum of appliances reaching mobile devices [3][8]. Such methods of identification nowadays emerge in banking systems [2][5]. Bank ATMs equipped with systems for client verification based on facial recognition methods are being tested. Despite the face detection and verification is implemented in many systems and was recently discussed in a number of publications, achieving human-level performance is still a challenge [10]. Robust results are often hard to accomplish in real conditions, because of the influence of such factors as illumination, head position or image quality [6]. Therefore, the evaluation of the algorithms performance in this specific working environment should be made.

* Multimedia Systems Department, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, {szczodry,andcz}at:sound.eti.pg.gda.pl

In this paper we present an evaluation of face detection algorithms in the conditions met in bank operating rooms. Videos were acquired in 3 bank branches, in 4 various operating room types during a normal bank activity. The face detection algorithm is investigated, since proper detection provides a base to further analysis, i.e. person identification. Persons participating in the tests were not specialists in image processing and behave as much naturally as possible, maintaining the frontal position to the camera.

The paper is organized as follows: the next section describes the concept of the banking client verification system and of the environment of the face image acquisition. In Section 3, the algorithms for face detection together with the methods of detection quality evaluation are introduced. The results are presented in Section 4. Last section concludes the paper with a discussion on the meaning of results achieved.

2. Bank client identity verification

In this section we present the concept and assumptions of the banking client visual verification system together with description of conditions met in bank operating rooms. Face detection is the one of the first phases of the further process (i.e. identification or verification). Therefore, the correct result at this stage should be provided.

Source of the face images is the camera mounted on the desk, in front of the customer. Acquisition of the image is provided by wide angle lens, since the position taken at the desk by customer can vary, as well as his/her height fluctuates. Since the procedure is supervised by a bank employee, there is no need for checking if the face is real one, and not a counterfeit one e.g. a photo.

From the client's assistant point of view, operation of the verification system should be as simple as possible. The process of face detection should be initiated manually and then run automatically. To provide a good quality image, the camera acquisition parameters such as exposure time or gain should adjust according to the lighting conditions. The processing unit should be maintained small, having a passive cooling in order to remain quiet.

Automatic detection of customer face should be robust to possible multiple face images presence in the background (other customers waiting, walking etc.). Therefore, false positive detections should be eliminated and selection of the face which distance is the shortest should be done.

A system for automatic face detection (being a part of a verification system) has been developed and tested employing video data acquired in bank operating rooms.

Each of four bank operating rooms was characterized by different light conditions and by diversified customers activity. Therefore, corresponding 4 groups of recording conditions, namely: C1 to C4, can be distinguished, 3 daylight (various intensity) with additional artificial light sources and one daylight (cloudy), only. A small industrial camera with USB 3.0 interface was providing the video source. Maximum camera exposure time was limited to 50 ms in order to maintain 20 fps. Frames were recorded at resolution of 2592×1440. The processing unit was a compact, Pico-ITX format board computer with 4 cores Intel Atom processor, so the image was scaled by factor 0.25 in order to reduce computational load.

Camera setup in all cases except C4 was such that the ceiling and wall light sources were visible. In the first light conditions group (C1) the camera was directed towards the

window at the distance of about 10 meters. A person could occlude the window. In the second group (C2), there was no direct influence of daylight in the frame, however numerous artificial light sources were present on the walls and ceiling. Additional visible light came from a tablet which was placed on the desk in the front of the investigated person. In the third group (C3), the daylight was prevailing, however not originated from the sunlight directly. In the fourth group (C4), the camera was directed towards a large window which constituted about 50% of background, while the weather was cloudy. Illustrations of the described conditions are provided in Figure 1 and Figure 2.

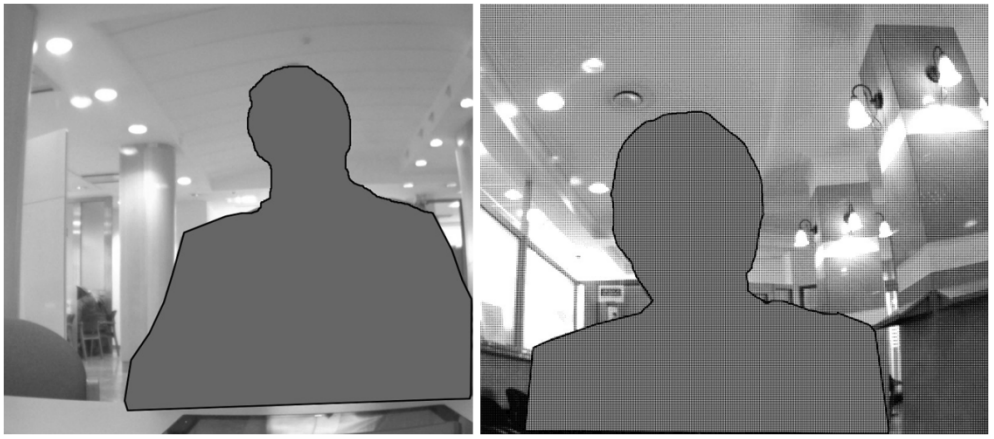

Figure 1. Camera view and person silhouette position examples for bank operating room 1 (left) and 2 (right)

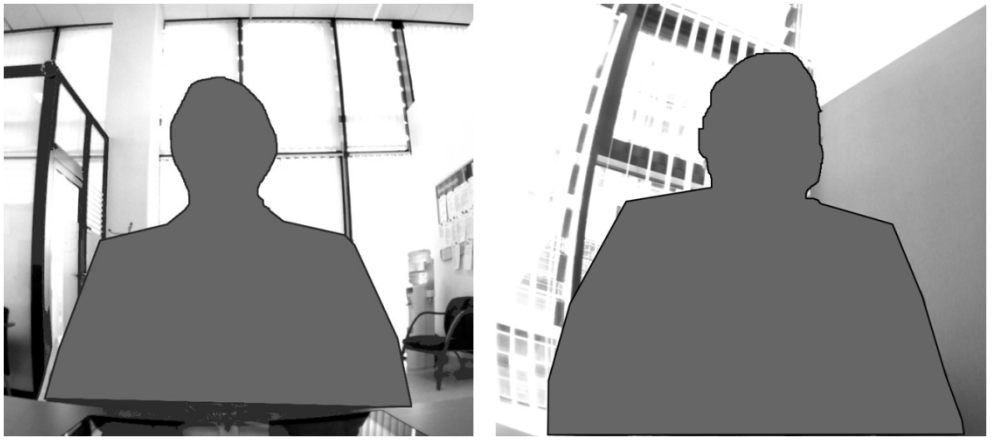

Figure 2. Camera view and person silhouette position examples for bank operating room 3 (left) and 4 (right)

The total number of unique individuals was 58 women and 20 men. Among those persons, 16 women and 10 men were wearing glasses. There were 10 men with full beard, among them 6 wearing glasses.

3. Detection and evaluation methods

Face detection was based on the Haar cascade classifier [11] used in the Viola-Jones algorithm available in OpenCV library [4]. It contains pre-trained classifiers for faces and eyes. The main parameters of the method which have the key influence to detection results are: scaling factor (SF) and minimum neighbors (MN). The scaling factor is responsible for matching the size of image in the multi scale representation to the size of the trained model. The second parameter value is a threshold for eliminating multiple detections which comes from the multi scale approach.

To fulfill the assumption of simplicity of software operation, and automatic detection of the client face in a wide angle view, an eye detection is performed on each region detected as a face. This approach eliminates many false positive matches, however it may cause problems (false negatives) in some difficult cases. After the acquisition ends, detected face images are sorted in the descending size order and then the largest ones are selected.

Each person participating in the survey was told that the face image is recorded and was asked to maintain a neutral face expression. For each investigated person the total number of 300 frames were acquired (lasting about 15 seconds). We have acquired images of 37 persons in the operating room 1 (conditions #1), 29 persons in the operating room 2 (conditions #2), 11 persons in the operating room 3, and 8 persons in the bank meeting room (conditions #4).

In the process of analysis, selected variations of face detection algorithm parameters were evaluated and compared by the resultant detection performance. The obtained measures are presented in the next section.

The performance of the face detection algorithm is considered at the pixel level. The method of the precision evaluation is based on the comparison of the algorithm output that is formed by rectangles (boxes) covering detected faces with manually marked 'ground truth' face positions in each frame. Annotations were created with a dedicated tool [12]. Metrics proposed in Mariano's et. al paper [7] (average detected box area precision, average detected box area recall) are used for the algorithm accuracy evaluation.

An average detected box area precision is defined for a frame t as follows:

$$P(t) = \frac{\sum_{D_i(t)} p(D_i(t))}{N_{D(t)}} \tag{1}$$

where

$$p(D_i(t)) = \frac{\left|D_i(t) \cap U_G(t)\right|}{\left|D_i(t)\right|}, \tag{2}$$

$D_i(t)$ - i-th detected object in frame t,
$N_{D(t)}$ - number of detected objects in frame t,
$U_G(t)$ - spatial union of the ground truth boxes in frame t,
operator $|\cdot|$ denotes number of pixels.

The metric denotes the average ratio of a detected object area that covers a ground truth object.

Next metric, namely the average object area recall, is defined for a frame *t* as:

$$R(t) = \frac{\sum_{G_i(t)} r(G_i(t))}{N_{G(t)}} \tag{3}$$

where

$$r(G_i(t)) = \frac{\left|G_i(t) \cap U_D(t)\right|}{\left|G_i(t)\right|} \tag{4}$$

$G_i(t)$ - *i*-th ground truth object in frame *t*,
$N_{G(t)}$ - number of ground truth objects in frame *t*,
$U_D(t)$ - spatial union of detected bounding boxes in frame *t*,
operator $|\cdot|$ denotes the number of pixels.

The metric denotes the average ratio of the total area of correct detections results to the ground truth objects total area.

Commonly used *F*1 metric provides a single measure of detection quality assessment. It combines the results values of precision (*P*) and recall (*R*), given by:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R} \tag{5}$$

4. Experiment results

Offline analysis of gathered video was performed for the set of parameter values. It was aimed to determine a parameter set that would serve best for each condition, preferably it should be set once and remain not changed for individual persons. The algorithm parameters are shown in Table 1.

Table 1. Configuration of algorithm parameters

Config#	SF	MN	Config#	SF	MN
1	1.1	2	9	1.9	2
2	1.2	2	10	2	2
3	1.3	2	11	2.5	2
4	1.4	2	12	3	2
5	1.5	2	13	1.1	3
6	1.6	2	14	1.1	4
7	1.7	2	15	1.1	5
8	1.8	2	16	1.1	6

Due to large number of possible combinations, it is difficult to present all results in the paper. Nevertheless, charts visualizing the obtained metrics values for selected cases are presented below. Results of precision/recall metrics were presented for selected 2 persons, men, who differ by beard and hair style and glasses presence. Person number one is

characterized by full beard and lush hair and not wearing glasses. Person No. 2 is a bald male, without a beard, wearing thick framed glasses.

Figure 3 and Figure 4 show precision and recall metrics of face detection algorithm outcomes for the person No. 1, in two lighting conditions. The best precision were denoted for configuration 10 and best precision for configuration 16 in both cases. Figure 5, Figure 6 and Figure 7 present the precision and recall metrics for person No. 2, in three various conditions. We can observe that the configuration 6 entails a very good precision in all cases, but low recall at the same time, meaning a small number of fine detections. Increasing recall is related to decreasing precision, as in case of parameter sets 13-16. For person No. 1, configuration 6 also provides very good precession and shows that appearance should have marginal influence on the detection result.

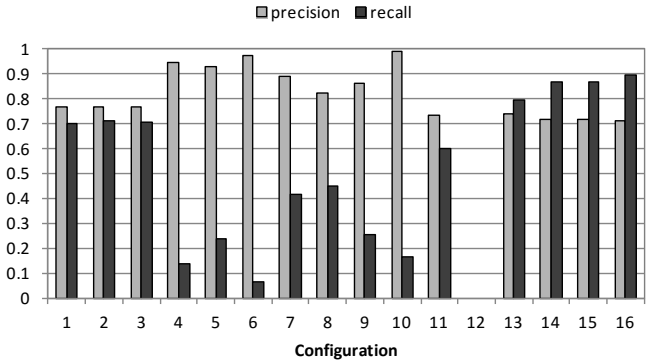

Figure 3. Precision and recall metrics for person No. 1, conditions C1

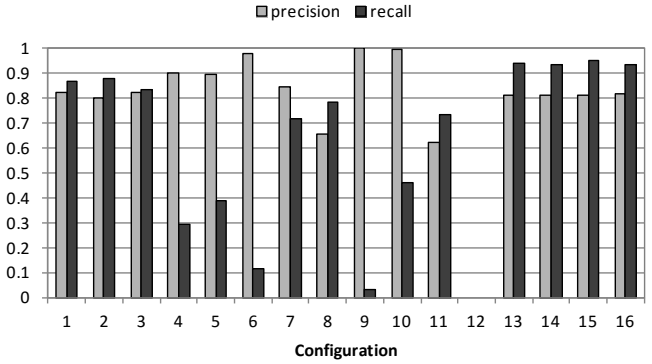

Figure 4. Precision and recall metrics for person No. 1, conditions C2

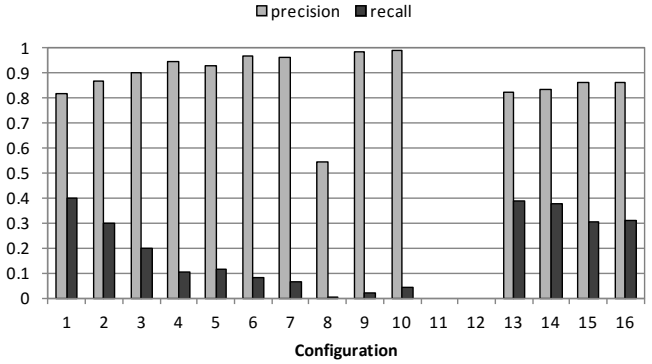

Figure 5. Precision and recall metrics for person No. 2, conditions C1

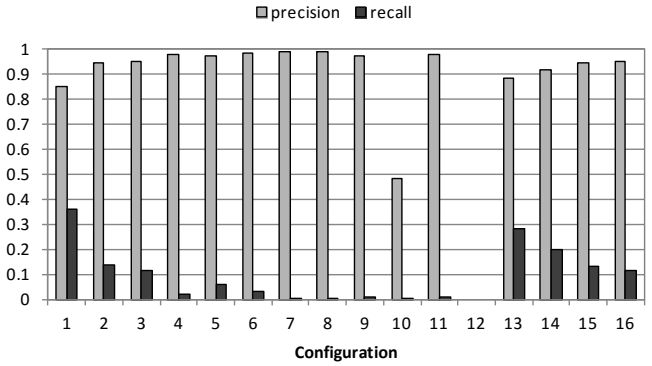

Figure 6. Precision and recall metrics for person No. 2, conditions C2

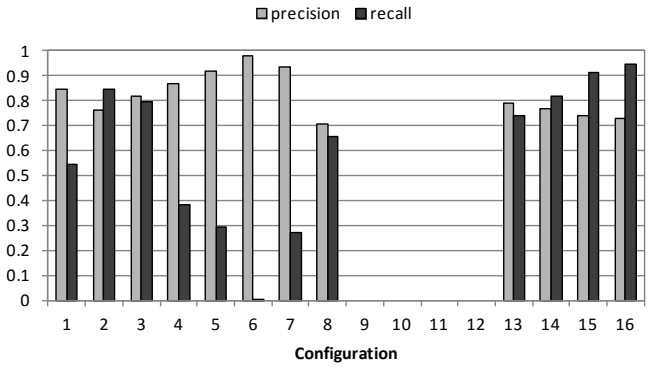

Figure 7. Precision and recall metrics for person No. 2, conditions C3

The results of *F*1 metrics for all configurations in all conditions are presented in Figure 8. The highest median values were observed for configurations 1-3 and 13-16.

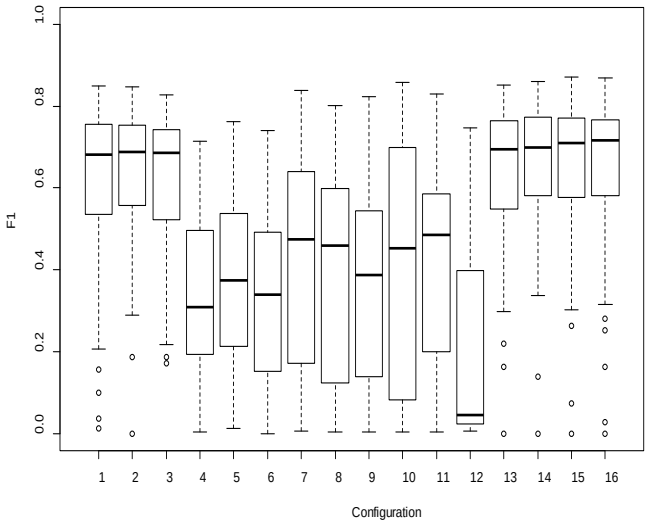

Figure 8. *F*1 metrics for all configurations in all conditions

Quality of detections at investigated bank operating rooms are presented in the form of a plot reflecting true positives to false positives ratio. Obtained ROC curves are depicted for each condition (C1-C4) in Figure 9 to Figure 12.

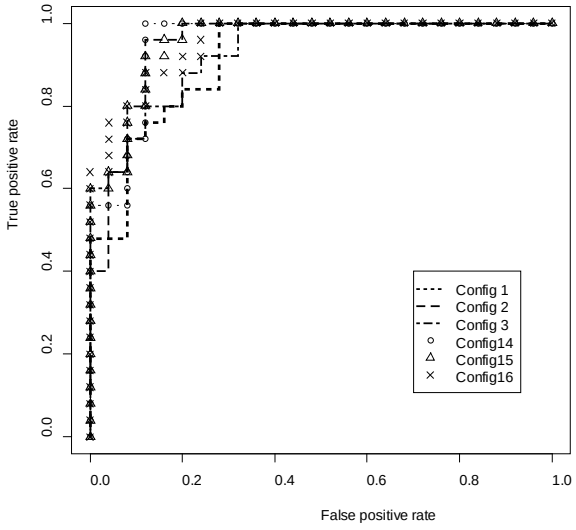

Figure 9. ROC curve for 6 best configurations, at the banking outlet C1

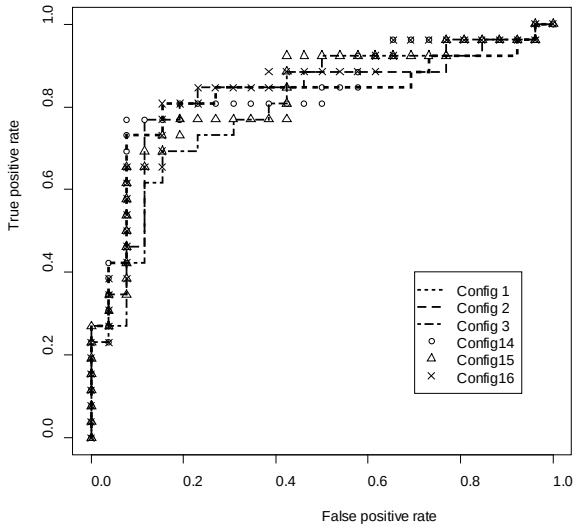

Figure 10. ROC curve for 6 best configurations, at the banking outlet C2

The worst detection rate, as expected, is observed for the banking outlet C4 as shown in Figure 12. The light conditions were unfavorable, and additional face illumination should be provided in that room for the proper system working.

In other difficult cases, i.e. untypical position of person, large skew of face, or critical proximity to the camera, the amount of properly detected faces might be small.

Similarly as in the previous observation [9], the group of last 4 configurations allow for the largest number of detected frames per video.

High precision and recall metrics values are observed for parameter configurations 13 to 16. *F*1 measure shows that the configurations 15 and 16 are the best ones in the case of the person 1 (*F*1 value of 0.79 for conditions C1 and 0.88 for C2). For the person 2 the highest *F*1 score was 0.82 for C3 at the configuration 16.

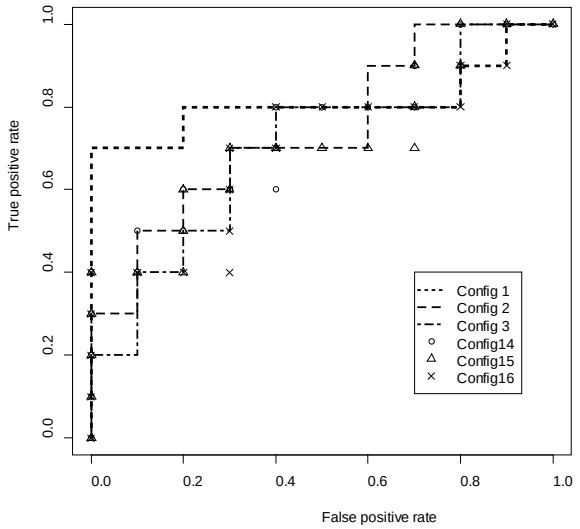

Figure 11. ROC curve for 6 best configurations, at the banking outlet C3

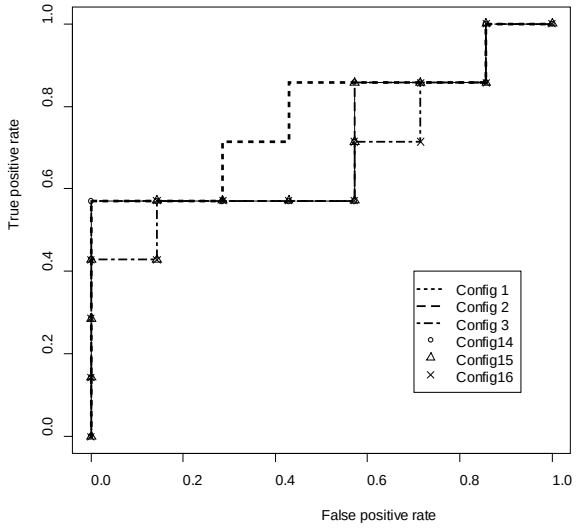

Figure 12. ROC curve for 6 best configurations, at the banking outlet C4

5. Conclusions

The results of experiments of face detection in real conditions met in bank operating rooms were presented in the paper. Obtained outcomes show that detection precision is good enough to make a vital part of the system for visual bank client verification. In real conditions many unpredictable factors have influence on the face image quality. The detection result is dependent on various locations of face, movements, positions of clients. Moreover, often we have to operate with face images which are different from the recommendations to be found in the norm ISO/IEC 19794-5.

The lack of necessary physical interaction with acquisition device increases the client comfort, however at the same time it makes detection or verification a more difficult in some cases.

Further improvement of the algorithm can include prediction of face position and its size in consecutive frames. Therefore, the future work will focus on the verification module included in the system.

Acknowledgment

This work was supported by the grant No. PBS3/B3/0/2014 Project ID 246459 entitled „Multimodal biometric system for bank client identity verification" funded by the Polish National Centre for Research and Development.

References

[1] Dalka P., Bratoszewski P., Czyzewski A., Visual lip contour detection for the purpose of speech recognition, *2014 International Conference on Signals and Electronic System* (*ICSES*), Poland, 2014.

[2] Górski G., System płatności mobilnych wykorzystujący biometryczną identyfikację użytkowników oraz infrastrukturę klucza publicznego, *PRZEGLĄD TELEKOMUNIKACYJNY* 8-9/2015 (in Polish).

[3] Hadid A., Heikkila J. Y., Silven O., Pietikainen M., Face and eye detection for person authentication in mobile phones, *2007 First ACM/IEEE International Conference on Distributed Smart Cameras*, Vienna, 2007, pp. 101-108.

[4] Laganiere R., *OpenCV Computer Vision Application Programming Cookbook*, Packt Publishing Ltd., 2014.

[5] Lupu C., Gaitan V. G., Lupu V., Security enhancement of internet banking applications by using multimodal biometrics, *IEEE 13th International Symposium on Applied Machine Intelligence and Informatics* (SAMI), 2015, Herl'any, 2015, pp. 47-52.

[6] Marciniak T., Chmielewska A., Weychan R., Parzych M., Dąbrowski A., Influence of low resolution of images on reliability of face detection and recognition, *Multimedia Tools and Applications*, **74**, 12, 4329–4349, 2015.

[7] Mariano V.Y. et al., Performance evaluation of object detection algorithms, *16th International Conference on Pattern Recognition, 2002. Proceedings*, vol. 3, pp. 965-969, 2002.
[8] Ren J., Kehtarnavaz N., Estevez L., Real-time optimization of Viola-Jones face detection for mobile platforms, *Circuits and Systems Workshop: System-on-Chip - Design, Applications, Integration, and Software*, 2008 IEEE Dallas, Dallas, TX, 2008, pp. 1-4.
[9] Szczodrak M., Czyżewski A., Face detection algorithms evaluation for the bank client verification; *Proc. Signal Processing, Algorithms, Architectures, Arrangements, and Applications*, pp. 186 - 190, Poznań, Polska, 21.9.2016 - 23.9.2016.
[10] Taigman Y., Yang M., Ranzato M., Wolf L., DeepFace: closing the gap to human-level performance in face verification, *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 1701-1708.
[11] Viola P., Jones M., Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001*. Vol. 1. IEEE, 2001.
[12] Vondrick C., Patterson D., Ramanan D., *Efficiently scaling up crowdsourced video annotation*, International Journal of Computer Vision (IJCV). June 2012.

Received 19.10.2016, Accepted 16.05.2017