

Liveness detection in remote biometrics based on gaze direction estimation

Krzysztof Adamiak, Dominik Żurek, Krzysztof Ślot
Lodz University of Technology
ul. Stefanowskiego 18/22, 90-924 Łódź, Poland
Email: kslot@p.lodz.pl

Abstract—The following paper presents a simple and fast liveness detection method based on gaze direction estimation under a challenge-response user authentication scenario. To estimate a line of sight, a procedure composed of several steps, including face and eye detection, derivation of gaze direction representation and subsequent classification, has been proposed. The proposed, novel gaze orientation descriptor is easy to compute and it provides sufficiently accurate estimates for the considered task. To assess a probability of genuine biometric trait presentation, recorded gaze direction responses induced by presentation of a randomly generated on-screen object, are matched against expected patterns.

I. INTRODUCTION

ONE OF the main threats that exist for unattended biometric authentication systems are so called 'presentation attacks', where a system is presented with a biometric artefact. The problem becomes especially severe when easy-to-spoof biometric traits are considered, such as e.g. fingerprints, face or iris images. To enable unattended (including remote, by means of popular mobile devices) user verification, biometric systems must cope with the stated problem. For this purpose, a methodology aimed at verification of biometric trait authenticity, referred to as liveness detection, has been developed.

Existing liveness detection approaches can be broadly categorized into two main groups: methods that exploit physiological properties of tissues and organs subject to analysis and various challenge-response schemes. Liveness detection is therefore trait-specific and many diverse ways for its assessment have been proposed so far. For example, several different tests are available for iris image authenticity verification, such as application of varying intensity illumination levels to check for iris physiological responses (this approach combines challenge-response scheme and physiology), analyzing presence of saccades or analyzing spectra of reflected light. Approaches considered for liveness testing in case of face-based recognition include blink detection [1], detection of eye-movements [2] detection of presence of facial expressions [3] or lip motion detection [4]. A natural means against presentation of photos provides 3D face recognition (however, it clearly becomes vulnerable if 3D masks are used). Fingerprint validity can be assessed by analyzing perspiration

processes. For some biometric traits, such as vein pattern structure, liveness detection is an intrinsic component of the adopted recognition methodology, as vein imaging (blood-flow imaging) is possible only for living tissues.

Widespread availability of mobile devices equipped with cameras and microphones brought up an interest in exploiting face images and voice in remote biometric authentication, despite their well-known limitations. One of the main problems that needs to be addressed in this scenario is clearly liveness detection. One of the most natural liveness detection schemes used in case of speaker recognition is a challenge-response procedure, where a speaker is prompted to utter some randomly generated text, and only after positive verification of the response, a biometric system proceeds with user-verification procedure. An application of a similar challenge-response scheme in video based face analysis have been proposed in [5], where a challenge requires a user to make voluntary blinks and mouth movements (opening and closing). Another interesting example of challenge-response scheme that utilizes gaze tracking and that is intended as a secure method for logging to computer systems has been proposed in [6]. A set of icons, which includes a randomly scattered subset of previously memorized ones, is displayed to a user, who is supposed to trace (with her/his eyes) a path, that defines a convex hull built upon known icons.

The presented paper proposes a simple liveness detection method that is based on verification of line-of-sight trajectory compliance (a response) with some expected pattern (a challenge). The pattern is defined by subsequent locations of a marker that gets displayed at random locations of a screen. The main element of the proposed method is a novel gaze direction estimation algorithm, which is computationally inexpensive, enabling its real-time application even on machines with limited computing power, such as mobile devices. Simplicity of the proposed gaze detection method results from its specific context: horizontal line-of-sight displacement evaluation is sufficient for execution of liveness detection procedure.

The proposed gaze detection method conforms to a general framework of the domain and includes two phases: derivation of eye-image representation that correlates well with gaze direction, followed by gaze direction evaluation. As widely available devices are considered for the method implementation, gaze tracking is performed using regular visible light cameras (the best performing gaze tracking

This work was supported by the Polish National Science Center under the research grant 2012/05/B/ST6/03647

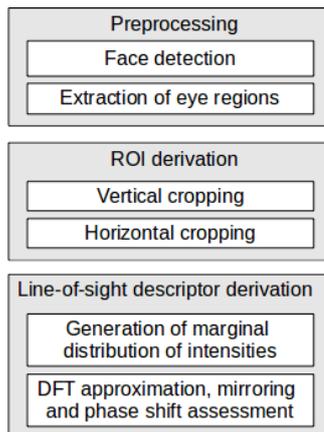


Fig. 1. Block diagram of gaze estimation procedure

methods exploit near infrared range and typically use infrared light sources [7]).

Several different eye image representations have been proposed for the purpose of gaze tracking applications. They belong to two broad categories: representations based on mutual locations of salient geometric eye features, such as inner and outer eye corners, iris/pupil centers, and appearance based representations (a broad review of relevant methods can be found e.g. in [8]). Gaze direction assessment exploiting features from the former group is basically a regression problem that can be solved using e.g. Support Vector Regression (SVR) [9] or neural networks [10]. For appearance-based descriptors, gaze direction assessment is made e.g. by computing between-region correlation coefficients [11] or using mean-shift algorithm [12]. The proposed algorithm for line-of-sight direction estimation is based both on salient feature detection and appearance-based eye modeling, and it is followed by regression based analysis, so it combines both of the presented general methodologies.

A structure of the presented paper is the following. The proposed gaze detection algorithm has been introduced in Section 2. Section 3 presents a background used for challenge-response liveness detection. Finally, experimental evaluation of the proposed method has been presented in Section 4.

II. GAZE ESTIMATION ALGORITHM

Block diagram of the proposed algorithm has been depicted in Fig. 1. A general idea of the proposed line-of-sight direction estimation is to confront information extracted from two eye images, where one of them is mirrored, so that deviations from a central fixation point (determined during a calibration phase of the procedure) get amplified.

The procedure begins with face detection, followed by eye-region detection, both performed using a well-known Viola-Jones algorithm [13]. The detected regions (containing the left and the right eye) are further refined by vertical and horizontal cropping, performed to increase processing speed and to facilitate subsequent analyses by eliminating complex yet unrelated structures, such as eyebrows. The resulting

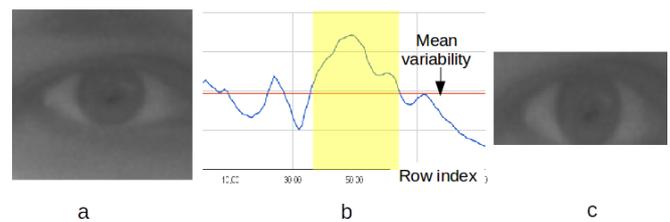


Fig. 2. Determination of vertical bounds for a rectangular image analysis window: an initial eye region produced by Viola-Jones algorithm (a), sample plot of horizontal intensity variance with the selected row interval (b) and the resulting analysis window (c)

regions of interest (ROI) become a domain for gaze direction assessment, which uses a quantitative descriptor that estimates eye disks horizontal offset from their 'neutral' position (forward gaze). The descriptor is derived from Fourier spectra of marginal distributions of vertical projections of eye image intensities and gaze direction is represented by a value of a phase shift between fundamental frequencies that approximate the considered functions. The presented approach is detailed in the following subsections.

A. ROI Derivation

An objective of vertical cropping of initial eye regions is elimination of irrelevant upper and lower image structures, such as e.g. eyebrows. To adjust initial eye regions vertically, we propose to analyze horizontal image variability. As eye images always contains regions of extreme intensities (white cornea versus black pupil), variability of image intensities along these rows is expected to dominate over the remaining regions. A criterion for selection of vertical bounds of a rectangular window that will be used for gaze analysis is based on analysis of horizontal variance projection function. We propose to extract from an initial eye region only the widest strip, composed of rows with gray-level variability above the average level, computed for the whole region (see Fig. 2).

An objective of horizontal cropping is to produce a normalized image analysis domain, where two landmarks: inner and outer eye-corners determine a system of reference. A use of eye corners as landmarks has important advantages (they are distinctive and separated by a fixed distance from each other, which offers a basis for pose estimation).

The proposed eye-corner detection method operates on vertically-cropped image eye-regions. To compensate for illumination variations, prior to further processing, eye region images are normalized in intensity. The algorithm begins with corner detection procedure, which seeks for image points with large contents variability. Two well-known methods that differ in the adopted decision criterion were examined to do the task: Harris [14] and Shi-Tomasi [15] detectors. Both methods were able to correctly detect all salient image points, including eye corners (see Fig. 3). However, as the Shi-Tomasi method favors features that are easier to track (this is an important aspect from the standpoint of computational efficiency of the algorithm), it has been selected in further analyses.

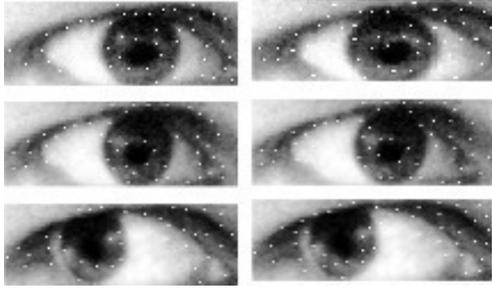


Fig. 3. Corner detection results in input images produced using Harris detector (left column) and Shi-Tomasi detector (right column).

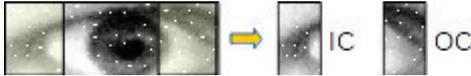


Fig. 4. Eye corner search domains (IC - inner corner search region, OC - outer corner search region).

As it can be seen from Fig. 3, corner detection procedure results in identification of a large set of salient points. To find eye corners, elements of this set are subject to subsequent analysis. Firstly, its domain gets restricted, so that inner and outer eye corners are sought only in feasible eye subregions, defined as boundary vertical bands of width set to 25% of eye image (Fig. 4).

To identify eye corners among a set of available salient points, a descriptor that summarizes image appearance within a square, 15×15 neighborhood around a salient point, has been generated. The neighborhood is divided into quadrants of size $r \times r$, $r = 8$ (with overlapping boundaries), and mean image intensity gradients are evaluated within each quadrant. A descriptor of a salient point $\mathbf{P}^{i,j}$, located in i -th row and j -th column of an image, is thus a collection of four vectors:

$$\mathbf{S}^{i,j} = [\mathbf{g}_{TL}^{ij}, \mathbf{g}_{TR}^{ij}, \mathbf{g}_{BL}^{ij}, \mathbf{g}_{BR}^{ij}] \quad (1)$$

where subscripts T,B,L,R label quadrants (top, bottom, left and right) and the mean gradient of pixel intensities $I(k,l)$ in a quadrant XY (\mathbf{g}_{XY}^{ij}) is given by:

$$\mathbf{g}_{XY}^{ij} = \frac{1}{r^2} \sum_{k,l \in XY} \nabla I(k,l) \quad (2)$$

Descriptor gradient vectors were finally normalized in length, so that they sum up to unity. Salient points are matched against eye-corner models. Four separate eye-corner models (inner and outer for the left and for the right eye), of the same structure as given by (1), were generated from a set of manually labeled images (five images per each corner - see Fig. 5). Descriptors derived for training images for particular eye corners were averaged, forming the corresponding eye-corner models $\mathbf{C}^{IL}, \mathbf{C}^{OL}, \mathbf{C}^{IR}, \mathbf{C}^{OR}$, where I,O,L,R denote inner, outer, left eye and right eye respectively. As it was the case for salient point descriptors, also eye-corner model gradients were analogously normalized in length. The derived models were matched against salient points that are present in

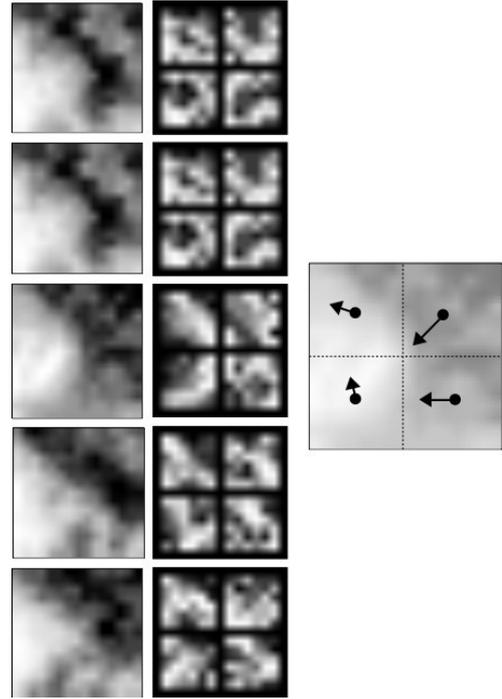


Fig. 5. Generation of eye-corner model: manually selected instances of inner left eye corner (left column), gradient magnitudes derived for quadrants (middle) and the resulting gradients (right).

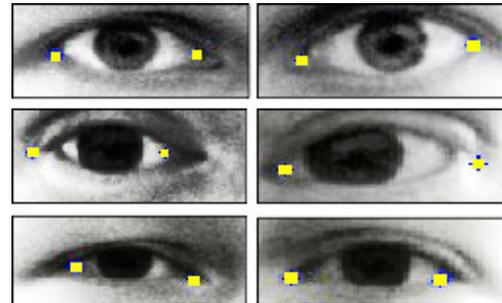


Fig. 6. Sample eye corner detection results performed for three different persons using the presented algorithm.

the corresponding eye-image bands (i.e. inner left eye-corner model was applied to salient points present in the 'IC' region of the left eye image etc.). A matching score was defined as a sum of dot products between components of a corner's α, β model descriptor and some considered salient point's descriptor:

$$F_{\alpha,\beta}^{ij} = \sum_{k=0}^3 \langle \mathbf{S}^{i,j}[k], \mathbf{C}^{\alpha,\beta}[k] \rangle \quad (3)$$

The score (3) gets maximized for salient point neighborhoods that match a particular eye-corner model. Sample results of eye-corner detection have been presented in Fig. 6.

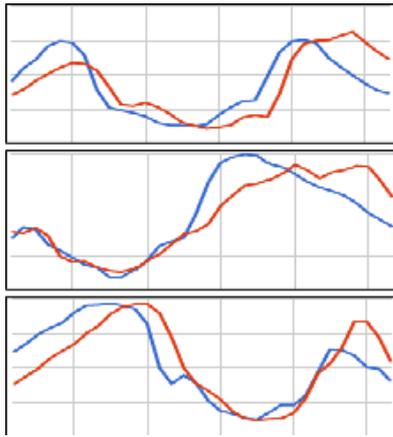


Fig. 7. Plots of image intensity distribution accumulated along a vertical axis for three different gaze directions: forward (top), left (middle) and right (bottom), for left (red) and right (blue) eyes.

B. Line-of-sight Direction Assessment

Regions of interest, derived for both eyes, provide domains that comprise information relevant for gaze direction assessment. To provide fast and accurate estimation of gaze direction, an appropriate descriptor that can be easily computed and that is robust against possible image artifacts, needs to be derived.

To meet the formulated goals, we decided to generate gaze direction descriptor based on differences in general appearance of left and right eye strips in horizontal direction. The appearance of a strip can be summarized using marginal distribution of strip pixel intensities accumulated in vertical direction. One can observe that a general shape of the resulting function is line-of-sight direction specific (see Fig. 7). Such a general shape can be easily quantified by using leading components of any of possible signal orthogonal transformations, such as e.g. Discrete Fourier Transform (DFT). DFT has been chosen as a basis for eye appearance representation and spectra of marginal distributions of vertically accumulated intensities, derived for different images (see Fig. 8), were analyzed. We found that the first periodic component appears to be an attractive means for summarizing eye appearance, as its phase shows good correlation with a gaze direction. Examples of approximation of marginal gray-level distributions by means of a fundamental component of its DFT decomposition have been shown in Fig. 9.

To amplify sensitivity of the representation, we decided to confront the approximation produced for marginal distribution derived for one of the eyes with an approximation produced for the mirrored marginal distribution of the other eye (see Fig. 9). Thus, a final descriptor of gaze direction is a phase difference between two first harmonics, where the first one approximates the marginal intensity distribution derived for the left eye and the second one approximates a mirrored distribution derived for the right eye.

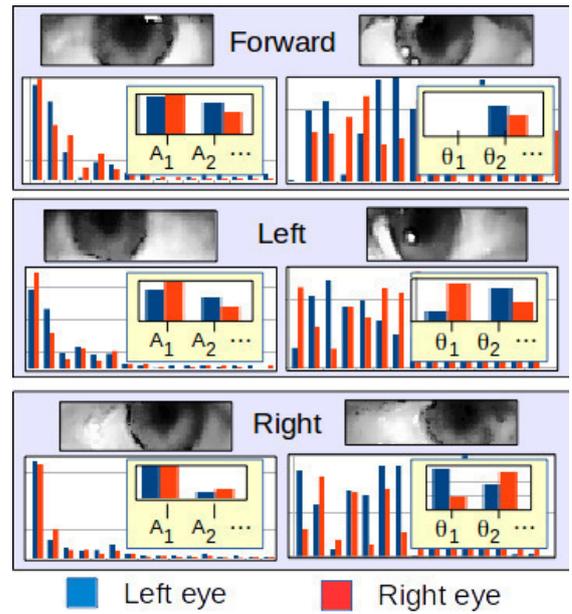


Fig. 8. Magnitude (left column) and phase (right column) spectra derived for marginal distributions of vertical image intensities for three gaze directions: forward (top), left (middle) and right (bottom) (magnitudes and phases of the two first periodic components have been enlarged for illustration clarity).

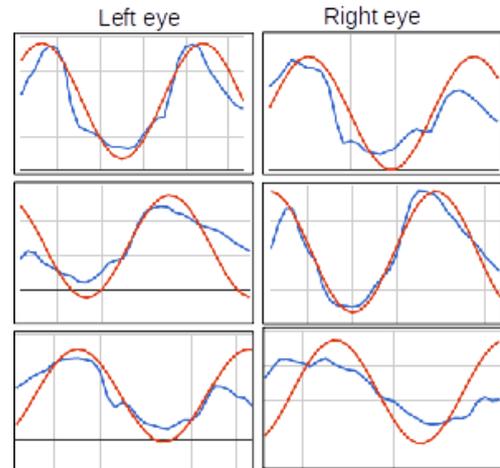


Fig. 9. Marginal distribution approximations using the first DFT harmonic for three different line-of-sight directions: forward (top), left (middle) and right (bottom).

III. LIVENESS DETECTION PROCEDURE

Challenge-response scheme has been used as a framework for liveness detection, where the challenge is an on-screen presentation of a marker (a circle) in time-varying locations and the expected response is a corresponding line-of-sight direction adjustment. Objects are presented at random locations and the system is attempting to determine the induced gaze direction, thus verifying a required reaction. If probability of correct gaze detection exceeds 0.5, one can expect that successive repetitions of the procedure will eventually provide

a required level of confidence that a user actually responds to a challenge. To derive a quantitative estimate of number of trials that are required for getting some predefined confidence level, we assume that the considered liveness detection scheme can be expressed in terms of a Bernoulli process. The main error in gaze direction estimation results from erroneous detection of eye corners, that can happen equally likely for any viewing angle, we assume that probabilities of correct gaze detection are the same, regardless of this angle. This justifies adoption of the Bernoulli scheme and allows to define the two outcomes required by the process in the following manner. A success occurs if a response (estimated line of sight) falls within an expected angular interval around actual position of a marker (we assume that an angular range of on-screen locations is evenly split into an even number of intervals). Otherwise, we consider that an outcome of an experiment is a failure. Given this framework, and given success and failure probabilities, we can estimate a number of challenge repetitions that is required to meet some predefined confidence level (or equivalently, liveness detection error probability). It can be shown [16] that probability of correct classification of at least $\frac{n+1}{2}$ elements of n -element sequence Q , given probability of correct entry classification $p > 0.5$:

$$p(Q, n) = \sum_{i=1}^{(n+1)/2} \binom{n}{\frac{n-1}{2} + i} p^{\frac{n-1}{2} + i} (1-p)^{\frac{n+1}{2} - i} \quad (4)$$

converges to unity as $n \rightarrow \infty$. Therefore, it is always possible to find such n that provides some desired confidence level T of an affirmative decision.

IV. EXPERIMENTAL EVALUATION OF THE PROCEDURE

The experimental setup used for proposed algorithm evaluation was the following. An application for generating a challenge draws a single marker at time-varying, seventeen equidistant discrete locations (a marker is a circle of a fixed size). Locations of the marker change every 2 seconds. During gaze direction detection accuracy tests marker locations were periodically updated in an oscillating manner. During liveness detection tests, marker locations were randomly selected. A user was situated in front of a screen at a fixed distance, so that marker is observed within a range of angles starting from -30 degrees to +30 degrees. A simple web camera, positioned centrally atop a screen was monitoring user's responses (challenge generation and image acquisition processes were synchronized). Each recorded frame was subject to a separate analysis (no object tracking mode was used, to provide more data for evaluation of all procedure steps).

Two databases were used throughout the experiments. The first one was prepared by the Authors and comprises 415 low resolution (640×480) test images of three subjects with manually labeled four eye corners and with labeled line-of-sight orientations. The second source of experimental material was a CAVE database [17]. 1176 high resolution ($5,184 \times 3,456$ pixels) images of 56 different persons, with known gaze directions, were used.

TABLE I
EYE CORNER DETECTION ACCURACY EVALUATION (COLUMN LABELS IDENTIFY A CORNER: OL - OUTER LEFT EYE CORNER, IL - INNER LEFT EYE CORNER, IR - INNER RIGHT EYE CORNER, OR - OUTER RIGHT EYE CORNER).

	All	OL	IL	IR	OR
Detected	96%	99%	100%	96%	96%
Identified	79.4%	86%	94%	93%	84%

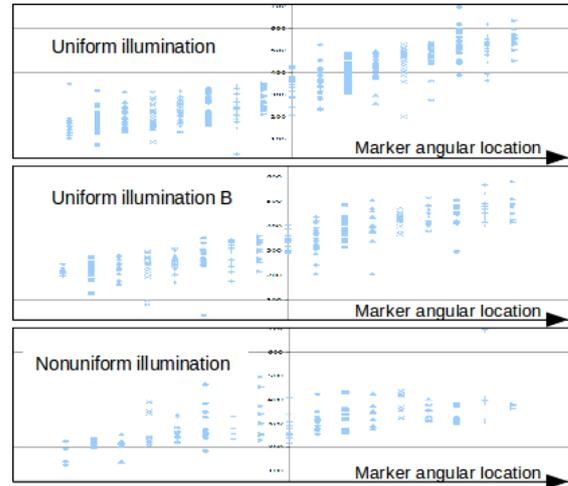


Fig. 10. Plots of estimated gaze directions with respect to actual ones for three different sets of experimental data.

The first phase of the evaluation was aimed at estimating accuracy of ROI derivation procedure. Results of eye-corner detection have been summarized in Table I. The first row shows performance of Shi-Tomasi algorithm - eye corner detection has been considered successful if any of produced salient points was sufficiently close to the considered landmark (within its 5×5 neighborhood). The second row of the table specifies performance of correct salient point identification. The first column of Table I indicates percentage of correct detection for all corners, whereas the remaining ones show scores for individual eye-corners.

The second set of experiments was concerned with evaluation of gaze detection accuracy. A set of three video sequences of a user asked to eye-track the oscillating marker were recorded and a functional relation between actual marker angular positions and positions calculated using the presented algorithm was derived. The sequences were differing in adopted illumination conditions: the first two sequences were taken under uniform illumination of different intensity, whereas in the third case a face was lit from aside. For every sequence a total of 510 frames were analyzed (an average of 30 frames per marker location). The results are summarized in Fig. 10, where plots show the computed marker locations against their actual locations. As it can be seen, there exist significant variations in gaze direction estimation, however one can identify angular intervals that can be exploited for liveness detection.

Given gaze-estimation results, it has been assumed that a marker will be displayed randomly at three different locations of a screen: two extreme positions and in the middle. An average probability of successful detection of gaze direction was evaluated to be $p \approx 0.66$. It follows from eqn. (4) that a number of required presentations of a marker necessary for obtaining a $T = 95\%$ level of confidence that a subject is actually following the marker equals 23.

The last part of experiments was concerned with evaluation of computational complexity of the proposed algorithm. The method was implemented in C++ programming language and executed on a desktop computer with i7, quad core processor, running at 2.4 GHz. The presented algorithm, excluding an initial phase of face detection and preliminary eye-region detection (both performed using Viola-Jones algorithm) took on average only 3 milliseconds to execute (the result was averaged for processing of 721 images). Although the aforementioned, initial preprocessing can be time consuming, one needs to note that it can be significantly accelerated if face tracking mode will be used for analysis of frames that follow the first one.

V. CONCLUSION

The proposed algorithm proves that liveness detection can be performed using line-of-sight estimation, by using a simple camera for image acquisition. Computational complexity of the procedure is low and we believe that it can be implemented on popular mobile device platforms. There exist several elements of the procedure that need to be explored to increase gaze-direction assessment accuracy, which is important to reduce a required duration of liveness detection procedure. The main directions of further exploration will be concerned with improving eye-corner identification performance (e.g. by applying multi-resolution analysis) and with modifying the adopted eye-image representation (e.g. by including more components of DFT decomposition of eye images).

ACKNOWLEDGMENT

The Authors wish to thank Dr Piotr Duch for his invaluable help in formulating the presented concepts.

REFERENCES

- [1] L. Sun, G. Pan, Z. Wu, S. Lao, "Blinking-Based Live Face Detection Using Conditional Random Fields," in *Int. Conf. on Biometrics*, pp. 252–260, 2007. <http://dx.doi.org/10.1007/978-3-540-74549-5-27>
- [2] Komogortsev, O. V.; Karpov, A.; Holland, C. D., "Attack of Mechanical Replicas: Liveness Detection With Eye Movements," *IEEE Transactions on Information Forensics and Security*, vol.10, no.4, pp. 716–725, 2015. <http://dx.doi.org/10.1109/TIFS.2015.2405345>
- [3] J. Li, Y. Wang, T. Tan, A.K. Jain, "Live face detection based on the analysis of Fourier spectra," in *SPIE Conf. on Biometric Technology for Human Identification*, vol. 5404, pp. 296–303, 2004. <http://dx.doi.org/10.1117/12.541955>
- [4] K. Kollreider, H. Fronthaler, M. I. Faraj, J. Bigun, "Real-time face detection and motion analysis with application in liveness assessment," *IEEE Transactions on Information Forensics and Security*, vol. 2 pp. 548–558, 2007. <http://dx.doi.org/10.1109/TIFS.2007.902037>
- [5] Singh, A. K.; Joshi, P.; Nandi, G. C., "Face recognition with liveness detection using eye and mouth movement," in *Int. Conf. on Signal Propagation and Computer Technology (ICSPCT)*, pp. 592–597, 2014. <http://dx.doi.org/10.1109/ICSPCT.2014.6884911>
- [6] D. Weinsshall, "Cognitive Authentication Schemes Safe Against Spyware," in *Proc. of IEEE Symposium on Security and Privacy*, pp. 300–306, 2006. <http://dx.doi.org/10.1109/SP.2006.10>
- [7] Ying Qi; Zhi-Liang Wang; Zhang Chuang, "A non-contact eye-gaze tracking system for human computer interaction," in *Int. Conf. Wavelet Analysis and Pattern Recognition*, vol.1, pp. 68–72, 2007. <http://dx.doi.org/10.1109/ICWAPR.2007.4420638>
- [8] D.W. Hansen, Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.32(3), pp. 478–500, 2010. <http://dx.doi.org/10.1109/TPAMI.2009.30>
- [9] Zhiwei Zhu; Qiang Ji; Bennett, K. P., "Nonlinear Eye Gaze Mapping Function Estimation via Support Vector Regression," in *Proc. of Int. Conf. on Pattern Recognition*, vol.1, pp. 1132–1135, 2006. <http://dx.doi.org/10.1109/ICPR.2006.864>
- [10] B.L. Nguyen, C. Tijus, F. Jouen, M. Molina, Y. Chahir, "Eye gaze tracking with free head movements using a single camera," in *Proc. of the 2010 Symposium on Information and Communication Technology*, pp. 108–113, 2010. <http://dx.doi.org/10.1145/1852611.1852632>
- [11] M. Betke, J. Gips, P. Fleming, "The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities", *IEEE Neural Systems and Rehabilitation Engineering* 10(1), pp. 1–10 (2002). <http://dx.doi.org/10.1109/TNSRE.2002.1021581>
- [12] T. Liu, C. Pang, "Eye-gaze Tracking Research Based on Image Processing", *Congress on Image and Signal Processing*, IEEE, pp. 176–180 (2008). <http://dx.doi.org/10.1109/CISP.2008.590>
- [13] P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol.1, pp. 511–518 2001. <http://dx.doi.org/10.1109/CVPR.2001.990517>
- [14] C. Harris, M. Stephens, "A combined corner and edge detector," in *Proc. of Fourth Alvey Vision Conference*, pp. 147–151, 1988. <http://dx.doi.org/10.5244/C.2.23>
- [15] Shi, C. Tomasi, "Good Features to Track," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 593–600, 1994. <http://dx.doi.org/10.1109/CVPR.1994.323794>
- [16] L. K. Hansen and P. Salamon, "Neural Networks Ensembles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 10, pp. 993–1001, 1990. <http://dx.doi.org/10.1109/34.58871>
- [17] B. A. Smith, Q. Yin, S. K. Feiner and S.K. Nayar, "Gaze Locking: Passive Eye Contact Detection for Human-Object Interaction", *ACM Symposium on User Interface Software and Technology (UIST)*, pp. 271–280, 2013.