A New Gaze Analysis Based
Soft-Biometric

Chiara Galdi[1], Michele Nappi[1], Daniel Riccio[2], Virginio Cantoni[3], and Marco Porta[3]

[1] Università degli Studi di Salerno,
via Ponte Don Melillo, 84084 Fisciano (Salerno) Italy
[2] Università degli Studi di Napoli Federico II,
via Cintia 21, 80126, Napoli
[3] Università degli Studi di Pavia,
via Ferrata 1, 27100 Pavia
chiara.galdi@gmail.com, mnappi@unisa.it, daniel.riccio@unina.it,
{virginio.cantoni,marco.porta}@unipv.it

Abstract. Soft Biometric traits are physical or behavioral human characteristics like skin color, eye color, gait, used by humans to distinguish their peers. However soft biometric characteristics lack in distinctiveness and permanence to identify an individual uniquely and reliably. In this paper a new Gaze Analysis based Soft-biometric (GAS) is investigated. The way an observer looks at a particular subject, was recorded with a remote eye tracker. Feature vectors were built for each observation and used for testing the system as a recognition system. The accuracy of the GAS system was assessed in terms of Receiving Operating Characteristic curves (ROC), Equal Error Rate (EER) and Cumulative Match Curve (CMC), and provided encouraging results.

Keywords: gaze analysis, soft biometrics, eye tracking.

1   Introduction

Most studies related to soft biometrics have to date considered the iris as a potential predictor for ethnicity and gender [1]. For example, Qiu et al. have used specific features in the texture of the iris to determine ethnicity—Asian vs. Caucasian [2]. A similar approach has been used to determine gender by Thomas et al. [3]. More recently, the problem of predicting both gender and ethnicity using iris patterns has been addressed by Lagree et al. in 2011 [9]. While undoubtedly related to soft biometrics, however, these studies do not actually aim at identification, but rather at recognizing a single character. On the other hand, in [10], iris color is actually used as a soft biometry for indexing, in order to obtain a subset of identities where actual recognition will be performed.

While all the above cited works relate to static aspects of the human eye, eye movements can also be exploited to infer biometrically significant anatomical characteristics of the oculomotor plant [11]. This paper, too, examines a dynamic aspect, with the purpose to assess the relevance of eye movement patterns as a soft biometry [12].

Reliable data about eye movements can be obtained through an eye tracker, that is a device (similar to an ordinary LCD monitor in its appearance) able to detect the user's gaze, usually when looking at a screen [4]. Eye movements occur as sudden, almost instantaneous, "saccades" (lasting less than 100 ms), followed by "fixation" periods of about 100-600 ms, during which the eye is almost still. According to the so-called "Eye-Mind Hypothesis" [5], there is a direct correspondence between the user's gaze and his or her point of attention. As demonstrated in some experiments [6], while it is possible to move one's attention without shifting the gaze, it is not possible to move one's gaze without shifting the attention, which is therefore strictly correlated to eye behavior. In particular, the vision process can occur both *overtly* and *covertly* [7], and it is just this last "vision modality" that is strictly related to a person's cognitive and psychological processes. Potentially, proper eye parameters can thus be exploited for a soft biometry which identifies, if not always the specific individual, at least groups of possible persons or classes of cognitive and emotional states.

The "scanpath" (i.e. the sequence of fixations acquired by an eye tracker during a vision process) obtained from the observation of images or other kinds of content in both overt and covert vision processes is one of the main sources of information for biometrics based on eye behavior. For example, simple forms of explicit authentication require the user to fixate some areas of the screen in sequence, thus creating a sort of visual "PIN". Implicit authentication, in which the user can watch wherever he or she wants on the screen (for example when observing a photograph), is much more complex. Although it is very difficult to find one-to-one correspondences between eye data and the subject, soft biometrics is however possible (see for instance [8]). The focus of this paper is specifically on verifying the conjecture that the way an individual looks at an image—specifically, at a head shot—might be a personal distinctive feature, albeit possibly a weak one. In particular, a new Gaze Analysis based Soft-biometric (GAS) is investigated.

2   Data Acquisition and Processing

A specific device, namely Tobii 1750 remote eye tracker, was used for data acquisition. It integrates all necessary components (camera, infrared lighting, etc.) into a 17'' LCD monitor (1280×1024 resolution). Data have been processed in order to have compact representations easier to compare with each other. Images employed in the experiment as stimuli, represent human faces. Each headshot has been segmented into significant areas of interest (AOIs). In order to attempt identification, a metric is needed to quantify distance between vectors. Even if several metrics have been investigated (Euclidean, Jaccard, Spearman, Cityblock, Cosine, and Correlation), we only report results for the Euclidean and Cosine distances, as they provided the best recognition accuracy.

2.1   Data Acquisition

In the Tobii device, five NIR-LED (Near Infra-Red Light Emitting Diodes) light eyes up producing reflection patterns. An image sensor records pupil position and

corneal-reflections to determine eyes position and the gaze point. With an accuracy of 0.5 degrees and relatively high freedom of movements, the system is ideal for real-use settings, where it would be intolerable to constrain users too much in their activities. For correct use of the system, at least one eye (better if both) must stay within the field of view of the infrared camera, which can be represented as a box with size 20×15×20 cm placed about 60 cm from the screen.

The sampling frequency of the device is 50 Hz (that is, the user's gaze coordinates are acquired 50 times a second). The Tobii ClearView gaze recording software was employed to define stimuli (still images, slideshows, videos, etc. to be presented to the subject), record and manipulate gaze data. The Tobii ClearView software acquires 50 raw gaze coordinates per second, from which fixations are then obtained (characterized by coordinates, duration and timestamps). For the purpose of our experiments, a fixation was considered a sequence of successive samplings detected within a circle with a 30 pixel radius, for a minimum duration of 100 ms. The ClearView software also allows to obtain two useful graphical depictions, namely *gazeplots* and *hotspots*. While a gazeplot displays the sequence of fixations of a user, in the form of circles with areas proportional to fixation times, a hotspot uses color codes to highlight those screen portions in which there are high concentrations of fixations – and consequently have been watched most. Gazeplot circles are numbered, thus clearly indicating the fixation order. A further output of the eye tracker is the *gaze replay*, which dynamically shows the evolution of fixations and *saccades* (fast eye movements) with time. A system output example is given in Figure 1.
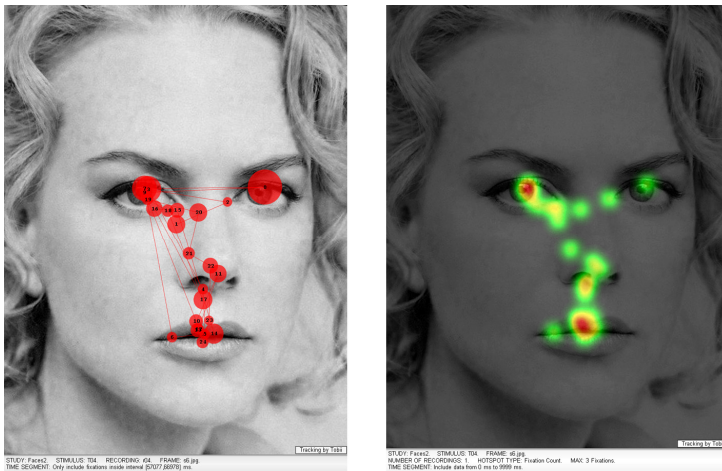

Fig. 1. A system output example: a gazeplot on the left and the corresponding hotspots on the right

2.2   Feature Extraction

As input images used as stimuli are all faces, it has been possible to define a limited set of AOIs which are always the same for all the observed photos, even if their

locations and dimensions can lightly change depending on the specific image. For this reason, GazePoint coordinates have been normalized and quantized according to the AOIs. Doing so, the size of the feature vector is fixed in advance and equals the number of AOIs. In particular, each AOI is assigned an element into the feature vector, which represents the total time spent looking at that AOI. It is worth noting that, at present, our feature vector does not take into account the number of visits for each AOI. If an observer spends some time on an AOI, quits and goes back on it later, only the total amount of time spent by the observer on that AOI is considered and computed by totaling up durations of both observations. Future enhancement of GAS strategy may involve the number of visits as an additional information to better characterize the observer behavior.

Figure 2 shows an example of a face sketch together with its AOIs marked with different colors and labeled with integer values.
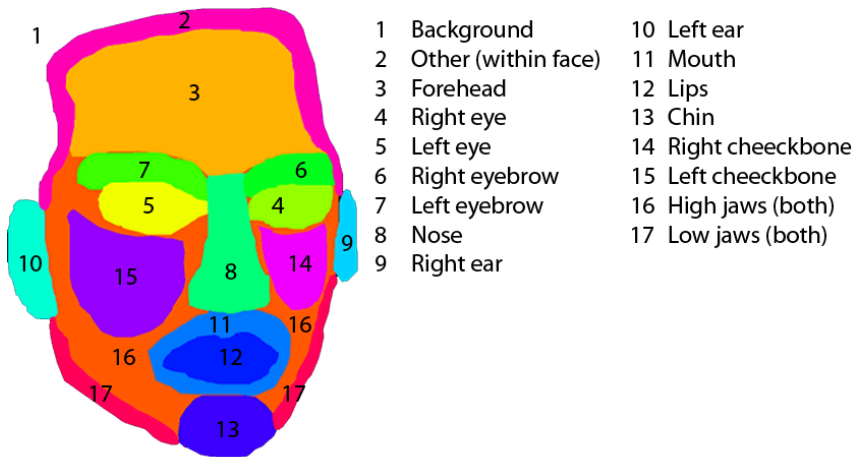

1	Background	10 Left ear
2	Other (within face)	11 Mouth
3	Forehead	12 Lips
4	Right eye	13 Chin
5	Left eye	14 Right cheeckbone
6	Right eyebrow	15 Left cheeckbone
7	Left eyebrow	16 High jaws (both)
8	Nose	17 Low jaws (both)
9	Right ear	

Fig. 2. A graphical sketch of the AOI mask

Doing so, feature vectors show a conveniently uniform representation that can be summed over observers as well as over subjects. That allows the fixation behavior of an observer to be modeled by averaging his or her feature vectors extracted from given set of face images used as training. In other words, the average feature vector over all training face images provides a description of the way that observer looks. In Figure 3, 16 feature vectors with a length of 17 AOIs are depicted for one of the observer involved in our experimentation (Observer 6). The bold line represents the average duration of fixations on each AOI computed over all 16 feature vectors. Different fusion strategies have been also investigated such as maximum, minimum and product rule; however the average value provided the best recognition accuracy in our experimental sessions.
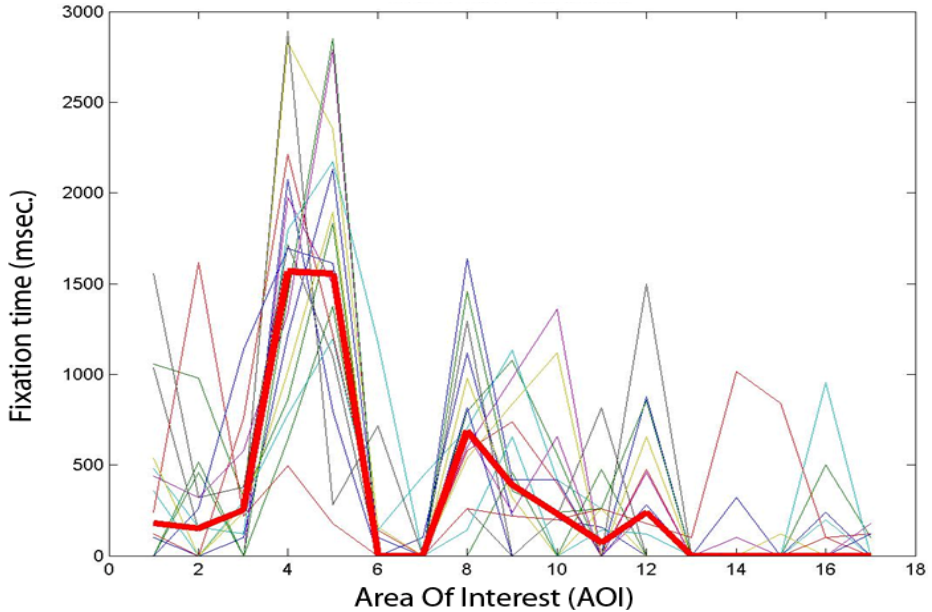
Fig. 3. Plots of feature vectors extracted from the gaze analysis of a given observer. Each vector refers to the observation of a single face image, while the (red) bold line represents the corresponding fixation model.

2.3   Matching

A small pool of face images that have been selected as training set is submitted to each observer and his/her fixation model is computed by averaging all corresponding feature vectors. A new observer is enrolled into the system by computing his/her fixation model, while the collection of all fixation models (one for each enrolled observer) constitutes the system gallery. When an observer asks for an authentication operation, a face image (different from that used for training) is submitted to him/her and the corresponding feature vector is then extracted. GAS computes the Euclidean distance between this vector and all fixation models into the system gallery. If the minimum distance is less than a fixed threshold δ, the system finds a match and the observer is associated with the identity corresponding to the fixation model that provided this minimum value. Otherwise the user is rejected by the system.

3   The Experimental Protocol

A total of 88 volunteer testers (59 males and 29 females) took part in the trials, subdivided into the following age groups: 17-18 (11 testers), 21-30 (39), 31-40 (7), 41-50 (15), 51-60 (7), 61-70 (8) and 71-80 (1). All participants reported normal or corrected-to-normal vision.

Prior to the beginning of the experiment, the observers were informed about the fact that some images, without specifying their kind, would appear on the eye trackers

display, interleaved with blank white screens with a small cross at their center. Each test was preceded by a short and simple calibration procedure, lasting about 10 seconds and consisting in following a moving circle on the screen. Participants were then instructed to look at the cross when the blank screen was displayed, and to freely watch wherever they wanted when the images were presented. The first blank screen was displayed for five seconds, while the others for three seconds. Each image was shown for ten seconds. 16 black-and-white pictures were used for the experiment, they contained close-up faces of 8 males and 8 females. Half of the faces (4 males and 4 females) were of famous persons (mostly actors and actresses), while the others were of people unknown to the testers.

The order of presentation of the 16 images was random. Behind the eye tracker there was a wall painted in neutral gray and the illumination of the room was uniform and constant. The tests were carried out in a quiet environment. On average, a single test session, including task explanation, device calibration etc., lasted a little more than five minutes.

36 of the 88 participants were involved in a second test session performed after the first one (with the same images), and 16 other participants were also involved in a third test session. The total number of tests carried out in the three sessions was therefore 140. Time intervals between the first and the second session, and between the second and the third session, ranged from a minimum of 5 days to a maximum of 9 days. As mentioned before, to compare observations Cosine and Euclidean metrics were used.

4   Experimental Results

The accuracy of the GAS system has been assessed in terms of Receiving Operating Characteristic curve (ROC), Equal Error Rate (EER) and Cumulative Match Curve (CMC). The ROC is a curve relating the Genuine Acceptance Rate and False Accepting Rate according to an acceptance threshold δ varying in the range [0,1]. The Equal Error Rate represents a sort of steady state for the system, as it corresponds to the point where False Acceptance Rate equals False Recognition Rate. The Cumulative Match Score at a rank n of a biometric identification system represents the likelihood that the correct identity is returned by the system among its top n answers. Thus the CMC is a curve representing the CMS with the rank ranging from 1 to N, where N is the number of enrolled subjects into the system gallery.

In all the following experiments, data acquired during the first session were used to enroll the observers, while observations captured during second and third sessions served as testing set.

In the first experiment only one feature vector has been used to build the fixation model of each observer, that is the fixation model coincides exactly with that feature vector. In particular, for each of the 16 face images in turn, the observation coming from the first session was considered as fixation model, while those acquired during second and third sessions were used as testing query. For each observer, results have been averaged over all 16 face images. Figure 4, shows the CMCs and ROCs obtained by averaging results over all the observers.
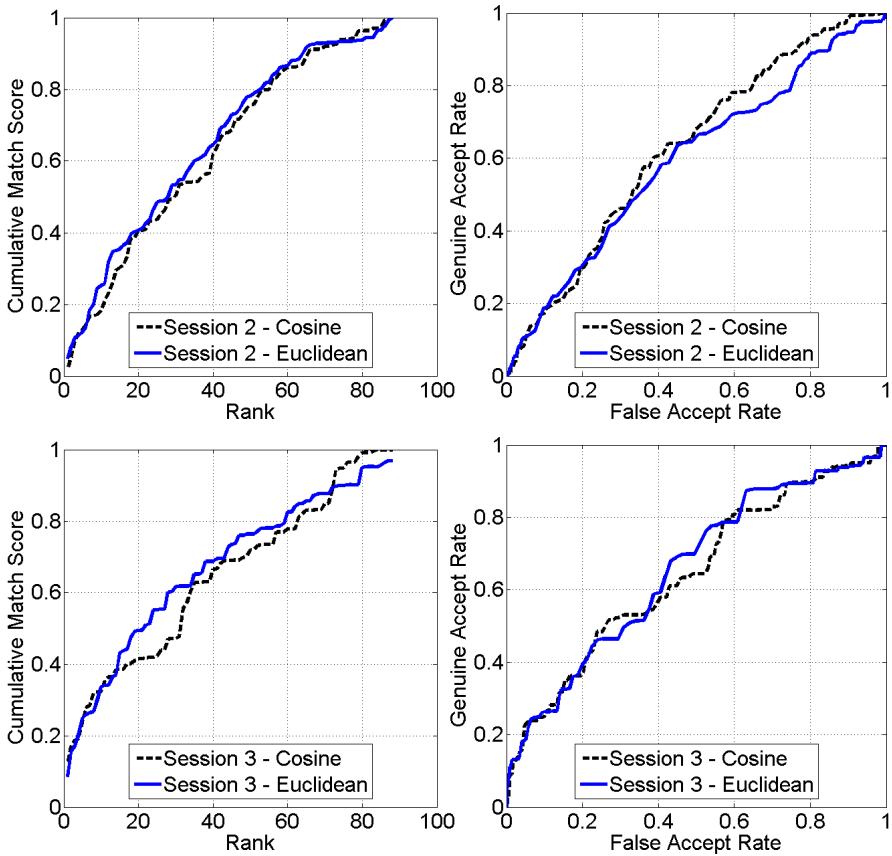
Fig. 4. CMC and ROC curves of the GAS system when only one feature vector is used to build the fixation model and the Euclidean/Cosine distance is used for matching

Looking at Curves in Figure 4 (upper left) and (upper right), it seems that Euclidean and Cosine distances reach comparable results. On the contrary, Figure 4 (lower left) and (lower right) show a small improvement in both CMC and ROC when the Euclidean distance is used. Session 2 and 3 mainly differ in the time gap from session 1 which was used to enroll the observers, thus pointing out the Euclidean distance to be more robust with respect to changes of feature vectors over time. This is also confirmed when looking at Equal Error Rates of the GAS system that are 0.404 and 0.415 for the Cosine and Euclidean distance respectively, when session 2 is used as probe and 0.432 and 0.355 when testing GAS with session 3. EERs also underline an improvement when using the Euclidean distance on session 3. For this reason, only the Euclidean distance is considered for next experiments.

In the second experiment, the fixation model of each user was built by using all 16 observations coming from the first session, while testing was performed by considering in turn one of the observations acquired during sessions 2 and 3

separately. Even in this case, results have been averaged over all face images belonging to the same session. Figure 5 reports the performance of GAS on session 2 and 3 when the Euclidean distance is used for matching.
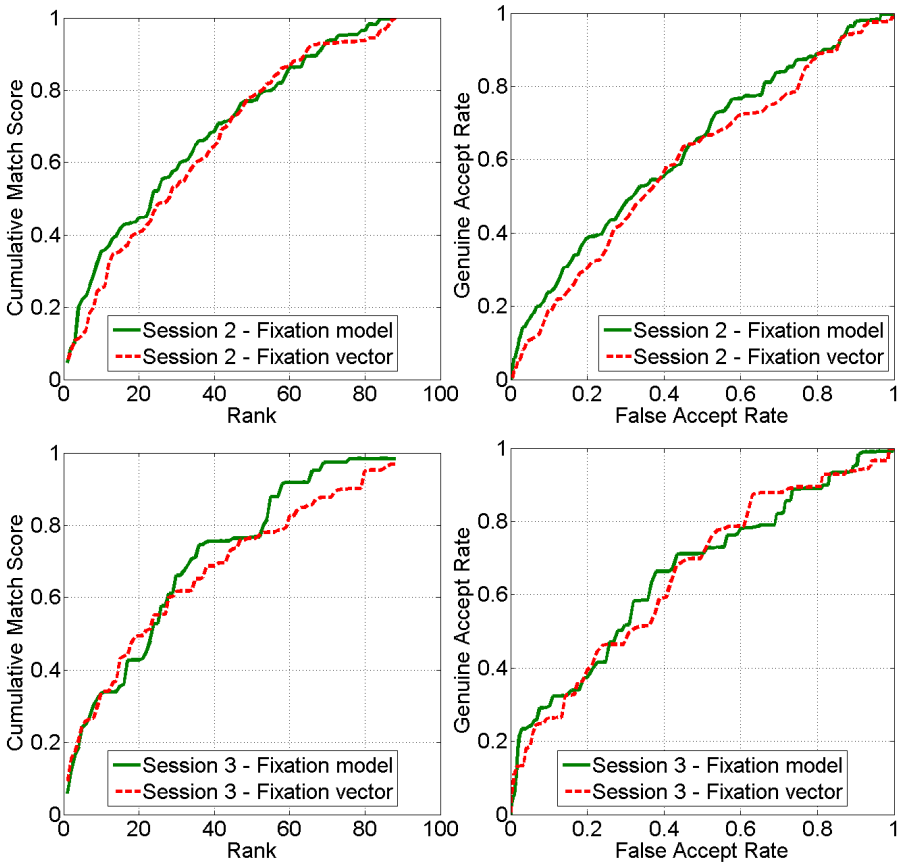

Fig. 5. CMC and ROC curves of the GAS system on session 2 and 3 when sixteen feature vectors (session 1) were used to build the fixation model and the Euclidean distance was used for matching

Figure 5 underlines that curves obtained exploiting the fixation model overcome those provided by a single feature vector in all cases. This is particularly noticeable for session 3 for which EER drops down from 0.408 (single feature vector) to 0.361 (fixation model). In other words, the fixation model better describes the way an observer looks a face image over time, since it is computed as the average of a pool of feature vectors that were used for training.

5   Conclusions

In this paper, the way an observer looks at a face image has been investigated as a potential new soft biometric trait by designing a new Gaze Analysis based

Soft-biometric system, namely GAS. Results obtained in terms of recognition accuracy confirm the feasibility of this line of research, even if many ways of improving the GAS architecture still remain to be tread. Increasing the number of AOIs by including some kind of "extra" points of interest (birthmarks, accessories, wrinkles) can take into account for further regions that are occasionally relevant in attracting the observer attention. The process of creation a mask of AOIs may also be automated by using a Facial Feature Point Detection algorithm, so that each fixation will be assigned with the closest Facial Feature Point. Last but not least, information coming from the temporal sequence of scanpaths and recurrent points of eye rest can be considered as additional features.

References

1. Porta, M., Ricotti, S., Jimenez Perez, C.: Emotional E-Learning through Eye Tracking. In: Proc. of the 2012 IEEE International Conference on Collaborative Learning, Marrakesh, Morocco, pp. 1–6 (2012)
2. Cantoni, V., Jimenez Perez, C., Porta, M., Ricotti, S.: Exploiting Eye Tracking in Advanced E-Learning Systems. In: Proceedings of the 13th International Conference on Computer Systems and Technologies (CompSysTech 2012), Rousse, Bulgaria, pp. 376–383 (2012)
3. Perego, E., Del Missier, F., Porta, M., Mosconi, M.: The Cognitive Effectiveness of Subtitle Processing. Media Psychology 13(3), 243–272 (2010)
4. Duchowski, A.T.: Eye Tracking Methodology – Theory and Practice, 2nd edn. Springer, London (2007)
5. Just, M.A., Carpenter, P.A.: Eye Fixations and Cognitive Processes. Cognitive Psychology 8, 441–480 (1976)
6. Poole, A., Ball, L.J.: Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In: Gahoui, C. (ed.) Encyclopedia of Human-Computer Interaction, pp. 211–219. Idea Group (2006)
7. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
8. Fookes, C., Maeder, A., Sridharan, S., Mamic, G.: Gaze Based Personal Identification. In: Wang, L., Geng, X. (eds.) Behavioral Biometrics for Human Identification: Intelligent Applications. IGI Global, Hershey (2010)
9. Lagree, S., Bowyer, K.: Predicting ethnicity and gender from iris texture. In: 2011 IEEE International Conference on Technologies for Homeland Security (HST), pp. 440–445 (November 2011)
10. Dantcheva, A., Erdogmus, N., Dugelay, J.-L.: On the reliability of eye color as a soft biometric trait. In: 2011 IEEE Workshop on Applications of Computer Vision (WACV), pp. 227–231 (January 2011)
11. Komogortsev, O.V., Karpov, A., Price, L., Aragon, C.: Biometric authentication via oculomotor plant characteristic. In: Proceedings of the IEEE/IARP International Conference on Biometrics (ICB), pp. 1–8 (2012)
12. Jain, A.K., Dass, S.C., Nandakumar, K.: Soft biometric traits for personal recognition systems. In: Proceedings of International Conference on Biometric Authentication, Hong Kong, pp. 731–738 (2004)