GANT: Gaze analysis technique for human identification

Virginio Cantoni [a], Chiara Galdi [b,*], Michele Nappi [b], Marco Porta [a], Daniel Riccio [c]

[a] Università degli Studi di Pavia, via Ferrata 1, 27100 Pavia, Italy
[b] Università degli Studi di Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (Salerno), Italy
[c] Università degli Studi di Napoli Federico II, via Cintia 21, 80126 Napoli, Italy

ABSTRACT

Anatomical biometric recognition is widely used in a large number of civilian and government applications, within well-tested biometric parameters. New sensors and matching algorithms have led to the deployment of soft biometrics, which may provide a fast and reliable identity finding procedure. These traits are physical or behavioral human characteristics like skin color, eye color, and gait, used by humans to recognize their peers, presenting distinctiveness and permanence to identify an individual uniquely and reliably. This paper regards a novel Gaze ANalysis Technique, namely GANT, exploiting a graph-based representation of fixation points obtained by an eye tracker during human computer interaction. The main goal is to demonstrate the conjecture that the way an individual looks at an image might be a personal distinctive feature, i.e. a soft biometric application. A novel dataset acquired through the Tobii 1750 remote eye tracker has been used to demonstrate GANT accuracy in soft biometry, in terms of Receiver Operating Characteristic Curve (ROC), Equal Error Rate (EER) and Cumulative Match Curve (CMC).

1. Introduction

Eye-based biometrics has received growing attention in the last few years. On the one hand, this may be due to the increasing availability of *eye trackers*, i.e. devices able to measure eye data such as gaze direction and pupil size. On the other hand, eye features and behaviors are more and more regarded as potentially safer authentication methods, especially when used in combination with traditional identity verification techniques.

In fact, most eye-based approaches pertain to the so-called "soft biometrics" category [1]. This means that, rather than finding a one-to-one matching between certain eye characteristics and a subject, these methods can provide a probability that specific features are associated to a certain person. Employed together with usual authentication solutions, such as those exploiting PINs or passwords, soft biometrics can increase security while not burdening the user with additional difficult-to-remember secret codes.

To date, the iris is probably the most studied eye feature for soft biometrics [2], used as a potential predictor for ethnicity (e.g. [3]), gender identification (e.g. [4]) or to obtain a small subset of identities where actual recognition will be performed (e.g. [5,6]). However, while iris-based identification techniques analyze static aspects of the human eye, an alternative approach is to use information connected to eye movements (measured through an eye tracker), to infer biometrically significant anatomical characteristics of the oculomotor plant [7].

Eye tracking technology has greatly improved in the last few years [8], and it is now possible to reliably find the user's gaze direction (typically on a screen) using devices almost indistinguishable from ordinary LCD monitors. Embedding gaze detection functionalities even into portable devices such as laptops and tablets (which has already been done—see for instance [9,10]), can enable the exploitation of eye movement data, making the so-called gaze-based identification a potentially ubiquitous biometric approach.

Eye movements occur as very fast, almost instantaneous, *saccades* (whose duration is usually less than 100 ms), alternated to *fixation* periods of about 100–600 ms (characterized by a relative stability of the eye). These movements usually take place in response to specific stimuli or mental processes. For example, research has been carried out to discover relationships between eye behaviors and emotional states (e.g. in the context of e-learning [11,12]) or to analyze the cognitive processes in a variety of tasks (e.g. in reading subtitles [13]). The so-called *Eye-Mind Hypothesis* [14], in particular, states that there is a direct correspondence between the user's gaze and his or her point of attention. Some experiments [15] have indeed demonstrated that while it is possible to move one's attention without shifting the

* Corresponding author. Tel.: +39 089 963345; fax: +39 089 963303.
E-mail addresses: virginio.cantoni@unipv.it (V. Cantoni), cgaldi@unisa.it (C. Galdi), mnappi@unisa.it (M. Nappi), marco.porta@unipv.it (M. Porta), daniel.riccio@unina.it (D. Riccio).

http://dx.doi.org/10.1016/j.patcog.2014.02.017

gaze, the opposite is more difficult. According to Hoffman [16], on the other hand, visual attention is always a little (100–250 ms) ahead of the eye, and when attention moves to a new position, the gaze is also moved there [17]. Apart from the different attention theories developed to date, however, what is certain is that attention is strictly correlated to eye behavior.

The vision process can occur both *overtly* and *covertly*, and it is just this last "vision modality" that is strictly connected to a person's cognitive and psychological processes [18]. Potentially, proper eye parameters can thus be exploited for soft biometry, at least to identify groups of possible persons or classes of cognitive and emotional states.

A *scanpath* is the sequence of fixations detected by an eye tracker during both overt and covert vision processes, and is also one of the main sources of information for eye-based biometrics. For instance, some kinds of authentication require the user to explicitly fixate specific screen areas in sequence, while others allow free observations of the displayed stimulus, such as a photograph or a video [19].

In this paper we examine dynamic aspects of eye behaviors, with the purpose to assess the relevance of eye movement patterns as a soft biometry. The purpose of our paper is specifically to verify the conjecture that the way an individual looks at an image—in particular, at a face—might be a personal distinctive feature, albeit possibly a weak one.

Another important aspect of this work is the creation of a database suitable for gaze analysis. It is the first of this kind and it is composed of the fixation point sequences recorded from the eye movements of individuals while they were looking at different images. As images used are all faces, with the same basic set of features, in the same canonical geometry, we believe that the differences between different observations were limited – because all people usually look approximately the same areas of the face, even if in a slightly different way – and we expect that using images very different from each other (such as landscapes and abstract images) will improve the recognition rate of the system.

The paper is structured as follows. Section 2 will briefly describe some representative works that exploit eye characteristics and activities for biometric purposes. Section 3 will describe the database composition and the methods used for feature extraction and matching. Section 4 illustrates the experimental protocol and results. Finally in Section 5 we present our conclusions, Section 6 lists the bibliography.

2. Related works

As said in the Introduction, different eye features and behaviors can be employed to implement biometric systems.

In a relatively old work, Kasprowski and Ober [20] analyzed eye movements (the gaze point coordinates) of subjects while following a jumping point on the screen. After collecting data, a set of features were extracted and analyzed through the Cepstrum technique (the inverse Fourier transform of the logarithm of the power spectrum of a signal). Different classifying algorithms were then used, namely k-Nearest Neighbor, Naïve Bayes, Decision Tree and Support Vector Machines. Bednarik at al. [21] exploited various kinds of eye data, including pupil sizes and their dynamics, gaze speeds and distances of infrared reflections on the eyes. After building a feature vector, its dimension was reduced using the Fast Fourier Transform and Principal Component Analysis techniques. Tests included different tasks, namely text reading, tracking of a moving cross and watching a static gray-level image. Deravi and Guness [22] recorded gaze data of testers while looking at some images for about 5 s each. Gaze durations, pupil positions, pupil sizes and gaze points were measured and then analyzed using the Forward Feature Selection, Backwards Feature Selection and Branch and Bound Feature Selection algorithms. Holland and Komogortsev [23] assessed the effects of eye tracking specifications and stimulus presentation on the biometric feasibility of complex eye movement patterns. Through two experiments, they examined the effects of varied stimulus type and varied spatial accuracy and temporal resolution. A third test was carried out using a low-cost eye tracker for cross-validation purposes. The authors found that, for biometric purposes, eye trackers with spatial accuracy of less than 0.5° and a sampling frequency greater than 250 Hz are recommended. The combination of eye behaviors and iris structure can of course provide better user recognition rates, although current eye trackers may not be fully suitable for detailed iris analysis. For instance, Komogortsev et al. [24] propose a biometric approach that exploits three different eye features, specifically the eye anatomical properties (represented by Oculomotor Plant Characteristics), the visual attention strategies (represented by Complex Eye Movement patterns) and the physical structure of the iris.

Like the study described in the present paper, the work by Rigas et al. [25] focuses on the free observation of face images. In each one of the eight test sessions carried out, participants watched 10 photos depicting human faces for 4 s each. Gaze positions were directly used to build eye trajectories. The similarity of spatial distributions of fixation points was quantified by means of a graph theoretic measure based on the multivariate generalization of Wald–Wolfowitz runs test. The obtained results indicate the existence of characteristic patterns that can be potentially exploited to discriminate persons.

Instead of static images, other investigations consider video-clips as stimuli. For example, Kinnunen et al. [26] propose a task-independent scenario in which short-term eye gaze direction is used to build feature vectors modeled by means of Gaussian mixtures. Eye movements, recorded while watching a 25 min long video, are described as a histogram of all angles the eye travels during a certain period. Liang et al. [27] present a video-based biometric identification model in which visual attention features, such as acceleration, geometric, and muscle properties, are obtained from gaze data and employed as biometric traits to recognize people. Experiments were carried out using a Back-Propagation neural network and a Support Vector Machine. Results showed that measuring video-based eye tracking data is a viable solution for biometric applications.

Eye tracking approaches are also interesting because they can allow the implementation of "intrusion detection" methods aimed at finding potential user behaviors different from usual: substantial discrepancies in eye activities may suggest that the current user is not the habitual one, thus alerting the system. For example, Holland and Komogortsev [28] present a variety of eye movement-based biometric features and their capability to correctly differentiate single persons. In particular, they investigate the scanpaths deriving from reading processes. Biedert et al. [29] base their intrusion detector on "learning effects", assuming that users become progressively more accustomed with certain tasks. In some experiments, student testers were asked to perform common tasks such as check for emails, read messages from their imaginary supervisor and perform specific actions connected with their hypothetical thesis work. Silver and Biggs [30] used a multi-modal approach considering both keystroke and eye-tracking biometrics. While keystroke information seemed to provide better results than eye tracking data, the combination of the two modalities is a promising method.

Unlike intrusion detection systems, ATM-like solutions require the user to explicitly perform certain eye tasks in order to authenticate. For instance, De Luca et al. [31] propose an authentication mechanism based on eye gestures, which derives from the idea that it is easier to remember complex shapes than long passwords or PINs. Gestures are obtained by moving the eyes in

specific ways, thus "drawing" patterns on the screen. Dunphy et al. [32] developed a software to vaguely simulate ATM verification. The user is presented with a sequence of five 3×3 grids containing nine faces each, and he or she has to recognize and select one face per grid (those chosen during an initial training phase). Weaver et al. [33] describe a system in which the user authenticates by looking at symbols on an on-screen keyboard. Instead of dwelling on the virtual keys or pressing a button to trigger the selection action, gaze points are grouped and automatically analyzed to find out the selected symbols.

Apart from the technical feasibility of gaze-based authentication, also the human factors coming into play in eye-controlled interfaces need to be considered. De Luca et al. [34], for example, assessed three different eye-based PIN-entry techniques (one of which based on gestures), finding that eye gaze interaction is an appropriate method for PIN entry, especially when element selection is performed by pressing a physical button. Brooks et al. [35] investigated some interface designs for eye movement biometric systems, focusing on task completion times, need for system recalibration and testers' usability perception.

3. Method

Data were acquired through the Tobii 1750 remote eye tracker, which integrates all its components (camera, infrared lighting, etc.) into a 17″ LCD monitor (1280×1024 resolution). As experimental stimuli, images representing human faces were used. Each face was segmented into significant areas of interest (AOIs).

3.1. Data acquisition

In the Tobii eye tracker, 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. 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 $20 \times 15 \times 20$ cm placed about 60 cm from the screen. The accuracy of the device is 0.5°.

The Tobii *ClearView* gaze recording software was employed to define stimuli (still images, slideshows, videos, etc. to be presented to

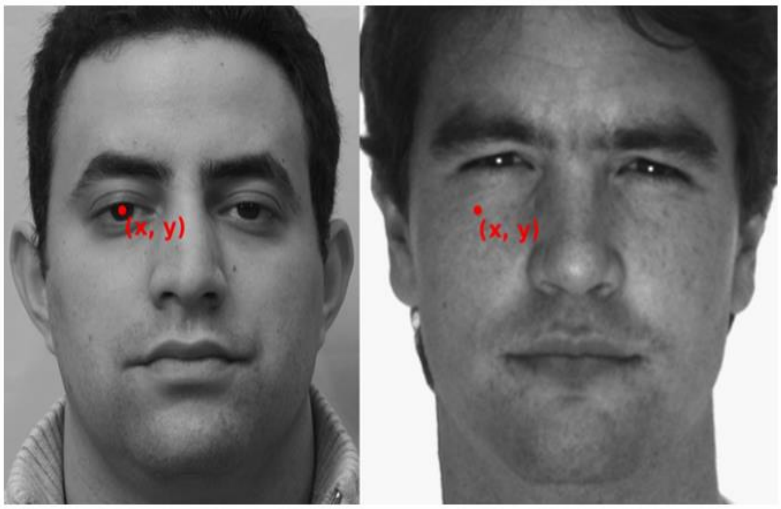

Fig. 1. In the left image, the red point at coordinates (280, 420) is exactly in the center of the right eye, while in the right image the point is on the right cheekbone.

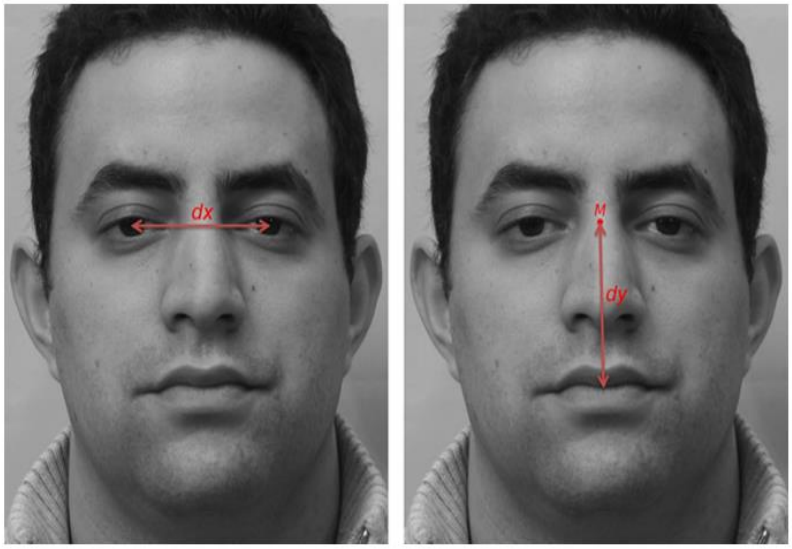

Fig. 2. *dx* (left) and *dy* (right) distances. *M* indicates the eyes' middle point.

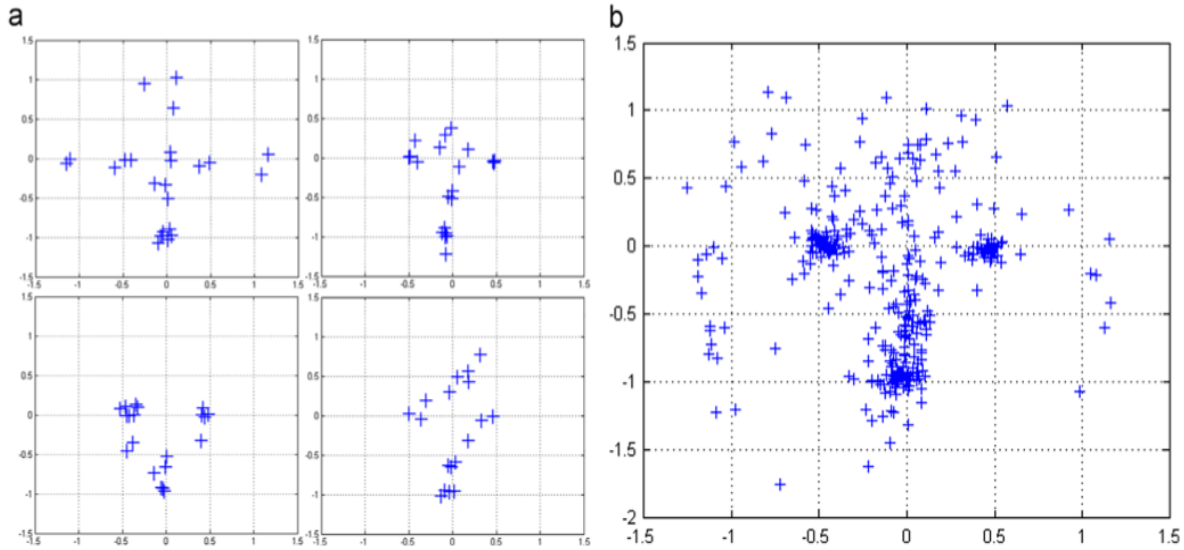

Fig. 3. Cloud of normalized fixation points of observer 15 in session 1: (a) observation points on four different face images; (b) merging of observation points coming from all the 16 observed face images.

the subject), as well as to record and manipulate gaze data. The system 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 as 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 sequence. A further output of the eye tracker is the *gaze replay*, which dynamically shows the evolution of fixations and saccades with time.

3.2. Data normalization

To analyze how a particular observer looks at faces, all his or her observations i.e. recordings of fixation point sequences while looking at each one of the 16 faces, are merged together. To do this, a normalization step is required, because in different face images a particular area of interest such as right or left eye and mouth, may be located in slightly different positions in the image. Even a deviation of only few pixels in the position of a face in an image with respect to another may imply that in the first image a fixation at coordinates (x, y) falls in the right eye area, while in the second the fixation falls in the right eyebrow area. Moreover, face images used in our experiments have little different sizes. In Fig. 1, for example, the point at coordinates $(x=280, y=420)$ corresponds to different face areas in different face images.

The normalization of fixation point coordinates is performed with respect to scale and shift, by means of an affine transformation. Let be dx the distance between the eyes and dy the distance between the mouth and the eyes' middle point (X_M, Y_M) (see Fig. 2), the scaling factors with respect to the horizontal and vertical directions are $s_x=1/dx$ and $s_y=1/dy$ respectively. The new normalized coordinates (x', y') are simply obtained as follows:

$$P'(x', y') = \begin{bmatrix} s_x & 0 \\ 0 & -s_y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -X_M \\ Y_M \end{bmatrix} = \begin{bmatrix} s_x x - X_M \\ Y_M - s_y y \end{bmatrix} \quad (1)$$

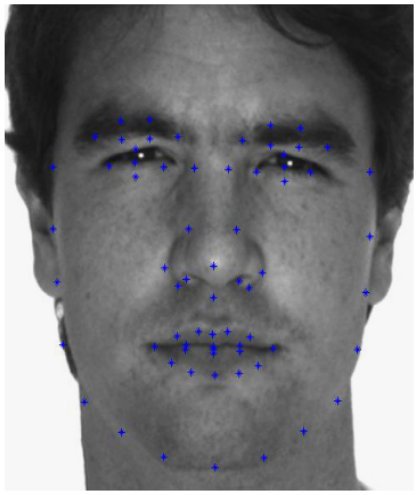

Fig. 4. STASM output on subject 3.

After normalization, the coordinates (x', y') will correspond to the same face area for all 16 subjects.

Once all coordinates of the fixation points are normalized for each of the 16 subjects observed, all fixation points of a given observer, in a single test session, are merged in a single plot. As an example, the result of the normalization phase for observer 15 is shown in Fig. 3. Now it is possible to analyze the cloud of fixation points for each observer and try to extract some features useful to distinguish an observer from another.

3.2.1. Face fiducial points detection

To automatically obtain the distances dx and dy and the coordinates of the eyes' middle point M, a face detector system presented in [36], namely, the extended Active Shape Model (STASM) algorithm, is used. First, faces are detected by a global face detector (Viola-Jones [37] or Rowley [38]), which extracts all regions of interest (ROI) from an image that contains at least one face. The detected ROIs are then submitted to the STASM algorithm, which searches for facial landmarks by minimizing a global distance between candidate image points and their homologues using a general model (shape model), which is

precomputed ("learned") over a wide set of training images. The algorithm locates 68 interest points. The precision of the location procedure depends on the amount of face distortion. For instance, the output of STASM on subject 3 is shown in Fig. 4.

Among the 68 points extracted by STASM we only need those representing the pupils' central points and the mouth's central point in order to compute *dx* and *dy* distances. Since the eyes' middle point *M* is not returned by STASM, its coordinates have to be computed starting from those of the pupils.

3.3. Feature extraction

For each observer in each session, we have now the set of all his or her fixation points. In our previous work "GAS – Gaze Analysis System", presented in [39], we subdivided faces into 17 ROIs such as right eye, left eye, forehead, lips, and background, and then a feature vector was built with each element corresponding to a ROI. The values in the feature vector corresponded to the sum of durations of fixations contained in the corresponding ROI. However, with the approach described in [39] we only analyzed the time spent by the observer in each ROI. Moreover, as each face is different from another, a specific ROI mask was built for each face. In this work, instead of using a different mask for each face, we decided to use a standard subdivision in ROI for all faces, that we will discuss later in Section 3.3.1.

The objective of this work is to obtain a fully automatic gaze analysis system capable to analyze which areas of a face the observer is used to look at, for how much time and in which order. The idea is to build a graph from the set of fixation points that contains information in the form of weights associated to nodes and arcs, related to the density of fixations and to the time spent in a specific area of the face, and to the observation path.

Another graph based approach for eye movement based biometric recognition is presented in [25]. The authors present a method based on the construction of a joint minimal spanning tree graph structure between a reference and a test sample of fixation points. The distance between the two samples is measured by a multivariate generalization of the Wald–Wolfowitz random runs test. However, with respect to our approach, only coordinates of fixation points are taken into account, while GANT, through the use of weights associated to nodes and arcs, analyzes other important distinctive features such as duration, density and trajectory of the fixation points.

3.3.1. Features graph

Because of the high number of points in the fixation cloud, they are first aggregated in a smaller number of nodes. In order to do

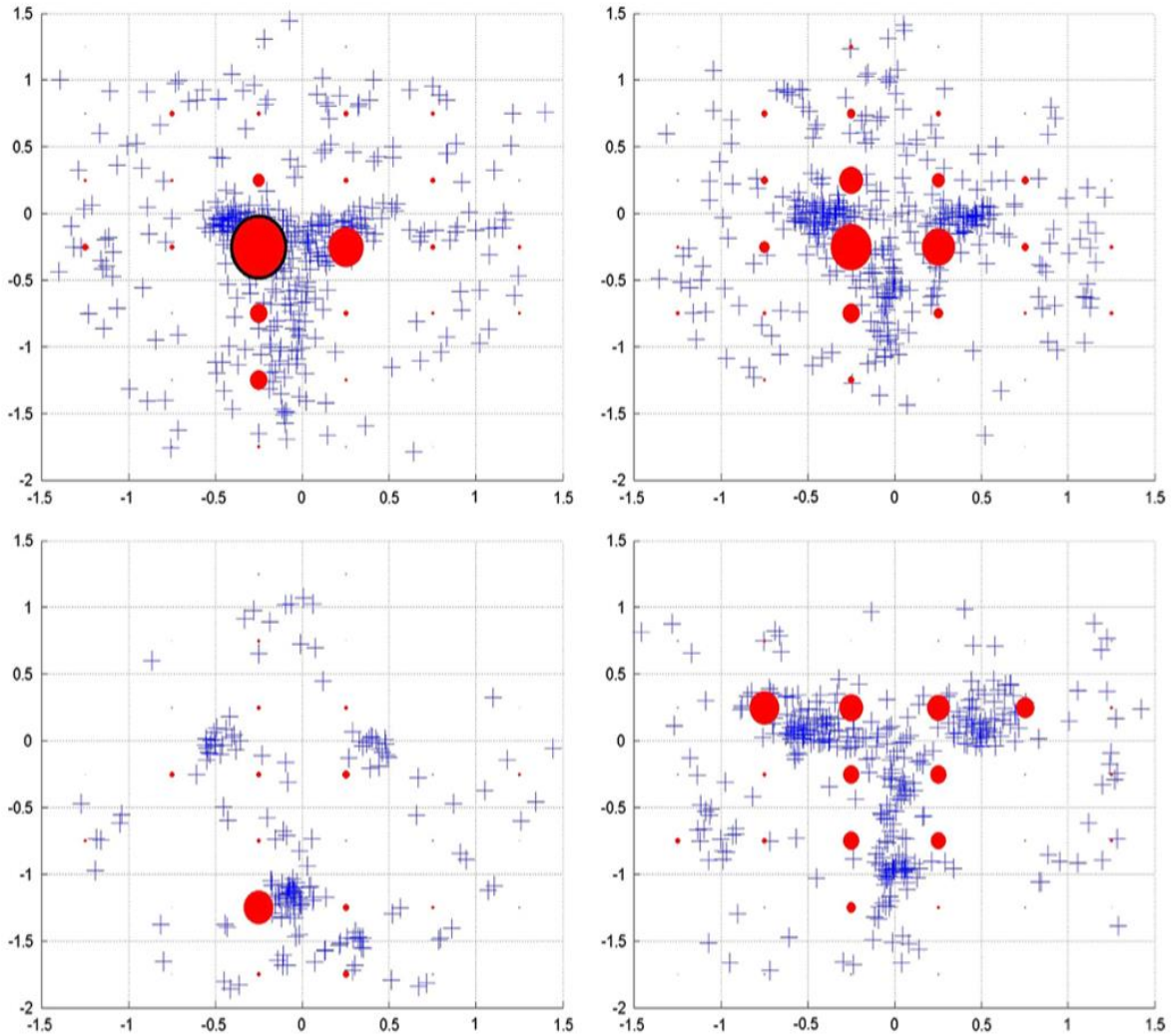

Fig. 5. Density graphs of four different observers: observer 10 (top-left); observer 41 (top-right); observer 56 (bottom-left); observer 86 (bottom-right). The size of red circles indicates the weight associated to the node. Black bordered circles indicate a weight > 95.

this, the cloud is subdivided using a grid. Since in our experiments we have observed that after normalization all fixation point x coordinates ranged in $[-1.5, 1.5]$ and y coordinates ranged in $[-2, 1.5]$, we have chosen a cell size of 0.5×0.5, obtaining a 7×6 grid (see Fig. 3). For each cell of the grid, we consider a node centered in that cell with a default weight equal to zero. Weights are then associated to each node based on:

- Density: the number of fixations in the corresponding cell;
- Duration: the sum of durations (in milliseconds) of each fixation in the corresponding cell.

Figs. 5 and 6 show some examples of density-graphs and duration-graphs.

The method used to create and associate weights to the arcs of the graph is a little more complex due to the fact that we have merged together different observations (i.e. observations of the 16 different subjects). We cannot build a unique path over all fixation points because they belong to different face images, so it is not correct to build an arc from the last node observed on an image and the first node observed on another image. For each face image, we want to increment the weight of an arc linking two nodes A and B each time the observer gaze passes from node A to node B and vice versa.

The arcs of the graph are defined as follows:

Let 0 be the initial weight for all arcs;
Let $f_{i,k}$ be the k^{th} fixation ($\mathbf{1} \leq k \leq \mathbf{n}$) of an observer on image i ($\mathbf{1} \leq i \leq \mathbf{16}$);
Let A, B be the first and second endpoints of an arc, A$|f_{i,1} \in$ A;
For each $f_{i,k}$ of the fixation sequence
    B$|f_{i,k+1} \in$ B;
    if B $<>$ A
        increment the weight of arc AB by one;
        A=B;
        k=k + 1;
if k=n
        break;
    endif
endif
endfor

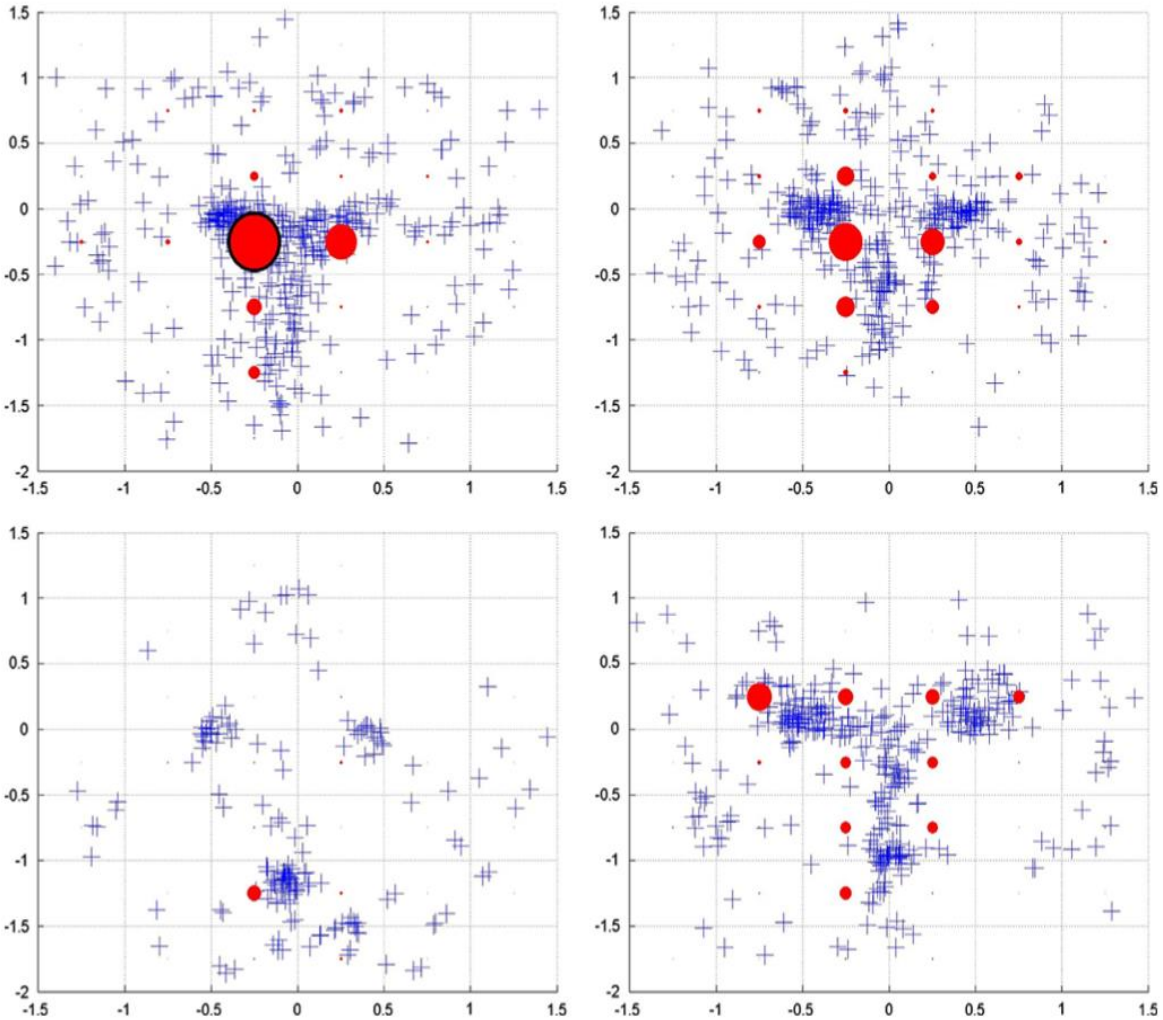

Fig. 6. Duration graphs of four different observers: observer 10 (top-left); observer 41 (top-right); observer 56 (bottom-left); observer 86 (bottom-right). The size of red circles indicates the weight associated to the node. Black bordered circles indicate a weight > 45,000 (ms).

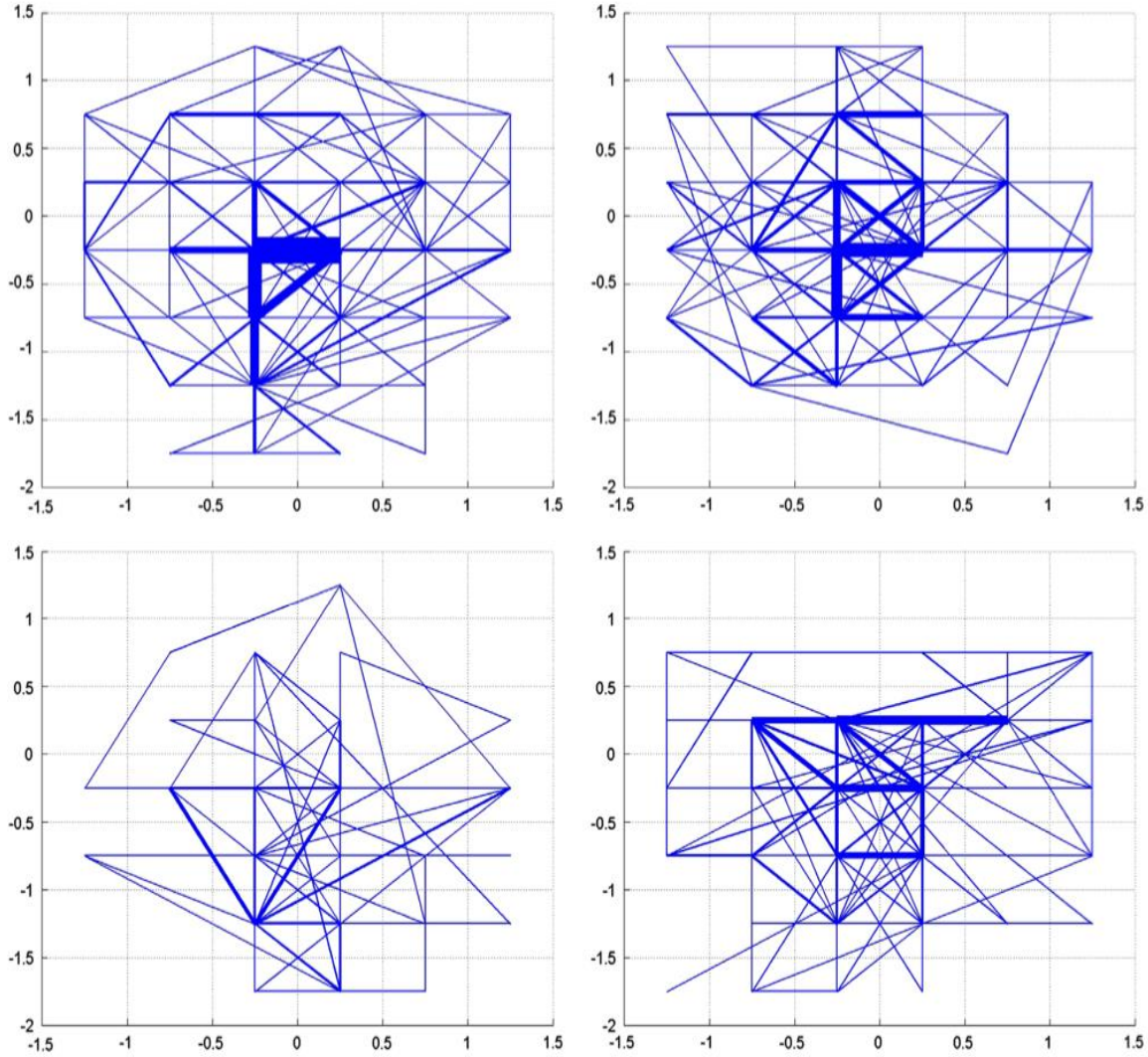

Fig. 7. Weighted arcs graphs of four different observers: observer 10 (top-left); observer 41 (top-right); observer 56 (bottom-left); observer 86 (bottom-right). The thickness of arcs specifies the associated weight.

Fig. 7 shows some examples of weighted arcs. The thickness of arcs specifies the associated weight: a ticker arc indicates that the gaze of the observer shifted more frequently between the two nodes connected by that arc. With this approach we do not consider the order in which nodes are observed, but, rather, which nodes are more closely related each other. This representation allows us to easily combine arc graphs, obtained from an observer of different face images, by summing up the weights of corresponding arcs, and also to make the comparison process described in the following easier.

3.4. Comparison

The graph obtained for each observer in each session represents the *observer fixation model*. Weights associated to nodes are described by two matrices 7×6 with each element corresponding to a node, one for densities and one for durations. To represent weighted arcs, an adjacency matrix 42×42 is used. To compare different observers based on densities, durations or arcs, the distance between couples of matrices of the same feature is measured through the Frobenius norm of the matrices' difference. Frobenius norm is a matrix defined as the square root of the sum of the absolute squares of its elements [40].

4. Experimental results

4.1. Experimental protocol

A total of 112 volunteer observers (73 males and 39 females) took part in the trials, subdivided into the following age groups: 17–18 (11 persons), 21–30 (58), 31–40 (9), 41–50 (16), 51–60 (8), 61–70 (9) and 71–80 (1). All participants reported normal or corrected-to-normal vision.

Prior to the beginning of the experiments, carried out in a quiet environment, participants were informed about the fact that some images, without specifying their kind, would appear on the eye tracker's display in full screen mode (to prevent the user from getting distracted during the gaze recording procedure). Face images were interleaved with blank white screens with a small cross at their center, to ensure a common starting location for stimulus exploration. The first blank screen was displayed for 5 s, while the others for 3 s. Each test was also preceded by a short and simple calibration procedure, lasting about 10 s 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. Each stimulus image was shown for 10 s.

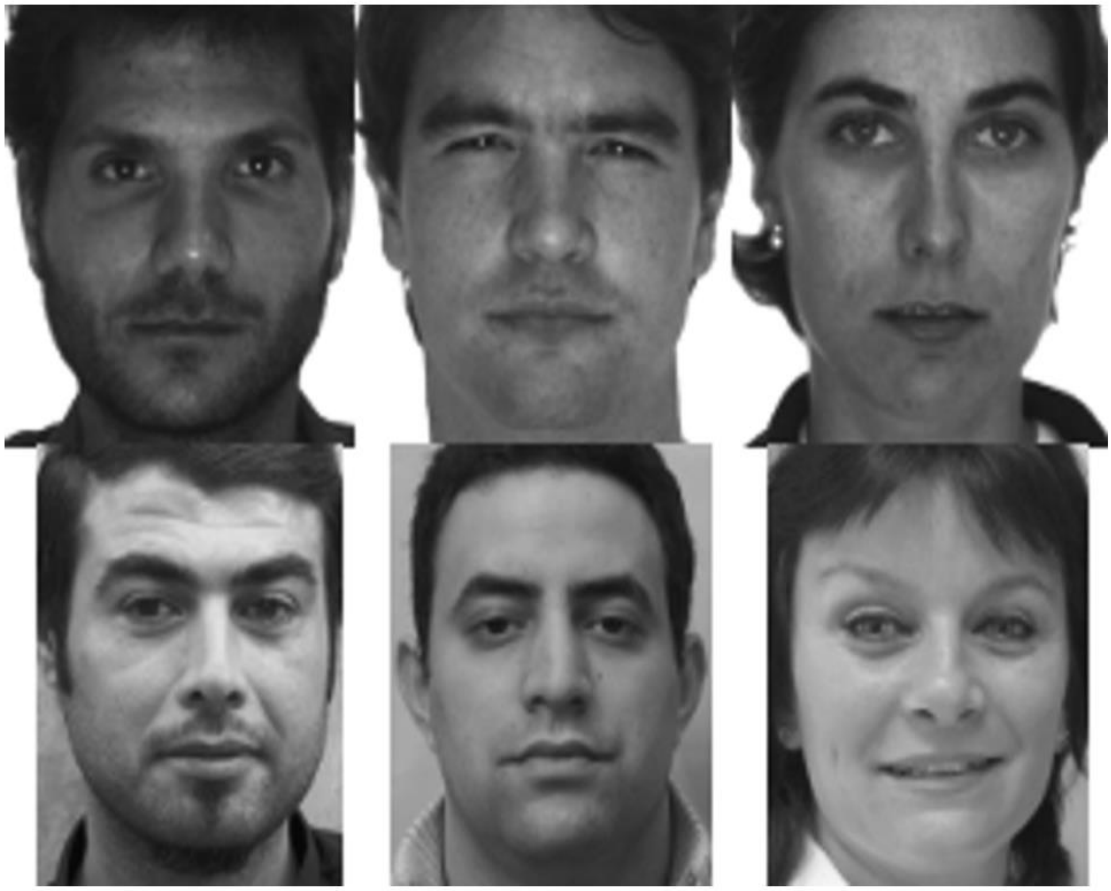

Fig. 8. Examples of stimuli.

Table 1
Test scheme for single feature experiments.

Feature	Experiment	Gallery	Probe
Density	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II

Table 2
Test scheme for combined feature experiments.

Features	Experiment	Gallery	Probe
Density–duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs–density	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Arcs–duration	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II
Density–duration–arcs	Experiment I	Gallery	Probe I
	Experiment II	Gallery	Probe II

Sixteen black-and-white pictures were employed in the experiments, which contained close-up faces of eight males and eight females. Half of the faces (four males and four females) were of famous persons (mostly actors and actresses), while the others were of people unknown to the observers. Images were mostly taken from the AR Face Database [41]. Examples of the stimulus image set are shown in Fig. 8.

Table 3
Single feature experiments results.

Experiment	EER	AUC	CMS(1)
Density I	0.3078	0.7893	0.2093
Density II	0.2785	0.7727	0.3043
Duration I	*0.2783*	*0.8032*	*0.1860*
Duration II	*0.2628*	*0.7960*	*0.3043*
Arcs I	0.3158	0.7330	0.1860
Arcs II	0.3484	0.7445	0.1739

The presentation order 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. Also, all images had similar gray-level distributions. On average, a single test session, including task explanation, device calibration, lasted a little more than 5 min.

A first set S_1 of tests was carried out with 88 participants. Of these, 36 were involved in a second session (with the same images), and 16 other participants were also involved in a third test session. One hundred forty tests were therefore carried out in the three sessions. 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.

A second set S_2 of tests, with 34 participants, was implemented after 1 year from S_1. Ten observers in this group had been involved in S_1 as well. Also in this case, three sessions were organized: 17

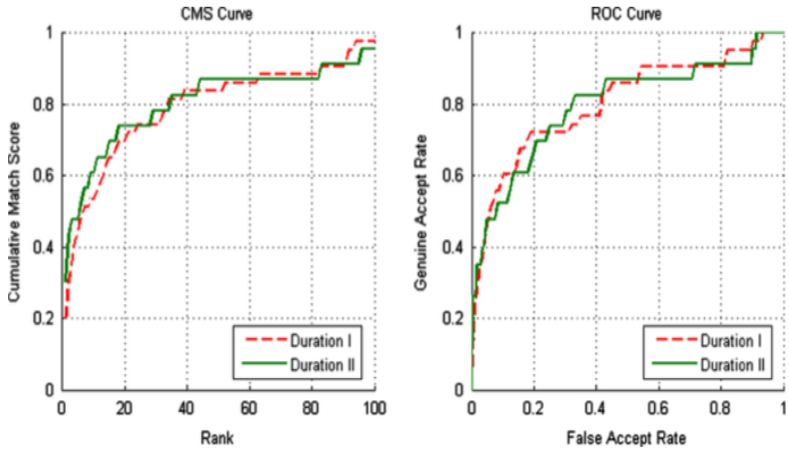

Fig. 9. Performance graphs for the duration feature.

Table 4
Combined features experiments results.

Experiment	EER	AUC	CMS(1)
Den-Dur I	0.2878	0.7917	0.1395
Den-Dur II	0.2901	0.7922	0.3043
Arc-Den I	0.3093	0.7824	0.1628
Arc-Den II	0.2984	0.7879	0.2174
Arc-Dur I	*0.2786*	*0.7892*	*0.1395*
Arc-Dur II	*0.2688*	*0.7928*	*0.3043*
Arc-Den-Dur I	0.2859	0.7905	0.1395
Arc-Den-Dur II	0.2911	0.7935	0.3043

observers out of 34 were involved in a second session (nine of whom had participated in S_1), and 13 took part in a third session (six of whom had participated in S_1). Sixty-four tests were therefore carried out in the three sessions. Time intervals between the first and the second session, and between the second and the third session, ranged from a minimum of 1 day to a maximum of 21 days.

4.2. Results

The accuracy of the GANT system has been assessed in terms of Area Under Curve (AUC – the area under the ROC curve), 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. CMS(1) indicates the value at rank 1, namely the Recognition Rate.

The fixation models have been subdivided into Gallery and Probe sets as follows:

- Gallery: the Gallery contains the fixation models obtained from data acquired in the first session (111 observers, where 87 observers comes from S_1 and 24 are the new observers coming from S_2. Observer 80 was discarded because the fixation points relative to one of the 16 images were missing);
- Probe I: the first probe set is composed of the fixation models obtained from data acquired in the second session (44 observers, where 36 comes from S_1 and eight are the new observers coming from S_2);
- Probe II: the second probe set is composed of the fixation models obtained from data acquired in the third session (23 observers, where 16 comes from S_1 and seven are the new observers coming from S_2).

We have carried out three kinds of experiments based on:

- Single features;
- Combined features;
- Weighted combined features.

For tests based on single features, we separately tested density, duration and arc features following the test scheme presented in Table 1.

For tests based on combined and weighted combined features, we tested all features combinations following the test scheme presented in Table 2. Before combining single features, scores have been normalized to the range [0, 1].

4.3. Single feature experiments

Results of the tests on single features show that the duration represents the best discriminating feature among the three, and in particular it has the best EER, AUC and CMS values. The results of the single feature experiments are summarized in Table 3.

The performance obtained when the duration is used as a discriminant feature is shown in Fig. 9.

4.4. Combined features experiments

We tested all the combinations of the three features that are density, duration and observation path. The final score of each combination is computed by averaging scores obtained by GANT with single feature. The results of this experiment are reported in Table 4.

The graphs related to the combination of features which achieved the best performance are presented in Fig. 10.

4.5. Weighted combined features experiments

We have finally tested all the weighted combinations of features in order to improve the performance of the system. Weights have been chosen proportionally to the results of single features in order that their sum was equal to 1. Higher weights have been given to scores relative to *duration* features which

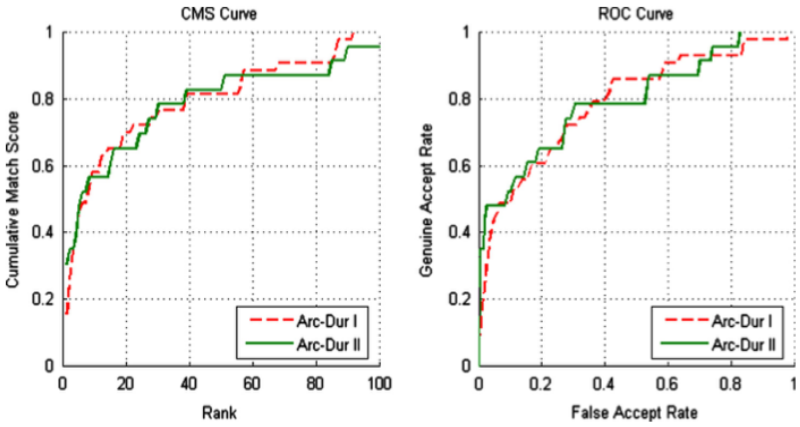

Fig. 10. Performance graphs for the experiment on the combination of arcs and duration features.

Table 5
Weighted combined features experiments results.

Experiment	EER	AUC	CMS(1)
Den-Dur I	0.2819	0.7924	0.1395
Den-Dur II	0.2933	0.7922	0.3043
Arc-Den I	0.3047	0.7828	0.1628
Arc-Den II	0.3038	0.7867	0.1739
Arc-Dur I	0.2803	0.7906	0.1395
Arc-Dur II	0.2706	0.7917	0.3043
Arc-Den-Dur I	*0.2719*	*0.7932*	*0.1395*
Arc-Den-Dur II	*0.2739*	*0.7930*	*0.3043*

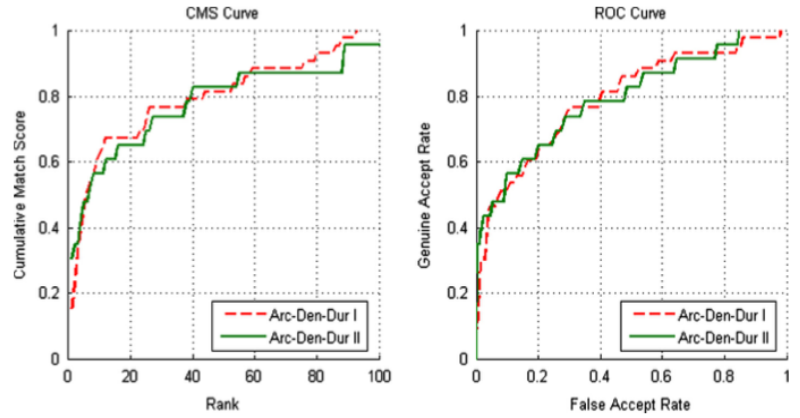

Fig. 11. Performance graphs for the experiment on the weighted combination of arcs, density and duration features.

performed better in our experiments, medium weights to scores relative to *density* features and lower weights to scores relative to arc features. The scores fusion has been performed through the weighted average. Results are reported in Table 5.

Fig. 11 shows the graphs of the performance of the combination of features Arcs, Density and Duration which obtained the best results.

4.6. Comparison with previous experiments

As mentioned in Section 3.3, in [39] we presented "GAS – Gaze Analysis System", and a first series of experiments on a part of the database (88 individuals in total, the set S_1 of tests described in Section 4.1), their performance graphs are shown in Fig. 12.

Table 6 shows the comparison, on the same dataset (88 observers), between the best results obtained in [39] (GAS) and the best results (duration feature) of the system presented in this work (GANT) in terms of EER, AUC and CMS. With GAS I and GANT I we indicate experiments in which the second session is used as Probe, while GAS II and GANT II indicate experiments in which the third session in used as Probe. With respect to the previous experiments, EER drops from 0.361 to 0.250 when using the third session as Probe set, while AUC and CMS(1) values have a significant improvement. The reason is that we provided a more robust representation of users' fixation model, increasing the number of AOIs and using the same standard subdivision in AOIs for all fixation sequences. Finally, in computing system performance we took into account users' observations over all the 16 faces, both for the Gallery and the Probe sets, while in the previous

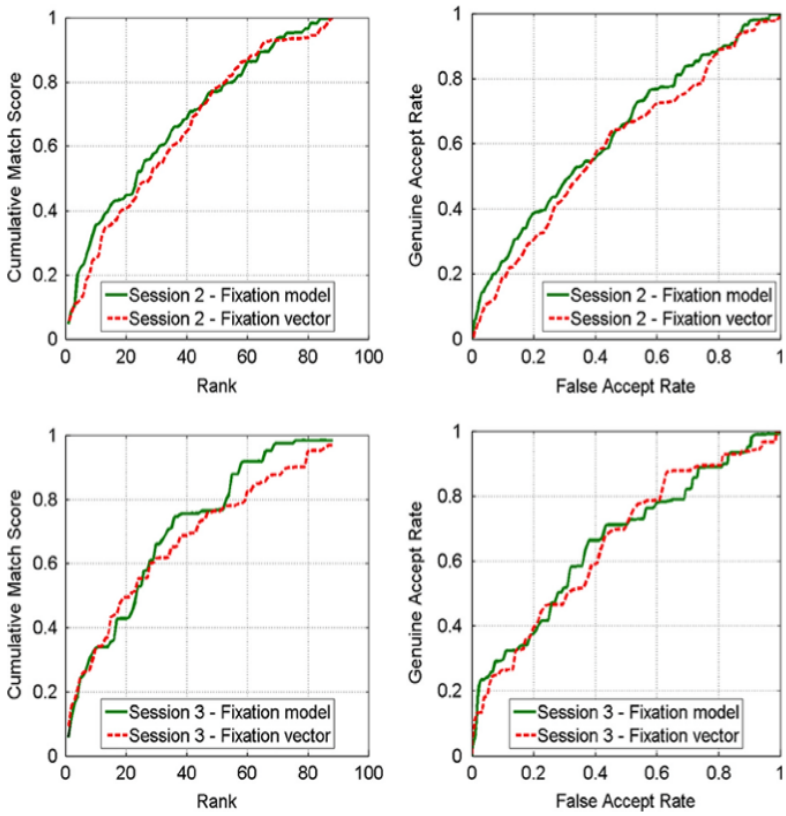

Fig. 12. GAS performance.

Table 6
Comparison between experiments presented in [39] (GAS) and best results of the system presented in this work (GANT).

Experiment	EER	AUC	CMS(1)
GAS I	0.424	0.6318	0.0476
GAS II	0.361	0.6617	0.0589
GANT I	0.224	0.8179	0.2286
GANT II	0.250	0.7901	0.3125

work testing was performed by considering, in turn, one of the 16 observations acquired during sessions 2 and 3 separately.

5. Conclusions

The recent interest of the research community in soft biometric led us to design and test a novel technique for gaze analysis, namely GANT. The GANT approach, applied to a wide dataset composed of 112 volunteer observers acquired through the Tobii 1750 remote eye tracker, verified the conjecture that the way an individual looks at an image might be a personal distinctive feature. However, significant improvements are required, specifically to allow the application of gaze analysis in large-scale identification scenarios. To further improve its performance, we plan to include GANT in a multi-biometric framework including iris and face, using a multi-view camera for acquisition.

Conflict of interest

None declared.

References

[1] A.K. Jain, S.C. Dass, K. Nandakumar, Soft biometric traits for personal recognition systems, in: Proceedings of the International Conference on Biometric Authentication, LNCS 3072, 2004, pp. 731–738.

[2] K.P. Hollingsworth, K.W. Bowyer, P.J. Flynn, The best bits in an iris code, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) 964–973.

[3] X. Qiu, Z. Sun, T. Tan, Global texture analysis of iris images for ethnic classification, in: Proceedings of the International Conference on Biometrics, 2006, pp. 411–418.

[4] V. Thomas, N. Chawla, K. Bowyer, P. Flynn, Learning to predict gender from iris images, in: Proceedings of the First IEEE International Conference on Biometrics: Theory, Applications, Systems (BTAS 2007), 2007, pp. 1–5.

[5] A. Dantcheva, N. Erdogmus, J.L. Dugelay, On the reliability of eye color as a soft biometric trait, in: IEEE Workshop on Applications of Computer Vision (WACV), 2011, pp. 227–231.

[6] K. Mock, B. Hoanca, J. Weaver, M. Milton, Poster: real-time continuous iris recognition for authentication using an eye tracker, in: Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS '12), 2012, pp. 1007–1009.

[7] O.V. Komogortsev, A. Karpov, L. Price, C. Aragon, Biometric authentication via oculomotor plant characteristic, in: Proceedings of the IEEE/IARP International Conference on Biometrics (ICB), 2012, pp. 1–8.

[8] A.T. Duchowski, Eye Tracking Methodology – Theory and Practice, second ed., Springer-Verlag, London, 2007.

[9] A. Eisenberg, Pointing With Your Eyes, to Give the Mouse a Break, New York Times, 2011 ⟨http://www.nytimes.com/2011/03/27/business/27novel.html?r=2⟩ (March 26).

[10] N. Westover, Tobii Brings Eye Control to Tablets, PCMAG.COM, September 19, 2013, retrieved from ⟨http://www.pcmag.com/article2/0,2817,2424565,00.asp⟩.

[11] M. Porta, S. Ricotti, C. Jimenez Perez, Emotional E-learning through eye tracking, in: Proceedings of the 2012 IEEE International Conference on Collaborative Learning & New Pedagogic Approaches in Engineering Education (EDUCON 2012), Marrakesh, Morocco, 2012, pp. 1–6.

[12] V. Cantoni, C. Jimenez Perez, M. Porta, S. Ricotti, Exploiting eye tracking in advanced E-learning systems, in: Proceedings of the Thirteenth International Conference on Computer Systems and Technologies (CompSysTech 2012), Rousse, Bulgaria, 2012, pp. 376–383.

[13] E. Perego, F. Del Missier, M. Porta, M. Mosconi, The cognitive effectiveness of subtitle processing, Media Psychol. 13 (3) (2010) 243–272.

[14] M.A. Just, P.A. Carpenter, Eye fixations and cognitive processes, Cogn. Psychol. 8 (1976) 441–480.

[15] A. Poole, L.J. Ball, Eye tracking in human-computer interaction and usability
Cahoui (Ed.), Encyclopedia
pp. 211–219.
nts, in: H. Pashler (Ed.),
9–154.
[17] H. Deubel, W.X. Schneider, Saccade target selection and object recognition: evidence for a common attentional mechanism, Vis. Res. 36 (1996) 1827–1837.
[18] L. Itti, C. Koch, A saliency-based search mechanism for overt and covert shifts of visual attention, Vis. Res. 40 (2000) 1489–1506.
[19] C. Fookes, A. Maeder, S. Sridharan, G. Mamic, Gaze based personal identification, in: L. Wang, X. Geng (Eds.), Behavioral Biometrics for Human Identification: Intelligent Applications, IGI Global, Hershey, PA, USA, 2010, pp. 237–263.
[20] P. Kasprowski, J. Ober, Eye movements in biometrics, in: Proceeding of the 2004 International Workshop on Biometric Authentication (BioAW 2004), 2004, pp. 248–258.
[21] R. Bednarik, T. Kinnunen, A. Mihaila, P. Fränti, Eye-movements as a biometric, in: H. Kalviainen, J. Parkkinen, A. Kaarna (Eds.), Image Analysis, LNCS, vol. 3540, Springer Berlin, Heidelberg, 2005, pp. 780–789.
[22] F. Deravi, S.P. Guness, Gaze trajectory as a biometric modality, Biosignals (2011) 128 (paper).
[23] C. Holland, O. Komogortsev, Complex eye movement pattern biometrics: the effects of environment and stimulus, IEEE Trans. Inform. Forensics Security 8 (12) (2013) 2115–2126.
[24] O.V. Komogortsev, A. Karpov, C.D. Holland, H.P. Proenca, Multimodal ocular biometrics approach: a feasibility study, in: Proceedings of the Fifth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '12), 2012, pp. 209–216.
[25] I. Rigas, G. Economou, S. Fotopoulos, Biometric identification based on the eye movements and graph matching techniques, Pattern Recognit. Lett. 33 (6) (2012) 786–792.
[26] T. Kinnunen, F. Sedlak, R. Bednarik, Towards task-independent person authentication using eye movement signals, in: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10), 2010, pp. 187–190.
[27] Z. Liang, F. Tan, Z. Chi, Video-based biometric identification using eye tracking technique, in: Proceedings of the 2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC), 2012, pp.728–733.
[28] C. Holland, O.V. Komogortsev, Biometric identification via eye movement scanpaths in reading, in: Proceedings of the 2011 International Joint Conference on Biometrics (IJCB 2011), 2011, pp. 1–8.
[29] R. Biedert, M. Frank, I. Martinovic, D. Song, Stimuli for gaze based intrusion detection, in: J.J. Jong Hyuk) Park et al. (eds.), Future Information Technology, Application, and Service, Lecture Notes in Electrical Engineering 164 (2012), pp. 757–763.
[30] D.L. Silver, A. Biggs, Keystroke and eye-tracking biometrics for user identification, in: Proceeding of the 2006 International Conference on Artificial Intelligence (ICAI 2006), 2006, Vol. 2, pp. 344–348.
[31] A. De Luca, R. Weiss, H. Hußmann, X. An, Eyepass – eye-stroke authentication for public terminals, in: CHI '08 Extended Abstracts on Human Factors in Computing Systems (CHI EA '08), 2008, pp. 3003–3008.
[32] P. Dunphy, A. Fitch, P. Olivier, Gaze-contingent passwords at the ATM, in: Proceedings of the Fourth Conference on Communication by Gaze Interaction (Communication, Environment and Mobility Control by Gaze, COGAIN 2008), 2008, pp. 59–62.
[33] J. Weaver, K. Mock, B. Hoanca, Gaze-based password authentication through automatic clustering of gaze points, in: Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2011), 2011, pp. 2749–2754.
[34] A. De Luca, R. Weiss, H. Drewes, Evaluation of eye-gaze interaction methods for security enhanced PIN-entry, in: Proceedings of the Nineteenth Australasian Conference on Computer-Human Interaction: Entertaining User Interfaces (OZCHI '07), 2007, pp. 199–202.
[35] M. Brooks, C.R. Aragon, O.V. Komogortsev, Perceptions of Interfaces for Eye Movement Biometrics, in: Proceedings of the International Conference on Biometrics (ICB), 2013, pp. 1–8.
[36] S. Milborrow and F. Nicolls, Locating facial features with an extended active shape model, in: Proceedings of the Tenth European Conference on Computer Vision (ECCV '08), 2008, pp. 504–513.
[37] P. Viola and M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001), 2001, pp. 511–518.
[38] H.A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 23–38.
[39] C. Galdi, M. Nappi, D. Riccio, V. Cantoni, M. Porta, A new gaze analysis based soft-biometric, in: Proceedings of the Fifth Mexican Conference on Pattern Recognition (MCPR 2013), Lecture Notes in Computer Science Volume 7914, 2013, pp. 136–144.
[40] G.H. Golub, C.F. Van Loan, Matrix Computations, third ed., Johns Hopkins, Baltimore, MD, 1996.
[41] A.M. Martinez and R. Benavente. The AR face database. CVC Technical Report #24, June 1998.

Virginio Cantoni is a Full Professor in Computer Engineering at the University of Pavia and former Dean of the Faculty of Engineering. In the period 2008–2011 he has been seconded to the Centro Linceo 'Beniamino Segre' of the Accademia dei Lincei. He is the founder and first Director of the University of Pavia's European School of Advanced Studies in Media Science and Technology and former Director of the Interdepartmental Centre for Cognitive Science. His research activity is concerned with pattern recognition and parallel architectures for image processing and computer vision. He has organized many International Conferences, Seminars and Workshops including a NATO Advanced Research Workshop on pyramidal systems for computer vision. An Expert and Project Reviewer for the EU Commission, he became a Fellow of the IAPR in 1994 and Fellow of the IEEE in 1997.

Chiara Galdi was born in 1988. She received the Laurea degree (cum laude) in Computer Science from the University of Salerno, Italy, in 2012. She is currently a Ph.D. student at the University of Salerno. Her main research interests include: biometrics, iris recognition under uncontrolled conditions.

Michele Nappi was born in Naples, Italy, in 1965. He received the Laurea degree (cum laude) in computer science from the University of Salerno, Salerno, Italy, in 1991, the M.Sc. degree in information and communication technology from the International Institute for Advanced Scientific Studies "E.R. Caianiello," Vietri sul Mare, Salerno, and the Ph.D. degree in applied mathematics and computer science from the University of Padova, Padova, Italy. He is currently an Associate Professor of computer science at the Università di Salerno. His research interests include pattern recognition, image processing, compression and indexing, multimedia databases, and visual languages. Dr. Nappi is a member of the International Association for Pattern Recognition.

Marco Porta is assistant professor at the University of Pavia, Italy. He received the Dr. Eng. (Master) degree in Electronic Engineering from the Polytechnic of Milan (Milan, Italy) in 1996 and the Ph.D. degree in Electronic and Computer Engineering from the University of Pavia in 1999. His research interests include eye tracking applications, vision-based perceptive interfaces, visual browsing of large collections of images, visual languages, e-learning, user-centered interface design and Human–Computer Interaction in general. He has published about 70 articles in international conference proceedings, journals and books, and has participated in several national and international projects.

Daniel Riccio He was born in Cambridge (U.K.), in 1978. He received the Laurea degree (cum laude) and the Ph.D. degree in computer science from the University of Salerno, Italy, in 2002 and 2006, respectively. He is an Assistant Professor at the University of Naples, Federico II, and a IEEE/IAPR member.