Pattern Recognition Letters 148 (2021) 114–120

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

User recognition based on periocular biometrics and touch dynamics

Andrea Casanova[a], Lucia Cascone[b], Aniello Castiglione[c], Weizhi Meng[d,*], Chiara Pero[b]

[a] Department of Mathematics and Computer Science, University of Cagliari, Cagliari, Italy
[b] Department of Computer Science, University of Salerno, Fisciano, Italy
[c] Department of Science and Technology, University of Naples Parthenope, Naples, Italy
[d] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark

ARTICLE INFO

Article history:
Received 20 October 2020
Revised 24 April 2021
Accepted 4 May 2021
Available online 20 May 2021

Keywords:
Touch dynamics
Eye movement
User recognition
Biometric identification
Behavioural biometrics

ABSTRACT

Web user behavioural recognition is the process by which web users are identified and distinguished through behavioural features. In this work, two sources of behavioural biometric data are analyzed for the development of this web user identification model, touch dynamics and the characteristics extracted from the periocular area related to the pupils, blinks and fixations. The approach adopted used to improve the overall performance of the multimodal biometric recognition system is based on a fusion at the Feature level to which different distance measure techniques (Euclidean, Bray-Curtis, Manhattan, Canberra, Chebyshev, Cosine) are applied to determine if the test sample belongs to the target subject. To further improve the system performance, we have applied multi-data processing methods such as Canonical Correlation Analysis (CCA) and Principal Component Analysis (PCA). The results obtained demonstrate the promise of these two different biometric traits and, above all, of their fusion. In fact, the fusion approach allows obtaining an accuracy higher than that of individual biometrics, reaching an accuracy of over 92%.

1. Introduction

Verification is the process of affirming that a claimed identity is correct by comparing the offered claims of identity with one or more previously enrolled templates. A synonym for verification is authentication. Identification occurs when a biometric system attempts to determine the identity of an individual. A biometric sample is collected and compared to all the templates in a database. Identification is close-set if the person is assumed to exist in the database. In this case, the system must determine if the person is in the database. A watchlist task is an example of open-set identification whereby the person may not have been enrolled in the database. Another common task, namely, negative identification attempts to check that the person has not been enrolled in the database before. This task is useful to avoid duplicate identities in the same database. Finally, recognition is a generic term that could imply either or both verification and identification. This term is generally avoided unless a broad coverage of both biometric applications is intended.

From the first days of their appearance on the Earth, human beings have distinguished each other by discriminating on their physical aspects and personal characteristics. All those information (i.e., voice, appearance, and so on) is processed by our brain that quickly comes to the decision if such subject is someone who is known or not. The methods that allow the automatic processing of those characteristics bound to human beings are known as biometric recognition. Those methods are able to process a single or a group of specific characteristics belonging to humans. Such characteristics are commonly categorized as behavioural and physiological biometrics. Behavioural biometrics tell more about the way of acting and living of an individual. Handwriting, speech tone, keystroke dynamics, gait [2] are some of the most common biometrics. All those aspects are strongly related to the habits and the psychology of an individual and, as a consequence, are subject to change over time since each of us tends to slightly modify our behaviour and habits. On the other hand, physiological biometrics, such as mouth, hands, eyes, face and so on are strictly bound to the physical aspect of an individual. Physiological traits are more steady over time even though some accidents may occur and alter them inadvertently. Typical examples of physiological biometrics can be considered how the veins are distributed into the retina's eye, the shape of the ear, the iris [3], the morphology of a face, the fingerprint of each finger and so on. Having in mind the above mentioned considerations, it is quite easy to discern that a system that adopts behavioural biometrics presents more rigid requisites comparing to the physiological ones due to the high variability of

* Corresponding author.
E-mail address: weme@dtu.dk (W. Meng).

https://doi.org/10.1016/j.patrec.2021.05.006
0167-8655/

its biometric aspects. In addition, it is well known that behavioural biometrics is the oldest and most accepted form of biometric traits (for example, the signature of a person is commonly and widely recognized as a form of behavioural biometric).

The yearly report published by the IDC (International Data Corporation) [18] showed that worldwide a total of 331.7 million smartphones were shipped in the second quarter of 2019 (2019Q2) even though there was a decline of 16.6% looking at the second quarter of 2020 (2020Q2) in which there were shipped 276.4 million smartphones. Apart from such a decrease, it is clear that the use of smartphones has surpassed that of PCs and so the researchers' attention is focusing on mobile devices [4]. In fact, modern smartphones have several sensors that can be used to derive actions performed by the device's owner and determine useful information such as location, rotation, acceleration, magnetic field, and so on. All those information may also be related somehow to the analysis of the behaviour of a user. Furthermore, almost all smartphones use touchscreens as a medium for interaction with the user. To study how people interact with a touchscreen it is possible to analyze *touch dynamics*. Those dynamics consider several aspects such as the time interval among keystrokes, the pressure of each touch, the movement of the way a user touches, the screen and so on. The last ones are not so easy to replicate and so can be considered as a distinctive element to identify users.

It is evident that due to the widespread use of smartphones, it is quite easy to access such biometric data. These communication devices are now increasingly embedded with different sensors. In addition, using such biometrics will not interfere with the common users' activities (even though it can be seen as a privacy mining task) since it is completely transparent to the user owning the device. In fact, the monitoring of the sensors as well as of the user behaviour when touching the touchscreen takes place in parallel. As said, the use of smartphones allows accessing of a large amount of behavioural and physiological properties that is an optimal solution for a system using biometric technologies. Smartphones are touched by users but there exists another way users interact with smartphones: their eyes. Thus, it might be straightforward to consider also the eyes as an additional biometric aspect, together with the touch dynamics.

In fact, in biometric systems eye movement is gaining more and more attention from the scientific community and many research studies are coming recently on that topic. It's a common feeling that gaze direction and eye contact have an important role when people communicate since they undergo several cognitive processes.

The driving idea of this work is to merge together (what is usually referred to as "Information Fusion") the eye movements and the touch dynamics to further improve the performance of the recognition systems by developing efficient and effective methodologies that use both types of biometrics. This approach also gives an additional advantage: it allows to gather data originating from various sources and transform it into a unique representation [10]. Furthermore, the strategy adopting such kind of fusion integrating both eye movements and touch dynamics results in more accurate recognition of the subjects.

This paper is arranged as below: in Section 2 we review some related works already published in literature; Section 3 presents the dataset used for the experimental analysis; Section 4 introduces the adopted methodology underlying the entire research; Section 5 illustrates the steps related to the experimental protocols including the data preparation together with implementation details; results and discussions are sketched in Section 6; at the end of the paper, Section 7 gives some concluding remarks and future directions.

2. Related work

2.1. Touch dynamics

There exist plenty of research papers dealing with the study of how people interact with smartphones and, specifically, how each user performs the phase of pressing the screen. This action can be measured and evaluated by looking at the velocity of each touch on the touchscreen of a device, being it a tablet, a smartphone or a laptop (supporting the touchscreen). The way each human being interacts with the touchscreen (and, in turn, with the device) can be considered a kind of (digital) signature that is strongly related to the human interacting with the system. Such signature can be seen as a distinctive and peculiar element identifying a specif individual. Historically, it was common knowledge that in II World War there were some experts that were able to discern if a message was typed by the same operator or not. This was possible by just analyzing the rhythms and the pace of how the text was written by a given person. This is usually considered the first documented case of the identification of a human by looking at the way he/she interacts with a keyboard. Clearly, the just presented scenario dates back to the last century and with the actual technologies, it is easier and easier to perform the same recognition thanks to the widespread availability of smartphones. In [36] authors show how the typing style (the so-called *Keystroke Analysis*) can be used as a way of recognizing users. In fact, in such research, the basic idea is to derive biometric characteristics of a user from the rhythm and pace a user press the keys. One of the first studies on this subject, dated back to 1980, was presented by Gaines et al. [15], where authors considered how 6 secretaries were used to type 3 different kinds of texts (with a length ranging from 300 up to 400 words) with the purpose of identifying those 6 persons.

Recently [13,20,35] presented several studies that were focused on the specific way of using a PC keyboard by different users. Also in this case the goal of all those researches was to identify users. Going on with the years, some researchers [7,12,26] targeted mobile devices equipped with hardware keyboards to derive the same information as the studies just mentioned. With the advent of smartphones, the researchers were directed to the use of touchscreens instead of physical keyboards. It is widely agreed that using the touch dynamics it is possible to derive a series of behavioural events such as multi-touch actions and touch movements (see Fig. 1) greater than the ones that are possible to derive from keystroke dynamics because the latter only adopts buttons as their unique method of input. However, Meng et al. [28] highlighted that both approaches (i.e., keystroke dynamics and touch) have several common aspects. In literature, there are many studies investigating the use of touch dynamics biometrics as a recognition method. Shen et al. [33] developed an authentication model based on dynamic touch actions, achieving an error rate between 1.72% and 9.01%. However, they indicated that the authentication accuracy can be improved when users perform specific tasks, such as browsing the web. Similarly, Meng et al. [27] presented SocialAuth, which analyzes users' touch in social networking applications. In the paper of [6], the authors introduced DIALERAUTH, a touch-based smartphone user authentication scheme. The system verifies users when they insert a 10-digit string. In the experimental phase, they adopted a neural network based approach with a True Acceptance Rate of 85.77%. Although several touch behavioural based biometric recognition schemes have been proposed, it is still an open challenge to devise a robust model, as a user's tactile actions are dynamic and therefore hard to model.

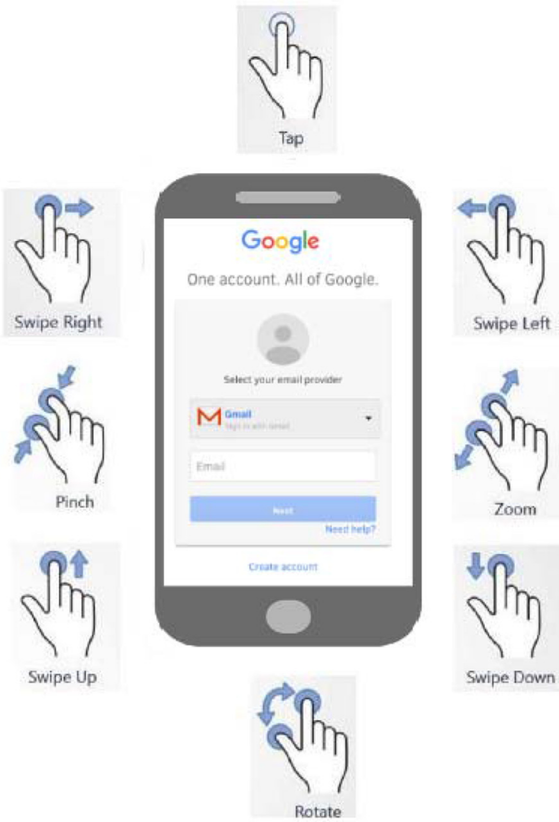

Fig. 1. Examples of using a touchscreen.

2.2. Periocular features

The use of the pupil as a discriminating biometric indicator is rather limited, although several studies highlight its potential in demographic classification tasks [11]. Cantoni et al. [8] are among the first to analyze the pupil size through ML techniques in order to obtain a classification by gender and age, highlighting interesting results. Current theory associates pupil dilation more with the emotional load aroused in the subject than with the positive or negative nature of the stimulus. That is, the emotional involvement with respect to an external stimulus largely depends on the propensities of the subject, on what an individual considers exciting or not. Therefore, pupillary variations were also analyzed to study participants' reactions to an experiment with respect to images of products they considered unpleasant, neutral or pleasant [29]. They analyzed the relationship between pupil diameter variability and the emotional state of participants in two different experiments under various types of workload. Sakamoto et al. [31] analyzed the relationship between pupil diameter variability and the emotional state of participants in two different experiments under various types of workload. In both experiments, a significant correlation was recorded between pupil size and emotional state.

From the existing works in the literature, it seems to be evident the existence of a continuum between a very low blink rate, relative to those performances that require high visual attention, and the increase in the frequency of blinks just before sleepiness and during boring tasks. Some scholars believe that blinking can provide useful information on central nervous system activation and fatigue levels. When the fatigue in carrying out an activity increases, performance decreases, resulting in an increase in the frequency and duration of blinks [25]. Conversely, fewer blinks have been associated with those activities that require more attention and concentration, this is because blink inhibition is needed to minimize information loss caused by visual perception disruptions, so in conditions that require a considerable attentional investment, the number of blinks is reduced [30]. Neuro-anatomical evidence supports the correlation between factors such as fear, anxiety, and alertness and the motor nuclei of the brainstem that control eye movements such as saccades. The time between two saccades is generally called fixation. [32] hypothesized the role of the first two ocular fixations for specific processes of memory recognition, i.e. familiarity and recollection. The experiments have shown that recollection benefits from the contribution of the second fixation, as opposed to familiarity, this is due to greater availability of inputs and for more information on the stimuli collected. In recent years, images of the periocular region have been exploited to recognize human identity as well, achieving approximately 80% accuracy with unconstrained image capture distances [19]. Recent studies demonstrated that the analysis of the pattern of the eye movements can be used to identify people in several application scenarios, including the online Web navigation. Those characteristics are interesting since it can be used also for other identification goals, such as face recognition as in Abate et al. [1] where the authors considered 14 periocular features, obtaining a recognition accuracy of a subject slightly lower than 80%. The identification of people can take place not only using the images of the periocular area but also by analyzing the data that can be extracted from an eye tracker. In the paper of [21], the authors decided to use eye tracker calibration data to identify users. They analyzed the data from three datasets with an accuracy of the identification ranged from 49% to 71%.

3. Data collection and description

The proposed work requires a dataset containing the above-mentioned biometric characteristics, i.e., a keystroke dynamics benchmark collected using a touch screen phone and eye movement patterns, including pupils dimension, blink, and fixation points. To the best of our state-of-art knowledge, currently, there are no available databases that simultaneously provide these two traits of an individual. Therefore, the RHU KeyStroke Dynamics Benchmark Dataset [14] and the GANT (Gaze ANalysis Technique) dataset [9] are combined into a multimodal database. From the fusion of both datasets, 19 individuals are selected, chosen according to age and gender, and for each subject, there are 18 acquisitions. So, the human touch dynamics features and the eye movements characteristics constitute the multimodal database adopted in the experimental phase. In the following subsections we present the two datasets adopted in the experimental phase.

3.1. RHU KeyStroke dynamics benchmark

The RHU Keystroke benchmark includes four "key event" features, namely Press-to-Press (*PP*), Press-to-Release (*PR*), Release-to-Press (*RP*) and Release-to-Release (*RR*) that store, respectively, the event time between: two key pressures, a key pressure and a key release, a key release and a key pressure and, finally, two key releases. In particular, it is possible to consider a sequence of typing events $N = (t_1, t_2 \ldots t_n)$, with $t_i = [t_i{}^p, t_i{}^a]$ where $t_i{}^p$ represents the press time and $t_i{}^a$ is defined as the release event. Based on these assumptions, it is possible to define these features as follows:

- $t_i{}^{pp} = t_i{}^p - t_{i-1}{}^p$ as *PP* time
- $t_i{}^{ap} = t_i{}^p - t_{i-1}{}^a$ as *PR* time
- $t_i{}^{pa} = t_i{}^a - t_{i-1}{}^p$ as *RP* time
- $t_i{}^{aa} = t_i{}^a - t_{i-1}{}^a$ as *RR* time

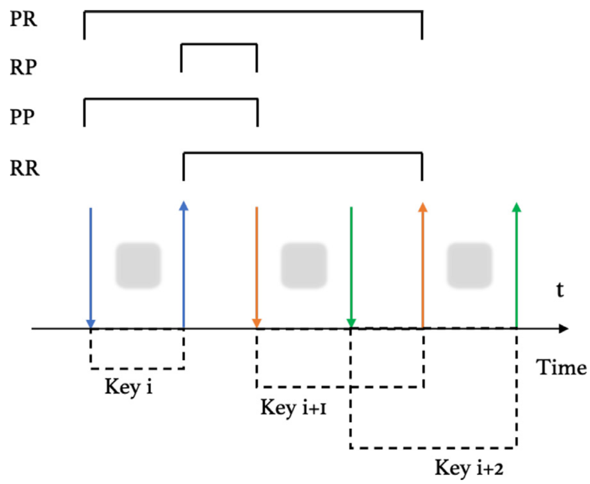

Fig. 2. Illustration of keystroke dynamics features in RHU Keystroke benchmark: Press-to-Press (*PP*), Press-to-Release (*PR*), Release-to-Press (*RP*) and Release-to-Release (*RR*).

Fig. 3. Some examples of images shown to participants during the data acquisition process in the GANT dataset. The first two columns show images of women and men, respectively. Above are images of strangers while below are images of famous people. Finally, the third column shows images of the two landscapes.

Fig. 2 illustrates the measurements that characterize a user's typing pattern in Rhu dataset. During the acquisition process, all participants typed the password: "rhu.university" 15 times spread on 3 different sessions temporal.

3.2. GANT dataset

In GANT dataset, information was collected through the Tobii 1750 eye tracker (1280 × 1024 screen resolution, 50 Hz sampling rate), adopting 16 human face images and 2 landscape images as experimental stimuli. The figures were interleaved with blank white screens with a cross at the center and showed randomly. The acquisition sessions were 2 and were held in 2012 and 2013. The order in which all images were shown is random. Fig. 3 illustrates some samples shown during the data acquisition process in GANT.

4. Fusion approach and multi-data processing methods

A uni-modal biometric system, i.e., one that exploits a single biometric feature, may experience several problems and show several limitations due to lack of data, poor quality of the collected information or, as in the case of soft biometric features, low discriminatory power. To overcome these problems, the choice of a multi-biometric system, i.e., a system that combines several biometric features, can help improve performance and consolidate the information obtained. Fusion can occur by considering different sources and at different levels. One of the most used and intuitive strategies is certainly the one concerning the fusion at the level of the features. Feature-level Fusion (FF) is the process of combining several feature vectors that are extracted separately from each biometric trait to make a single one. It is believed that this fusion strategy is more discriminative and can achieve better performance than others (e.g., decision-level fusion) because the new vector has richer information than the single original input ones. In their work, Bokade and Kanphade [5] proposed a multi-modal system based on a FF approach for biometric authentication by analyzing three biometric traits such as face, palm print and ear. The performance improvement is evident. Several methods have been proposed in the literature for the fusion of features [38]. The two most famous fusion strategies are:

- Serial feature fusion is certainly one of the most common and intuitive strategies. The idea is in fact to simply concatenate the sets of feature vectors into one vector. So, if for example, there are two vectors of dimensions p, q as input, then, after the fusion, there will be a vector of dimension equal to $(p+q)$.
- Parallel feature fusion, a method based on a complex vector. For example, if there are two input vectors of features, x and y, they are combined into a complex vector $z = x + iy$ where i is the imaginary unit.

After feature extraction, the feature selection and fusion process often relies on Principal Component Analysis (PCA). This is a famous strategy adopted for reducing the dimensionality of the feature space. In recent years there has been a growing interest in features fusion based also on Canonical Correlation Analysis (CCA) [17]. CCA-based methods have become popular as there have been several improvements in the performance of fusion systems.

4.1. Canonical correlation analysis

In statistics, Canonical Correlation Analysis (CCA) is one of the most widely used multi-data processing methods. The idea behind this method is to study the mutual linear relationships between two sets of variables. Suppose two vectors of variables from two different modalities $X = (X_1, X_2, \dots, X_n)$ and $Y = (Y_1, Y_2, \dots, Y_m)$ with mean vectors μ_x and μ_y and covariance matrices Σ_x and Σ_y, respectively. The purpose is to maximize the correlation between these two vectors. Let Σ be the overall covariance matrix, this contains all the information on relationships between pairs of characteristics:

$$\Sigma = \begin{pmatrix} \Sigma_x & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_y \end{pmatrix} \quad (1)$$

where $\Sigma_{xy} = E[(X-\mu_x)(Y-\mu_y)^T]$ is the covariance between X and Y and $\Sigma_{yx} = \Sigma_{xy}^T$. $E(\cdot)$ is the expectation operator, T is the transpose operation. Define new variables $X^* = W_x^T X$ and $Y^* = W_y^T Y$ via linear combinations of X and Y, the purpose is to maximize the pair-wise correlations across the two sets:

$$corr(X^*, Y^*) = \frac{cov(X^*, Y^*)}{var(X^*) \cdot var(Y^*)} = \frac{W_x^T \Sigma_{xy} W_y}{\sqrt{W_x^T \Sigma_x W_x} \cdot \sqrt{W_y^T \Sigma_y W_y}} \quad (2)$$

X^* and Y^* are the so-called canonical variates. The contribution of the correlation coefficient is maximized using Lagrange multipliers maximizing the covariance between X^* and Y^* and respecting the unit variance constraints $var(X^*) = var(Y^*) = 1$. Therefore, after a series of transformations what remains to be solved are the following two characteristic equations:

$$\begin{cases} \Sigma_x^{-1}\Sigma_{xy}\Sigma_y^{-1}\Sigma_{yx}\hat{W}_x = R^2\hat{W}_x \\ \Sigma_y^{-1}\Sigma_{yx}\Sigma_x^{-1}\Sigma_{xy}\hat{W}_y = R^2\hat{W}_y \end{cases} \tag{3}$$

where $\hat{W}_x$ and $\hat{W}_y$ are the eigenvectors and R^2 is the diagonal matrix of eigenvalues. The two matrices have a number of non-zero eigenvalues equal to $d = rank(\Sigma_{xy}) \leq \min(m, n)$. So, the matrices W_x e W_y are composed of the ordered eigenvectors corresponding to the non-zero eigenvalues. The Feature-level fusion is performed on X^* and Y^*. As we have already said, there are two Feature-level fusion rules for these vectors, concatenation or summation [37]. The canonical variates are fused into a new vector in this way, respectively:

$$Z_1 = [X^*, Y^*] = [W_x^T X, W_y^T Y] \tag{4}$$

or

$$Z_2 = X^* + Y^* = W_x^T X + W_y^T Y \tag{5}$$

where Z_1 and Z_2 are known as Canonical Correlation Discriminant Features.

4.2. Principal component Analysis

Feature-level fusion has an important role when multiple features are used in the process of user authentication. However, in many real-world applications, the number of samples is usually less than the number of features. The covariance matrices Σ_x and Σ_y in fact in this way are singular and not invertible. A solution to overcome the intrinsic dimensionality of fused features is to consider a Principal Component Analysis (PCA) strategy. PCA is a linear transformation for all those applications where it is necessary to analyze a huge amount of data. It is used in a way to highlight similarities and differences between data without too much information loss. Let $X \in R^N$ a vector of N dimensions. Once the covariance matrix Σ_x of X has been calculated, the algorithm factorizes it as follows:

$$\Sigma_x = B\Lambda B^T \tag{6}$$

where $\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ is the diagonal matrix of eigenvalues of Σ_x with diagonal elements in decreasing order ($\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$) while $B = (b_1, b_2, \ldots, b_N)$ is the eigenvectors matrix. Hence, a certain subset $k < N$, of eigenvectors is selected as basis vectors and the data X are projected onto this new basis:

$$Y = P^T X \tag{7}$$

where $P = (b_1, b_2, \ldots, b_k)$. The new Y vector captures the most relevant features of the original data. The essence of PCA is that it is possible to reduce the size of the data and preserve as much variance as possible.

5. Experimental protocols

The architecture of the proposed multimodal biometric system is shown in Fig. 4. The input of each module is represented by the data belonging to the corresponding dataset. Then, the Feature-level fusion techniques were applied. Subsequently, 70% of the resulting dataset was used to extract the identification pattern for each user and the remaining samples made up the test set, ensuring that at least one acquisition per subject was included. Finally, to evaluate the dissimilarity or similarity score between the test samples and the training samples of the 19 subjects, several distance measure techniques were calculated.

5.1. Data preparation

In this work as mentioned in Section 3, the contributions of two datasets have been fused, one in which the biometric characteristics of touch are extracted and one in which the data of the periocular area are extracted. To organize and analyze a large amount of data as efficiently as possible, various statistical indices have been extracted such as the mean, quantiles, minimum, maximum, standard deviation. The goal is to express and synthesize the position of a frequency distribution by means of a real value representative of the phenomenon as a whole, summarizing the aspects considered most important. Therefore, for the RHU KeyStroke Dynamics benchmark data set for each acquisition, the “key event” characteristics have been summarized through these indices. For the GANT dataset instead, in order to have greater consistency, also with the other dataset, and considering that the information extracted from a single acquisition could be not sufficiently representative of the analyzed phenom, only those subjects for which there were 3 acquisition sessions were selected. As in fact, it has been clarified also in Section 2, for example also the memory could affect these values. Then, for each image observed (18, in total) from the subjects, the information related to pupil size, frequency and duration of blink and fixation was first extracted and then the indexes were used to synthesize them with respect to the 3 acquisitions. Subsequently, coherently to the gender and to the registry component were associated the various subjects between the two datasets and, at each acquisition of the touch were made to correspond the information related to an image in a random way.

5.2. Detector implementation

To solve our authentication problem, we ask ourselves how to find the similarity between the models of the same subject and highlight the distance between those of different subjects? The notion of distance is the most important basis for classification. The right choice of distance measurement is one of the most decisive steps for determining the correct classification of subjects. The correctness of the classification is mainly influenced by two factors. The first is the extraction of an adequate set of characteristics from the data set, the second is the creation of a discriminating identification pattern for every subject. As mentioned in Section 3, for the 19 subjects under examination there are 18 acquisitions. Once 30% of the samples have been selected, and ensured that at least one acquisition per subject has been extracted, these samples make up what we call the test set. The remaining acquisitions are then used to obtain the identification pattern for each subject. Two different strategies are applied. Once the acquisitions are selected, both the mean and the median are calculated for each subject, obtaining two datasets, X_1 and X_2, respectively.

To match, the distance between the subject extracted from the test set and the 19 identification patterns in the created data set $X_i (i = 1, 2)$, is calculated. The identifier of the vector to which the shortest distance from the analyzed one is associated is the one selected as searched and this distance is assigned as a score. In this work, 6 different detectors were implemented to measure the similarity between a subject’s modeled behaviour and a new sample. Therefore, taking an element from the test set $u = (u_1, \ldots, u_n)$ and one from the set containing the reference patterns $v = (v_1, \ldots, v_n)$, where n is the number of features considered, the following distances were calculated:

- Euclidean: the Euclidean distance is the most commonly used distance in all applications. It is the most obvious way of representing the distance between two points. The Euclidean distance between two points is then equal to the length of the

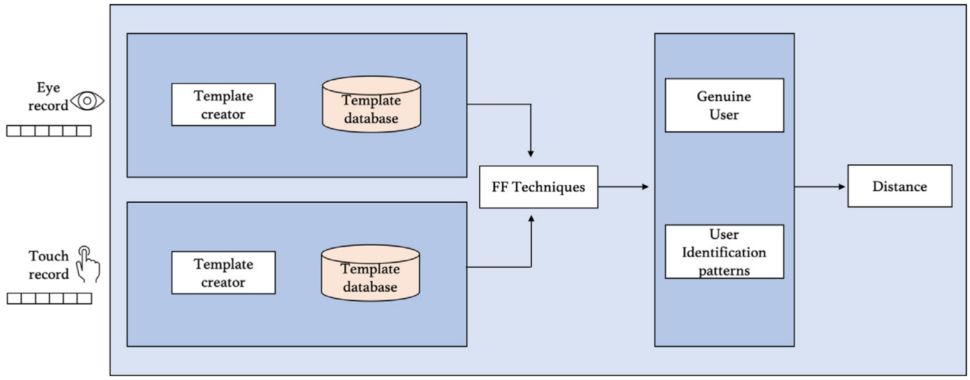

Fig. 4. Architecture of the proposed multimodal biometric system.

Table 1
Performance of feature-level fusion (FF) strategies with parameters (in bold the best result).

Method	GANT	RHU	GANT + RHU	CCA + GANT + RHU	PCA + GANT + RHU
mean	79,59%	59,18%	89,90%	91,84%	88,76%
median	77,55%	62,24%	**92,85%**	89,90%	90,82%

Table 2
Performance of feature-level fusion (FF) strategies in multimodal system with parameters and metrics.

Dataset + Method	Metrics	Accuracy
GANT + RHU (*median*)	Manhattan	92,85%
GANT + RHU (*median*)	Bray-Curtis	90,81%
CCA + GANT + RHU (*concatenation, mean*)	Bray-Curtis	91,84%
CCA + GANT + RHU (*sum, mean*)	Cosine	91,84%
PCA + GANT + RHU (*median*)	Euclidean	90,82%

segment joining them [16]:

$$distance = (\sum_{i=1}^{n}(u_i - v_i)^2)^{1/2}. \tag{8}$$

- Bray-Curtis: the Bray Curtis dissimilarity is used to quantify the differences in species populations between pairs of samples. It is mainly used in ecology, biology and medicine [34]. It is always a number between 0 and 1. If 0, the two sites share all the same species; if 1, they don share any species:

$$distance = \frac{\sum_{i=1}^{n}|u_i - v_i|}{\sum_{i=0}^{n}|u_i + v_i|}. \tag{9}$$

- Manhattan: the Manhattan distance is also known as city block distance. It is preferred over the Euclidean distance metric when the data size increases. It calculates the distance it would take to get from one data point to another if you followed a grid path [24]:

$$distance = \sum_{i=1}^{n}|u_i - v_i|. \tag{10}$$

- Canberra: the Canberra distance is very sensitive for values close to 0, where it is more sensitive to proportional than absolute differences [22]. This is most evident in the higher dimensional space, respectively an increasing number of variables. In turn, it is less affected by high-value variables than the Manhattan distance. Being a very sensitive measure, it is applicable to identify deviations from normal patterns:

$$distance = \sum_{i=1}^{n}\frac{|u_i - v_i|}{|u_i| + |v_i|}. \tag{11}$$

- Chebyshev: the Chebyshev distance is also called maximum value distance or chessboard distance. The distance between two vectors is the largest of their differences along dimension coordinates [16]:

$$distance = \max_i |u_i - v_i|. \tag{12}$$

- Cosine: the Cosine similarity is generally used as a metric to measure distance when the magnitude of vectors is not important. Its geometric meaning is rather intuitive as it is the angle between two vectors. It is one of the most used measures to represent the relationship between two sets [23]:

$$distance = \frac{\sum_{i=1}^{n} u_i \cdot v_i}{\sqrt{\sum_{i=1}^{n} u_i^2} \cdot \sqrt{\sum_{i=1}^{n} v_i^2}}. \tag{13}$$

6. Results and discussions

In order to analyze the performance of the proposed multimodal biometric recognition system, several experiments were performed and compared with each other. The results were evaluated in terms of Recognition Accuracy (%), i.e the ratio between the true positives and the true negatives for the total subjects adopted in the experimental phase. So, aimed at exploring the suitable Feature-level fusion techniques, different methodologies were performed on touch dynamics and eye movements databases respectively, also considering their fusion and their combination with the aid of techniques such as CCA and PCA (mentioned in Section 4). We also applied two different strategies, mean and median. For each method and dataset analyzed in the experimental phase, the Table 1 shows the best recognition rates. The Table 2 illustrates some detectors, used to measure the similarity between a subject's behaviour and a new sample, which produced the best pattern recognition accuracy. By comparing the results, the multi-biometric system using Manhattan distance with median provided the best recognition accuracy, equal to 92,85%, than that using other fusion strategies. As we can see, great results were also obtained with the following measurements: BrayCurtis, Cosine and, finally, Euclidean.

7. Conclusions and future directions

In this work, we presented an approach for integrating features extracted from the periocular area together with human touch behaviour to improve the overall performance of a multimodal biometric recognition system. Feature-level fusion (FF) approaches and different distance measure techniques were applied on eye and touch biometric traits, in order to determine if the test sample belongs to the target subject. The comparison was also extended with the analysis and application of CCA and PCA methods. The obtained results demonstrate the promise of these two different biometric traits and, consequently, of their fusion. Our method is conducive to pattern recognition and increases the accuracy of identification. In the future, we will further investigate new methodologies to improve recognition scores and evaluate the overall performance of our system.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A.F. Abate, L. Cimmino, F. Narducci, C. Pero, Biometric face recognition based on landmark dynamics, in: DASC 2020, IEEE, 2020, pp. 607–611, doi:10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00105.

[2] P. Barra, C. Bisogni, M. Nappi, D. Freire-Obregón, M. Castrillón-Santana, Gait analysis for gender classification in forensics, in: G. Wang, M.Z.A. Bhuiyan, S. De Capitani di Vimercati, Y. Ren (Eds.), Dependability in Sensor, Cloud, and Big Data Systems and Applications, Springer Singapore, Singapore, 2019, pp. 180–190.

[3] S. Barra, C. Bisogni, M. Nappi, S. Ricciardi, F-FID: fast fuzzy-based iris denoising for mobile security applications, Multimed. Tools Appl. 78 (10) (2019) 14045–14065, doi:10.1007/s11042-019-7156-9.

[4] S. Barra, M. Marras, G. Fenu, Continuous authentication on smartphone by means of periocular and virtual keystroke, in: International Conference on Network and System Security, Springer, 2018, pp. 212–220.

[5] G. Bokade, R. Kanphade, Secure multimodal biometric authentication using face, palmprint and ear: a feature level fusion approach, 2019.

[6] A. Buriro, B. Crispo, S. Gupta, F. Del Frari, DIALERAUTH: a motion-assisted touch-based smartphone user authentication scheme, in: Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy, 2018, pp. 267–276.

[7] P. Campisi, E. Maiorana, M. Bosco, A. Neri, User authentication using keystroke dynamics for cellular phones, Signal Process. IET 3 (2009) 333–341, doi:10.1049/iet-spr.2008.0171.

[8] V. Cantoni, L. Cascone, M. Nappi, M. Porta, Demographic classification through pupil analysis, Image Vis. Comput. 102 (2020).

[9] V. Cantoni, C. Galdi, M. Porta, D. Riccio, L. De Maio, R. Distasi, M. Nappi, Gaze recording - free observation of human faces, 2020b, https://doi.org/10.17632/3kn4jdd4kf.2

[10] A. Casanova, L. Cascone, A. Castiglione, M. Nappi, C. Pero, Eye-movement and touch dynamics: a proposed approach for activity recognition of a web user, in: 2019 15th International Conference on Signal-Image Technology Internet-Based Systems (SITIS), 2019, pp. 719–724.

[11] L. Cascone, C. Medaglia, M. Nappi, N. F., Pupil size as a soft biometrics for age and gender classification, Pattern Recognit. Lett. (2020), doi:10.1016/j.patrec.2020.10.009.

[12] N. Clarke, S. Furnell, Authenticating mobile phone users using keystroke analysis, Int. J. Inf. Sec. 6 (2007) 1–14, doi:10.1007/s10207-006-0006-6.

[13] H. Crawford, Keystroke dynamics: characteristics and opportunities, in: 2010 Eighth International Conference on Privacy, Security and Trust, IEEE, 2010, pp. 205–212.

[14] M. El-Abed, M. Dafer, R. El Khayat, RHU keystroke: a mobile-based benchmark for keystroke dynamics systems, in: 2014 International Carnahan conference on security technology (ICCST), IEEE, 2014, pp. 1–4.

[15] R.S. Gaines, W. Lisowski, S.J. Press, N. Shapiro, Authentication by Keystroke Timing: Some Preliminary Results, Technical Report, Rand Corp Santa Monica CA, 1980.

[16] S. Gultom, S. Sriadhi, M. Martiano, J. Simarmata, Comparison analysis of k-means and k-medoid with ecluidience distance algorithm, chanberra distance, and chebyshev distance for big data clustering, IOP Conf. Ser. 420 (2018) 012092, doi:10.1088/1757-899x/420/1/012092.

[17] M. Haghighat, M. Abdel-Mottaleb, W. Alhalabi, Discriminant correlation analysis for feature level fusion with application to multimodal biometrics, in: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2016, pp. 1866–1870.

[18] IDC, Smartphone Vendor Market Share, https://www.idc.com/promo/smartphone-market-share/vendor, 11 (2020).

[19] V. Ipe, T. Thomas, Periocular recognition under unconstrained image capture distances, Commun. Comput. Inf. Sci. 1365 (2021) 175–186.

[20] M. Karnan, A. Muthuramalingam, N. Krishnaraj, Biometric personal authentication using keystroke dynamics: a review, Appl. Soft Comput. 11 (2011) 1565–1573, doi:10.1016/j.asoc.2010.08.003.

[21] P. Kasprowski, Identifying users based on their eye tracker calibration data, 2020. 10.1145/3379157.3391419

[22] G. Lance, W.T. Williams, Mixed-data classificatory programs I - agglomerative systems, Aust. Comput. J. 1 (1967) 15–20.

[23] H. Liao, Z. Xu, Approaches to manage hesitant fuzzy linguistic information based on the cosine distance and similarity measures for HFLTSs and their application in qualitative decision making, Expert Syst. Appl. 42 (12) (2015) 5328–5336.

[24] M. Malkauthekar, Analysis of euclidean distance and manhattan distance measure in face recognition, in: IET Conference Publications, 2013, pp. 503–507.

[25] L. McIntire, R. Mckinley, C. Goodyear, J. McIntire, Detection of vigilance performance using eye blinks, Appl. Ergon. 45 (2013), doi:10.1016/j.apergo.2013.04.020.

[26] I.V. McLoughlin, et al., Keypress biometrics for user validation in mobile consumer devices, in: 2009 IEEE 13th International Symposium on Consumer Electronics, IEEE, 2009, pp. 280–284.

[27] W. Meng, W. Li, L. Jiang, J. Zhou, SocialAuth: designing touch behavioral smartphone user authentication based on social networking applications, in: IFIP International Conference on ICT Systems Security and Privacy Protection, Springer, 2019, pp. 180–193.

[28] W. Meng, Y. Wang, D. Wong, S. Wen, Y. Xiang, TouchWB: touch behavioral user authentication based on web browsing on smartphones, J. Netw. Comput. Appl. 117 (2018), doi:10.1016/j.jnca.2018.05.010.

[29] Q.-X. Qu, F. Guo, Can eye movements be effectively measured to assess product design?: gender differences should be considered, Int. J. Ind. Ergon. 72 (2019) 281–289.

[30] T. Sakai, H. Tamaki, Y. Ota, R. Egusa, S. Inagaki, F. Kusunoki, M. Sugimoto, H. Mizoguch, EDA-based estimation of visual attention by observation of eye blink frequency, Int. J. Smart Sens. Intell. Syst. 10 (2017) 296–307, doi:10.21307/ijssis-2017-212.

[31] K. Sakamoto, S. Aoyama, S. Asahara, H. Mizushina, H. Kaneko, Relationship between emotional state and pupil diameter variability under various types of workload stress, in: B.-T. Karsh (Ed.), Ergonomics and Health Aspects of Work with Computers, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 177–185.

[32] C. Schwedes, D. Wentura, The relevance of the first two eye fixations for recognition memory processes, Memory 27 (2019) 1–15, doi:10.1080/09658211.2019.1567789.

[33] C. Shen, Y. Zhang, X. Guan, R. Maxion, Performance analysis of touch-interaction behavior for active smartphone authentication, IEEE Trans. Inf. Forensics Secur. 11 (2015), doi:10.1109/TIFS.2015.2503258. 1–1

[34] W. Silva, D. Jasbick, R. Wilson, P. Azevedo-Marques, A. Traina, L. Santos, A. Jorge, D. De Oliveira, M. Bedo, A two-phase learning approach for the segmentation of dermatological wounds, in: Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2019, pp. 343–348.

[35] P.S. Teh, A. Teoh, S. Yue, A survey of keystroke dynamics biometrics, Sci. World J. 2013 (2013) 408280, doi:10.1155/2013/408280.

[36] P.S. Teh, N. Zhang, A.B.J. Teoh, K. Chen, A survey on touch dynamics authentication in mobile devices, Comput. Secur. 59 (C) (2016) 210–235, doi:10.1016/j.cose.2016.03.003.

[37] J. Yang, J.-Y. Yang, D. Zhang, J.-F. Lu, Feature fusion: parallel strategy vs. serial strategy, Pattern Recognit. 36 (6) (2003) 1369–1381.

[38] Y.-H. Yuan, Q.-S. Sun, Q. Zhou, D.-S. Xia, A novel multiset integrated canonical correlation analysis framework and its application in feature fusion, Pattern Recognit. 44 (5) (2011) 1031–1040.