2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

An Empirical Analysis of Ensemble Systems in Cancellable Behavioural Biometrics: a Touch Screen Dataset

Marcelo Damasceno and A.M.P Canuto

Abstract— **This paper presents an experimental analysis of a revocable biometric verification problem using ensemble systems. Behavioural Biometric-based systems are a future emergent area on identification, verification and access control systems of users. However, there is still progress to be done in this field, specially related to system security and acceptable results for practical use. Cancellable Biometrics is a alternative solution to the security problem of biometric data. This technique consists of applying transformation functions to biometric data in order to protect the original characteristics of biometric template. In this case, if biometric template has compromised, a new representation of original biometric data can be generated. Although cancellable biometrics were proposed to solve privacy concerns, this concept raises new issues, becoming the authentication problem more complex and difficult to solve. Thus, more effective authentication structures are needed to perform these tasks. This work aims to investigate the use of ensemble systems in cancellable behavioural biometric system used by million people (touchscreen devices). Apart this, we also present an empirical analysis, comparing the ensemble structures with single classification algorithms.**

I. Introduction

The most used method to identify/verify users in authentication systems still is through username and password [1]. Unfortunately, the username-password approach presents some problems regarding security and reliability on storage and use of this personal information. The main reason for this is related to the level of security endowed by such methods, since the personal information on which they rely can be easily misplaced, shared, or stolen. In addition, the use of same username and password for different services on the Internet leads to the stress to remember secure, long and complex passwords, which is not an easy task and may cause the use of the same password to different authentication systems.

Biometrics can be considered as the science of establishing the identity of a person using his/her anatomical and/or behavioural traits. Biometric traits have a number of desirable properties with respect to their use as an authentication token, such as reliability, convenience, universality, among others. Thus, these human characteristics have been used to user verification and/or identification in several authentication systems throughout the world. In this way, problems such as to remember complex passwords are minimized using biometric characteristics [2].

Marcelo Damasceno is a lecturer in Federal Institute of Rio Grande do Norte and PhD student in the Department of Informatics and Applied Mathematics in the Federal University of Rio Grande do Norte, Brazil who A.M.P Canuto is his supervisor (email: marcelo.damasceno@ifrn.edu.br, anne@dimap.ufrn.br).

Biometrics can be broadly divided in two classes: Physical and Behavioural. Physical biometrics are usually related to the body characteristics, such as face, fingerprint, iris recognition, hand geometry, among others. Unlike physical biometrics, behavioural biometrics are related to user behaviour/actions [3]. These biometrics use behavioural patterns, such as gait or typing and use them in the authentication systems. The behavioural biometrics is considered as non-intrusive. In other words, personal collection is usually not perceived by users.

The biometric-based authentication systems have some concerns which must be addressed. The main issue concernes about security of biometric authentication systems [4]. The security is important for biometrics-based authentication systems because the biometric is permanently associated with a user. In case of a user biometric has compromised, all authentication systems that use this specific user biometric are also compromised. Therefore, the possibility of revoking or cancelling a biometric, if compromised, is a required feature of these systems. Thus, the use of cancellable biometrics have been increasingly adopted to address such issues [4].

The idea of cancellable biometric approach is transform or intentionally distort the original biometric data. Thus, the distorted biometric data (cancellable) are used for user authentication/identification instead of biometric on its original format. Unfortunately, the use of distorted data usually decreases the performance of biometric-based system, due to the complexity level of transformed biometric data is generally higher than original data. Therefore, it is important for a biometric system to support a good trade-off between discrimination capability and non-invertibility (high security) when using cancellable transformations in any biometric modality. Thus, more effective authentication structures are needed to perform these tasks. The use of cancellable biometrics is widely reported in the literature related to physical biometrics use [5, 6]. However, very little has been done to apply cancellable transformations in behavioural biometrics.

As a contribution to this important topic, this paper investigates the performance of different ensemble structures in the context of cancellable behavioural biometrics, more specifically a touch-screen dataset. Therefore, we aim to analyse the performance of these pattern recognition structures in both original and transformed biometric space. The main aim of this article is to analyse the ensemble performance using cancellable data in a behavioural biometric context. A expected result is increase the performance, in accuracy and safety, of user verification, theme increasingly important in current days.

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

This paper is divided into eight sections and it is organized as follows. Section I showed the Introduction, Section II describes the subject background of this paper, while the ensemble systems are described in Section III. The cancellable transformations are described in Section IV and the behavioural biometric TouchAnalytics is discussed in Section V. The Experimental Methodology is explained in Section VI, while the Section VII discusses the results obtained in the experiments. Finally, Section VIII presents the final remarks of this paper.

II. BackGround

A. Behavioural Biometrics

Behavioural biometric is based in the identification and verification of users using activity patterns which can be measured, for example, walking, talking, gesturing, signature recognition and computer usage patterns [3, 7].

Recently, many researchers have been studying behavioural biometrics due to several advantages over physical biometrics. Advantages as, data collection methods can be carried out without the user knowledge and the use of relatively inexpensive technologies like keyboard, mouse and webcam.

The use of biometrics arrises a security problem of biometric traits. Security issues as how the user will change their physical characteristics or computer use patterns in case of being stolen or lost. Therefore, it is important to provide ways to revoke a biometric when stolen or lost.

B. Cancellable Biometrics

Biometric data is user-dependent since the biometric characteristics is unique and it is hard to be changed in case of being stolen or lost. It is difficult, sometimes impossible, to change or to adapt fingerprint or gait characteristics. One of the main problems of security faced by biometric system users is an unauthorized copy of the stored data [8]. Hence, the biometric templates must be stored in a protected way using a protection scheme that possess the following four properties [4].

- Diversity: the secure template must not allow cross-matching across databases, thereby ensuring the user's privacy.
- Revocability: it should be easy to revoke a compromised template, reissuing a new one based on the same biometric data.
- Security: it must be computationally hard to obtain the original biometric template from the secure template. This property prevents an adversary from creating a physical spoof of the biometric trait from a stolen template.
- Performance: the biometric template protection scheme should not degrade the recognition performance of a biometric system.

Unfortunately it is difficult to define a template protection that can satisfy all these characteristics due to the trade-off among them. In this way, several template protection methods have been proposed in the literature [8]. In [4], these methods were broadly divided in two approaches, which are:

1) Biometric cryptosystem: Some public information about the biometric template is stored and it is usually referred to as helper data. Biometric cryptosystems are also known as helper data-based methods. Biometric cryptosystems can be further classified as key binding and key generation systems depending on how the helper data is obtained;
2) Cancellable transformation: In this case, a transformation function (f) is applied to the biometric template (T) and only the transformed template ($f(T)$) is stored in the database. The original template T is hard to be obtained from $f(T)$, which means, the transformed template $f(T)$ can be disposable in case of security issues.

How the authentication process of cancellable biometric templates is performed in transformed space is the main difference between these two approaches, whereas for most biometric cryptosystems, the storage of public biometric information is necessary, in which is applied to retrieve or generate keys (helper data). In this case, the biometric comparisons are performed indirectly through the verification of key validities.

The feature transformation schemes can be further categorized as salting and non-invertible transformations. In the first case, the transformation function f is invertible, while f is (as implied in the name) either non-invertible or hard to invert in the second case. The use of a one-way function, f, that is *easy to compute* (in polynomial time) but *hard to invert* (given $f(x)$, the probability of finding x in polynomial time is small) is the main purpose of using non-invertible transformations. This paper focus on the use of non-invertible transformation functions. Hereafter, the terms transformation function or cancellable transformation will be taken as referring to the non-invertible case.

III. Ensemble Systems

One way to combine the results of different classifiers in biometric data is through the use of ensemble systems, also known as multi-classifier systems or fusion of experts. These systems exploit the idea that different classifiers can offer complementary information about patterns, thereby improving the effectiveness of the overall recognition process [9].

Figure 1 presents a general structure of an ensemble system, which is composed of a set of N individual classifiers (IC_n), organized in a parallel way. The individual classifiers receive the input patterns and send their output to a combination process which is responsible for providing the final output of the system. The individual classifiers may have different subsets of attributes or not.

There are two main issues to consider in the design of an ensemble, which are: the ensemble components, and the combination methods which will be used. In relation to the first issue, the ensemble members are chosen and executed. The appropriate choice of a set of individual classifiers is

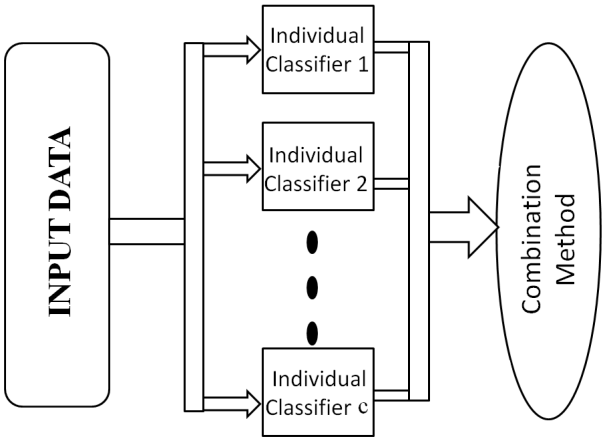

Fig. 1. An illustration of the general framework of an ensemble system

fundamental to the overall performance of an ensemble. The ideal situation would be to choose a set of base classifiers with uncorrelated errors - which would be combined in such a way as to minimize the effect of these failures. In other words, the individual classifiers should be diverse among themselves. Depending on its particular structure, an ensemble can be realised using two main approaches: heterogeneous and homogeneous. The first approach combines different types of classification algorithms as individual classifiers. In contrast, the second approach combines classification algorithms of the same type.

Once a set of individual classifiers has been created, the next step is to choose an effective way of combining their outputs, which is a typical decision-level fusion method. According to [10], the possible ways of combining the outputs of N classifiers in an ensemble depend on which information we obtain from each individual classifier (IC_n). The combination of the outputs diverge from the simplest approach using class labels or rank values to the utilization of more elaborate information, such as support degree D [10].

A. Learning Strategies in Ensemble Systems

The composition of a set of identical classifiers in a ensemble system do not add no gain in terms of performance. Thus, it is important to emphasize that diversity plays an important role in the design of accurate and well-generalized ensembles [10]. The ideal situation, in terms of combining classifiers, would be a set of classifiers with uncorrelated errors (diversity). Diversity in ensemble systems can be reached by using different parameter settings, different training datasets and different classifier types. A stardard way to promote diversity is through the use of learning strategies, also known as learning architectures or simply architectures, that provide different datasets for the individual classifiers of an ensemble system. Most common architectures are: Bagging [11], Boosting [12], Stacking [13] and Voting [14].

B. Ensemble Systems for Cancellable Biometrics

Several transformation functions have been reported for different biometric modalities, face [15], signature [16, 17], fingerprint [5, 6], iris [18], voice [19], among others. However, most of them use single classification/matching algorithms.

In the context of ensemble systems, Canuto et al. [8, 20] applied ensemble systems to cancellable biometrics and achieved very promising results, demonstrating that the use of ensemble systems improves the accuracy of cancellable biometrics. However, they used only physical biometric modalities in the identification process using ensemble systems . Unlike these studies, this paper applies ensemble systems in the context of behavioural biometrics for user verification task.

IV. Cancellable Transformations

The non-invertible transformation functions transform the biometric data in a way that it is computationally hard to get the original form [4]. The application of these functions distorts the original data arising some undesired consequences as high variance, consequently making the user verification more difficult. Thus, verification and authentication systems which use distorted data must provide better performance than systems using original data.

The literature reports that ensemble systems offer better performance than single classifiers [10]. Therefore, in this work we analyse the performance of ensemble systems in cancellable behavioural data. It is applied four transformation functions to the data. The transformation function chose were Interpolation, BioHashing, BioConvolving and Double Sum. All the transformation functions are better described in [21].

A. Interpolation

This technique is based on polynomial interpolations. It consists in generating a new biometric model by extracting function points resulting from the attribute interpolation process. The attributes compose the original biometric model.

Although it is simple, this algorithm makes the inversion of the transformed function difficult, generating a reasonable level of security to the system. Therefore, it is very efficient in satisfying two of the main requirements for transformation techniques, which are simplicity and efficiency at the same time. The following steps describe how a transformation function based on interpolation is applied to a biometric model.

1) Given the original biometric model $\Gamma \in \mathbb{R}^n$, where n is the number of attributes of Γ. A function $f(x)$ is obtained through interpolation of the attributes of the model. It is created one function $f(x)$ for each user. Therefore, in order to have approximate functions of the discretized data, it is important to use a polynomial function with a significant degree g, usually given by the greater degree supported by the system;
2) Within range of the function domain $x \in \mathbb{R} \mid 0 < x \leq n$, a vector of random numbers

is generated for all biometric data $\delta \in \mathbb{R}^d$. δ consists of uniformly distributed pseudo-random numbers, where d is the dimension. The number of coefficients d is given empirically and can interfere in the behaviour of the model (usually $d = n$);

3) The coefficients x of the vector δ are then individually applied to function $f(x)$, generating as output the transformed model M. The revocability of this model depends exclusively on the creation of a new random vector, in case the current vector is corrupted or stolen. The interpolation technique can be adapted to any biometric modality, since they provide a biometric feature vector. Also, the interpolation model depends on some variables, such as the size d (step 2) of the random vector and the degree g of the polynomial interpolator.

B. BioHashing

BioHashing technique has originally used in other biometric modalities, such as fingerprint, palm and face [21]. BioHashing algorithm is characterized by transforming original biometric into a non-invertible binary sequence. This invertible binary sequence is based in a inner product between the original biometric vector and each pseudo-random orthonormal vector $o_i \in \mathbb{R}^n \mid i = 1, ..., m$. Each o_i is obtained using the Gram-Schmidt algorithm with original biometric data as input.

BioHashing technique has originally used in other biometric modalities, such as fingerprint, palm and face [22]. In [23], the authors started the adaptation of BioHashing to iris data. In this work we use the original BioHashing algorithm but in future we will use the adaptation developed by our group, described in [8]. The original algorithm works as:

1) Given the original biometric model $\Gamma \in \mathbb{R}^n$, with n being the number of attributes of Γ, a set of pseudo-random vectors $p_i \in \mathbb{R}^n \mid i = 1, ..., m$ is generated, where m is the number of vectors of dimension n and $m \leq n$;
2) The Gram-Schmidt algorithm is applied in $p_i \in \mathbb{R}^n \mid i = 1, ..., m$ to obtain m orthonormal vectors $o_i \in \mathbb{R}^n \mid i = 1, ..., m$;
3) The inner product between the original biometric vector Γ and each pseudo-random orthonormal vector o_i is calculated, $\langle \Gamma \mid o_i \rangle \mid i = 1, ..., m$, where $\langle \bullet \mid \bullet \rangle$ indicates the inner product operation.
4) It is then created a m-bit Biohashing model through a binary discretization of the values obtained in the inner products, $b = b_i \mid i = 1, ..., m$, where:

$$b_i = \begin{cases} 1, & \text{if } \langle \Gamma \mid o_i \rangle \leq \tau. \\ 0, & \text{if } \langle \Gamma \mid o_i \rangle > \tau. \end{cases} \quad (1)$$

with τ being an empirically determined threshold. In this work we choose $\tau = 0.5$.

The performance of Biohashing function is exclusively dependent on the variable m, which is the number of pseudo-random orthonormal vectors. The number of vectors must be determined empirically and corresponds to the number of attributes for each instance of the transformed dataset.

C. BioConvolving

BioConvolving method was originally proposed for signature [21]. In this method, the transformed functions are created through linear combinations of sub-parts of the original biometric template Γ. Basically, this method divides each original biometric sequences into W non-overlapping segments, according to a randomly selected transformation key d. Then, the transformed functions are obtained by performing a linear convolution between the obtained segments. A general description of BioConvolving is presented as follows.

1) Define the number of segments to divide the original biometric model (W).
2) Select randomly a number $(W - 1)$ of values d_j. The selected numbers have to be between 1 and 99 and they must be ranked in an increasing order. The selected values are arranged in a vector $d = [d_0, \ldots, d_W]$, having kept $d_0 = 0$ and $d_W = 100$. The vector d represents the key of the employed transformation.
3) Convert the values d_j into b_j based on the following function $b_j = round((\frac{d_j}{100} * n)), j = 0, \ldots, W$, where n is the number of attributes and $round$ represents the nearest integer;
4) Divide the original sequence $\Gamma \in \mathbb{R}^n$, into W segments $\Gamma(q)$ of length $N_q = b_q - b_{q-1}$ and which lies in the interval $[b_{q-1}, b_q]$ of the attributes;
5) Apply the linear convolution of the functions $f(\Gamma(q)), q = 1, \ldots, W$ to obtain the transformed function $f = \Gamma(1)*, \ldots, \Gamma(W)$.

A transformed function is therefore obtained through the linear convolutions of parts of original sequence Γ. Due to the convolution operation in step (5), the length of the transformed functions is equal to $K = N - W + 1$, being therefore almost the same of original data. A final signal normalization is applied, oriented to obtain transformed functions with zero mean and unit standard deviation. Different transformed results can be obtained from the same original functions, simply varying the size of segments W or the values of the parameter key d. According to [24], the BioConvolving security approach is based on the fact that a blind deconvolution problem is needed in order to retrieve the original template. Moreover, in [24] it was also shown for signature that even if multiple transformed templates are stolen, it is not possible to retrieve the original template.

D. Double Sum

Double Sum cancellable transformation is a simple method and it consists of summing the attributes of original biometric model with two other attributes of the same sample. In other words, each attribute of original biometric model is transformed into the sum of three attributes randomly chosen. In this case, even if an impostor has access to transformed

data, it will not be possible to define original data from transformed one. The functioning of Double Sum transformation is described as follows:

1) For each sample of original biometric model $\Gamma \in \mathbb{R}^n$, where n is the number of attributes of Γ, we generate a random vector $L \in \mathbb{R}^n$ based on security key k that it can be a general one or it can be defined for each user. This random vector is responsible for distributing the attributes of the biometric samples. In other words, the original biometric model is re-organized based on the random vector L generating a intermediate biometric trait β;
2) Define two vectors $\{C_1$ and $C_2 \in \mathbb{R}^n\}$, where the elements of these vectors, $\{(c_1(i)$ and $c_2(i)) \in \mathbb{N}^* | (c_1(i)$ and $c_2(i)) \leq n\}$, are randomly defined choosing a number between 0 and n. This random choice is also made using two security keys k, one for each vector. These vectors contain the position of the attributes to be summed. We decided to use two vectors because the generation of two sets of data doubles the security of the cancellable transformation, since it is necessary two deconvolution process in order to obtain the original data. This deconvolution requires a solution of a double system with n variables;
3) The attributes of the reorganized biometric data, β, are summed with the attributes whose positions are defined by the random vectors $C_1(i)$ and $C_2(i)$ defined in the previous item. Thus the transformed model is obtained according to the following equation.

$$T(i) = \beta(i) + \Gamma(c_1(i)) + \Gamma(c_2(i)). \quad (2)$$

The double sum method can be considered as non-invertible, since the number of possible combination is very high and it is based on the number of attributes n of the biometric model Γ. Equation 3 calculates the number of possible combination and it is extremely high as well as the computational time needed to perform the inverse process. Using Equation 3 can observe that the solution of a double system of n variable, with $n > 50$ is almost impossible to solve in a feasible processing time.

$$C_s = n!^3 \quad (3)$$

The revocability of this method is guaranteed by the security key. This key is responsible for the reorganization of the original biometric data and for the choice of the original data attributes that will be summed. In case of being lost or stolen, a new transformation model can be created using a different security key k. In this paper, we are using the same security key k and the dimension of the transformed model is the same of the original dataset.

V. TouchAnalytics

Nowadays, data security on smartphones is a evident concern. Accordingly, people and companies are interested in improve the security of their data. The main security issues related with smartphone are user authentication and data theft. Thus, we focus this paper in authentication process.

Currently, the authentication process in a touch-screen device as smartphones is based on pin-numbers (number sequences) or combination of drawings. The length of a pin-number is the main security fault, because pin-numbers are usually composed of 4 numbers, what makes the combination space small. In contrast, the combination of drawings is a useful and intuitive user interface used in the authentication process. Basically, a user connects graphic elements to generate a path and this path is used as password. Unfortunately, the drawing method has some problems. These problems are discussed in [25, 26].

Data from touchscreen interaction is behavioural biometric used, which represents a combination of strokes collected from smartphones usage. This dataset, called TouchAnalytics, was collected by Frank et al. [2]. The authors present how the data was collected, processed and some initial results. This section highlights the most important aspects of this paper.

The TouchAnalytics dataset is composed of 30 attributes and all of them are derived from strokes obtained by 41 users. A stroke is a trajectory encoded as a sequence of vectors $s_n = (x_n, y_n, t_n, p_n, A_n, o_n^f, o_n^{ph}), n \in 1, 2, ..., N$ with location x_n, y_n , time stamp t_n , pressure on screen p_n, area A_n occluded by the finger, finger orientation o_n^f and phone orientation o_n^{ph} (landscape or portrait). The attributes have information about area covered, stroke pressure, direction, velocity and acceleration.

Strokes are composed by horizontal and scrolling (vertical) movements. However, the dataset was divided in two parts: the first with horizontal strokes and the second with scrolling ones. This division was made to verify some similarities in user direction patterns (horizontal and vertical movements).

Moreover, as we use a verification process, the dataset was binarized and a different dataset was created for each user. In other words, the biometric data was splited by user. Strokes belonging to the corresponding user is considered positive ('1') and the others are negative ('0').

As a result of the binarization transform, we have a huge number of negative examples and few positives examples, featuring an imbalanced dataset. Imbalanced datasets produces biased classifiers in the prevalent class. This problem was resolved using a lab-made tool that takes into consideration the number of negative classes and the number of positive examples. $T = \frac{N_p}{N_{nc}}$ is the number of negatives instances that will be randomly selected in each N_{nc} negative class. Where N_p is the number of positive instances and N_{nc} is the number of negative classes. Thus, the number of negatives instances will be $N_{cn} * T$. The selected instances are placed together with the positive instances, which becomes a dataset with the same number of negative and positive instances.

In Frank et al. [2], the authors presented initial results using three different scenarios, which are:

1) Inter Session: The goal is to authenticate users across

multiple sessions performed in the same day.

2) Inter Week: The goal is to authenticate users after in two different weeks (the period of time between these two sessions is one week).
3) Intra Session: All the user data was used in the process, time independently. In this scenario, we used a 10 fold cross-validation process.

In this paper, we will use only the Intra Session experiment.

In [2], they present some results using *k*-NN and SVM (Support Vector Machine) classifiers. In addition, the results use Equal Error Rate (EER) metric which informs when the false acceptance and the false rejection become equal. The EER can be obtained from the ROC curve when the false positive rate (false acceptance) is equal to false rejection rate $(1 - \text{true positive rate})$. According to Frank et al. [2], the mean EER ranges from 0% to 4% across all sessions: Inter Session, Inter Week and Intra Session. The mean EER in Intra Session are 0%. It seems that, within one session, most users do not considerably change their touch behaviour. Inter Session EER reaches from 2% to 3% and Inter Week EER reaches from 0% to 4%. These results indicate that behavioural biometrics (touch data) have good perspectives in practical use. Therefore, we decided to investigate how machine learning ensembles behaved using cancellable data, such as it was promising using original data. In [7] was developed an analysis of *k*-NN and SVM methods using the same 3 experiments (Inter Session, Inter Week and Intra Session) defined by Frank et al. [2].

VI. Empirical Analysis

An empirical analysis is conducted in order to validate the use of ensemble systems in cancellable behavioural biometrics. This investigation will use only the Intra Session experiment. The main reason is that this scenario is the most suitable one to apply elaborated pattern recognition structures as ensemble systems. In addition, for simplicity we decided to use different ensemble structures only in one scenario.

We applied three ensemble structures, in which two of them are emerged from two learning strategies (Bagging and Stacking) and the third one is a traditional ensemble structure using the majority voting as combination method. First of all, for the structures generated by Bagging, we applied two different ensemble sizes, six and twelve individual classifiers. As we noticed, the performance delivered by both structures are very similar, for the other two approaches (Stacking and Voting), we use ensembles with 6 individual classifiers.

We used two different classification algorithm as combination methods for ensembles generated by Stacking, which are: *k*-NN and Logistic Regression. In addition, for the heterogeneous structures (Bagging and Voting), we use SVM and *k*-NN as individual classifiers in half-by-half proportion.

The 10-fold cross-validation methodology was used in empirical analysis. Thus, all results presented in this paper refer to the mean over 10 different test sets. In addiction, an initial investigation was conducted in order to define the values of parameters used by supervised learning algorithms.

The Mann-Whitney statistical test is applied to compare the results from different learning methods. The Mann-Whitney is a statistical pairwise used to compare two samples (set of results). This paper compares EER of ensemble systems applied in cancellable data versions with the EER achieved in original data. For this test, the confidence level is 95% ($\alpha = 0.05$).

The classification algorithms of this investigation were extracted from WEKA [1] package. In general, the algorithms were used with the following parameters: *k*-NN with k = 5; and SVM with polynomial kernel.

VII. Results and Discussion

Tables I and II show the EER values and standard deviation of the ensemble systems in each cancellable transformation, for horizontal and scrolling traits (as described in Section V), respectively.

Homogeneous Bagging with 6 and 12 classifiers, Stacking with 6 classifiers, using *k*-NN and SVM classifiers with *k*-NN as combination method (Stacking_*k*-NN_SVM_*k*-NN (Stack_*k*S*k*)) and Logistics function as combination method (Stacking_*k*-NN_SVM_Logistics (Stack_*k*SL)) and finally, the Voting structure were the ensemble systems used.

The Bagging_6_*k*-NN (Bag_6_*k*) and the Bagging_6_SVM (Bag_6_S) are the bagging structures with 6 *k*-NN and SVM classifiers respectively. Bagging_12_*k*-NN (Bag_12_*k*) and Bagging_12_SVM (Bag_12_S) are the bagging structures with 12 *k*-NN and SVM classifiers.

In this analysis, we also carried out a statistical test, comparing two-by-two ensembles with and without cancellable transformations (columns 3, 4, 5 and 6 against the original dataset in column 2). In Tables I and II, the bold numbers represent the cases where the use of cancellable transformation caused an statistical increase in the accuracy of ensemble systems. In addition, the shaded cells indicate when these improvements is not statistically different (the performance of the original dataset and the transformed ones are statistically similar).

TABLE I
Median Results using Scrooling Traits

Method	Original	Interpolation	BioHashing	BioConvol.	Double Sum
Bag_6_*k*	7.6 ± 4.8	8.9 ± 5.4	32.3 ± 12.6	$\mathbf{3.3 \pm 10.7}$	8.7 ± 5.5
Bag_12_*k*	7.4 ± 4.9	8.6 ± 5.1	32.4 ± 12.4	$\mathbf{3.2 \pm 10.8}$	8.6 ± 5.4
Bag_6_S	9.2 ± 6.4	12.4 ± 8.2	32.4 ± 19	$\mathbf{2.3 \pm 7.8}$	11.9 ± 8.1
Bag_12_S	9.2 ± 6.4	12.3 ± 8	31.2 ± 15.5	$\mathbf{2.1 \pm 7.8}$	11.7 ± 8.3
Stack_*k*S*k*	7.8 ± 5.1	10 ± 6.3	32.3 ± 13.1	$\mathbf{3.4 \pm 10.6}$	10 ± 6.5
Stack_*k*SL	7.2 ± 4.7	9 ± 5.5	32.7 ± 12.7	$\mathbf{3.4 \pm 10.9}$	9.1 ± 5.8
Voting	8.9 ± 6.4	10.9 ± 6.7	32.6 ± 12.6	$\mathbf{3.6 \pm 11}$	11.4 ± 7.5

As can be observed from Tables I and II, ensembles generated by Interpolation and Double Sum functions have similar statistical results when compared with results achieved by the Original datasets, for all ensemble structures. These transformed datasets usually delivered higher EER than the original dataset, but they are not statistically different.

[1] http:www.cs.waikato.ac.nz/ml/WEKA

TABLE II
MEDIAN RESULTS USING HORIZONTAL TRAITS

Method	Original	Interpolation	BioHashing	BioConvol.	Double Sum
Bag_6_*k*	8 ± 4.5	10.6 ± 6.3	32.8 ± 9.8	$\mathbf{0.1 \pm 0.3}$	8.7 ± 4.8
Bag_12_*k*	7.7 ± 4.3	10.3 ± 6.2	32.8 ± 10.2	$\mathbf{0.1 \pm 0.3}$	8.6 ± 4.6
Bag_6_S	11.1 ± 7.3	16.1 ± 9	34.4 ± 17.7	$\mathbf{0.4 \pm 0.5}$	13.3 ± 8.4
Bag_12_S	11.7 ± 7.8	16.1 ± 9.2	33.5 ± 26	$\mathbf{0.3 \pm 0.4}$	13.1 ± 8.5
Stack_*k*S*k*	8.5 ± 4.9	12.1 ± 7.3	34 ± 9.5	$\mathbf{0.2 \pm 0.4}$	10.7 ± 6.3
Stack_*k*SL	7.7 ± 4.5	10.8 ± 6.7	33.1 ± 10	$\mathbf{0.2 \pm 0.4}$	9.6 ± 5.4
Voting	9.7 ± 5.8	13.7 ± 7.1	33.5 ± 9.7	$\mathbf{0.2 \pm 0.4}$	11.9 ± 6.9

Still in Tables I and II, we can see that when using BioConvolving the EER values are statistical better than EER from Original data, for both strokes directions and for all ensemble structures. We can conclude that the use of ensemble systems in behavioural cancellable biometrics do not deteriorate the EER results, when comparing with the EER results achieved by Original data. Our only exception was BioHashing transformation that achieves the worst EER values, when compared with results of the Original dataset, using both scrolling and horizontal strokes and all ensemble structures.

Therefore, these results show that we can use ensemble systems and cancellable transformation in cancellable behavioural biometrics instead of the original data, without deteriorating the performance of the biometric-based authentication systems. This is an important result because the cancellable characteristics offer interesting features to behavioural biometrics context as biometric revocability in case of some security issue.

Comparing the accuracy of different ensemble structures, it can be seen that the accuracy of Bagging structures was higher than the other two structures. This is not an expected result since we believe that the use of heterogeneous structures would lead to an increase in the diversity level of the ensemble systems and, as a consequence, in the accuracy of these systems.

Our previous work [7] analysed the use of a single classifier (*k*-NN and SVM) using the same transformation functions in all Frank's experiments (Inter Session, Inter Week and Intra Session). We can conclude, using these results, that the use of ensemble structures improves the results using Interpolation, BioConvolving and Double Sum functions in scrolling strokes compared with results achieved in our previous work [7]. The BioHashing dataset was the only case in which similar results were obtained, in the scrolling stroke datasets. Using the horizontal strokes, we archive better results than the results showed in [7], for all transformed datasets (Interpolation, BioHashing, BioConvolving and Double Sum).

The results of this paper support the literature when it states that ensemble systems are more powerful than single classifiers. As a future work we will focus on parameter optimization to improve the results achieved with BioHashing dataset. In addition, we believe that the use of soft and multimodal behavioral biometrics can increase even further the performance of the biometric-based systems.

VIII. FINAL REMARKS

In this paper, we performed a comparative analysis of well-known ensemble structures applied to cancellable behavioural biometrics. The touchscreen dataset, provided by [2], on its IntraSession scenario was used in this paper. In this investigation, four cancellable transformations (Interpolation, BioHashing, BioConvolving and Double Sum) were applied to this dataset in order to validate the importance and perspectives of cancellable behavioural biometrics.

The Interpolation and Double Sum results were statistical similar to Original results. The mean EER of Original dataset varies from 7.4% to 11.7%, while in Interpolation dataset, the EER varies between 8.6% and 16.1%. In Double Sum dataset, EER varies from 8.6% to 13.3%. In contrast, BioConvolving provided the best results, being statistically better than results obtained by Original dataset. The mean ERR of BioConvolving dataset varies from 0.1% and 3.60%, and it was statistically superior than all other datasets, for all ensembles structures. The results obtained by BioConvolving are very promising and indicate that the use of cancellable behavioural biometrics can have a positive effect in biometric-based authentication systems. In addition, we have observed that the results achieved in this paper are better than in our previous paper [7], and this shows that the use of ensembles methods are better than using single classifiers.

Through this analysis, we have demonstrated that the use of a transformation function usually provides similar or better performance than the original biometric data, except in BioHashing function. In addition, the use of cancellable behavioural biometrics data brings great opportunities for research, providing the advantages of behavioural biometrics and the security of cancellable biometrics. We can see that simple touch movements, even transformed (distorted), can be used as source of user verification.

As a future work, in order to improve the results obtained in this paper, we will use different classification algorithms, such as MultiLayer Perceptrons. In addition, we can apply optimization technique in order to optimize the cancellable transformation parameters, mainly for BioHashing function. Finally, we can enhance diversity in ensemble systems through the combination of transformation functions, leading to the multi-biometrics context.

REFERENCES

[1] W. Jackson, "Antisec hackers claim theft of military e-mails from booz allen," Internet, Julho 2011, acessado em Novembro de 2011. [Online]. Available: http://gcn.com/articles/2011/07/11/antisec-booz-allen-hack-military-emails.aspx

[2] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication," in *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, 2013, pp. 136–148. [Online]. Available: http://www.mariofrank.net/touchalytics/

[3] K. Revett, *Behavioral Biometrics : a Remote Access Approach*. John Wiley & Sons, Ltd, 2008.

[4] A. K. Jain, K. Nandakumar, and A. Nagar, "Biometric template security," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 113:1–113:17, January 2008. [Online]. Available: http://dx.doi.org/10.1155/2008/579416

[5] C. Lee and J. Kim, "Cancelable fingerprint templates using minutiae-based bit-strings," *Journal of Network and Computer Applications*, vol. 33, no. 3, pp. 236 – 246, 2010.

[6] A. Nagar, K. Nandakumar, and A. K. Jain, "A hybrid biometric cryptosystem for securing fingerprint minutiae templates," *Pattern Recogn. Lett.*, vol. 31, pp. 733–741, June 2010.

[7] M. Damasceno and A. M. Canuto, "An Empirical Analysis of Cancellable Transformations in a Behavioral Biometric Modality," in *13th Conference on Hybrid Intelligent Systems*, 2013.

[8] A. M. Canuto, F. Pintro, and J. ao C. Xavier-Junior, "Investigating fusion approaches in multi-biometric cancellable recognition," *Expert Systems with Applications*, vol. 40, no. 6, pp. 1971–1980, 2013.

[9] A. M. P. Canuto, M. Abreu, L. Oliveira, J. C. X. Jr., and A. Santos, "Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles," *Patt Recogn Letters*, vol. 28, no. 4, pp. 472–486, 2007.

[10] L. I. Kuncheva, *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, Inc, 2004.

[11] L. Breiman, "Stacked Regressions," *Machine Learning*, vol. 24, pp. 49–64, 1996.

[12] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," 1996.

[13] D. H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, p. 241260, 1992.

[14] T. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems*, ser. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2000, vol. 1857, pp. 1–15.

[15] T. Boult, "Robust distance measures for face-recognition supporting revocable biometrics token," in *7thInt. Conf. Autom. Face and Gesture Recog*, 2006, p. 560 to 566.

[16] L. Nanni, E. Maiorana, A. Lumini, and P. Campisi, "Combining local, regional and global matchers for a template protected on-line signature verification system," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3676 – 3684, 2010.

[17] E. Maiorana, P. Campisi, and A. Neri, "Template protection for dynamic time warping based biometric signature authentication," in *Int conf on Digital Signal Processing*, ser. DSP'09. IEEE Press, 2009, pp. 526–531.

[18] S. Kanade, D. Petrovska-Delacretaz, and B. Dorizzi, "Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data," in *IEEE Conf on Computer Vision and Pattern Recognition(CVPR)*, 2009, pp. 120 –127.

[19] W. Xu and M. Cheng, "Cancelable voiceprint template based on chaff-points-mixture method," in *Computational Intelligence and Security, 2008. CIS '08. International Conference on*, vol. 2, 2008, pp. 263 –266.

[20] A. M. P. Canuto, M. C. Fairhurst, F. Pintro, J. C. X. Junior, A. F. Neto, and L. M. G. Gonalves, "Classifier ensembles and optimization techniques to improve the performance of cancellable fingerprint," *International Journal of Hybrid Intelligent Systems*, vol. 8, no. 3, pp. 143–154, 2011.

[21] F. Pintro and A. Canuto, "An Experimental Study of Ensemble Systems on Cancellable Iris," in *2012 Brazilian Symposium on Neural Networks*, 2012.

[22] A. B. Teoh, Y. W. Kuan, and S. Lee, "Cancellable biometrics and annotations on BioHash," *Pattern Recognition*, vol. 41, no. 6, pp. 2034–2044, 2008.

[23] L. Nanni and A. Lumini, "Empirical tests on biohashing," *Neurocomputing*, vol. 69, no. 16-18, pp. 2390–2395, 2006.

[24] E. Maiorana, M. Martinez-Diaz, P. Campisi, J. Ortega-Garcia, and A. Neri, "Template protection for hmm-based on-line signature authentication," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008, pp. 1–6.

[25] M. Martinez-Diaz, C. Martin-Diaz, J. Galbally, and J. . F. ierrez, "A comparative evaluation of finger-drawn graphical password verification methods," in *12th International Conference on Frontiers in Handwriting Recognition*, 2010.

[26] X.-Y. Liu, H.-C. Gao, L.-M. Wang, and X.-L. Chang, "An enhanced drawing reproduction graphical password strategy," *Journal of Computer Science and Technology*, vol. 26, no. 6, pp. 988–999, 2011.