An evaluation of a three-modal hand-based database to forensic-based gender recognition

DA COSTA ABREU, Marjory <http://orcid.org/0000-0001-7461-7570> and GONCALVE, Julliana Caroline

Available from Sheffield Hallam University Research Archive (SHURA) at:

https://shura.shu.ac.uk/25634/

An evaluation of a three-modal hand-based database to forensic-based gender recognition

Julliana Caroline Gonçalves de A.S.M, Márjory Da Costa-Abreu[1]

[1] Federal University of Rio Grande do Norte (UFRN)
Rio Grande do Norte, BR.

juhcaroline@ppgsc.ufrn.br, marjory@dimap.ufrn.br

Abstract. *In recent years, behavioural soft-biometrics have been widely used to improve biometric systems performance. Information like gender, age and ethnicity can be obtained from more than one behavioural modality. In this paper, we propose a multimodal hand-based behavioural database for gender recognition. Thus, our goal in this paper is to evaluate the performance of the multimodal database. For this, the experiment was realised with 76 users and was collected keyboard dynamics, touchscreen dynamics and handwritten signature data. Our approach consists of compare two-modal and one-modal modalities of the biometric data with the multimodal database. Traditional and new classifiers were used and the statistical Kruskal-Wallis to analyse the accuracy of the databases. The results showed that the multimodal database outperforms the other databases.*

1. Introduction

Biometrics refers to the use of physical and behavioural characteristics to identify a person. Characteristics like gender, age, ethnicity can provide information that helps to determine the identity. This type of biometric is labeled *soft-biometrics* [Miguel-Hurtado et al. 2016]. Traditional systems that combine hard-biometrics (physical characteristics) within unimodal soft-biometrics can still suffer problems like non-universal biometrics traits, or insufficient accuracy caused by noisy data. The fusion of multiple sources of biometrics helps to minimise these problems, contributing to the flexibility in user authentication and providing the search on a large biometric database computationally efficient [Jagadiswary and Saraswady 2016].

Thus, we propose a feature-level fusion of three hand-based databases (handwritten signature, keyboard dynamics and touch screen dynamics) for gender recognition and analyse the accuracy of the databases by comparing the multimodal performance with three two-modal databases (keytouch, sigkey and sigtouch) and with the three single modalities.

The paper is organised as follows: Section 2 will present a background to understand modalities performance in similar approaches. Section 3 will present the feature acquisition process. Section 4 will present our results for feature-level fusion by the use of classification algorithms and applying a statistical test and, finally, Section 5 will present our final remarks about this work.

2. Hand-based behavioural biometrics data: gender recognition

The sections 2.1, 2.2 and 2.3 will present the most recent works focusing on the feature-level fusion of keystroke dynamics and handwritten signature.

2.1. Keyboard Dynamics

The modality was first investigated back in the 1970' and started with the popularisation of personal computers.

[Idrus et al. 2013] proposed a benchmark test suite using a database containing biometrics like keystroke dynamics and other soft biometrics traits. A total of 110 participated in the experiment, and 100 samples were collected per user. The Support Vector Machine was used to classify the data and gave as a result of a recognition rate ranging between 66.4% and 78.2%.

In [Tsimperidis et al. 2018], the proposed work aimed to find which of keystroke dynamics features were most proper for the gender classification task. For this, a new dataset of 117 participants was collected using information collected during the daily user's usage of their computers. A total of 10.000 features were collected, and the classifiers like Support Vector Machine, Multilayer Perceptron Neural Network and Radial Basis Function network were used. The results obtained an accuracy of over 95%.

2.2. Touch screen Dynamics

Despite the mobile devices being now more popular than the keyboard devices, this new modality still presents many challenges. The work [Buriro et al. 2016] aimed to investigate the estimation of soft attributes on smart mobile devices. For this, a scheme was designed to estimate age, gender and operating-hand using time-based keystroke features collected when the user entered his/her secret PIN/password. The proposed method obtained 82.5% of accuracy for gender recognition.

The software and protocol proposed by [Miguel-Hurtado et al. 2016] to use swipe gesture data to predict the user's sex was done in similar modalities such as keystroke and mouse biometrics. First, the BestFirst feature selection technique was used and then, the Naive Bayes, Logistic Regression, Support Vector Machine and Decision Tree classification algorithms. The preliminary results showed the possibility of sex recognition with a 78% accuracy rate using two different swipe directions.

2.3. Online Handwritten Signature

The growing tendency to use devices to authenticate individuals justifies the works presented as follows.

In [Bhattacharyya et al. 2018], a feature extraction method was used to gender discrimination. Hindi off-line handwritten signatures were considered as an input for the proposed framework which had two parts. First, features such as roundness, kurtosis, mean, standard deviation were collected. Then, the features were divided into two sets with the Euler numbers. The Support Vector Machine(SVM) and Back Propagation Neural Network (BPNN) were used to classify the data. An improvement of 4.7% in gender classification was obtained in the training dataset when the Euler number is added as a feature.

[AbdAli and Putz-Leszczynska 2014] presented the age and gender influence on verification factors. A database containing 40 signatures from 50 users where 20 were genuine signatures and 20 were forgery signatures for each user. A total of 32 features

were collected, and then the Minimum Redundancy Maximum Relevance (MRMR) feature selection method was used. An algorithm called Global Classifier based on the universal forgery idea, was used to distinguish forgeries and genuine signatures. The results obtained an Equal Error Rate (EER) of 7% for the female gender and 6.25% for the male gender.

3. Methodology

As we can see from the previous section, no work fuses keystroke, touch screen and handwritten signature focusing on gender recognition. The experiments were carried out with a Brazilian hand-based behavioural biometrics database [Da Silva et al. 2016] and consists of an online handwritten signature as well as keyboard keystroke dynamics and touch screen keystroke dynamics.

For the experiment, the data were collected from 76 individuals, being 55 male and 21 female. To collect both keyboard and touch screen data, the individuals typed the same mix of 20 common words in Brazilian Portuguese, including English cognates. The reason for choosing each word can be found in [Da Silva et al. 2016], and the list of the words in Table 1.

america	jesus	homem	zoom
internet	mouse	carro	selfie
coisa	felicidade	porque	ultimo
normal	pequeno	case	mulher
fazer	primeiro	video	cuba

Table 1. Keystroke Dynamics Words

Sigdist	VELXZERO	TOTALTIME	TIME3
AVXV	AVPRESS	YSIZE	DUR2
VEL1Y	PIXELCENX	XSIZE	DUR1
VEL2	PTD	INITDIR	DIST1
VEL4X	HWRATION	VELCOR	DCHANGE
VEL5	SET	PIXELCENY	DUR4
VELYZERO	AVYV	VEL1X	VEL3
VEL4Y	VEL6	TIME5	DUR3

Table 2. Handwritten Signature Features

To obtain three samples of each user, we have considered three occurrences of the same digraph in different words and two or three occurrences according to the proximity of keys in the keyboard. A total of 14 digraphs were added in the keystroke database: ME, ER, RI, IC, CA, IM, IR, SE, MO, OO, DE, EL, RM and UE. From each digraph, we extracted the RP-latency feature and hold time of both keys in a physical keyboard. For the virtual keyboard were extracted RP-latency and hold time of the digraph's second key.

The samples of the handwritten signature database were collected by asking the user to sign their full names three times. The signature features were obtained considering the x and y coordinates, the pressure of the pen and time in each point of the signature. In total, 34 features were considered to form the database. The formulas can be found in [Da Silva et al. 2016], and the features are shown in Table 2.

For analysing the impact of database fusion, we use three classifiers which is the same used in [Da Silva et al. 2016] (KNN, SVM, MLP) can be found on Weka toolkit [Hall et al. 2009], the WiSARD [Rangel et al. 2016] and LibSVM [Chang and Lin 2011]. Recent biometric recognition works [Cotta et al. 2018], [Lusquino Filho et al. 2018] indicated a significant performance of these algorithms.

For each classifier, the stratified percentage split (60%) was used due to the number of classes of the problem[Hall et al. 2009]. To test our hypothesis we used the nonparametric statistical test *Kruskal-Wallis* [Bewick et al. 2004]. This test verifies if, be-

tween the samples of the same group, there is statistical evidence that these samples come from the same distribution.

4. Results

Since our goal is to investigate the three-modal dataset accuracy for gender recognition, firstly, we have tested the five previously cited classifiers in seven different configurations, four combinations (sigkeytouch = handwritten signature + keyboard + touch, keytouch = keyboard + touch, sigkey = handwritten signature + keyboard and sigtouch = handwritten signature + touch) and three single modalities (keyboard, touchscreen and handwritten signature). Each classifier were executed 30 times, for all configurations and a confidence value of 95% (p-value = 0.05) was used.

Table 3 presents the average accuracy of the classifiers for the configurations.

CLASSIFIER	SIGKEYTOUCH	KEYTOUCH	SIGKEY	SIGTOUCH	KEYBOARD	SIGNATURE	TOUCH
SVM	79.12	67.03	64.83	71.42	**83.51**	75.82	68.13
LibSVM	79.12	63.73	63.73	63.73	79.12	79.12	79.12
KNN	81.31	70.32	62.63	65.93	75.82	80.21	73.62
MLP	69.45	59.34	67.14	**70.76**	69.35	**72.74**	67.91
WiSARD	81.31	64.83	67.03	64.83	76.92	79.12	76.92

Table 3. Average classifiers accuracy rates in percentage for all the modalities

According to the results presented in Table 3, we can observe that excepting for keyboard, sigtouch and signature databases using MLP and SVM classifiers, the databases had lower average performance compared to sigkeytouch database, indicating that sigkeytouch database has a better accuracy of recognition.

Observing the keyboard, sigtouch and signature databases, we verify that users 108 and 109 of signature database and 13 users (44, 45, 51, 55, 56, 61, 64, 65, 81, 88, 97, 102 and 110) of touch database users may have contributed to the sigkeytouch classifier performance decrease, adding redundant data.

The Table 4 shows the p-value for SKT (sigkeytouch) *versus* the other configurations (KT - keytouch, SK - sigkey, ST - sigtouch, KBD - keyboard, SIG - signature and TC - touch).

	P-value					
Classifier	KT vs. SKT	SK vs. SKT	ST vs. SKT	KBD vs. SKT	SIG vs. SKT	TC vs. SKT
SVM	<**0.0001**	<**0.0001**	<**0.0001**	0.115	0.115	<**0.0001**
LibSVM	<**0.0001**	<**0.0001**	<**0.0001**	1	1	1
KNN	<**0.0001**	<**0.0001**	<**0.0001**	**0.002**	0.115	<**0.0001**
MLP	<**0.0001**	0.063	0.164	0.595	**0.011**	0.094
WiSARD	<**0.0001**	<**0.0001**	<**0.0001**	<**0.0001**	0.109	<**0.0001**

Table 4. P-value for SKT versus the other configurations

As we mentioned before, the p-value is used to indicate the statistical difference; in this case, a p-value < 0.05 implies different accuracies between two databases. From this, only databases with this p-value are used. The average of each database is compared to the average of other databases and the number of times that the average is higher ($+$) and equal ($=$) than the other averages is counted. The process is repeated for all databases.

	Databases						
	SIGKEYTOUCH	KEYTOUCH	SIGKEY	SIGTOUCH	KEYBOARD	SIGNATURE	TOUCH
+	**59.94%**	3.33%	3.33%	26.64%	26.64%	53.28%	29.97%
=	**39.96%**	26.94%	36.63%	36.63%	**49.95%**	**39.96%**	**43.29%**
TOTAL ACCURACY	**99.9%**	30.27%	39.96%	63.27%	73.59%	93.24%	73.26%

Table 5. Accuracy in percentage for each database when the database accuracy was higher than the others (+), the same as the others (=) and the total accuracy in percentage for each database

For the Table 5 we can observe that for the + case the sigkeytouch and signature had the highest accuracies 59.94% and 53.28% while the databases presented the same accuracy 39.96% for the = case. However, even a similar scenario between the sigkeytouch and signature databases, the multimodal database had the highest total accuracy (99.9%).

5. Final Remarks

In this work, we analysed the impact of the multimodal hand-based database for gender recognition. For this, the multimodal database was compared with two-modal and one-modal modalities. We have tested five classifiers and confirmed the accuracy enhancement by applying a statistical test.

The classifiers presented a varied performance being the SVM and WiSARD the two with the best overall performance. Also, we observed that there was no significant variation among the classifier performances for the multimodal database, excepting for MLP classifier.

For the statistical test, it is possible to see that keyboard, signature and touch databases have the same accuracy or even higher than the sigkeytouch database for the = case. However, the total accuracy of the multimodal database is still higher than the others, recognising the gender in 99.9%. The feature space dimensionality increase due to the fusion of the biometric sources and the overlap reduction of the feature distributions due to the dissimilarity of individuals explain this behaviour.

We understand the importance of our work to the analyses of the impact of the three-modal biometric database for gender recognition. As future work, we aim to investigate the impact of feature selection on the accuracy of the databases in order to know the influence of each modality set of features on gender recognition.

Acknowledgments

References

AbdAli, S. and Putz-Leszczynska, J. (2014). Age and gender-invariant features of handwritten signatures for verification systems. In *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2014*, volume 9290, page 929021. International Society for Optics and Photonics.

Bewick, V., Cheek, L., and Ball, J. (2004). Statistics review 10: further nonparametric methods. *Critical care*, 8(3):196.

Bhattacharyya, S., Sarkar, T., et al. (2018). Euler number based feature extraction technique for gender discrimination from offline hindi signature using svm & bpnn classifier. In *2018 Emerging Trends in Electronic Devices and Computational Techniques (EDCT)*, pages 1–6. IEEE.

Buriro, A., Akhtar, Z., Crispo, B., and Del Frari, F. (2016). Age, gender and operating-hand estimation on smart mobile devices. In *2016 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE.

Chang, C.-C. and Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27.

Cotta, K. P., Ferreira, R. S., and França, F. M. (2018). Weightless neural network wisard applied to online recommender systems. In *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 348–353. IEEE.

Da Silva, V. R., Silva, J. C. G. d. A., and Da Costa-Abreu, M. (2016). A new Brazilian hand-based behavioural biometrics database: Data collection and analysis. In *The 7th IET International Conference on Imaging for Crime Detection and Prevention (ICDP-16)*, page 1.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.

Idrus, S. Z. S., Cherrier, E., Rosenberger, C., and Bours, P. (2013). Soft biometrics database: A benchmark for keystroke dynamics biometric systems. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, pages 1–8. IEEE.

Jagadiswary, D. and Saraswady, D. (2016). Biometric authentication using fused multi-modal biometric. *Procedia Computer Science*, 85:109–116.

Lusquino Filho, L., França, F. M., and Lima, P. M. (2018). Near-optimal facial emotion classification using a wisard-based weightless system. In *ESANN*.

Miguel-Hurtado, O., Stevenage, S. V., Bevan, C., and Guest, R. (2016). Predicting sex as a soft-biometrics from device interaction swipe gestures. *Pattern Recognition Letters*, 79:44–51.

Rangel, F., Faria, F., Lima, P. M. V., and Oliveira, J. (2016). Semi-supervised classification of social textual data using wisard. *ESANN. Citeseer*.

Tsimperidis, I., Arampatzis, A., and Karakos, A. (2018). Keystroke dynamics features for gender recognition. *Digital Investigation*, 24:4–10.