Human-Assisted Signature Recognition based on Comparative Attributes

Derlin Morocho
Electric and Electronic Department
Universidad de las Fuerzas Armadas ESPE
Sangolquí, Ecuador
dmorocho@espe.edu.ec

Aythami Morales, Julian Fierrez, Ruben Vera-Rodriguez
BiDA Lab, Biometrics and Data Pattern Analytics Laboratory
Universidad Autónoma de Madrid
Madrid, España
{aythami.morales, julian.fierrez, ruben.vera}@uam.es

***Abstract*— This work analyzes the performance of comparative attributes labeled by humans for handwritten semi-automatic signature recognition. Despite the large deployment of automatic systems, humans have still an important role in many tasks related to handwritten signature recording and verification. How humans can help to improve these processes is a primary aim of different research lines. Comparative attributes try to exploit the abilities of humans to extract discriminant information of the signatures. Instead of absolute attributes (e.g. is this stroke vertical?), the comparative attributes offer richer responses (e.g. how vertical is this stroke?). In this work we present a new semi-automatic signature labeling interface inspired by Forensic Document Examiners (FDE). Fifteen comparative attributes have been labeled by 21 laymen, where each one carries out the labeling of 28 signatures from 130 users of the publicly available corpus BiosecurID database. Through the manual labeling, a new Bio-HSL (Biometric-Handwritten Signatures Labelling) database is generated, which contains 4,968,600 signature attributes. The results show that comparative attributes outperform absolute attributes for semi-automatic signature recognition with Equal Error Rates ranging from 5.5% for random comparisons to 21.2% for simulated forgeries.**

***Keywords*— signature recognition, biometrics, human in the loop, forensics.**

I. INTRODUCTION

Biometric recognition is a very broad field of research, including a large number of research area, most of them focused on the analysis and evaluation of human physiological and behavioral traits for automatic recognition applications.

The handwritten signature is a popular biometric modality because of its social acceptance in the legal and commercial fields, and it is one of the oldest methods to certify the identity of an individual or to give authenticity to legal documents [1]. As shown in Fig. 1, the applications of signature recognition are diverse and include: banking, product sales, parcel/courier delivery, and public notary among other. In most of these applications, the human tasks are mainly related to recording and verifications are usually done offline by Forensic Document Examiners (FDE) only in case of complaints or disputes.

What human actions can assist Automatic Signature Verification Systems (ASV) is the main question investigated in the present work.

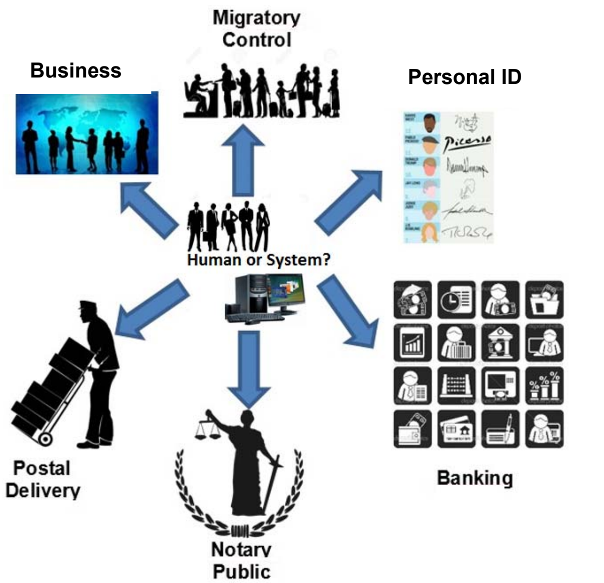


Fig. 1. Applications of signature recognition.

Layman's intervention in ASVs can be done in multiple levels or phases of the biometric system as shown in Fig. 2. Possible layman interventions include: quality assessment to eliminate poor quality samples, feature annotation, sample sorting, support decision, among others. However, there is a large room for research on this area. The performance of the layman in the different signature verification tasks has been undervalued [2] and recent studies suggest their capacity to improve automatic systems [3].

The performance of FDE in signature identification tasks has been studied in recent works [2,4-6]. The performance of experts is comparable or superior to the performance of Automatic Signature Verifiers (AVS). The large experience and specialized training of FDE increase their ability to recognize forged signatures. However, experts only act upon selected cases in which there are important concerns related to the identity of the signer, usually legal prosecutions. In most of the cases, people without signature recognition experience are in charge of the signature collection process.

2379-2140/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDAR.2017.373

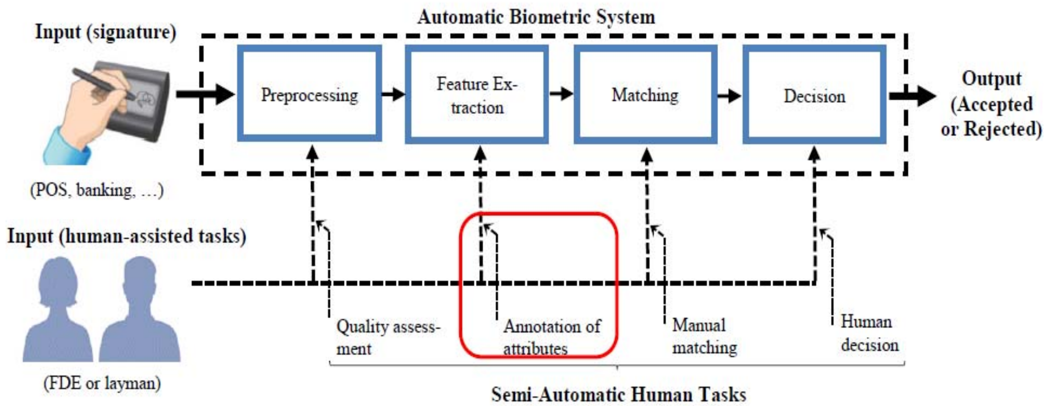


Fig. 2. Human-assisted signature recognition basic scheme [9].

Their task is mainly related with the supervision of the correct acquisition of the signature samples, without specific actions related to signature verification or forgery detection. The performance of people without FDE experience has attracted the interest of researchers in last years [3,7-9]. Human performance was measured by analyzing the opinion of the laymen based on visual comparisons from a genuine set of signatures and unmarked samples including genuine and counterfeit signatures using crowdsourcing platforms.

These studies have made it possible to establish a human performance baseline, but there is a lack of knowledge in factors that influence human scores. Previous studies [9,10,11] suggest that laymen perform better labeling than classifying. The labeling of signatures attributes inspired on FDE has a more guided component that allows to improve the inherent capacity of laymen to detect forgeries. In [11], researchers propose signature attributes labeled by laymen as features for Support Vector Machines classification algorithms. The results reported in [11] with attributes labeled by 11 laymen ranged from Equal Error Rates of 6.89% for random comparisons to 24.22% for simulated forgeries. The attributes proposed in [11] were based on absolute labels related with features of the signature. As an example, the shape of the signature was categorized into 6 different labels: rounded strokes, vertical strokes, horizontal strokes, calligraphic model, vertical and horizontal strokes, or unknown. However, the signature complexity makes difficult to categorize certain samples. The presence of flourish and text, together with the high subjectivity of the task, make the labeling task quite challenging.

The present work explores the human performance in signature recognition based on comparative attributes instead of absolute attributes. The idea is to replace the absolute labels by ranges. For example, a signature is not labeled as horizontal or vertical. Instead of that, comparative attributes try to label how vertical or horizontal is it.

The main contributions of this work are twofold:

i) The evaluation of comparative attributes labeled by laymen as features for semi-automatic signature recognition.

ii) A new database of signature attributes including more than 4 million labels made by 21 different laymen.

The rest of the work is organized as follows: Section II summarizes related works. Section III describes the proposed comparative attribute-based signature recognition. Section IV reports the experiments and results. Finally, Section V summarizes the conclusions and future works.

II. RELATED WORKS

Research studies in the field of biometrics applied to human-assisted systems allow us to identify the real abilities of the human for recognition in comparison with the capabilities of the automated systems [3,6,7,9,12-16]. The use of human annotations in semi-automatic biometric recognition systems has provided encouraging results in the literature. The annotation of attributes performed by humans has emerged as a way to improve automatic systems in face recognition [8,12,16,19], soft biometrics [16-19] or signature [10,11].

In [16-19], soft biometrics were used for the description of human faces and body attributes. The idea underlying those work focuses on the fact that people naturally use labels and physical attribute estimates to describe other people. The results presented in [16] concluded that the absolute body descriptions to identify individuals resulted in an accuracy of identification of under 50%, because the absolute labels proved to be a bad form of description, bound to subjectivity and interference. On the other hand, the comparative labels proved to be less subjective than the traditional forms of description and are preferred by the majority of specialists, obtaining for this particular work an accuracy up to 95%.

Research works developed in [9-11] explored the recognition of human-assisted signature, including performance of humans in signature recognition based on absolutes attributes labeled by laymen. Their results suggest the potential of these recognition schemes in applications involving human intervention. In comparison with automatic algorithms based on static (offline) and dynamic (online) features [20], the attribute-based matching showed its potential to complement existing automatic approaches.

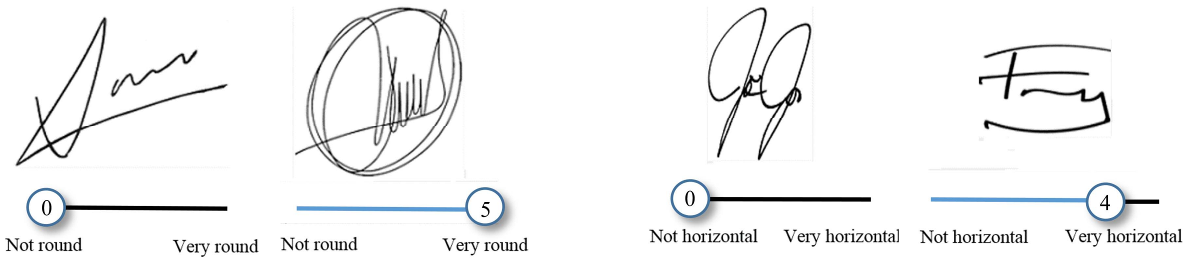


Fig 3. Example of the two different attributes: Roundness (left) and Horizontality (right).

III. Experiments

The experiments explore the potential of comparative labels for signature recognition and their utility in semi-automatic signature recognition schemes.

A. Comparative Attributes

Human performance obtained in [3,7,9,10,12] suggests that laymen find it difficult to correctly recognize the authenticity of signatures. However, it is well accepted that FDE can achieve competitive results based on their specialized training and experience. Here, we propose to analyze comparative attributes of signatures inspired by FDE training, which are manually labeled by laymen.

In this work we evaluate the performance of 17 signature attributes divided into: 11 comparative attributes with labels ranging from 1 to 5 (see Fig. 3), 2 absolute attributes with binary labels, and 4 measures. The attributes, inspired in FDE and previous works [9,11], include features related to the shape, punctuation, retouches and loops, both for the flourish and the text. A handwritten signature labeling application has been developed, including those attributes and the image of one signature each time. Table I and Table II show the attributes analyzed in this work (clearer visual examples of similar attributes can be found in [11]). The human intervention has a large subjectivity caused by the personal perception and experience of each annotator. Therefore, the application offers a brief introduction to each of the attributes for each of the laymen.

B. Experimental Protocol

The signature database used is the BiosecurID corpus [21], which comprises 28 signatures from 130 different signers acquired in 4 different sessions. The 28 signatures are divided into 16 genuine signatures and 12 simulated forgeries. The annotation of the attributes was made by 21 M.Sc. students (from Universidad de las Fuerzas Armadas ESPE in Ecuador) without any previous experience on FDE analysis. No information about the authenticity (genuine or imitation) of the samples was provided to the annotator and all signatures were analyzed separately and randomly.

TABLE I. Comparative Attributes (C1-C5) And Absolute Attributes (A1-A2)

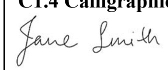
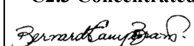
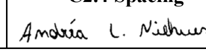
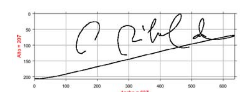
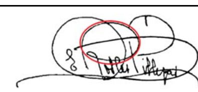
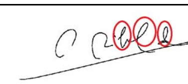
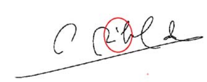
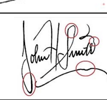
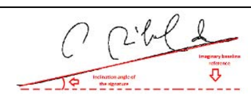
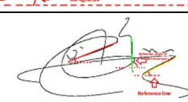
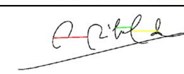
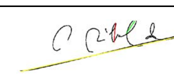
Shape of the signature C1	**C1.1 Vertical** 	**C1.2 Round** 
	C1.3 Horizontal 	**C1.4 Calligraphic** 
Graph Order C2	**C2.1 Clear** 	**C2.2 Confused** 
	C2.3 Concentrated 	**C2.4 Spacing** 
Aspect Ratio C3		
Flourish Loops C4		
Text Loops C5		
Punctuation Marks A1		
Signature Retouch A2		

TABLE II. Measured attributes (D1-D4)

Orientation D1	
Character Tilting D2	
Character Spacing D3	
Stroke Lengths D4	

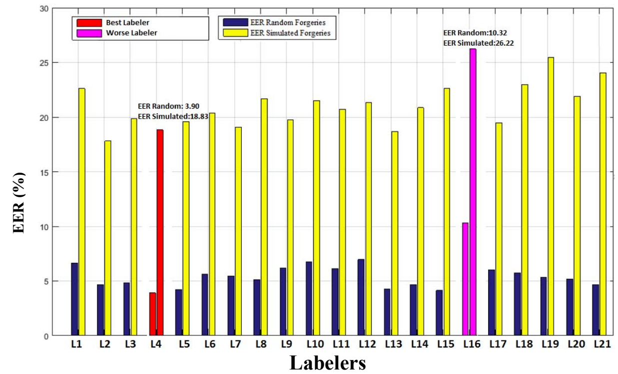


Fig. 4. Layman performance: The best and the worst annotator are highlighted with different color

A new Bio-HSL (Biometric-Handwritten Signature Labels) database, which contains labeled data by those 21 laymen, was generated through the labeling application. Each layman has labeled 13 attributes (C1-C5,A1,A2) plus 10 measures (D1 plus 3 measures for each D2, D3 and D4) for each of the 3,640 signatures in the database (130×28). Therefore, the database comprises 1,758,120 attributes: 23 attributes × 28 signatures per signer × 130 signers × 21 laymen. The experiments are divided into two categories:

- **Random Forgery**: the model of the user is evaluated using genuine samples from other users (different to the owner) as impostor attacks (simulation of users who try to spoof the identity of other users with their own signature).
- **Simulated Forgery**: also known as skilled forgeries, the model of the user is evaluated using imitations made by other users (with different level of skill, see the database description for details [21]).

For the experimental protocol we have employed the protocol proposed in [11] for a fair comparison of the results. We use the 4 genuine samples of the first session as training set. The distance between the training matrix (matrix with 4×23 attributes) and a given signature (vector with 1×23 attributes) is calculated using the Manhattan distance proposed in [11]:

$$d = \sum_{i=1}^{M} |f_i - \bar{g}_i| / \sigma_i \quad (1)$$

where f_i is the i-th feature of the given signature, $\bar{g}$ is the average of the training matrix for the measure features or the mode of the training matrix for the comparative attributes, M is the number of attributes (23 in our case) and σ_i is the standard deviation of the i-th feature from the training matrix. The scores are them normalized using the hyperbolic tangent [9]:

$$d' = \frac{1}{2}\left(\tanh\left(0.01\left(\frac{d-\mu_d}{\sigma_d}\right)\right)+1\right) \quad (2)$$

where μ_d and σ_d are the mean and standard deviation of the scores obtained by cross-validation with training samples.

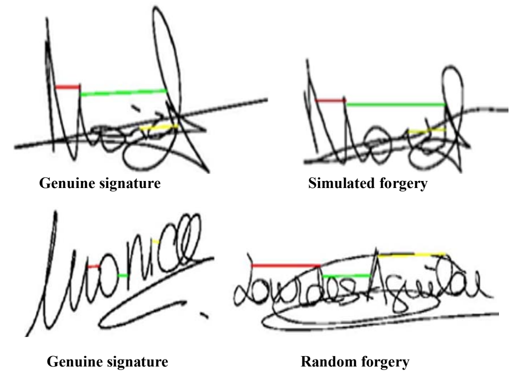


Fig. 5. Analysis of attribute D3 for genuine signatures versus simulated forgery (upper) and random (down).

C. Results

The performance in terms of Random EER and Simulated EER allows us to determine the potential of comparative attributes. Fig. 4 shows the human (layman) performance obtained according the protocol explained in previous sections. The results show the performance of each of the laymen in terms of average EER for random and simulated signatures. The results show that 38% of the laymen present a Random EER under 5%, where the best EER is 3.90% and the worse EER is 10.32%. For simulated forgeries we have that 52% of layman have a Simulated EER of less than 21% with the best EER of 18.83% and the worst EER of 26.22%.

The analysis of the results suggest that the characteristics measured are more discriminating for random signatures than for simulated signatures, Fig 5 shows some examples of D3 measure for random and simulated forgeries.

Table III shows the results obtained averaging all the annotator performances and the results obtained in previous works using the same database and experimental protocol. The table includes the performance of two state-of-the-art ASV systems based on both online/dynamic information (Dynamic Time Warping) and offline/static information (Local Directional Patterns and Support Vector Machine).

The results suggest the superior performance of comparative attributes against absolute attributes. However, there is a large gap with automatic systems, especially with online ASV. It is important to highlight that performance of attribute-based methods is highly related with the laymen. The results showed in Table III are averaged from 21 laymen performances in the case of comparative labels and 11 laymen in the case of absolute attributes (different laymen in each case). The differences in the performances (1% in random comparison and 3% for simulated forgeries) are statistically significant for the 384,930 scores calculated for the random comparisons (12×130×21+129×130×21) and the 65,520 scores calculated for the simulated forgeries (12×130×21×2).

TABLE III. PERFORMANCE (EER%) OF THE **PROPOSED ATTRIBUTES** USED FOR ASV = AUTOMATIC SIGNATURE VERIFICATION

System	EER [%]	
	Random	Simulated
ASV based on online features [11]	1.9	6.9
ASV based on offline features [11]	4.72	20.27
Semi-Automatic Absolute Attributes [11]	6.89	24.22
Semi-Automatic Comparative Attributes	**5.57**	**21.20**

IV. CONCLUSIONS AND FUTURE WORK

This work explores the recognition of human-assisted signature based on comparative attributes. The results suggest the potential of these recognition schemes in applications involving human interventions. The results suggest that comparative labels offer more discriminant information than absolute attributes, even if the uncertainty is greater (more labels increases the number of possible responses). However, there is plenty of room for further research in this area and the number of open questions is large: What is the consistency of the annotated attributes applied to another database? Can the comparative attributes improve the automatic recognition in combined schemes? What is the stability of the annotators for different number of signatures?

ACKNOWLEDGMENT

A.M. is supported by a Juan de la Cierva Incorporacion (JdC-I) contract (IJCI-2015-24742) from the Spanish MINECO. This work was partially funded by CECABANK and project CogniMetrics (TEC2015-70627-R) from Spanish Government (MINECO/FEDER).

REFERENCES

[1] D. Impedovo and G. Pirlo “Automatic signature verification: The state of the art”, IEEE Trans. on Systems, Man, and Cybernetics (Part C), vol. 38(5), pp. 609-635, 2008.

[2] M. I. Malik, M. Liwicki, A. Dengel, B. Found, “Man vs. Machine: a comparative analysis for forensic signature verification”, Proc. of the 16th Int. Graphonomics Society Conf., 2013, pp. 9-13.

[3] D. Morocho, M. Proaño, D. Alulema, A. Morales, J. Fierrez, “Signature Recognition: Human performance analysis vs. automatic system and feature extraction via crowdsourcing”, Proc. Mexican Conf. on Pattern Recognition, Springer International Publishing, 2016, pp. 324-334.

[4] L. S. Oliveira, E. Justino, C. Freitas, R. Sabourin, “The graphology applied to signature verification”, Proc. 12th Conf. of the Int. Graphonomics Society, 2005, pp. 286-290.

[5] J. Coetzer, B.M. Herbst, J.A. Du Preez, “Off-line signature verification: A comparison between human and machine performance”, Proc. 10th Int. Workshop on Frontiers in Handwriting Recognition, 2006, pp. 481-485.

[6] M. I. Malik, M. Liwicki, A. Dengel, “Part-based automatic system in comparison to human experts for forensic signature verification”, Proc. Int. Conf. on Document Analysis and Recognition, 2013, pp. 872-876.

[7] D. Morocho, A. Morales, J. Fierrez, R. Tolosana, “Signature recognition: establishing human baseline performance via crowdsourcing”, Proc. 4th Int. Workshop on Biometrics and Forensics, 2016.

[8] L. Best-Rowden, S. Bisht, J. Klontz, A. K. Jain. “Unconstrained Face Recognition: Establishing Baseline Human Performance via Crowdsourcing”, Proc. of Int. Conf. on Biometrics, 2014, pp. 1-8.

[9] D. Morocho, A. Morales, J. Fierrez and R. Vera-Rodriguez, “Towards human-assisted signature recognition: improving biometric systems through attribute-based recognition”, IEEE Int. Conf. on Identity, Security and Behavior Analysis, 2016.

[10] D. Morocho, J. Hernandez-Ortega, A. Morales, J. Fierrez, J. Ortega-Garcia, “On the evaluation of human ratings for signature recognition”, IEEE Int. Carnahan Conf. on Security Technology, 2016.

[11] A. Morales, D. Morocho, J. Fierrez, R. Vera, “Signature authentication based on human intervention: performance and complementarity with automatic systems”, IET Biometrics, vol. 6(4), pp. 307-315, 2017.

[12] D. Reid and M. Nixon, “Human Identification using Facial Comparative Descriptions”, Proc. Int. Conf. on Biometrics, 2013.

[13] N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. I. Khalil, M. N. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. Roure-Alcobe, J. Fabregas, M. Faundez-Zanuy, J.M. Pascual-Gaspar, V. Cardeñoso-Payo, C. Vivaracho-Pascual, “Biosecure signature evaluation campaign (BSEC2009): evaluating online signature algorithms depending on the quality of signatures”, Pattern Recognition, vol. 45, pp. 993-1003, 2012.

[14] M. I. Malik, M. Liwicki, L. Alewijnse, W. Ohyama, M. Blumenstein, B. Found, “ICDAR2013 competitions on signature verification and writer identification for on- and offline skilled forgeries (SigWiComp2013)”, Proc. of Int. Conf. on Document Analysis and Recognition, Tunisia, 2013, pp.1108-1114.

[15] N. Kumar, A. C. Berg, P. N. Belhumeur, S. K. Nayar. “Describable visual attributes for face verification and image search”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 33(10), pp. 1962-1977, 2011.

[16] D. Reid, M. Nixon, S. Stevenage, "Soft Biometrics; Human Identification Using Comparative Descriptions", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 36(6), pp. 1216-1228, 2014.

[17] D. Martinho-Corbishley, M. S. Nixon, Mark, J. N. Carter, “Soft Biometric Recognition from Comparative Crowdsourced Annotations”, Proc. Int. Conf. on Imaging for Crime Prevention and Detection, London, UK, 2015, pp. 1-6.

[18] A. K. Jain, K. Nandakumar, X. Lu, U. Park, “Integrating Faces, Fingerprints, and Soft Biometric Traits for User Recognition”, Proc. of Biometric Authentication Workshop, LNCS 3087, 2004, pp. 259-269.

[19] P. Tome, J. Fierrez, R. Vera-Rodriguez, M. Nixon, “Soft Biometrics and their Application in Person Recognition at a Distance”, IEEE Trans. on Information Forensics and Security, vol. 9(3), pp. 464-475, 2014.

[20] J. Galbally, M. Diaz-Cabrera, M.A. Ferrer, M. Gomez-Barrero, A. Morales, J. Fierrez, “On-Line Signature Recognition Through the Combination of Real Dynamic Data and Synthetically Generated Static Data”, Pattern Recognition, vol. 48, pp. 2921-2934, 2015.

[21] J. Fierrez, J. Galbally, J. Ortega-Garcia, M. R. Freire, F. Alonso-Fernandez, D. Ramos, D. T. Toledano, J. Gonzalez-Rodriguez, J. A. Siguenza, J. Garrido-Salas, E. Anguiano, G. Gonzalez-de-Rivera, R. Ribalda, M. Faundez-Zanuy, J. A. Ortega, V. Cardeñoso-Payo, A. Viloria, C. E. Vivaracho, Q. I. Moro, J. J. Igarza, J. Sanchez, I. Hernaez, C. Orrite-Uruñuela, F. Martinez-Contreras, J. J. Gracia-Roche, "BiosecurID: A Multimodal Biometric Database", Pattern Analysis and Applications, vol. 13(2), pp. 235-246, 2010.