First International Fingerprint Liveness Detection Competition—LivDet 2009*

Gian Luca Marcialis[1], Aaron Lewicke[2], Bozhao Tan[2], Pietro Coli[1], Dominic Grimberg[2], Alberto Congiu[1], Alessandra Tidu[1], Fabio Roli[1], and Stephanie Schuckers[2]

[1] University of Cagliari - Department of Electrical and Electronic Engineering – Italy
{marcialis,roli,pietro.coli}@diee.unica.it
[2] Clarkson University - Department of Electrical and Computer Engineering – USA
{lewickat,tanb,sschucke}@clarkson.edu

Abstract. Fingerprint recognition systems are vulnerable to artificial spoof fingerprint attacks, like molds made of silicone, gelatin or Play-Doh. "Liveness detection", which is to detect vitality information from the biometric signature itself, has been proposed to defeat these kinds of spoof attacks. The goal for the LivDet 2009 competition is to compare different methodologies for software-based fingerprint liveness detection with a common experimental protocol and large dataset of spoof and live images. This competition is open to all academic and industrial institutions which have a solution for software-based fingerprint vitality detection problem. Four submissions resulted in successful completion: Dermalog, ATVS, and two anonymous participants (one industrial and one academic). Each participant submitted an algorithm as a Win32 console application. The performance was evaluated for three datasets, from three different optical scanners, each with over 1500 images of "fake" and over 1500 images of "live" fingerprints. The best results were from the algorithm submitted by Dermalog with a performance of 2.7% FRR and 2.8% FAR for the Identix (L-1) dataset. The competition goal is to become a reference event for academic and industrial research in software-based fingerprint liveness detection and to raise the visibility of this important research area in order to decrease risk of fingerprint systems to spoof attacks.

Keywords: Fingerprint, biometrics, spoofing, liveness detection, anti-spoofing protection, security.

1 Introduction

The widespread use of personal verification systems based on fingerprints has shown security vulnerabilities. Among the others, it is well-known that a fingerprint verification system can be deceived by submitting artificial reproductions of fingerprints

* LivDet 2009 Group is constituted by several Ph.D and under graduate students which con-tributed to the data set collection and LivDet09 web site managing (http://prag.diee.unica.it/ LivDet09).

P. Foggia, C. Sansone, and M. Vento (Eds.): ICIAP 2009, LNCS 5716, pp. 12–23, 2009.
© Springer-Verlag Berlin Heidelberg 2009

made up of silicon or gelatin to the electronic capture device. These images are then processed as "true" fingerprints.

A suggested solution to combat the use of artificial fingers in fingerprint verification is known as "liveness detection". In this, a standard verification system is coupled with additional hardware or software modules aimed to certify the authenticity of the submitted fingerprints. Whilst the hardware-based solution are the most expensive, the software-based ones attempt to measure liveness from characteristics of images themselves by simply integrating image processing algorithms. The problem of liveness detection is treated as a two-class classification problem (live or fake). An appropriate classifier is designed in order to extract the probability of the image vitality given the extracted set of features.

In order to assess the main achievements of the state of the art in fingerprint liveness detection, the Department of Electrical and Electronic Engineering of the University of Cagliari, in cooperation with the Department of Electrical and Computer Engineering of the Clarkson University, is proud to announce the first edition of the Fingerprint Liveness Detection Competition 2009 (LivDet 2009), which is held in the context of 15th International Conference on Image Analysis and Processing (ICIAP 2009). LivDet 2009 is open to all academic and industrial institutions which have a solution for software-based fingerprint vitality detection problem.

The goal of the competition is to compare different methodologies for software-based fingerprint liveness detection with a common experimental protocol and data set. As a reference event for academic and industrial research, the competition will raise the visibility of this important research area. The competition is not defined as an official system for quality certification of the proposed solutions, but rather, it hopes to impact the state of the art in this crucial field—security in biometric systems.

Each participant has been invited to submit its algorithm in a Win32 console application. The performance has been evaluated by utilizing a very large data set of "fake" and "live" fingerprint images captured with three different optical scanners. The performance rank has been compiled and the "best" algorithm has won the "Best Fingerprint Liveness Detection Algorithm Award" at ICIAP 2009. A Special Session of ICIAP 2009 has been devoted to present and discuss the experimental results.

In this paper, we summarize the competition characteristics and the final results achieved from the algorithms submitted by participants. Section 2 describes the problem of fingerprint spoofing. Section 3 is devoted to the competition results. Section 4 concludes the paper with some discussions on reported results. An appendix has been added after references in order to describe algorithms submitted by participants.

Fig. 1. Consensual method - the person puts his finger on a soft material

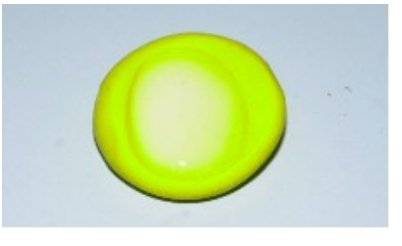

Fig. 2. Consensual method. The negative impression.

2 Background

The duplication of a fingerprint (also named "fingerprint spoofing") has remote origin from science fiction novels of the beginning of the twentieth century. In these last years this question is the focal point of numerous research groups, both academic and industrial. The first spoofing studies date back to 2000 and 2002 [1-2]. These works showed the possibility of the fingerprint reproduction and the defrauding of a biometric system. The steps to create spoof images are as follows: (1) The user puts his finger on a soft material to form the mold (Play Doh, dental impression material, plaster, etc.), see Figure 1. The negative impression of the fingerprint is fixed on the surface. (2) Silicone liquid or another similar material (wax, gelatin, etc) is poured in the mold or pressed in the mold (e.g., Play Doh), see Figure 2. When the liquid is hardened the spoof is formed, see Figure 3. This is the process was used to collect images from silicon, gelatin, and Play-Doh for the competition dataset.

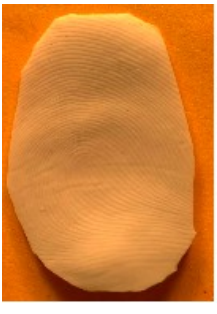

Fig. 3. Consensual method - the stamp with the reproduction of the pattern

It is also possible to create fingerprint spoofs without cooperation with a latent print left by an unintentional user. The latent print is enhanced, photographed and printed in negative on a transparency. A mold is created by etching a printed circuit board and the spoof is formed by dripping a liquid (e.g., silicon, gelatine or wax) on the board.

When faced with this threat, a biometric device must decide if the finger on the acquisition sensors is from the authorized user present at the time of capture. In other words, the recognition process must be upgraded with an added function for detecting the “vitality” (or “liveness”) of the submitted biometric. Due to the difficulty of the task, the first goal is to achieve good “liveness” detection rates when the consensual method is applied. It is worth noting that this method results in the best quality replica and images, since the subject is “consensual”.

Liveness detection can be performed by adding some additional hardware to the capture device (e.g. for checking blood pressure, or heartbeat, which are not present in a “fake” finger), thus increasing their cost. Another solution is to integrate into standard fingerprint sensors additional algorithms which are able to detect the “liveness” degree from the captured image. They are so-called “software-based” approaches.

For software-based liveness, the question is: Are there biometric "liveness" measurements which can be extracted from captured images? Software-based liveness is the topic of this competition.

Several algorithms for detecting fingerprint liveness have been proposed [3-5], but the main problem is to understand how these algorithms may impact a fingerprint verification system when integrated. In particular, the objective of this competition is to evaluate various approaches' performance by a shared and well-defined experimental protocol, in order to assess the state of the art in the field on a common database.

3 Experimental Protocol and Evaluation

Due to the wide variety of current liveness detection algorithms, the competition defines some constraints for the submitted algorithms:

1) Methods must output, for each image, a "liveness degree" ranging from 0 to 100 (e.g. posterior probability of "true" class).
2) A training set of fake and live fingerprint images will be made available to each participant, freely downloadable from the LivDet site after the participant registration. These images are a subset (25%) of the entire data set.
3) Each submitted algorithm, as a Win32 console application, must follow the input and output sequence required.
4) Each submitted algorithm is tested using a withheld dataset that is the remaining 75% of the entire data set.

3.1 Participants

The competition is open to academic and industrial institutions. Each user receives, after his registration and information about the competition rules, a password to enter into the site and managing his personal information and uploaded files. In order to finalize the registration, it is necessary to submit a license agreement of the data set use. The filled and signed agreement is sent through fax and by mail. Once a signed consent form is obtained, a link for downloading the training set is given. Each participant gives their preference on whether to enter as anonymous. All results will be included in the final report. Results published at LivDet 2009 cannot be used for commercial purposes. The goal of the competition is merely to establish a baseline of the state-of-the-art.

3.2 Data Set

The data set for the final evaluation is constituted of three sub-sets, which contain live and fake fingerprint images from three different optical sensors. Table 1 lists the scanners we used for data collection and the image numbers in the total database.

Images have been collected by a consensual approach, as described in the Background section, using different materials for the artificial reproduction of the fingerprint (gelatin, silicone, play-doh). The downloadable training set is 25% of the above data. At the end of the competition, the entire data set will be made available by signing an appropriate license agreement. Fig. 4 shows example fake fingerprint images from the three optical scanners.

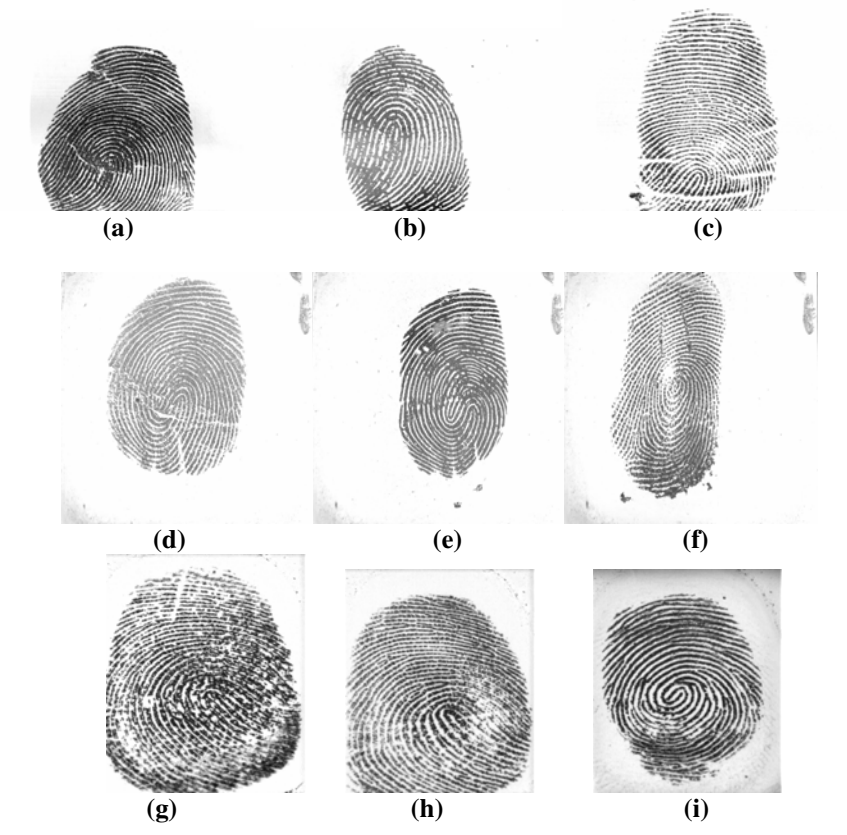

Fig. 4. Examples of fake fingerprint images, from Crossmatch: (a) Play-Doh, (b) gelatin, (c) silicone; from Identix: (d) Play-Doh, (e) gelatin, (f) silicone; from Biometrika: (g)-(i) silicone

Table 1. Fingerprint sensors and data collection for LivDet 2009

DATABASE	Scanners	Model No.	Resolution (dpi)	Image size	Live Samples	Fake Samples
Dataset #1	Crossmatch	Verifier 300 LC	500	480x640	2000	2000
Dataset #2	Identix	DFR2100	686	720x720	1500	1500
Dataset #3	Biometrika	FX2000	569	312x372	2000	2000

3.3 Algorithm Submission

Each submitted algorithm must be a Win32 console application with the following list of parameters:

LIVENESS_XYZ.exe [ndataset] [inputfile] [outputfile]

Each parameter, specified in Table 2, and related to the data set configuration and must be set before submission. Each user can configure his algorithm by the training set available after registration. Only Win32 console applications with the above characteristics will be accepted for the competition. Participants may publish also the source code of their algorithm, but this is not mandatory.

Table 2. Formats of submission requirements

Arguments	Description
LIVENESS_XYZ.exe	It is the executable name, where XYZ is the identification number of the participant. LIVENESS_XYZ.exe Format : Win32 console application (.exe)
[ndataset]	It is the identification number of the data set to analyse. Legend: 1=Crossmatch, 2=Identix, 3=Biometrika
[inputfile]	Txt file with the List of images to analyse. Each image is in RAW format (ASCII)
[outputfile]	Txt file with the output of each processed image, in the same order of inputfile. The output is a posterior probability of the live class given the image, or a degree of "liveness" normalized in the range 0 and 100 (100 is the maximum degree of liveness, 0 means that the image is fake). In the case that the algorithm has not been able to process the image, the correspondent output must be -1000 (failure to enroll).

3.4 Performance Evaluation

The parameters adopted for the performance evaluation will be the following:

Evaluation per sensor

- *Frej_n*: Rate of failure to enroll for the sub-set *n.*
- *Fcorrlive_n*: Rate of correctly classified live fingerprints for sub-set *n.*
- *Fcorrfake_n*: Rate of correctly classified fake fingerprints for sub-set *n.*
- *Ferrlive_n*: Rate of misclassified live fingerprints for sub-set *n.*
- *Ferrfake_n*: Rate of misclassified fake fingerprints for sub-set *n.*
- *ET*: Average processing time per image
- *MAM*: Max. Allocated Memory while the algorithm is running.

Overall evaluation

- *Frej*: Rate of failure to enroll.
- *Fcorrlive*: Rate of correctly classified live fingerprints.

- *Fcorrfake_n*: Rate of correctly classified fake fingerprints.
- *Ferrlive_n*: Rate of misclassified live fingerprints.
- *Ferrfake_n*: Rate of misclassified fake fingerprints.

3.5 Declaration of the Winner

The winner will be awarded by simple averaging the overall classification errors on the three sensors. Only one winner will be awarded. The declaration will be made during the social dinner of ICIAP 2009. A Special Session of ICIAP 2009 will be devoted to present and discuss the performance of the proposed algorithms.

4 Results and Discussion

Four algorithm submissions successfully completed the competition at the time of submission of this paper: Dermalog Identification Systems GmbH (Dermalog), Biometric Recognition Group - ATVS at Universidad Autonoma de Madrid (ATVS), Anonymous (industry) and Anonymous 2 (academic). Details regarding the submitted algorithm from ATVS is given in the Appendix.

The rate of misclassified spoof fingerprints (*Ferrfake_n*) is given in Figure 5 and Table 3 and the rate of misclassified live fingerprints (*Ferrlive_n*) is given in Figure 6 and Table 3. The best results are for Dermalog algorithm on Identix dataset with 2.7% *Ferrlive* and 2.8% *Ferrfake.* Dermalog, ATVS, Anonymous, and Anonymous 2 achieved an average *Ferrfake* of 5.4%, 9%, 16%, and 16.0% respectively. Dermalog, ATVS, and Anonymous achieved an average *Ferrlive* of 20%, 30%, 33%, and 13.2% respectively.

For the majority of the algorithms, the liveness values are clustered around 0 or 100, so changing the threshold has little effect or no effect on *Ferrfake* and *Ferrlive.* Therefore we set the threshold to an arbitrary value of 50 to denote liveness. However, the results given by Anonymous 2 range fairly evenly between 40 and 60 so changing the threshold impacts the results. To determine a reasonable threshold, ROC curves of *Ferrfake* vs *Ferrlive* were generated for each data set for Anonymous 2 algorithm only. From these, thresholds were selected which minimizes both *Ferrfake* and *Ferrlive* simultaneously, resulting in a threshold of 72.9 for Identix, 63.9 for Crossmatch, and 73.9 for Biometrika.

Ferrlive for Biometrica was unexpectedly high for all algorithms. We hypothesize that this is related to the number of distinct live subjects in the training dataset, as well as the fact that the images were collected in a single session where as the other devices were collected over multiple sessions, as seen in Table 4. Both Crossmatch and Identix had over 35 and 63 unique individuals, respectively, in the training set, while Biometrica had only 13. In addition, for Crossmatch and Identix only four images per subject (2 fingers, 2 images) were collected during a single visit, while for Biometrica, all 40 images were collected during one visit. This highlights the importance of including a large number of unique individuals, as well as multiple visits for the

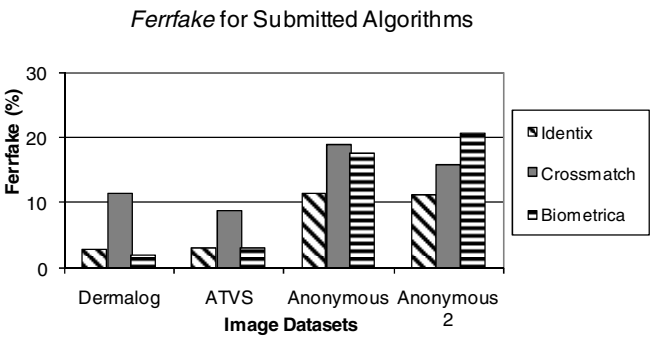

Fig. 5. Rate of misclassified spoof fingerprints (*Ferrfake*_n) for submitted algorithms, (left to right) Dermalog, ATVS, Anonymous, and Anonymous 2, for each of the image sets (Identix, Crossmatch, and Biometrica)

training live images. This creates a training dataset that is representative of the variability of live finger such that a liveness algorithm generalizable to unseen images can be developed.

Table 3. Rate of misclassified live fingerprints (*Ferrlive*_n) and rate of misclassified spoof fingerprints (*Ferrfake*_n) (%) for submitted algorithms (Dermalog, ATVS, Anonymous, Anonymous 2) for each dataset (Identix, Crossmatch, Biometrica), as well as average for each algorithm

	Datasets							
Submitted	**Identix**		**Crossmatch**		**Biometrica**		**Average**	
Algorithms	***Ferrlive***	***Ferrfake***	***Ferrlive***	***Ferrfake***	***Ferrlive***	***Ferrfake***	***Ferrlive***	***Ferrfake***
Dermalog	2.7%	2.8%	7.4%	11.4%	74.1%	1.9%	20.1%	5.4%
ATVS	9.8%	3.1%	8.8%	20.8%	71.7%	3.1%	30.1%	9.0%
Anonymous	15.2%	11.5%	27.1%	18.9%	56.0%	17.6%	32.8%	16.0%
Anonymous2	9.8%	11.3%	14.4%	15.9%	15.6%	20.7%	13.2%	16.0%

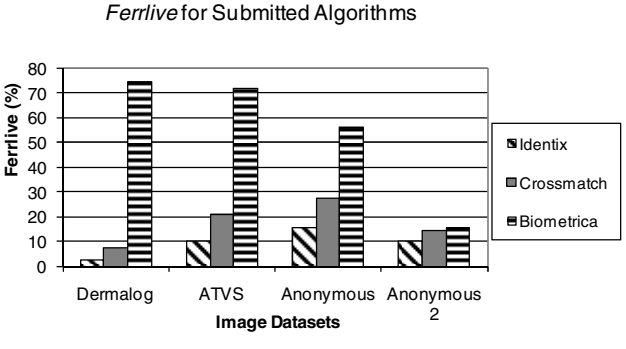

Fig. 6. Rate of misclassified live fingerprints (*Ferrlive*_n) for submitted algorithms, (left to right) Dermalog, ATVS, Anonymous, and Anonymous 2 for each of the image sets (Identix, Crossmatch, and Biometrica)

Table 4. Number of unique subjects in training and tests, as well as the average number of images per subject. It should also be noted that Identix and Crossmatch were collected over multiple visits, while Biometrica was collected during a single visit.

Scanners	# of Training Subjects	# of Testing Subjects	Aver Images / subject
Identix	35	125	18.75
Crossmatch	63	191	15.75
Biometrika	13	37	40.0

Failure to enroll rates are shown in Table 5. Some algorithms utilized this feature, most likely as a quality check. Images that had a failure to enroll were not included in the calculation of *Ferrlive* and *Ferrspoof*.

An alternative method of accounting for failure to enrol is to consider a false reject for a live individual an error; whereas for a spoof image, a failure to enrol would be a successful rejection of a spoof images, i.e., not an error. The overall classification error rates which considers all errors, as well as failure to enroll as described above, are given in Table 6.

Average processing time per image for each algorithm is shown in Table 7. The anonymous algorithm had the shortest processing time with average of 0.07 seconds per image. The algorithms that had longer processing time likely reflected that the algorithms were created in Matlab and compiled as an executable. Anonymous 2 algorithm had run-time problems such that elapsed time could not be estimated.

Table 5. Failure to enroll rates for submitted algorithms

		Data Sets		
		Identix	**Crossmatch**	**Biometrica**
Submitted Algorithms	**Dermalog**	0.9%	1.1%	0.0%
	ATVS	0.0%	0.0%	0.0%
	Anonymous	2.0%	2.2%	1.3%
	Anonymous 2	0.0%	0.0%	0.0%

Table 6. Overall classification error which considers rate of spoof and live classification errors, as well as errors where live images resulted in a failure to enroll

		Datasets			
		Identix	**Crossmatch**	**Biometrica**	**Average**
Submitted Algorithms	**Dermalog**	3.6%	10.5%	37.9%	17.3%
	ATVS	6.5%	14.8%	71.7%	31.0%
	Anonymous	14.2%	23.7%	37.2%	25.0%
	Anonymous 2	10.5%	15.2%	18.1%	14.6%

Table 7. Average elapsed time per image (in seconds)

Elapsed Time (s)		*Data Sets*		
		Identix	**Crossmatch**	**Biometrica**
Submitted Algorithms	**Anonymous**	0.12	0.07	0.07
	Dermalog	0.94	0.56	0.28
	ATVS	46.95	50.04	10.24

5 Conclusions

In summary, LivDet 2009 is the first international public competition for software-based fingerprint liveness detection. Entries were submitted from four participants demonstrating the state-of-the art in fingerprint liveness. Best results achieved ~2.5% error. It is hoped that this first competition of fingerprint liveness detection is followed by a number of competitions such that further improvement in algorithm performance is encouraged. In particular, our expectation is that the proposed experimental protocol, data sets, and algorithm results, may become a standard reference point for the research community, such that increased algorithm performance can be achieved. An effective liveness detection algorithms is a key component to minimize the vulnerability of fingerprint systems to spoof attacks.

Acknowledgements

We would like to thank the funding support from Center for Identification Technology Research (CITeR) for the success of this project.

We also thank the ICIAP 2009 chairs, Mario Vento, Pasquale Foggia and Carlo Sansone, who kindly hosted the competition.

References

1. Ligon, A.: An investigation into the vulnerability of the Siemens id mouse Professional Version 4 (2002), http://www.bromba.com/knowhow/idm4vul.htm
2. Matsumoto, T., Matsumoto, H., Yamada, K., Hoshino, S.: Impact of artificial gummy fingers on fingerprint systems. In: Proceedings of SPIE, vol. 4677, Optical Security and Counterfeit Deterence Techniques IV, Yokohama
3. Tan, B., Schuckers, S.: A new approach for liveness detection in fingerprint scanners based on valley noise analysis. Journal of Electronic Imaging 17, 11009. SPIE (2008)
4. Chen, Y., Jain, A.K., Dass, S.: Fingerprint deformation for spoof detection. In: Biometric Symposium 2005, Cristal City, VA (2005)
5. Coli, P., Marcialis, G.L., Roli, F.: Fingerprint silicon replicas: static and dynamic features for vitality detection using an optical capture device. International Journal of Image and Graphics 8(4), 495–512 (2008)
6. Lim, E., Jiang, X., Yau, W.: Fingerprint quality and validity analysis. In: Proc. International Conference on Image Processing, ICIP, vol. 1, pp. 469–472 (2002)
7. Chen, Y., Dass, S., Jain, A.: Fingerprint quality indices for predicting authentication performance. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) AVBPA 2005. LNCS, vol. 3546, pp. 160–170. Springer, Heidelberg (2005)
8. Chen, T., Jiang, X., Yau, W.: Fingerprint image quality analysis. In: Proc. International Conference on Image Processing, ICIP, vol. 2, pp. 1253–1256 (2004)
9. Hong, L., Wan, Y., Jain, A.: Fingerprint imagen enhancement: Algorithm and performance evaluation. IEEE Trans. on Pattern Analysis and Machine Intelligence 20, 777–789 (1998)

A Appendix: BRG-ATVS Submission to LivDet: System Description

Javier Galbally, Fernando Alonso, Julian Fierrez, and Javier Ortega

Biometric Recognition Group–ATVS, EPS, Universidad Autonoma de Madrid,
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
{javier.galbally,fernando.alonso,julian.fierrez,
javier.ortega}@uam.es

The proposed approach presented to LivDET uses a new parameterization based on quality measures for a software-based solution in fingerprint liveness detection. This novel strategy requires just one fingerprint image to extract the necessary features in order to determine if the finger presented to the sensor is real or fake. This fact shortens the acquisition process and reduces the inconvenience for the final user. A general diagram of the liveness detection system is shown in Fig. A. In the first step the fingerprint is segmented from the background. Once the useful information of the total image has been separated, ten different quality measures are extracted which will serve as the feature vector that will be used in the classification. Prior to the classification step, the best performing features are selected depending on the sensor that is used in the acquisition. Once the final feature vector has been generated the fingerprint is classified as real (generated by a living finger), or fake (coming from a gummy finger).

A.1 Feature Extraction

The parameterization used to solve the liveness detection problem comprises ten quality-based features. Image quality can be assessed by measuring one of the following properties: ridge strength or directionality, ridge continuity, ridge clarity, integrity of the ridge-valley structure, or estimated verification performance when using the image at hand. In the following, we give some details about the quality measures used in this paper. We have implemented several measures that make use of the above mentioned properties for quality assessment:

Ridge-strength measures. Orientation Certainty Level (Q_{OCL}) [6], Energy concentration in the power spectrum (Q_E) [7].

Ridge-continuity measures. Local Orientation Quality (Q_{LOQ}) [8], Continuity of the Orientation Field (Q_{COF}) [6]**.**

Ridge-clarity measures. Mean (Q_{MEAN}) and standard deviation (Q_{STD}) values of the gray level image. Local Clarity Score (Q_{LCS1} and Q_{LCS2}) [8]. Amplitude and variance of the sinusoid that models ridges and valleys (Q_A and Q_{VAR}) [9].

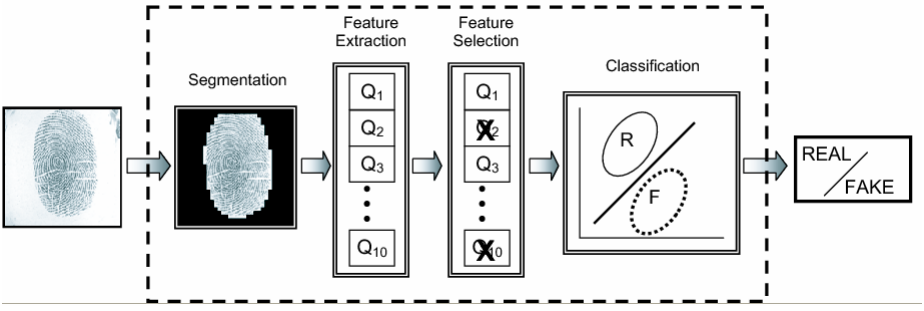

Fig. A. General diagram of the liveness detection system presented in this work

A.2 Feature Selection

Due to the curse of dimensionality, it is possible that the best classifying results are not obtained using the set of ten proposed features, but a subset of them. As we are dealing with a ten dimensional problem there are $2^{10} - 1 = 1{,}023$ possible feature subsets, which is a reasonably low number to apply exhaustive search as feature selection technique in order to find the best performing feature subset. This way we guarantee that we find the optimal set of features out of all the possible ones.

A.3 Classifier

We have used Linear Discriminant Analysis (LDA) as classifier. All the parameterized samples of a certain dataset are used to fit the two normal distributions representing each of the classes (real and fake). The sample being classified (which was left out of the training process) is assigned to the most probable class.