Iris Liveness Detection Methods in Mobile Applications

Ana F. Sequeira[1,2], Juliano Murari[3] and Jaime S. Cardoso[1,2]
[1]*INESC TEC (formerly INESC Porto), Porto, Portugal*
[2]*Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*
[3]*Universidade Federal de S. Paulo, São Paulo, Brazil*

Keywords: Biometrics, Iris, Liveness Detection, Fake Database, Handheld Device.

Abstract: Biometric systems are vulnerable to different kinds of attacks. Particularly, the systems based on iris are vulnerable to direct attacks consisting on the presentation of a fake iris to the sensor trying to access the system as it was from a legitimate user. The analysis of some countermeasures against this type of attacking scheme is the problem addressed in the present paper. Several state-of-the-art methods were implemented and included in a feature selection framework so as to determine the best cardinality and the best subset that conducts to the highest classification rate. Three different classifiers were used: Discriminant analysis, K nearest neighbours and Support Vector Machines. The implemented methods were tested in existing databases for iris liveness purposes (Biosec and Clarkson) and in a new fake database which was constructed for evaluation of iris liveness detection methods in the mobile scenario. The results suggest that this new database is more challenging than the others. Therefore, improvements are required in this line of research to achieve good performance in real world mobile applications.

1 INTRODUCTION

Biometric systems can offer several advantages over classical security methods as they rather identify an individual by what he is instead of based on something he knows or possesses. However, in spite of its advantages, biometric systems have some drawbacks, including: i) the lack of secrecy (e.g. everybody knows our face or could get our fingerprints), and ii) the fact that a biometric trait cannot be replaced (no new iris can be generated if an impostor "steals" it). Furthermore, biometric systems are vulnerable to external attacks which could decrease their level of security. Concerning these vulnerabilities we find in the literature (Galbally et al., 2007) an analysis of the eight different points of attack on biometric recognition systems previously identified (Ratha et al., 2001). These points are illustrated in Fig. 1.

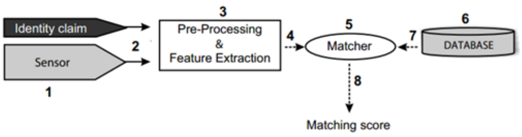


Figure 1: Architecture of an automated biometric verification system. Possible attack points are numbered from 1 to 8, from (Galbally et al., 2007).

These attacks are divided into two main groups: direct and indirect attacks.

- *Direct Attacks:* the first vulnerability point in a biometric security system is the possibility to generate synthetic biometric samples (for instance, speech, fingerprints or face images) in order to fraudulently access a system. These attacks at the sensor level are referred to as direct attacks. It is worth noting that in this type of attacks no specific knowledge about the system operation is needed (matching algorithm used, feature extraction, feature vector format, etc). Furthermore, the attack is carried out in the analogue domain, outside the digital limits of the system, so the digital protection mechanisms (digital signature, watermarking, etc.) can not be used.
- *Indirect Attacks:* this group includes all the remaining seven points of attack. Attacks 3 and 5 might be carried out using a Trojan Horse that bypasses the feature extractor, and the matcher respectively. In attack 6 the system database is manipulated (a template is changed, added or deleted) in order to gain access to the application. The remaining points of attack (2, 4, 7 and 8) are thought to exploit possible weak points in the communication channels of the system, extract-

Sequeira A., Murari J. and Cardoso J..
Iris Liveness Detection Methods in Mobile Applications.
DOI: 10.5220/0004691800220033
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 22-33
ISBN: 978-989-758-009-3

ing, adding or changing information from them. In opposition to direct attacks, in this case the intruder needs to have some information about the inner working of the recognition system and, in most cases, physical access to some of the application components (feature extractor, matcher, database, etc.) is required.

Among the different existing biometric traits, iris has been traditionally regarded as one of the most reliable and accurate. This fact has led researchers to pay special attention to its vulnerabilities and in particular to analyze to what extent their security level may be compromised by spoofing attacks. These attacks may consist on presenting a synthetically generated iris to the sensor so that it is recognized as the legitimate user and access is granted. The most common and simple approaches are those carried out with high quality iris printed images (Ruiz-Albacete et al., 2008). However, other more sophisticated threats have also been reported in the literature such as the use of contact lenses (Wei et al., 2008).

The development of iris liveness detection techniques is crucial for the deployment of iris biometric applications in daily life. The evolution in the use of mobile devices in our society also raises the urge for liveness solutions in the mobile biometric field. To pursue this goal there is also a need for suitable databases in which new methods can be tested.

In this work we implemented state-of-the-art methods conceived to deal with spoofing attacks in iris recognition, in particular, the use of printed images and contact lenses. The proposed method comprises a feature selection method, in order to determine the best cardinalities and respective subset of features, with the use of state-of-the-art classifiers. This framework intended to achieve the best classification rates with only the "necessary" number of features Two existing databases were tested, one comprising samples of printed iris images and another comprising images of eyes with contact lenses. Taking in account the results obtained and the characteristics of the databases available and the new trend of performing biometric recognition in mobile scenarios, we constructed a new fake iris database. This database comprises printed copies of the original images (after being printed, the images were acquired with the same device and in similar conditions as the original ones). We found this new database to be more challenging than the others.

This paper is organized as follows. In section 2 the concept of liveness detection in an iris recognition system is presented. In section 3, we explain the algorithms implemented. In section 4 is presented the database constructed with fake printed images for testing liveness detection methods in iris recognition. In section 5, the dataset of images is presented in 5.1, the methodology used is presented in 5.2 and the results and their discussion are presented in 5.3. Finally, in section 6 we draw some conclusions and sketch some ideas for future works.

2 IRIS LIVENESS DETECTION

The problem of liveness detection of a biometric trait can be seen as a two class classification problem where an input trait sample has to be assigned to one of two classes: real or fake. The key point of the process is to find a set of discriminant features which permits to build an appropriate classifier which gives the probability of the sample vitality given the extracted set of features (Galbally et al., 2012b).

Biometric recognition systems are vulnerable to be spoofed by fake copies (Daugman, 2004), for instance, fake finger tips made of commonly available materials such as clay and gelatine. Iris is no exception. There are potential threats for iris-based systems, the main are (He et al., 2009):

- Eye image: Screen image, Photograph, Paper print, Video signal.
- Artificial eye: Glass/plastic etc.
- Natural eye (user): Forced use.
- Capture/replay attacks: Eye image, IrisCode template.
- Natural eye (impostor): Eye removed from body, Printed contact lens.

The feasibility of some attacks have been reported by some researchers (Daugman, 1998; Daugman, 2004; Lee et al., 2005) who showed that it is actually possible to spoof some iris recognition systems with printed iris and well-made colour iris lens. Therefore, it is important to detect the fake iris as much as possible (He et al., 2009).

Several liveness detection methods have been presented through the past recent years. In fact, anti-spoofing techniques were presented that use physiological properties to distinguish between real and fake biometric traits. This is done in order to improve the robustness of the system against direct attacks and to increase the security level offered to the final user. Iris liveness detection approaches can broadly be divided into: i)software-based techniques, in which the fake irises are detected once the sample has been acquired with a standard sensor (i.e., features used to distinguish between real and fake eyes are extracted from the iris image, and not from the eye itself), and ii) hardware-based techniques, in which some specific device is added to the sensor in order to detect particular properties of a living iris such as the eye hippus (which is the permanent oscillation that the eye pupil presents even under uniform lighting conditions) or the pupil response to a sudden lighting event (e.g., switching on a diode) (Galbally et al., 2012b). According to this author, even though hardware-based approaches usually present a higher detection rate, the software-based techniques have the advantage of being less expensive (as no extra device in needed), and less intrusive for the user (very important characteristic for a practical liveness detection solution). In general, a combination of both type of anti-spoofing schemes would be the most desirable approach to increase the security level of biometric systems. (Galbally et al., 2012b)

In this work we focus on software based techniques since these are more easily and affordable applicable in real-world applications.

In the literature we found that the methods of liveness detection may be classified into four categories based on the physical features of biometric and liveness data and the timing of measurement (Une and Tamura, 2006). In this framework, the biometric data are used in the iris recognition and the liveness data are used in the liveness detection. We can itemize the four categories:

- Perfect matching model: Both biometric and liveness data are simultaneously obtained from the same physical feature.
- Simultaneous measuring model: Biometric and liveness data are simultaneously obtained from different physical features.
- Same biometric measuring model: Biometric and liveness data are obtained from the same physical feature with different timings.
- Independent measuring model: Biometric and liveness data are obtained from different features with different timings.

The ideal configuration of liveness detection for biometrics recognition is represented by the perfect matching model with the highest ability to distinguish between live and fake irises (Kanematsu et al., 2007).

The potential of quality assessment to identify real and fake iris samples acquired from a high quality printed image has previously been explored as a way to detect spoofing attacks (Galbally et al., 2012b). Some quality based features have been used individually for liveness detection in traits such as iris (Kanematsu et al., 2007; Wei et al., 2008) or face (Li et al., 2004). A strategy based on the combination of several quality related features has also been used for spoofing detection in fingerprint based recognition systems (Galbally et al., 2012a) as well as in iris liveness detection (Galbally et al., 2012b). In this latter work, a set of quality measures are used as iris liveness detection features to aid the classification of fake or real iris images included in a framework of feature selection. We find in literature that works concerning the quality of iris images are often the starting point to iris liveness detection techniques. One example is the assessment of the iris image quality based on measures like occlusion, contrast, focus and angular deformation (Abhyankar and Schuckers, 2009), other is the use of texture analysis of the iris (He et al., 2007), among others like, for example, the analysis of frequency distribution rates of some specific regions of iris (Ma et al., 2003).

The way forward seems to be the development of techniques for iris liveness detection that work well independently of the particular characteristics of the databases available nowadays. It is required to develop and improve methods as well as to construct new databases in less constrained conditions.

3 IMPLEMENTED METHODS

Some of the measures are obtained from the entire eye image but others are extracted only from the iris region, therefore a segmentation step is required. We choose to make the segmentation process manually, in order to ensure reasonable accuracy. The manual segmentation is done by marking three differents points in the image. The first point is the eye centre, i.e., we consider a single centre for both pupil and iris. The second point is marked in the pupil border and the third in the iris border. With these points is possible to determine the iris and pupil radius and then approximate the contours as two concentric circles. With the manual segmentation's information we are able to map the regions of interest which will be eventually used by the liveness detection algorithms.

3.1 Algorithm 1 - High Frequency Power

The *High Frequency Power* algorithm, which provides feature 1, works on the whole image and measures the energy concentration in the high frequency components of the spectrum using a high pass convolution kernel of 8x8. The application of this convolution is a good Fourier Transform approximation and works as high frequency spectral analysis, which can be considered an estimator of focus (Daugman, 2002). The focus of a real iris, as it is a 3D volume, is different from a fake iris focus, which has a 2D surface. For more details on the method see (Galbally et al., 2012b).

3.2 Algorithm 2 - Local Contrast

The *Local Contrast* algorithm, which provides feature 2, is based on bounding box that involves the iris and the pupil. The bounding box is divided in blocks of $P \times P$ and for each block it is applied the *Fast Fourier Transform* (FFT) algorithm to extract the medium power frequencies, which better represents the contrast. The final value is given by the number of blocks with medium values (between 20 and 60) divided by the total number of blocks. This algorithm was inspired in an occlusion estimation technique (Abhyankar and Schuckers, 2009) and it was adapted for contrast estimation for iris liveness detection in (Galbally et al., 2012b) where more details can be found.

3.3 Algorithm 3 - Global Contrast

The *Global Contrast* algorithm, which provides feature 3, explores the fact that parts extremely bright or dark of the image are not useful and can be considered as noise. Thus, pixels near medium value (128 in 8-bit image) are considered of best contrast (Abhyankar and Schuckers, 2009). In order to quantify the contrast, the original pixels values are normalized between 0 and 25 (Figure 2). Original pixels near medium value will get higher values in the normalized scale, as well as very low and very high values (< 10 and > 245) are normalized to 0. This measure was presented in (Abhyankar and Schuckers, 2009) and it was adapted for global contrast estimation for iris liveness detection in (Galbally et al., 2012b) where more details can be found.

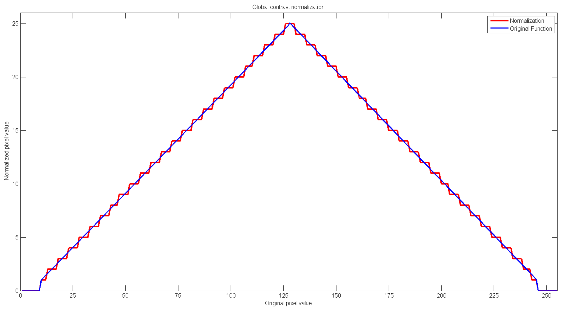


Figure 2: Normalization function of the algorithm 3.

3.4 Algorithm 4 - Frequency Distribution Rates

The *Frequency Distribution Rates* algorithm consists in different mathematical combinations of three different parameters which consider respectively the power of the low (F_1), medium (F_2), and high (F_3) frequencies (computed according to the 2D Fourier Spectrum) from two iris subregions in the horizontal direction. This subregions are illustrated in Figure 3. Each subregion is subdivided in three circular concentric region, which determine the three different frequencies, i.e, for the first subregion, F_1^1 refers to the central circle, F_2^1 refers to the middle circular ring and F_3^1 refers to the outer circular ring, as depicted in Figure 4. The final F_1 is given by the average between the two regions: $F_1 = \frac{F_1^1 + F_1^2}{2}$. The same is done to F_2 and F_3. More details on the method can be found in (Ma et al., 2003; Galbally et al., 2012b).

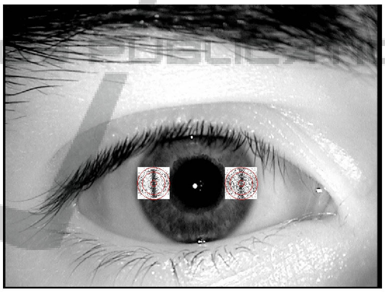


Figure 3: Example of the subregions used in the algorithm 4 (Galbally et al., 2012b).

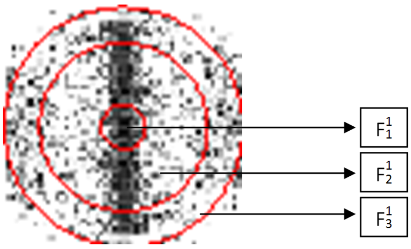


Figure 4: One of the regions of interest subdivided to calculate the frequencies (Galbally et al., 2012b).

With the three final frequencies we extract seven different combinations, represented in Table 1 (Ma et al., 2003; Galbally et al., 2012b).

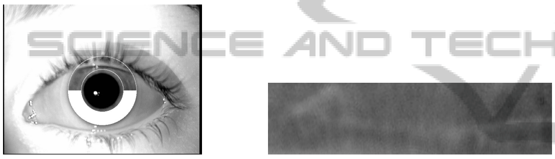
3.5 Algorithm 5 - Statistical Texture Analysis

The *Statistical Texture Analysis* algorithm was developed as a contact lens countermeasure. The outer portion of the colour contact lens (corresponding to re-

Table 1: Extracted measures from the final frequencies.

Features no.	Combination
4	$F_1 + F_2 + F_3$
5	$F_2/(F_1 + F_3)$
6	F_3
7	F_2
8	F_1
9	$(F_1 + F_2)/F_3$
10	$(F_1 * F_2)/F_3$

gions closer to outer circle) provides the most useful texture information for fake iris detection since this section of the fake iris is insensitive to the pupil dilation (He et al., 2007). The region of interest is the lower part of the iris in order to minimize the occlusion by the eyelashes and eyelids, which in general occurs in the upper iris portion. In order to achieve invariance to translation and scale, the region of interest is further normalized to a rectangular block of a fixed size $W \times H$ (Figure 5).



(a) Original (b) Normalized

Figure 5: Region of interest used in the algoritm 5 (He et al., 2007).

After the normalization, the GLCM (Gray Level Co-occurence Matrix), one of the most proeminent approaches used to extract textural features (Haralick et al., 1973), is calculated. Four measures are extracted: the mean (μ) and standard deviation (σ), direct from the normalized region of interest, and the contrast (con) and the energy (e) from the GLCM matrix. These measures will provide features 11 to 14 and its values are given, respectively, by the equations below:

$$\mu = \frac{1}{W*H}\sum_{i=1}^{H}\sum_{j=1}^{W} I(i,j) \quad (1)$$

$$\sigma = \sqrt{\frac{1}{W*H}\sum_{i=1}^{H}\sum_{j=1}^{W}(I(i,j)-\mu)^2} \quad (2)$$

$$con = \sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2 P(i,j) \quad (3)$$

$$e = \sum_{i=1}^{N}\sum_{j=1}^{N} P(i,j)^2 \quad (4)$$

Where I denotes the normalized iris image, W is the width of the normalized iris image, H is the height of the normalized iris image. P is the co-occurrence matrix and N denotes the dimension of the co-occurrence matrix. For more details on the method see (He et al., 2007).

3.6 Feature Selection

The algorithms implemented originated 14 different features. Due to this dimensionality it is possible that the best classification results are not obtained using all the features, but a subset of them. It is convenient to search for the optimum number and set of features. To exhaustively test all possibilities is not feasible. Therefore we use the "Sequential Forward Floating Selection" (SFFS) (Pudil et al., 1994) to perform feature selection. The SFFS is basically a combination of search methods such as "Plus-l-Minus-r" (Stearns, 1976) and Sequential Forward Search (SFS) (Whitney, 1971). The appearance of "floating" comes from the fact that the values l and r are not fixed, i.e., they can "float". Another aspect is the dominant direction of search, including (*forward*) or excluding (*backward*) characteristics (Pudil et al., 1994). We use the *Mahalanobis* distance as criterion function. The SFFS has shown to be competitive when compared to other selection techniques (Jain and Zongker, 1997).

3.7 Classification

The classification results were obtained using three classification methods: Discriminant Analysis (DA), k-Nearest Neighbour (kNN) and Support Vector Machine (SVM).

4 MobBIO*fake*: IRIS IMAGES CAPTURED WITH A HANDHELD DEVICE

The MobBIO*fake* database was constructed upon the MobBIO Multimodal Database (Blind Ref, 2013). The MobBIO Multimodal Database comprises the biometric data from 105 volunteers. Each individual provided samples of face, iris and voice. The equipment used for the samples acquisition was an Asus Transformer Pad TF 300T, with Android version 4.1.1. The device has two cameras, one frontal and one back camera. The camera used was the back camera, version TF300T-000128, with 8 MP of resolution and autofocus.

The iris images were captured in two different lighting conditions, in a room with both natural and artificial sources of light, with variable eye orientations and occlusion levels, so as to comprise a larger variability of unconstrained scenarios. Each volunteer contributed with 16 images (8 of each eye) with a 300×200 resolution. Some examples of iris images are depicted in Figure 6.

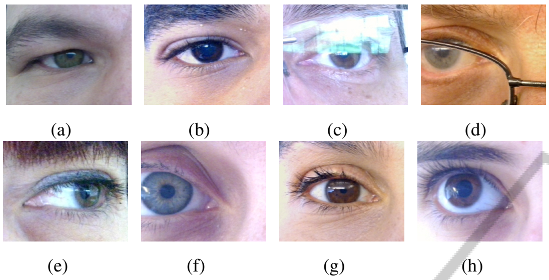


Figure 6: Iris images from MobBIO database illustrating different kinds of noise: a) Heavily occluded; b) Heavily pigmented; c) Glasses reflection; d) Glasses occlusion; e) Off-angle; f) Partial eye; g) Reflection occlusion and h) Normal.

MobBIO*fake*

The MobBIO*fake* is composed by a subset of 800 iris images from MobBIO and its corresponding fake copies, in a total of 1600 iris images. The fake samples were obtained from printed images of the original ones captured with the same handheld device and in similar conditions. From the original dataset of images

The aim of constructing such a database is, on one hand, to fulfil the necessity of databases and, on the other hand to broad the acquisition conditions of the images. The number and variety of databases for iris liveness detection is somewhat limited so the fact that these images were captured with a portable device and are *RGB* images come as a novelty and makes it possible to evaluate liveness methods in this new upcoming scenario.

The construction of the MobBIO*fake* upon the MobBIO iris images subset comprised several steps. The images of each volunteer were joined in a single image, as shown in Figure 7.

A preprocessing (contrast enhancement) was applied using GIMP software (GIMP, 2008) to the image. This enhancement is believed to improve the quality of the fake sample (Ruiz-Albacete et al., 2008). After this, the images were printed in a professional printer using high quality photographic paper. At this point we were able to capture the images. Each individual image (a image of one single eye) was acquired using the same portable device and in similar lighting conditions as the original ones were captured, as illustrated in Figure 8.

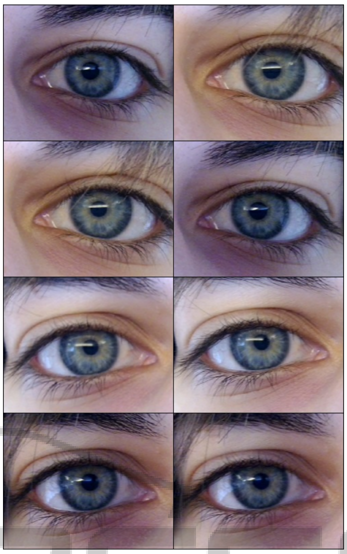


Figure 7: *MobBIOfake* construction: joint images of one volunteer.

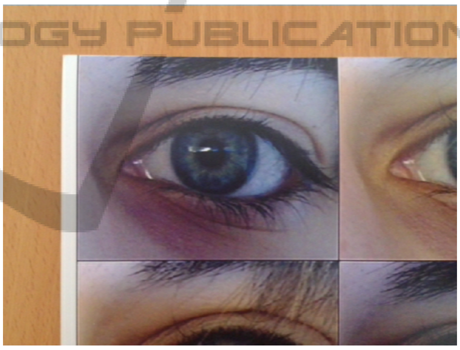
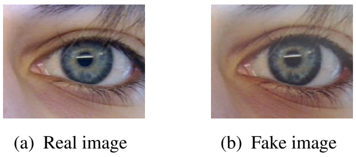


Figure 8: *MobBIOfake* construction: fake samples acquisition.

Finally, the individual eye images were cropped and resized to fix dimensions. An example of a real image and its copy is depicted in Figure 9.



(a) Real image (b) Fake image

Figure 9: Corresponding real and fake images of *MobBIO*.

5 EXPERIMENTAL SETUP

5.1 Datasets

The implemented methods were tested in three datasets. One was a database constructed within our work, the MobBIO*fake*, comprised of 800 iris images and its corresponding fake copies, captured with the same portable device and in similar conditions as the original ones. The description of the construction of this dataset was detailed in section 4.

The other two databases, described below, are the Biosec database (Fierrez et al., 2007), composed by real iris images and the corresponding fake printed images; and the Clarkson database (S. Schuckers and Yambay, 2013) comprising real iris images and fake ones obtained by the use of contact lenses.

Biosec

The Biosec database was created at the Polytechnic University of Madrid (UPM) and the Universitat Politècnica de Catalunya (UPC). The images were acquired in an office room with a large table for *hardware* of biometric recognition system and two chairs, one for the donor and one for the supervisor of the acquisition process. Environmental conditions such as lighting and noise, were not controlled to simulate a real situation (Fierrez et al., 2007). To construct the false database original images are pre-processed and printed on paper using a commercial printer. Then the printed images were presented to the iris sensor, obtaining the fake copy. This study considered different combinations of pre-processing, printing equipment and paper type (Ruiz-Albacete et al., 2008). Therefore, the database used consists of real and fake iris images and follows the same structure as the *original* database. Biosec dataset comprises a total of 1600 images: 800 real images and its corresponding 800 fake samples. All images are in greyscale and its dimensions are 640 × 480 (Galbally et al., 2012b). The two eyes of the same individual are considered as different users. The acquisition of both real and fake samples were made using the sensor *LG IrisAccess EOU3000* (Ruiz-Albacete et al., 2008).

Clarkson

The subset of Clarkson database that we used was made available under request and contains 270 real iris images and 400 fake iris images. The fake samples are images of eyes with contact lenses comprising 14 models of contact lenses. There are two different lighting conditions in the database, which was acquired by video (capturing 100 frames and with variation of focus). The Clarkson database was made available to participants of the *LivDet-2013* competition (S. Schuckers and Yambay, 2013) .

5.2 Methodology

The proposed method is depicted in Figure 10.

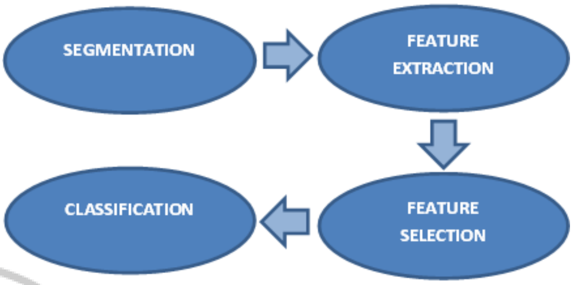


Figure 10: Steps of the proposed method.

The first step of the method is the *segmentation*, in this case it was made manually so as to purge our results from the errors associated with automatic methods of iris segmentation. Although it has to be noted that in a real world application this step needs to be necessarily automatized.

The second step was the *feature extraction*. This comprises the application of the methods described in subsections 3.1, 3.2, 3.3, 3.4 and 3.5.

Next step was the *feature selection*. This comprises the application of the method *Sequential Forward Floating Search*, described in subsection 3, before applying the classifiers to evaluate the proposed methods. We ran the SFFS to obtain the best subset for each cardinality from $\aleph = 2$ to $\aleph = 12$ features. This range is defined by the selection method when considering a set amount of 14 features.

The last step was the *classification*. We used three state-of-the-art classifiers: Discriminant Analysis (DA), k-Nearest Neighbours (kNN) and Support Vector Machines (SVM). For each cardinality, ($\aleph = 2, ..., 12$), the results of classification were obtained calculating the average of the results of 50 runs for classification of the images based on the corresponding best $\aleph$ features. The results were obtained by randomly dividing, in each run, the 1600 samples in two sets: 1000 samples for training and 600 for testing. The parameter k in kNN was optimized using cross-validation, and tested in the interval $[1, 20]$ by steps of 1. For the SVM, we used a polynomial kernel and also used cross-validation for optimization of the parameters. It was performing a ”grid-search” on the parameters of the models. Exponentially growing sequences of C were tested: $C = 2^N$ with N varying between -1 and 15. For the polynomial degree, d, values tested were: $d = 1, 2, 3, 4, 5$.

For the evaluation of the accuracy of the features extracted in discriminating between fake and real images, we used the Equal Error Rate (EER). The EER is obtained when the false acceptance rate (FAR) and the False Rejection Rate (FRR) are equal. For the classification results we use the *missclassification rate* averaged over the 50 runs.

5.3 Experimental Results and Discussion

In this section we present the results obtained by the proposed method for iris liveness detection. The algorithms applied return a set of 14 different features.

The first step was to analyse individually each of the 14 different features, for each image dataset. By the analysis of the histogram obtained for fake and real images we can from that moment have a hint about which features will be good discriminative between fake and real images. For each histogram, the *threshold* obtained considering equal error rate (EER) allow us to determine the minimum error associated with that feature, for the considered dataset. In Table 2 are shown the minimum error values associated with each feature for each database.

Table 2: Minimum error associated with each feature for each database.

	Associated Error (%)		
Feature no.	*Biosec*	*MobBIOfake*	*Clarkson*
1 (alg1)	31.3	31.8	35.4
2 (alg2)	17.2	31.2	**26.3**
3 (alg3)	21.1	**21.9**	26.7
4 (alg4)	15.1	40.8	31.5
5 (alg4)	43.6	27.9	36.0
6 (alg4)	14.4	42.9	31.5
7 (alg4)	15.6	33.0	31.3
8 (alg4)	15.2	30.8	31.9
9 (alg4)	39.9	26.4	36.3
10 (alg4)	15.8	29.3	31.8
11 (alg5)	22.9	27.2	32.2
12 (alg5)	17.2	35.7	37.8
13 (alg5)	**13.2**	35.7	39.1
14 (alg5)	25.8	29.3	37.9

It is clear from the Table 2 that each dataset has a variable behaviour concerning the features obtained. Simply observing the minimum errors (emphasized in the table) we may conclude that the “best” feature for one database is not necessarily the best for any of the others.

To enlighten a bit more how the discriminative power of each feature was analysed we show in Figure 11 the best and worse feature for each database. This histograms illustrate clearly the efficiency of each feature in discriminating real images from fake images. For some features the lines for fake and real images are well separated while for others this lines are too much coincident compromising the separability between the two classes.

The next step was to perform feature selection as to avoid possible redundancies in the set of features. Reducing the number of features to the strictly necessary will improve the computational efficiency of the method. In Table 3 are shown the best subset of features for each cardinality, from 2 to 12, for each database.

Again we observe the diversity of the results obtained for each database. Another relevant aspect is the combinations of features, in some cases we observe that features that individually do not have a good performance when combined provide the best subsets. This fact reinforces the pertinence of using a method for feature selection.

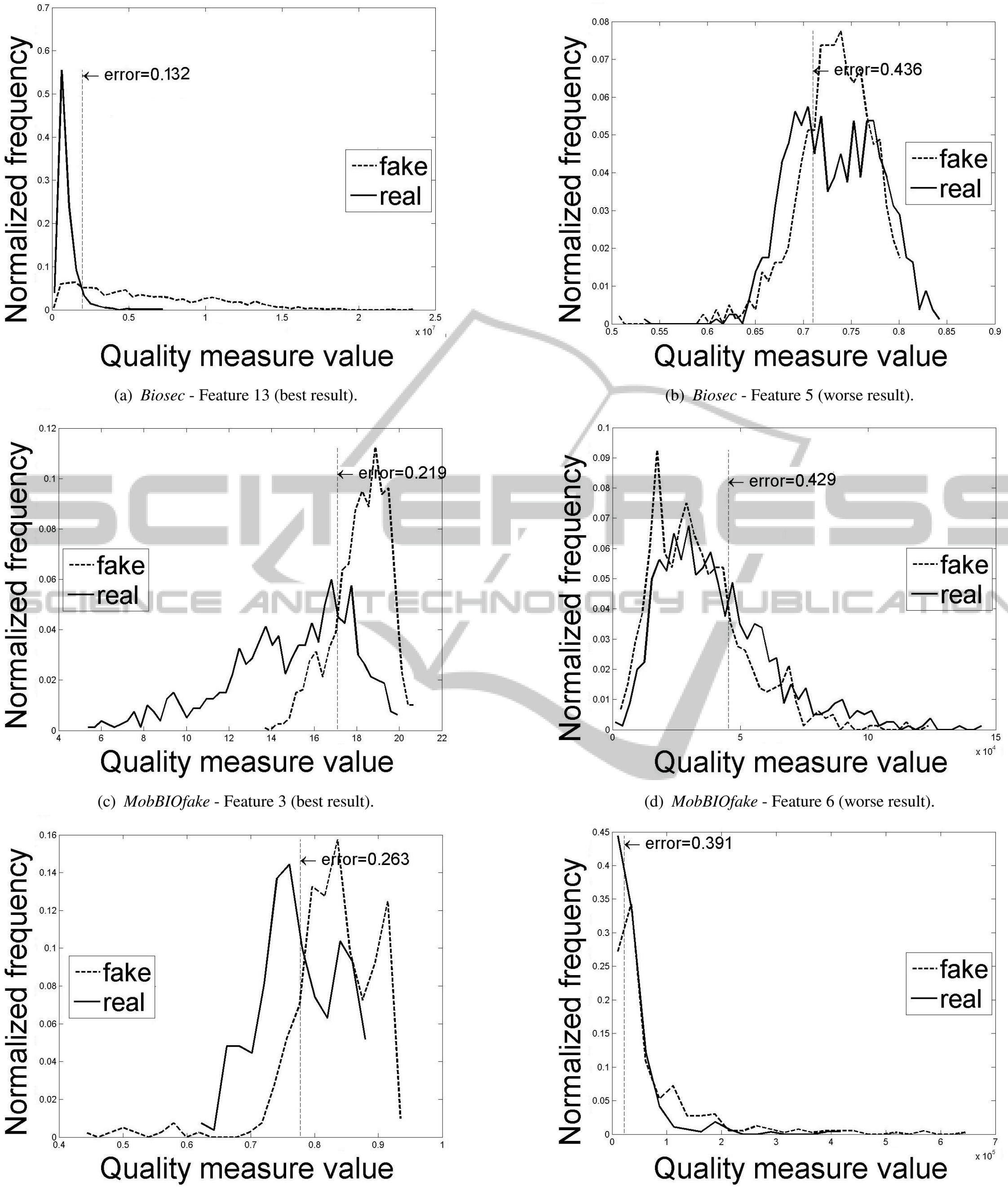
Finally, Tables 4, 5 and 6 show the classification results for each cardinality, for each database, obtained using the best subset determined by the feature selection (averaged over the 50 runs).

Table 4: Classification results for *Biosec* (classification errors in %).

ℵ	DA		kNN		SVM	
	μ	σ	μ	σ	μ	σ
2	10.24	0.99	10.34	1.23	10.16	1.17
3	4.36	0.92	4.43	0.83	4.68	0.67
4	0.52	0.24	0.76	0.27	0.77	0.34
5	1.14	0.48	0.89	0.34	0.78	0.29
6	0.85	0.38	0.56	0.27	0.54	0.26
7	0.92	0.32	**0.40**	0.20	0.57	0.31
8	0.93	0.30	0.47	0.25	0.56	0.33
9	1.80	0.47	1.11	0.34	0.87	0.28
10	1.28	0.39	0.46	0.28	0.73	0.26
11	1.30	0.28	0.40	0.24	0.52	0.33
12	1.68	0.59	**0.37**	0.24	0.50	0.26

Table 5: Classification results for *MobBIOfake* (classification errors in %).

ℵ	DA		kNN		SVM	
	μ	σ	μ	σ	μ	σ
2	18.03	1.31	16.52	1.47	17.29	1.13
3	29.29	1.54	17.34	1.25	20.69	1.83
4	17.50	1.42	**12.62**	1.18	14.36	0.94
5	18.03	1.30	13.00	1.26	14.18	1.20
6	18.29	1.44	12.82	1.02	14.33	1.15
7	18.88	1.35	13.27	1.35	14.55	1.27
8	18.31	1.11	13.52	1.21	13.74	1.28
9	18.34	1.28	14.44	1.25	14.11	2.22
10	17.58	1.38	13.92	1.19	13.39	1.01
11	17.15	1.35	14.51	1.44	12.53	1.39
12	17.25	1.12	14.45	1.21	**12.50**	1.21



(a) *Biosec* - Feature 13 (best result).

(b) *Biosec* - Feature 5 (worse result).

(c) *MobBIOfake* - Feature 3 (best result).

(d) *MobBIOfake* - Feature 6 (worse result).

(e) *Clarkson* - Feature 2 (best result).

(f) *Clarkson* - Feature 13 (worse result).

Figure 11: Histograms for the best result/smallest minimum error (left) and worse result/biggest minimum error (right) for each database.

The overall best results were obtained for Biosec database and the worst overall results were obtained for MobBIO*fake*. This is not a surprising result since we expected this latter to be a more challenging database due to its characteristics. It was notorious from the study of features individually that this database presented the worse results. We interpret this fact as a sign that new databases were needed for the research of liveness in new scenarios.

Comparing the classifiers, we conclude that DA led to worse results. This fact is also not surprising since this classifier may be considered simpler than

Table 3: Best subset of features for each cardinality, for each database.

$\aleph$	Subset of features		
	Biosec	*MobBIOfake*	*Clarkson*
2	[1 6]	[3 10]	[3 14]
3	[1 2 11]	[5 8 10]	[9 11 14]
4	[1 2 6 11]	[3 5 8 10]	[3 9 11 14]
5	[1 2 6 11 13]	[3 4 7 8 10]	[2 3 9 11 14]
6	[1 2 6 11 12 13]	[3 4 7 8 9 10]	[1 2 3 9 11 14]
7	[1 2 5 6 11 12 13]	[3 4 5 7 8 9 10]	[1 2 3 9 11 12 14]
8	[1 2 5 6 7 11 12 13]	[3 4 5 7 8 9 10 13]	[1 2 3 5 9 11 12 14]
9	[1 2 5 6 7 9 10 11 13]	[3 4 5 7 8 9 10 12 13]	[1 2 3 5 9 11 12 13 14]
10	[1 2 5 6 7 9 10 11 12 13]	[3 4 5 7 8 9 10 11 12 13]	[1 2 3 4 5 6 9 11 12 14]
11	[1 2 3 5 6 7 9 10 11 12 13]	[2 3 4 5 7 8 9 10 11 12 13]	[1 2 3 4 5 6 9 11 12 13 14]
12	[1 2 3 5 6 7 9 10 11 12 13 14]	[1 2 3 4 5 7 8 9 10 11 12 13]	[1 2 3 4 5 6 7 9 11 12 13 14]

Table 6: Classification results for *Clarkson* (classification errors in %).

$\aleph$	DA		kNN		SVM	
	μ	σ	μ	σ	μ	σ
2	29.25	2.48	18.86	2.38	21.63	2.65
3	23.38	2.25	18.38	2.06	16.29	2.55
4	20.15	2.61	10.64	1.85	9.20	2.16
5	17.53	2.47	7.82	1.90	7.03	1.62
6	15.74	2.08	8.89	1.71	7.45	2.10
7	14.36	2.01	8.32	2.16	**6.77**	1.41
8	14.55	1.88	9.50	1.63	7.57	1.75
9	12.99	2.49	8.88	1.74	6.92	2.22
10	11.03	1.92	7.86	1.65	5.89	1.86
11	11.02	2.11	7.17	1.32	5.74	1.56
12	14.33	3.17	7.51	1.82	**5.69**	1.65

the others. The kNN achieved the overall best results.

Now, analysing each database *per se*, we observe for the Biosec database that the best average classification rate was obtained with kNN. In terms of the cardinality of features, we note that the best average result, 0.37%, obtained with a subset of 12 features, is followed closely by the value 0.4% with only a cardinality of 7. And this again encourages the use of feature selection since the computational time and complexity is lowered if we lower the number of features.

Concerning the MobBIO*fake*, undoubtedly the classification errors obtained are higher than the other databases, what is not unexpected as we already referred. The best average results were obtained with the SVM classifier, 12.50% , but corresponding to a high cardinality, 12. Not very far form this value we find a subset with much lower cardinality, 4, for the kNN, with an average error of 12.62%.

Analysing the Clarkson results, we note that the combination of features improved considerably the results when compared with the performance of the features individually. The best average result was obtained with SVM, 5.69%, this value is not as good as the Biosec best result but is better than the MobBIO*fake* one, but unfortunately it reffers to a subset of high cardinality, 12. However, we may find a 3^{rd}-best value with a cardinality of 7.

6 CONCLUSIONS AND FUTURE WORK

The actuality of the iris liveness detection topic is unquestionable. As the field of application of iris recognition broads, to embrace the demands of a society highly dependant on mobile and portable devices, the necessity of improving the security urges. To achieve new methods it is also necessary to explore new scenarios of image acquisition and this leads to the necessity of adequate, freely, public available databases.

In this work, we constructed a new database for iris liveness detection purposes with images acquired in unconstrained conditions and with a handheld device. This database was tested for state-of-the-art methods and the results were compared with the results obtained for two existing and tested databases. The MobBIO*fake* database proved itself to be more challenging and our results may not be considered satisfactory but lead to a new more challenging scenario.

Published works present methods tested with existing databases which achieve excellent results, (0% error classification rate). However, we note that some of these methods are closely connected with the particular database characteristics. The results with this database did not achieve that excellent accuracy, but we consider this justifiable by the fact that we avoided the use of methods strongly dependent on the images used, such as ratios of iris and pupil radius or areas, among others.

For future work, we foresee the necessity of improving the existing methods and develop new ones more suitable to the new imaging scenarios. Another aspect to invest is the segmentation step which preferably should be automatic, however, the iris segmentation problem constitutes by itself a whole new set of challenges.

We participated in an iris liveness competition, the "LivDet Competition 2013" (Clarkson University and of Technology, 2013a), held as part of the IEEE BTAS 2013[1]. We applied this methodology combined with an automatic segmentation method (Monteiro et al., 2013; Monteiro et al., 2014) and achieved the first place[2].

ACKNOWLEDGEMENTS

The first author would like to thank the Fundação para a Ciência e Tecnologia (FCT) - Portugal for the financial support for the PhD grant with reference SFRH/BD/74263/2010. The second author would like to thank the National Council for Scientific and Technological Development (CNPq) - Brazil.

REFERENCES

Abhyankar, A. and Schuckers, S. (2009). Iris quality assessment and bi-orthogonal wavelet based encoding for recognition. *Pattern Recognition*, 42(9):1878 – 1894.

Blind Ref, B. R. (2013). Reference removed for blind review.

Clarkson University, N. D. U. and of Technology, W. U. (2013a). Liveness Detection-iris competition 2013. IEEE BTAS 2013. http://people.clarkson.edu/projects/biosal/iris/.

Clarkson University, N. D. U. and of Technology, W. U. (2013b). Liveness Detection-iris competition 2013. IEEE BTAS 2013. http://people.clarkson.edu/projects/biosal/iris/results.php.

Daugman, J. (1998). Recognizing people by their iris patterns. *Information Security Technical Report*, 3(1):33–39.

Daugman, J. (2002). How iris recognition works. In *International Conference on Image Processing*, volume 1, pages I–33 – I–36.

Daugman, J. (2004). Iris recognition and anti-spoofing countermeasures. In *7-th International Biometrics conference*.

Fierrez, J., Ortega-Garcia, J., Torre Toledano, D., and Gonzalez-Rodriguez, J. (2007). Biosec baseline corpus: A multimodal biometric database. *Pattern Recognition*, 40(4):1389–1392.

Galbally, J., Alonso-Fernandez, F., Fierrez, J., and Ortega-Garcia, J. (2012a). A high performance fingerprint liveness detection method based on quality related features. *Future Generation Computer Systems*, 28(1):311–321.

Galbally, J., Fierrez, J., and Ortega-Garcia, J. (2007). Vulnerabilities in biometric systems: attacks and recent advances in liveness detection. *DATABASE*, 1(3):4.

Galbally, J., Ortiz-Lopez, J., Fierrez, J., and Ortega-Garcia, J. (2012b). Iris liveness detection based on quality related features. In *5th IAPR International Conference on Biometrics (ICB)*, pages 271–276. IEEE.

GIMP, G. (2008). Image manipulation program. *User Manual, Edge-Detect Filters, Sobel, The GIMP Documentation Team*.

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions*, (6):610–621.

He, X., An, S., and Shi, P. (2007). Statistical texture analysis-based approach for fake iris detection using support vector machines. In *Advances in Biometrics*, pages 540–546. Springer.

He, X., Lu, Y., and Shi, P. (2009). A new fake iris detection method. In *Advances in Biometrics*, pages 1132–1139. Springer.

Jain, A. and Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 19(2):153–158.

Kanematsu, M., Takano, H., and Nakamura, K. (2007). Highly reliable liveness detection method for iris recognition. In *SICE, 2007 Annual Conference*, pages 361–364. IEEE.

Lee, E., Park, K., and Kim, J. (2005). Fake iris detection by using purkinje image. In *Advances in Biometrics*, volume 3832 of *Lecture Notes in Computer Science*, pages 397–403. Springer Berlin / Heidelberg.

Li, J., Wang, Y., Tan, T., and Jain, A. K. (2004). Live face detection based on the analysis of fourier spectra. In *Defense and Security*, pages 296–303. International Society for Optics and Photonics.

Ma, L., Tan, T., Wang, Y., and Zhang, D. (2003). Personal identification based on iris texture analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 25(12):1519–1533.

Monteiro, J. C., Oliveira, H. P., Sequeira, A. F., and Cardoso, J. S. (2013). Robust iris segmentation under unconstrained settings. In *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 180–190.

Monteiro, J. C., Sequeira, A. F., Oliveira, H. P., and Cardoso, J. S. (2014). Robust iris localisation in challenging scenarios. In *CCIS Communications in Computer and Information Science*. Springer-Verlag.

Pudil, P., Novovičová, J., and Kittler, J. (1994). Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125.

Ratha, N. K., Connell, J. H., and Bolle, R. M. (2001). An analysis of minutiae matching strength. In *Audio-and*

[1] http://www.btas2013.org/

[2] http://people.clarkson.edu/projects/biosal/iris/results.php

Video-Based Biometric Person Authentication, pages 223–228. Springer.

Ruiz-Albacete, V., Tome-Gonzalez, P., Alonso-Fernandez, F., Galbally, J., Fierrez, J., and Ortega-Garcia, J. (2008). Direct attacks using fake images in iris verification. In *Biometrics and Identity Management*, pages 181–190. Springer.

S. Schuckers, K. Bowyer, A. C. and Yambay, D. (2013). *Liviness Detection - Iris Competition 2013*. http://people.clarkson.edu/projects/biosal/iris/.

Stearns, S. D. (1976). On selecting features for pattern classifiers. In *Proceedings of the 3rd International Joint Conference on Pattern Recognition*, pages 71–75.

Une, M. and Tamura, Y. (2006). liveness detection techniques. *IPSJ Magazine*, 47(6):605–608.

Wei, Z., Qiu, X., Sun, Z., and Tan, T. (2008). Counterfeit iris detection based on texture analysis. In *ICPR 2008. 19th International Conference on Pattern Recognition.*, pages 1–4. IEEE.

Whitney, A. W. (1971). A direct method of nonparametric measurement selection. *Computers, IEEE Transactions*, 100(9):1100–1103.