From the Iriscode to the Iris: A New Vulnerability of Iris Recognition Systems

Javier Galbally[a], Arun Ross[b], Marta Gomez-Barrero[a], Julian Fierrez[a], Javier Ortega-Garcia[a]

[a]Biometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid.
C/ Francisco Tomas y Valiente 11, 28049 Madrid. Spain.
[b]Integrated Pattern Recognition and Biometrics Lab (i-PRoBe), West Virginia University.
PO Box 6222, Morgantown, WV 26506. USA.

Abstract

A binary iriscode is a very compact representation of an iris image, and, for a long time, it has been assumed that it did not contain enough information to allow the reconstruction of the original iris. The present work proposes a novel probabilistic approach to reconstruct iris images from binary templates and analyzes to what extent the reconstructed samples are similar to the original ones (that is, those from which the templates were extracted). The performance of the reconstruction technique is assessed by estimating the success chances of an attack carried out with the synthetic iris patterns against a commercial iris recognition system. The experimental results show that the reconstructed images are very realistic and that, even though a human expert would not be easily deceived by them, there is a high chance that they can break into an iris recognition system. Furthermore, as the proposed reconstruction methodology is able to generate not just one, but a large amount of iris-like patterns with iriscodes which fall within the intra-class variability of a genuine user, the proposed method has other potential applications including enrollment improvement or individuality studies.

Keywords: Image reconstruction, Biometric systems, Iris recognition, Binary iriscode, Security, Individuality.

Email addresses: `javier.galbally@uam.es` (Javier Galbally), `arun.ross@mail.wvu.edu` (Arun Ross), `marta.barrero@uam.es` (Marta Gomez-Barrero), `julian.fierrez@uam.es` (Julian Fierrez), `javier.ortega@uam.es` (Javier Ortega-Garcia)

1. Introduction

Although being relatively young compared to other mature and long-used security technologies, biometrics have emerged in the last decade as a pushing alternative for applications where automatic recognition of people is needed. Certainly, biometrics are very attractive and useful for the final user: forget about PINs and passwords, you are your own key [1, 2]. However, we cannot forget that as any technology aimed to provide a security service, biometric systems are exposed to external attacks which could compromise their integrity [3]. Thus, it is of special relevance to understand the threats to which they are subjected and to analyze their vulnerabilities, in order to prevent possible attacks and develop adequate countermeasures which increase their benefits for the final users.

Among the different existing biometric traits, iris has been traditionally regarded as one of the most reliable and accurate [4]. After some preprocessing steps in which the iris is localized, segmented and normalized, the vast majority of iris recognition systems perform some type of filtering operation in order to generate the final binary template (i.e., iriscode) which is stored in the enrollment phase. Then, in the authentication phase, matching is performed between iriscodes applying some specific measure that operates at bit level such as the very widely used Hamming distance [5, 6].

The iriscode has been adopted as a *de facto* standard by most iris-based systems, as it is a very efficient and compact representation of the specific discriminative characteristics contained within a person's iris pattern. As such, it has been a common belief in the biometric community that binary templates do not comprise enough information in order to reconstruct the original iris image from them [7]. Furthermore, iriscodes have been proven to be unique and random for *real* iris images [8].

However, some recent studies have arisen different concerns regarding the soundness of these widely spread believes [9, 10]. Are iriscodes really impossible to be reversed engineered in order to recover the original iris pattern from them? Is it possible to generate different *synthetic* iris-like patterns which yield iriscodes very similar to one given? In summary, can we generate *synthetic* images that match a specific binary template deceiving this way iris recognition systems?

In the present work we address all these questions proposing a novel probabilistic approach based on genetic algorithms for the generation of iris-like synthetic patterns whose corresponding iriscodes match that of a genuine user. Two main goals are pursued:

- On the one hand, explore whether the reconstructed images produced by the new method may be used to carry out attacks against state-of-the-art and commercial systems (e.g., injecting the reconstructed sample in the communication channels or manufacturing a fake printed iris). This will also serve as validation for the new technique.

- On the other hand, determine if it is possible to generate different synthetic patterns with very similar iriscodes to one given.

The second objective defies the individuality of binary templates: synthetic images visually different to an original sample may produce iriscodes which fall within the intra-class variability of the genuine user, belonging this way to the same identity according to iris recognition systems. In this new scenario, further questions that fall out of the scope of the present work may be posed regarding the security of iris-based algorithms. Can iris recognition be trusted solely to the matching of iriscodes? Should some type of image-based recognition approach be added in order to detect synthetic images with similar binary templates but visually different? What other countermeasures could be developed to repel the attacks?

The work has been carried out from a computer-based perspective. This means that our goal is not to generate iris images that could fool a human expert, but that are considered as genuine by automatic iris recognition systems. Even so, different strategies to make the synthetic patterns look as realistic as possible are also explored in the experimental part of the article presenting statistical results regarding the perception that people have of the reconstructed images appearance.

In order to follow a fully reproducible experimental protocol which permits the comparison of the results with future studies, experiments are carried out on two publicly available databases. Furthermore, the iris recognition systems used for development and testing are well known and commercial state-of-the-art systems which may be easily obtained by any interested party.

The rest of the article is structured as follows. Iris recognition is briefly summarized in Sect. 2. The novel probabilistic iris reconstruction algorithm is presented in Sect. 3. Then, the databases and iris recognition systems used in the experimental protocol are described in Sect. 4. In Sect. 5 we present and analyze the development and validation results of the algorithm. Conclusions are finally drawn in Sect. 6.

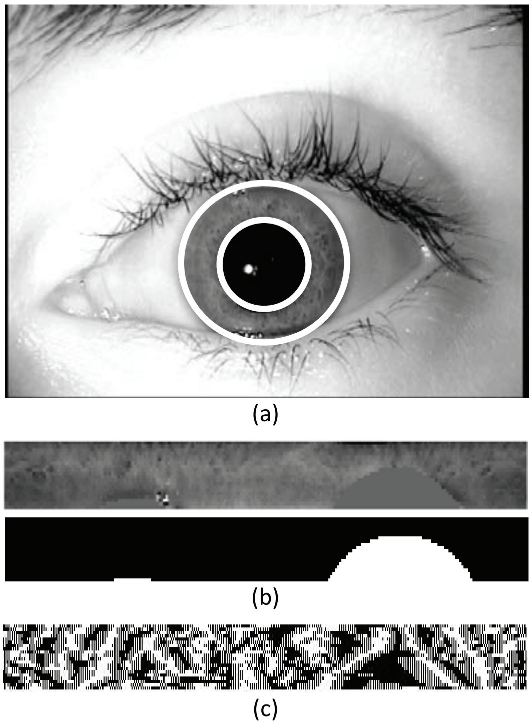

Figure 1: Example of the segmentation (a), the normalization and occlusion mask (b), and the encoding (c) stages followed by most iris recognition systems.

2. Summary of Iris Recognition

The objective of this section is to briefly summarize those aspects of the very complex iris recognition problem which are directly related to the present study and which are essential for the correct understanding of the work. For a more comprehensive, descriptive and self-contained review on automatic iris recognition the reader is referred to [8, 11, 12, 13, 14, 15].

Common iris recognition systems comprise five different stages: iris acquisition, iris location and segmentation, normalization, encoding and matching. As has been mentioned before, the main objective of this work is to reconstruct an iris pattern from its encoded binary template. Thus, although the acquisition and segmentation tasks may be very challenging under certain scenarios (e.g., distance acquisition, uncontrolled lighting conditions, eye deviation, etc.) they are not relevant to this study and will not be treated here.

- **Normalization**. Once the iris has been segmented, the vast majority of iris recognition systems transform the round-like iris pattern in cartesian coordinates to a normalized rectangular image of fixed dimensions in polar coordinates. These are the type of images that will be reconstructed using the algorithm described in this work. The normalization process may be reversed and the normalized iris patterns can be merged again into the original eye images (of the same or of a different user).

- **Encoding**. Although a greater diversity of methods have been reported in this stage compared to the normalization one, most of them use some type of filtering strategy (being the Gabor filters the most widely used) prior to a quantization of the filtered output phase that generates the final binary representation of the iris image (i.e., the iriscode).

Finally, the matching is performed between two iriscodes using in general some bitwise operator such as the Hamming distance. In most cases, in the segmentation stage, a mask showing the occluded areas of the iris (e.g., by the eyelids or eyelashes) is also given as output. Then the matching score is only computed on the "useful" bits of the iriscode.

In Fig. 1 an example of the normalization and encoding stages is shown. The original iris image appears on top (a) with the two white circles marking the boundaries of the segmented iris. Its corresponding normalized image together with the mask showing the occluded areas (b) and its iriscode (c) are shown below.

3. The Reconstruction Method

To give generality to the problem being addressed, some mathematical notation is introduced in this section. For the particular case of iris image reconstruction, in the following, $\mathbf{B}$ represents the compromised iriscode of the user whose iris image is being reconstructed, $\mathbf{I_R}$ represents the reconstructed *normalized* iris image which is a solution to the problem, $\mathbf{B_R}$ its associated iriscode and δ is the matching threshold that determines if two iris images are coming from the same subject.

Problem statement. Consider a $R \times C$ dimensional matrix $\mathbf{I_R}$ of real values, which is divided into $H \times L$ square blocks of dimension $R/H \times C/L$, with $H \leq R$ and $L \leq C$. This matrix is mapped by some unknown function $\mathcal{F}$ to a binary matrix $\mathbf{B_R}$ (i.e., $\mathbf{B_R} = \mathcal{F}(\mathbf{I_R})$) of dimensions $K \times W$ (being in general K a multiple of R and W a multiple of C).

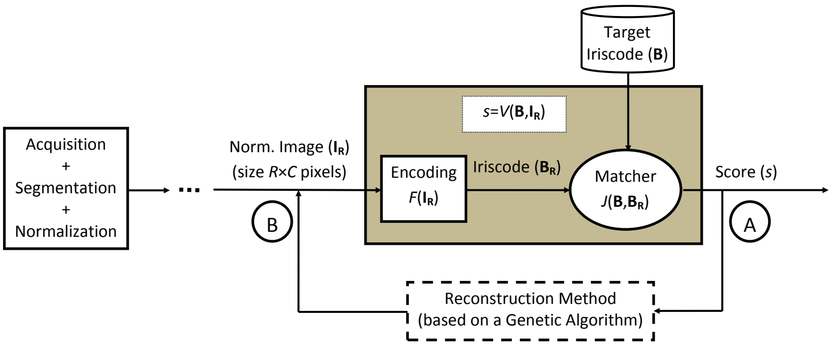

Figure 2: General diagram of the scheme followed in the present work. A detailed diagram of the reconstruction approach (dashed rectangle) is given in Fig. 3 where points A and B show, respectively, the input and output of the algorithm.

Consider the problem of finding an $\mathbf{I_R}$ matrix such that, its associated $\mathbf{B_R}$ matrix (unknown), produces a similarity score (s) bigger than a certain threshold δ, when it is compared to a *known* binary matrix $\mathbf{B}$ according to some unknown matching function $\mathcal{J}$, i.e., $\mathcal{J}(\mathbf{B}, \mathbf{B_R}) > \delta$.

For clarity, we will define a new function $\mathcal{V}$ as: $\mathcal{V}(\mathbf{B}, \mathbf{I_R}) = \mathcal{J}(\mathbf{B}, \mathcal{F}(\mathbf{I_R})) = \mathcal{J}(\mathbf{B}, \mathbf{B_R}) = s$

Assumptions. Let us assume: That we have access to the evaluation of the function $\mathcal{V}(\mathbf{B}, \mathbf{I_R})$ for several trials of $\mathbf{I_R}$.

Algorithm. The problem stated above may be solved using a genetic algorithm to optimize the similarity score given by the system, according to the general diagram shown in Fig. 2. Genetic algorithms, which have shown a remarkable performance in optimization problems [16], are a heuristic search tool that iteratively applies certain rules inspired in natural evolution to a population of individuals (possible solutions) according to a given fitness function which has to be optimized. At each generation (i.e., iteration) the algorithm evolves towards better solutions. In this particular case:

- The fitness value associated to each individual (normalized iris image) is the matching score, $s = \mathcal{V}(\mathbf{B}, \mathbf{I_R})$.

- Usually genetic algorithms operate with individuals which are binary vectors. For this problem, the genetic algorithm has been modified to work with matrices of real values (i.e., $\mathbf{I_R}$) where each of the $H \times L$ blocks in which they are divided represents a gene of the individual.

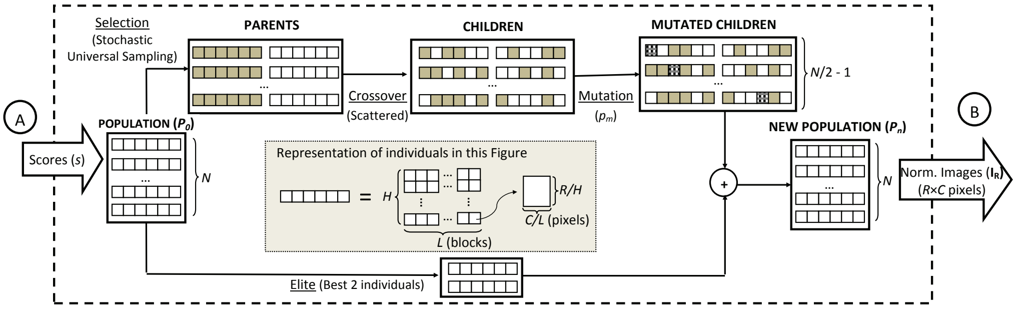

Figure 3: Diagram of the probabilistic method proposed in the present work for the reconstruction of iris images from their iriscode. Points A and B (input and output of the reconstruction algorithm respectively) may be seen for reference in Fig. 2. As it is shown in the shaded chart in the center of the figure, although individuals are represented as vectors for simplicity, strictly they are matrices of size $R \times C$ pixels divided into $H \times L$ blocks.

As can be seen in Fig. 3, the steps followed by the reconstruction algorithm are:

1. Generate an initial population P_0 with N individuals of size $R \times C$ (i.e., dimensions of the normalized iris images), and divide each of the individuals into $H \times L$ rectangular blocks.
2. Compute the similarity scores s^i of the individuals ($\mathbf{I_R}^i$) of the population P_0, $s^i = \mathcal{V}(\mathbf{B}, \mathbf{I_R}^i)$, with $i = 1, \ldots, N$.
3. Four rules are used at each iteration to create the next generation P_n of individuals from the current population:
   (a) **Elite**: the two individuals with the maximum similarity scores are kept unaltered for the next generation.
   (b) **Selection**: certain individuals, the *parents*, are chosen by stochastic universal sampling [17]. This way, the individuals with the highest fitness values (similarity scores) are more likely to be selected as parents for the next generations: one subject can be selected 0 or many times. From the original N individuals, only $N - 2$ are eligible (as the best two are the elite) from which $N/2 - 1$ *fathers* and $N/2 - 1$ *mothers* are chosen.
   (c) **Crossover**: parents are combined to form the $N - 2$ *children* of the next generation following a scattered crossover method: a random binary matrix of size $H \times L$ is created and the genes (blocks) of the child are selected from the first parent where the value of the random matrix is 1, and from the second when it is 0 (viceversa for the second child).

(d) **Mutation**: random changes are applied to the blocks of the new children with a mutation probability p_m. When a block is selected for mutation it is changed for the equivalent block in the individual of the population with the highest fitness value.

4. Redefine $P_0 = P_n$ and return to step 2.

Stopping criteria. The algorithm stops when: i) the best fitness score of the individuals in the population is higher than the threshold δ (i.e., the image has been successfully reconstructed), ii) the variation of the similarity scores obtained in a number of generations is lower than a previously fixed value, or iii) when the maximum number of generations is reached.

Important notices. There are different important characteristics of the reconstruction method presented above that should be highlighted as they differentiate it from other previously published iris reconstruction techniques [10]:

- Due to the probabilistic nature of the four rules being applied, the algorithm produces different solutions at each execution, even when its initialization and parameter values are the same. This enables the reconstruction of more than one normalized iris images ($\mathbf{I_R}$) with very similar iriscodes ($\mathbf{B_R}$) to the target one ($\mathbf{B}$).
- The algorithm does not need to know the mapping function $\mathcal{F}$ between the normalized iris images ($\mathbf{I_R}$) and their corresponding iriscodes ($\mathbf{B_R}$).
- The algorithm does not need to know the matching function $\mathcal{J}$.
- The algorithm does not need to know the function $\mathcal{V}$, just its outcome to given inputs.
- No *real* iris images are involved in the reconstruction process. As will be explained in Sect. 4 the initial population P_0 is taken from a database of fully *synthetic* iris images.

4. Experimental Protocol: Databases and Systems

As it is shown in Fig. 4 the experimental protocol is divided into a development and a validation stage, where two different databases and two different iris recognition systems have been used in order to ensure totally unbiased results. All of them are publicly available so that the results obtained in this study are fully reproducible and may be compared with future similar works.

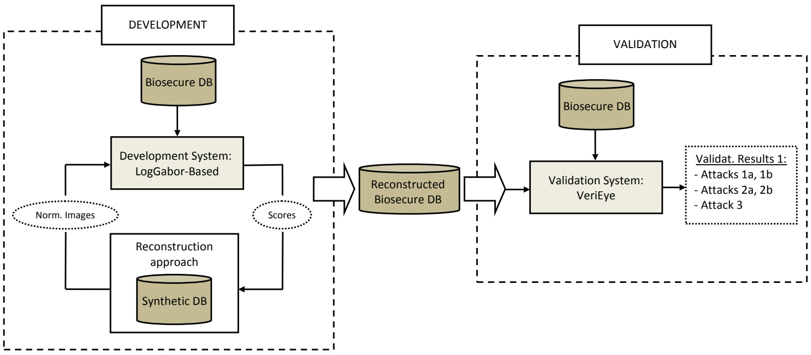

Figure 4: Diagram of the experimental protocol followed in the present work. The databases and systems used are highlighted with a darker shade. The protocol is described in Sects. 4 and 5.

4.1. Databases

Two databases, one of real samples and one of synthetic samples, are used in the experiments. The iris images to be reconstructed are taken from the real database (Biosecure DB), while the synthetic dataset (SDB) is used for the initialization of the reconstruction algorithm (see Fig. 4).

As was described in Sect. 3, the reconstruction method proposed in the present work needs a set of iris images for its initialization. This pool of initial samples is taken from a database of fully synthetic iris images for two main reasons: on the one hand, to avoid any possible overlap between the reconstructed images and those used in the reconstruction process (which could lead to overoptimistic results), and, on the other hand, to avoid having any real images involved in the reconstruction method.

- **The real database: Biosecure DB**. The real images to be reconstructed in the experiments are taken from the iris subcorpus included in the Desktop Dataset of the multimodal BioSecure database [18], which comprises voice, fingerprints, face, iris, signature and hand of 210 subjects, captured in two time-spaced acquisition sessions. This database was acquired thanks to the joint effort of 11 European institutions and has become one of the standard benchmarks for biometric performance and security evaluations [19]. It is publicly available through the BioSecure Foundation[1].

[1] http://biosecure.it-sudparis.eu/AB

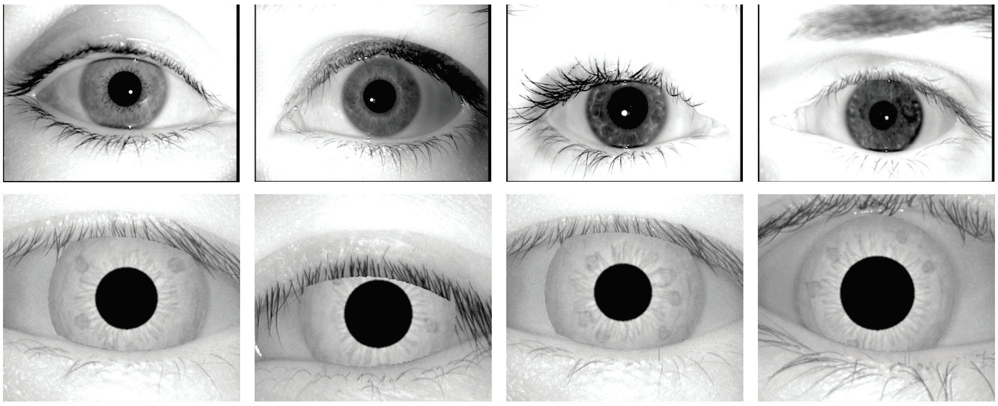

Figure 5: Typical examples of the iris images that can be found in the two databases used in the experiments: real (top) and synthetic (bottom).

The database consists of three datasets captured under different acquisition scenarios, namely: i) Internet Dataset (DS1, captured through the Internet in an unsupervised setup), ii) Desktop Dataset (DS2, captured in an office-like environment with human supervision), and iii) the Mobile Dataset (DS3, acquired on mobile devices with uncontrolled conditions). The iris subset used in this work includes four grey-scale images (two per session) per eye, all captured with the Iris Access EOU3000 sensor from LG. In the experiments both eyes of each subject have been considered as separate users, leading this way to $210 \times 2 \times 4 = 1,680$ iris samples.

- **The synthetic database: SDB**. Being a database that contains only fully synthetic data it is not subjected to any legal constraints and is publicly available through the CITeR research center[2].

  The synthetic irises are generated following the method described in [20], which is divided in two stages. In the first stage, a Markov Random Field model is used to generate a background texture representing the global iris appearance. In the next stage, a variety of iris features such as radial and concentric furrows, collarette and crypts, are generated and embedded in the texture field. The database includes seven grey-scale images of 1,000

[2] http://www.citer.wvu.edu/

different subjects.

Typical examples of the eye images that can be found in Biosecure DS2 (top) and SDB (bottom) are shown in Fig. 5. We can observe that, as was our intention in order to avoid biased results, the samples in both datasets are totally different.

4.2. Iris recognition systems

Two different iris recognition systems are used in the experiments (see Fig. 4). The first one, which consists of totally accessible software modules, is used as development system for the reconstruction of the iris images. The second one, totally different to the previous, is used in the validation stage in order to determine if the reconstructed images are recognized as authentic by systems following different approaches in the encoding stage to that used for development.

- **Development: LogGabor filters-based** [21]. For the development stage, where the real iris images are reconstructed, a modified version of the iris recognition system developed by L. Masek [21] is used. This system was selected for several reasons: i) it is publicly available and its source code may be freely downloaded[3], ii) although its performance is certainly lower than that of current state-of-the-art iris recognition systems, it is widely used in many iris related publications to give baseline results, and iii) it is divided in independent software modules which permit the access to the matching score (requirement of the proposed reconstruction method).

  The different stages involved in iris recognition (described in Sect. 2) are implemented following a *classical* approach: i) *segmentation*, the method proposed in [22] is followed, modelling the iris and pupil boundaries as circles; ii) *normalization*, a technique based on Daugmans rubber sheet model that maps the segmented iris region into a 2D array is used [11]; iii) *feature encoding*, produces a binary template of $20 \times 480 = 9,600$ bits and the corresponding noise mark (representing the eyelids areas) by filtering the normalized iris pattern with 1D Log-Gabor wavelets and quantizing the filtered output to four levels (i.e., two bits) according to [11]; and iv) *matching*, a modified Hamming distance that takes into account the noise mask bits is used.

[3] www.csse.uwa.edu.au/pk/studentprojects/libor/sourcecode.html

- **Validation: VeriEye** [23]. For the validation experiments the VeriEye commercial system marketed by Neurotechnology[4] is used as a final test to determine the attacking potential of the reconstructed iris images. The motivation for its selection is twofold, i) on the one hand, it has proven an outstanding reliability being ranked among the top performing systems in the NIST Iris Exchange (IREX) independent evaluation in 2009 [24], and, ii) on the other hand, being a commercial system it works as a black-box for the user, who has no knowledge of the algorithms used at any of the stages of the iris recognition process (being a commercial system its implementation details are industrial secrets). This way, the results are ensured to be totally unbiased and not due to a specific adaptation of the reconstruction algorithm to a given validation system.

5. Results: Performance

In addition to avoid biased results, the experimental framework has been designed to evaluate the performance of the reconstruction algorithm and its degree of compliance with the main objectives set in this work: i) if the iris images reconstructed following the proposed method are able to compromise the security of iris recognition systems (main goal of the present work), ii) if the reconstruction scheme is able to produce different iris-like patterns with an iriscode very similar to one given (secondary goal of the present work).

For this purpose, as was already introduced in Sect. 4.2, two totally different iris recognition systems were used: one in the development stage and the other one for validation purposes (see Fig. 4).

5.1. Development experiments: LogGabor filters-based system

The objectives of this first set of experiments are: i) to reconstruct the real iris images in Biosecure DB starting from their iriscodes, and ii) to fix the values of the different parameters involved in the reconstruction algorithm.

In order to achieve these two goals, one sample of each of the 420 users present in the Biosecure DB (right and left irises of 210 subjects) were randomly selected and their iriscode computed according to the publicly available iris recognition system developed by Masek [21]. The dimensions of the normalized iris images produced by this system are $R \times C = 20 \times 240$ and the size of their corresponding binary templates $K \times W = 20 \times 480$ (i.e., each pixel is coded with two bits).

[4] http://www.neurotechnology.com/verieye.html

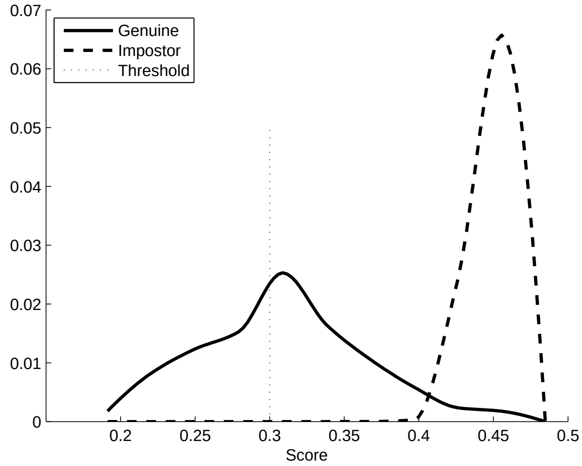

Figure 6: Genuine and impostor score distributions of the iris recognition system used in the development experiments. The selected positive matching threshold is marked with a vertical dotted line, $\delta = 0.3$.

This same system was then used to generate 5 different reconstructed images of each binary template using the algorithm proposed in the present contribution (i.e., the algorithm was applied 5 times to reconstruct each iriscode), thus leading to a database of $5 \times 420 = 2,100$ reconstructed iris images (named Reconstructed Biosecure DB in Fig. 4).

In order to determine the positive matching threshold δ at which an iriscode is considered to have been successfully reconstructed, the iris recognition system performance was evaluated on the Biosecure DB. Genuine scores were computed matching the first sample of each user to the other 3 images of that same user (i.e., $420 \times 3 = 1,260$ genuine scores), while impostor scores were generated comparing the first iris of each user to the first sample of the remaining users in the database (i.e., $420 \times 419 = 175,980$ impostor scores). The two sets of similarity values are depicted in Fig. 6, where the selected positive matching threshold has been highlighted with a vertical dotted line. We can observe that, below that value, $\delta = 0.3$, the probability of having an impostor score is almost zero. Thus, two iris images producing such a similarity score may be considered to come from the same user.

Extensive experiments were undertaken to determine the most effective parameter values of the reconstruction algorithm, finally finding a good operating point for: population size $N = 80$, mutation probability $p_m = 0.003$, and block

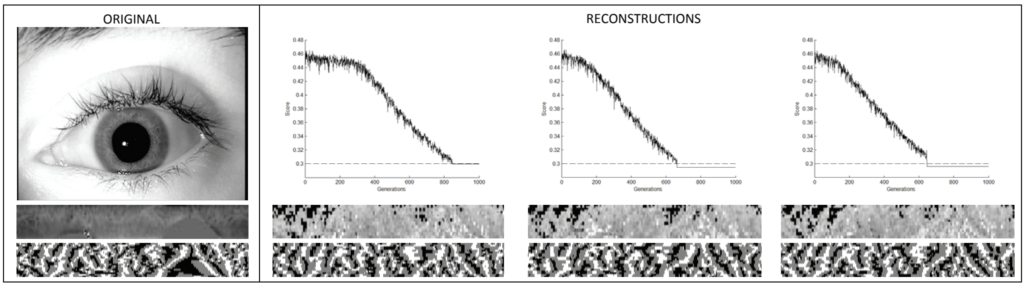

Figure 7: Three example executions (right) of the reconstruction algorithm for the same original image (left). For the reconstruction samples the evolution of the score through the generations is shown on top (positive matching threshold marked with a horizontal dashed line), with the final reconstructed normalized image and its corresponding iriscode below.

size $R/H \times C/L = 2 \times 2$ pixels (i.e., each normalized image is divided into $H \times L = 10 \times 120$ blocks).

In Fig. 7 three different reconstructions of an original normalized iris image are shown. We can see that, although the reconstructed patterns do not look like the original one and that the forming blocks may be distinguished, their corresponding iriscodes are all very similar among themselves and present a high degree of resemblance with the original. On top of each reconstructed image the evolution of the score through the iterations is shown. The optimization process may be clearly distinguished, reaching a lower value at each generation. Marked with a horizontal dashed line is the positive matching threshold $\delta = 0.3$.

5.2. Validation experiments: VeriEye

The iris images reconstructed in the development stage are used to test the vulnerabilities of the VeriEye iris recognition system (see the validation chart in Fig. 4). As mentioned in Sect. 4.2, this system operates as a black-box, that is, given an input, it returns an output with no information about the algorithms followed to get that final result. Several remarks have to be made regarding the inputs and outputs of VeriEye:

- *Inputs*. Normalized iris samples in polar coordinates are not accepted by VeriEye. The input to the system has to be an image containing a *circular* iris in cartesian coordinates. For this reason, in order to attack the system, all the reconstructed irises were transformed into desnormalized images such as those shown in Fig. 8.

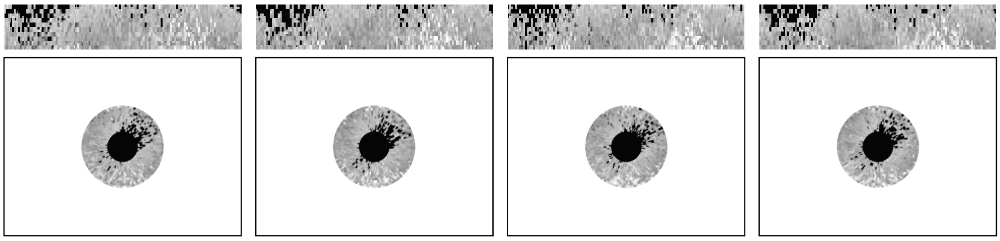

Figure 8: Four reconstructed iris images in polar coordinates (top) all recovered from the same original iris, and their corresponding desnormalized images in cartesian coordinates used to attack the VeriEye commercial system (bottom).

- *Outputs*. The system output is only a similarity score in case of a positive matching. When the matching threshold is not reached, a 0 is returned, making this way more difficult the possibility of a hill-climbing attack [25]. In case that an error occurs during the recognition process (most likely at the segmentation stage), a negative value is returned.

The performance of the attacks is measured in terms of its Success Rate (SR), which is defined as the percentage of successful attacks (A_s) out of the total carried out (A_T), i.e., SR $= A_s/A_T \times 100$. The key factor to compute the SR is to define what constitutes an attack and when it is considered to be successful. In the experiments, three representative attacks will be taken into account in order to estimate the performance of the reconstructed iris images:

1. **Attack 1**: 1 reconstruction *vs* 1 real. In this case the attack is carried out on a 1 on 1 basis. That is, one reconstructed image is matched against one real image and, if the resulting score exceeds the fixed matching threshold, the attack has been successful. Two possible scenarios may be distinguished in this case depending on the real image being attacked:
   (a) The real image being attacked is the original sample from which the synthetic images were reconstructed. In this scenario the total number of attacks performed which will be used to compute SR$_{\mathrm{1a}}$ is $A_{T\mathrm{1a}} = 420 \times 5 = 2,100$.
   (b) The real image being attacked is one of the other three samples of the same user present in the Biosecure DB. For this experiment the total number of attacks performed which will be used to compute SR$_{\mathrm{1b}}$ is $A_{T\mathrm{1b}} = 420 \times 3 \times 5 = 6,300$.

FAR	**SR** (%) - VeriEye					
	SR_{1a}	SR_{1b}	SR_{2a}	SR_{2b}	SR_3	**Average**
0.1%	81.2	66.7	96.2	92.8	96.7	**86.7**
0.05%	79.2	63.4	96.2	91.4	95.2	**85.1**
0.01%	77.3	60.9	95.2	90.9	93.8	**83.6**
0.0001%	69.0	49.1	92.8	82.8	82.9	**75.3**

Table 1: SR of the different attacking scenarios considered against VeriEye at the four operating points tested.

2. **Attack 2**: 5 reconstructions *vs* 1 real. In this case all five reconstructions are matched against the real sample. The attack is successful if at least one of the synthetic images is able to access the system. This represents the most likely attack scenario analyzed in other related vulnerability studies [9]: the iriscode of the legitimate user is compromised and the intruder makes different reconstructions of the iris to try to break the system. The attacker will gain access if any of the reconstructions gets a positive score. The same two scenarios as in attack 1 can be considered here, being the total number of attacks carried out in each of them $A_{T2\mathrm{a}} = 420$ and $A_{T2\mathrm{b}} = 420 \times 3 = 1,260$. The resulting success rates will be noted as SR_{2a} and SR_{2b}, respectively.
3. **Attack 3**: 5 reconstructions *vs* average (4 real). It is a common practice in many biometric recognition systems to match the test sample against several stored templates and return as final score the average of all the matchings. To emulate this scenario each reconstructed iris image is matched against the four samples of the real user available in the Biosecure DB. The attack is successful if the average of the four matchings *of any of the five reconstructions* is higher than the given operating threshold. Thus, in this case, the total number of attacks performed in order to compute SR_3 is $A_{T3} = 420$.

In general, the success chances of an attack are highly dependent on the False Acceptance Rate (FAR) of the system. Thus, the vulnerability of the system to the attacks with the reconstructed images is evaluated at three operating points corresponding to: FAR=0.1%, FAR=0.05%, and FAR=0.01%, which, according to [26], correspond to a low, medium and high security application, respectively. For completeness, the system is also tested at a very high security operating point corresponding to FAR$\ll$0.01%.

As was mentioned before, this commercial system does not return impostor scores (i.e., they are always 0) which means that its FAR may not be statistically computed on a given database. In order to fix the threshold for the different operating points, a deterministic equation is given in the documentation enclosed with the system.

During the experiments, the system was unable to segment (i.e., reported an error) 1.4% of the real images in the Biosecure DB. This means that, for these cases, a sample from the legitimate user would have not been able to access the system. Thus, the highest SR that may be reached by the attacks is 98.6%. Moreover, 0.5% of the reconstructed images were not correctly segmented (these are computed as unsuccessful attacks).

Several observations can be made from the results of the validation experiments carried out on VeriEye shown in Table 1:

- The high performance of the reconstruction algorithm is confirmed, reaching an average SR of around 85% for the three usual operating points considered and over 95% for the most likely attacking scenario (i.e., SR_{2a}).

- Even for an unrealistically high security point (i.e., FAR=0.0001%), the reconstructed images would have, on average, almost 75% chances of breaking the system.

- As expected, it is more probable that the synthetic samples are positively matched to the original image from which they were reconstructed than to other real images of the same user (see the decrease in the SR between SR_{1a} *vs* SR_{1b} and between SR_{2a} *vs* SR_{2b}).

- Even so, the reconstructed images still present a high probability of breaking the system even when the stored templates are not the one from which they were recovered (average SR of SR_{1b} and SR_{2b} around 75%).

- Furthermore, for the case of using several real samples of the user for verification (SR_3), the reconstructed samples are still able to access the system for around 94% of the attempts in the usual operating points, and for over 80% in the extremely high operating point tested.

- When the SR of attacks 1 and 2 are compared (i.e., 1vs1 and 1vs5, respectively) an increase of around 27% may be observed on average when several reconstructions of the iris image are available. These results prove the higher attacking potential of the probabilistic reconstruction approach compared to deterministic algorithms that can only generate one iris image from each iris code.

- Besides, a new possible vulnerability of iris recognition applications has been raised, as the tested system positively matches images with a black circle in the middle and a white background (such as the ones showed in Fig. 8) that should by no means be recognized as an eye image.

6. Conclusions

The results retrieved from the systematic experiments carried out using the proposed reconstruction approach have shown that it is unlikely to deceive a human expert with the generated fake samples. This does not seem to be simply due to a limitation of the method but rather to a lack of information in the iriscodes themselves that does not allow to fully reconstruct the complete original greyscale iris pattern.

However, this work has demonstrated, as it was its main objective, that iriscodes do have sufficient information to generate synthetic iris-like images with very similar binary templates to the genuine pattern. The experimental findings presented in the article have shown that an eventual attack against iris recognition systems using such reconstructed images would have very high success chances provided that we can present the fake samples to the system.

The experimental findings have also shown the ability of the proposed probabilistic approach to reconstruct many synthetic samples from one given iriscode (second goal of the work). This fact not only increases drastically its attacking success rate compared to methods that can only generate one synthetic sample from each binary template, but also opens the possibility to other potential applications which fall outside the security field on which this paper is focused:

- The probabilistic approach may be used to synthetically increase the amount of available data of a subject (i.e., the number of training samples) in order to improve the performance of iris recognition systems [27, 28].

- Biometric samples are personal data and different privacy concerns have arisen regarding their distribution and protection [29]. The proposed reconstruction method is able to generate synthetic iris patterns totally different to the original (see Fig. 7) which are, nevertheless, positively matched to the user's identity. This means that the synthetic samples may be considered as an alternative representation of the user's identity and, as such, they may be stored as enrollment templates avoiding this way possible privacy issues as the genuine trait would only be used for testing.

- As mentioned above, the reconstructed samples may be considered as a new iris pattern class which, although is visually different from that of the genuine user (see Fig. 7), represents the same identity (i.e., iriscode). This way, the reconstruction method may be useful to conduct experiments on the individuality of the iris trait [30, 31, 32]. Given that we can generate synthetic templates with an iriscode which falls within the intra-class variability of a different class (i.e., user), can we consider iriscodes really unique?

Furthermore, the work has arisen the need of including in commercial iris recognition systems some verification strategy to check that the samples presented to the system are those of an eye and not some other type of simple iris-like looking image.

It may be argued that, for attacks such as the one considered in this work to be successful, the first condition is that the original user's template falls in the wrong hands. This may be difficult, yet possible, in classic biometric systems where the enrolled templates are kept in a centralized database. In this case, the attacker would have to extract the information from the database or intercept the communication channel when the stored template is released for matching.

However, nowadays, Match-on-Card (MoC) applications in which the matching is performed inside a smartcard where the enrolled template of the user is also stored, are rapidly growing due to several appealing characteristics such as their scalability and privacy (you carry the only copy of your biometric data) [33]. Furthermore, biometric data is being stored in many official documents such as the new biometric passport [34], some national ID cards [35], or the US FIPS-201 Personal Identity Verification inititatives (PIV) [36] and the ILO Seafarers Identity Card Program [37]. In spite of the clear advantages that these type of distributed systems present, templates are more likely to be compromised as it is easier for the attacker to have physical access to the storage device and, as has already been proven [38], fraudulently obtain the information contained inside. This makes MoC systems potentially more vulnerable to the type of threat described in this article.

In either case, centralized or MoC systems, the present work has proven that attacks using reconstructed iris images constitute a real threat, stressing out the importance of equipping automatic recognition systems with all the necessary measures against it. These may include two complementary approaches:

- **Prevention**. Aimed to avoid the users' templates being compromised, for example by securely storing biometric data or protecting the communication channels through encryption [39].

- **Protection**. Aimed to minimize the probabilities of the attack of breaking into the system should a template be compromised. This would be the case of biometric-based countermeasures to detect synthetic from real iris images such as the liveness-detection techniques [40].

Research works such as the one presented in this article pretend to bring some insight into the difficult problem of biometric security evaluation through the systematic study of biometric systems vulnerabilities so that effective countermeasures that minimize the effects of the detected threats may be developed, in order to increase the confidence of the final users in this thriving technology.

Acknowledgment

This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Challenge (TEC2009-11186) from Spanish MEC, TABULA RASA (FP7-ICT-257289) and BEAT (FP7-SEC-284989) from EU, and *Cátedra UAM-Telefónica*.

Javier Galbally was awarded with a fellowship "José Castillejo" from the Spanish MEC in order to carry out this work.

Arun Ross was supported by grants from the Center for Identification Technology Research (CITeR) and National Science Foundation (NSF CAREER Award IIS 0642554).

Javier Galbally would also like to thank Arun Ross as the main coordinator of the Integrated Pattern Recognition and Biometrics Laboratory (i-PRoBe) at the West Virginia University for hosting him during the development of this research work.

References

[1] A. K. Jain, A. Ross, S. Pankanti, Biometrics: a tool for information security, IEEE Trans. on Information Forensics and Security 1 (2) (2006) 125–143.

[2] J. Wayman, A. Jain, D. Maltoni, D. Maio, Biometric systems. Technology, design and performance evaluation, Springer, 2005.

[3] B. Schneier, The uses and abuses of biometrics, Communications of the ACM 48 (1999) 136.

[4] A. Jain, P. Flynn, A. Ross (Eds.), Handbook of biometrics, Springer, 2008.

[5] J. Daugman, How iris recognition works, in: Proc. IEEE Int. Conf. on Image Processing (ICIP), 2002, pp. I.33 – I.36.

[6] D. Monro, S. Rakshit, D. Zhang, DCT based iris recognition, IEEE Trans. on Pattern Analysis and Machine Intelligence 29 (2007) 586–595.

[7] International Biometric Group, Generating images from templates, White paper (2002).

[8] J. Daugman, Probing the uniqueness and randomness of iris codes: Results from 200 billion iris pair comparisons, Proceedings of the IEEE 94 (2006) 1927–1935.

[9] R. Cappelli, D. Maio, A. Lumini, D. Maltoni, Fingerprint image reconstruction from standard templates, IEEE Trans. on Pattern Analysis and Machine Intelligence 29 (2007) 1489–1503.

[10] S. Venugopalan, M. Savvides, How to generate spoofed irises from an iris code template, IEEE Trans. on Information Forensics and Security 6 (2011) 385–394.

[11] J. Daugman, How iris recognition works, IEEE Trans. on Circuits and Systems for Video Technology 14 (2004) 21–30.

[12] J. Daugman, The importance of being random: Statistical principles of iris recognition, Pattern Recognition 36 (2003) 279–291.

[13] J. Daugman, New methods in iris recognition, IEEE Trans. on Systems Man and Cybernetics - Part B: Cybernetics 37 (2007) 1167–1175.

[14] J. Daugman, Iris Recognition, Springer, 2008, Ch. 4, pp. 71–90.

[15] K. Bowyer, K. Hollingsworth, P. Flynn, Image understanding for iris biometrics: a survey, Computer Vision and Image Understanding 110 (2008) 281–307.

[16] D. E. Goldberg, Genetic Algorithms in Search Optimization and Machine Learning, Addison Wesley, 1989.

[17] J. E. Baker, Reducing bias and inefficiency in the selection algorithm, in: Proc. Int. Conf. on Genetic Algorithms and their Application (ICGAA), L. Erlbaum Associates Inc., 1987, pp. 14 – 21.

[18] J. Ortega-Garcia, J. Fierrez, F.Alonso-Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J.-L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B.Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. W. R. Ng, M. Fairhurst, J. Hennebert, A. Humm, M.Tistarelli, L. Brodo, J.Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S.-K. Pavani, A. Frangi, L. Akarun, A.Savran, The multi-scenario multi-environment BioSecure multimodal database (BMDB), IEEE Trans. on Pattern Analysis and Machine Intelligence 32 (2010) 1097–1111.

[19] A. Mayoue, B. Dorizzi, L. Allano, G. Chollet, J. Hennebert, D. Petrovska-Delacretaz, F. Verdet, Guide to biometric reference systems and performance evaluation, Springer, 2009, Ch. BioSecure multimodal evaluation campaign 2007 (BMEC 2007), pp. 327–372.

[20] S. Shah, A. Ross, Generating synthetic irises by feature agglomeration, in: Proc. IEEE Int. Conf. on Image Processing (ICIP), 2006, pp. 317–320.

[21] L. Masek, P. Kovesi, Matlab source code for a biometric identification system based on iris patterns, Master's thesis, School of Computer Science and Software Engineering, University of Western Australia (2003).

[22] V. Ruiz-Albacete, P. Tome-Gonzalez, F. Alonso-Fernandez, J. Galbally, J. Fierrez, J. Ortega-Garcia, Direct attacks using fake images in iris verification, in: Proc. COST 2101 Workshop on Biometrics and Identity Management (BioID), Springer LNCS-5372, 2008, pp. 181–190.

[23] Neurotechnology. [link].
URL `http://www.neurotechnology.com/verieye.html`

[24] P. Grother, E. Tabassi, G. W. Quinn, W. Salamon, IREX I: Performance of iris recognition algorithms on standard images, Tech. rep., National Institute of Standards and Technology (2009).

[25] J. Galbally, C. McCool, J. Fierrez, S. Marcel, On the vulnerability of face verification systems to hill-climbing attacks, Pattern Recognition 43 (2010) 1027–1038.

[26] ANSI-NIST, ANSI x9.84-2001, biometric information management and security (2001).

[27] M. E. Munich, P. Perona, Visual identification by signature tracking, IEEE Trans. on Pattern Analysis and Machine Intelligence 25 (2003) 200–217.

[28] J. Galbally, J. Fierrez, M. Martinez-Diaz, J. Ortega-Garcia, Improving the enrollment in dynamic signature verification with synthetic samples, in: Proc. IAPR Int. Conf. on Document Analysis and Recognition (ICDAR), 2009.

[29] S. Prabhakar, S. Pankanti, A. K. Jain, Biometric recognition: security and privacy concerns, IEEE Security and Privacy 1 (2003) 33–42.

[30] S. Pankanti, S. Prabhakar, A. K. Jain, On the individuality of fingerprints, IEEE Trans. on Pattern Analysis and Machine Intelligence 24 (2002) 1010–1025.

[31] R. Bolle, S. Pankanti, J. Connell, N. Ratha, Iris individuality: a partial iris model, in: Proc. of Int. Conf. on Pattern Recognition, 2004, pp. 927–930.

[32] K. Hollingsworth, K. Bowyer, P. Flynn, The best bits in an iris code, IEEE Trans. on Pattern Analysis and Machine Intelligence 31 (2009) 964–973.

[33] C. Bergman, Advances in Biometrics: sensors, algorithms and systems, Springer, 2008, Ch. Match-on-card for secure and scalable biometric authentication, pp. 407–422.

[34] ICAO, ICAO document 9303, part 1, volume 2: Machine readable passports - specifications for electronically enabled passports with biometric identification capability (2006).

[35] Government of Spain. [link].
URL `http://www.dnielectronico.es/`

[36] NIST, NIST Special Publication 800-76, Biometric Data Specification for Personal Identity Verification, Feb. 2005.

[37] ILO, ILO SID-0002, Finger Minutiae-Based Biometric Profile for Seafarers Identity Documents, Intl Labour Organization. (2006).

[38] J. van Beek, ePassports reloaded, in: Black Hat USA Briefings, 2008.

[39] U. Uludag, S. Pankanti, S. Prabhakar, A. K. Jain, Biometric cryptosystems: issues and challenges, Proc. of the IEEE 92 (2004) 948–960.

[40] Z. Wei, X. Qiu, Z. Sun, T. Tan, Counterfeit iris detection based on texture analysis, in: Proc. IAPR Int. Conf. on Pattern Recognition (ICPR), 2008.