

Securing Iris Recognition Systems Against Masquerade Attacks

Javier Galbally^a, Marta Gomez-Barrero^a, Arun Ross^b, Julian Fierrez^a and Javier Ortega-Garcia^a

^aBiometric Recognition Group - ATVS, EPS, Universidad Autonoma de Madrid. SPAIN
Email: {javier.galbally, marta.barrero, julian.fierrez, javier.ortega}@uam.es; ^bIntegrated Pattern
Recognition and Biometrics Lab (i-PRoBe), West Virginia University. US
Email: arun.ross@mail.wvu.edu

ABSTRACT

A novel two-stage protection scheme for automatic iris recognition systems against masquerade attacks carried out with synthetically reconstructed iris images is presented. The method uses different characteristics of real iris images to differentiate them from the synthetic ones, thereby addressing important security flaws detected in state-of-the-art commercial systems. Experiments are carried out on the publicly available Biosecure Database and demonstrate the efficacy of the proposed security enhancing approach.

Keywords: Iris recognition, Inverse Biometrics, Security, Vulnerabilities, Countermeasures

1. INTRODUCTION

Biometrics is the science of recognizing individuals based on their physical and behavioral traits such as fingerprints, face, iris, gait and voice.¹ Tremendous efforts have been directed towards improving the matching accuracy of biometric systems as assessed by various performance metrics.² However, other aspects of a biometric system are relatively under explored. In particular, only recently have researchers focused on the possibility of reconstructing synthetic biometric images that can potentially be injected into a biometric system thereby undermining its utility.

Among the various biometric traits that have been studied in the recent past, iris is commonly believed to be reliable and accurate - a claim that has been reinforced with large-scale experiments on operational datasets.³ The templates used by most iris-based systems are the so called *IrisCodes*, which are binary representations of the iris pattern.⁴ Since an IrisCode is an extremely compact representation of the iris, it has been a common belief in the biometric community that binary templates do not divulge enough information to reconstruct the original iris image from them⁵ (i.e., the template extraction scheme has been traditionally considered to be a oneway function).

However, this belief has been recently questioned in the literature by researchers who have explored the reversibility of IrisCodes.^{6,7} In particular, in,⁷ a probabilistic reconstruction method based on genetic algorithms was presented and used to evaluate the vulnerabilities of a commercial iris recognition system (VeriEye, by Neurotechnology*) against several masquerade attacks. The attacks were carried out by matching the synthetically reconstructed iris images to the original patterns, showing a lack of resistance of the system against this type of threat.

In the present work we address the security issues disclosed in⁷ by proposing effective countermeasures to detect the synthetic iris patterns reconstructed from a real IrisCode. Thus, the main objective of the work is to develop a reliable solution to an actual vulnerability flaw present in commercial biometric applications in order to enhance the level of security offered to the final users.

The rest of the paper is structured as follows. The iris reconstruction method is briefly summarized in Section 2. The results of the attacks are presented in Section 3. The novel two-stage protection approach is described in Section 4. Conclusions are finally drawn in Section 5.

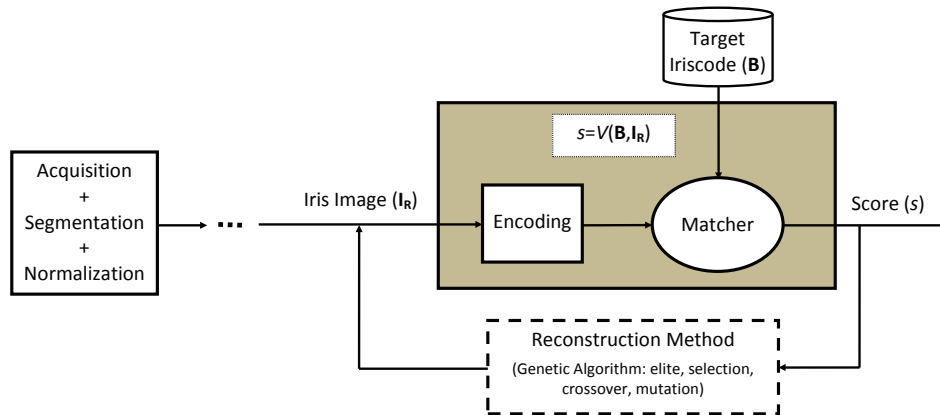


Figure 1. General diagram of the probabilistic reconstruction scheme based on genetic algorithms proposed in.⁷

2. THE IRIS RECONSTRUCTION APPROACH

The challenging reverse engineering problem of reconstructing an iris pattern from its IrisCode, was solved in⁷ using a probabilistic approach based on genetic algorithms. These algorithms, which have shown a remarkable performance in optimization problems,⁸ are heuristic search tools that iteratively apply certain rules inspired by 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 a better solution.

For the particular problem considered here, the fitness value associated to each individual (iris image) is the matching score s^i that results when comparing each of the possible solutions in the population (i.e., synthetic iris images, \mathbf{I}_R^i) at each iteration with the target iriscode (\mathbf{B}) being reconstructed (i.e., $s^i = \mathcal{V}(\mathbf{B}, \mathbf{I}_R^i)$), where $i = 1, \dots, N$, being N the size of the population. A general diagram of the reconstruction approach is shown in Fig. 1.

The matching scores to be optimized are computed using the iris recognition system developed by Masek,⁹ which is publicly available[†] and used in many iris-related publications to obtain baseline results.

In order to accomplish the optimization task, the genetic algorithm uses four rules at each iteration to create the next generation of individuals from the current population, namely: *i*) **Elite**, the two individuals with the maximum similarity scores are kept unaltered for the next generation; *ii*) **Selection**, certain individuals, the *parents*, are chosen by stochastic universal sampling,¹⁰ so that the individuals with the highest fitness values (similarity scores) are more likely to be selected as parents for the next generations; *iii*) **Crossover**, parents are combined to form the *children* of the next generation following a scattered crossover method;⁸ *iv*) **Mutation**, random changes are applied to the genes (basic forming components of the individual) of the new children with a certain mutation probability.

The algorithm stops when: *i*) the best fitness score of the individuals in the population is higher than a certain 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.

A complete description and evaluation of the reconstruction algorithm may be found in.⁷ In that work it is demonstrated that the proposed scheme is able to generate not just one, but many synthetic iris patterns that have very similar IrisCodes as the real one.

3. THE PROBLEM: ATTACKING THE SYSTEM

In order to determine the robustness of iris recognition systems to attacks carried out with synthetic samples reconstructed according to the approach described in Section 2, different experiments are carried out using the iris subset included in the Biosecure DB DS2.¹¹

*<http://www.neurotechnology.com/verieye.html>

†www.csse.uwa.edu.au/pk/studentprojects/libor/sourcecode.html

Table 1. Success Rates (SR) of the attacks against VeriEye with the two types of images considered at the four operating points tested.

Type	FAR (%)			
	0.1	0.05	0.01	0.0001
Raw (SR %)	96.2	96.2	95.2	92.8
Embedded (SR %)	97.1	97.0	95.8	93.0

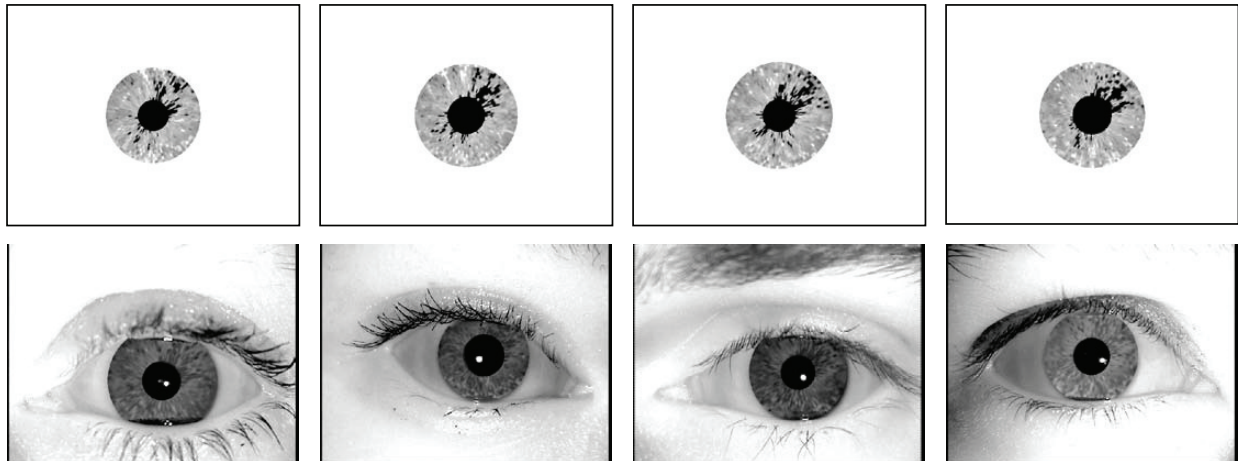


Figure 2. Typical examples of the two types of images used in the attacks: raw reconstructed iris samples (top, referred to as Rec-Raw), and reconstructed iris samples embedded in a real eye image (bottom, referred to as Rec-Embedded).

This iris subset includes four grey-scale images per eye captured in two separate sessions (two per session), all acquired with the Iris Access EOU3000 sensor from LG in an office-like environment with human supervision. In the experiments, the two eyes of each of the 210 subjects included in the database have been considered as separate classes (users). In total, the iris subset comprises of $210 \times 2 \times 4 = 1,680$ samples.

For the vulnerability assessment experiments, one sample from each of the 420 users present in the dataset was randomly chosen. Then, the proposed reconstruction scheme was used to generate five synthetic iris patterns from each of the selected real irises, resulting in a database of $5 \times 420 = 2,100$ reconstructed iris images.

The reconstructed synthetic images are then used to attack the VeriEye commercial system at three realistic operating points corresponding to $\text{FAR}=0.1\%$, $\text{FAR}=0.05\%$, and $\text{FAR}=0.01\%$. For completeness, the system is also tested at a very high security operating point corresponding to $\text{FAR} \ll 0.01\%$.

The performance of the attacks is assessed according to their Success Rate (SR), which is defined as the ratio of the number of successful attacks to the total number of attacks performed. An attack is successful when *any* of the five reconstructed iris patterns is positively matched to the original sample by the verification system being tested (i.e., VeriEye).

Two different types of images were used for the attacks, *i*) the raw reconstructed images (see Fig. 2, top), and *ii*) the reconstructed samples after being embedded in a real eye image (see Fig. 2, bottom). These two sets of images are referred to as Rec-Raw and Rec-Embedded, respectively. The Success Rates (SR) obtained by both types of images for the different operating points considered are shown in Table 1.

Several observations can be made from the results shown in Table 1: *i*) the reconstructed images represent a real threat to iris recognition systems (even commercial ones), reaching a SR of over 95% (on average, for the operating points considered); *ii*) even for an unrealistically high security point (i.e., $\text{FAR}=0.0001\%$), the reconstructed images would have over 90% chances of breaking the system; *iii*) the tested system positively matches very simple iris-like looking images that should by no means be recognized as an eye image (raw images with a black circle in the middle and a white background shown in Fig. 2 top).

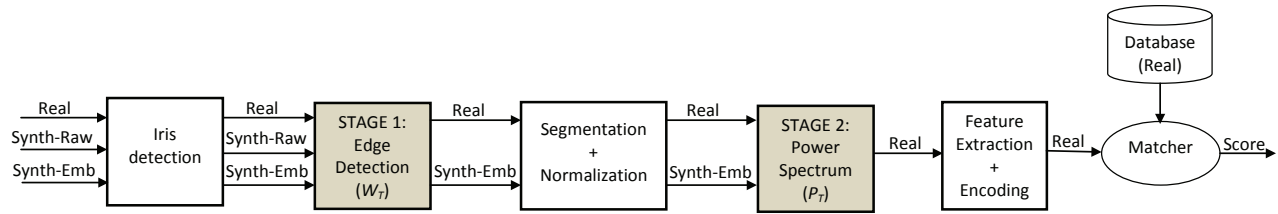


Figure 3. General diagram of the two-stage protection approach proposed in the present work.

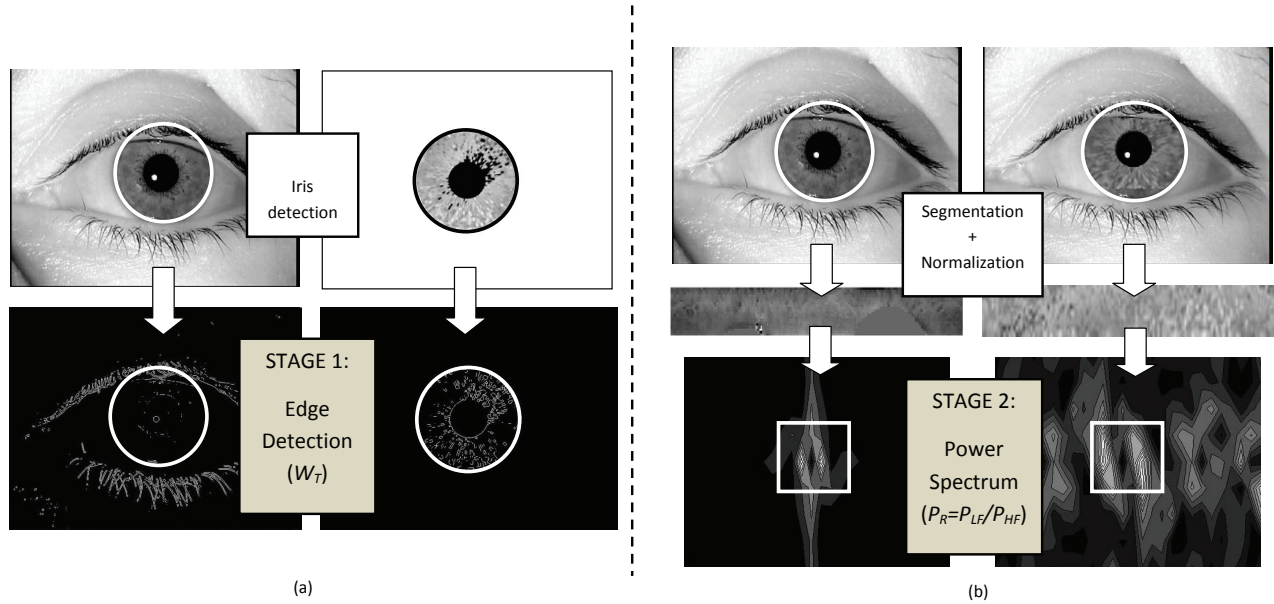


Figure 4. Example of Stage 1 (a) and Stage 2 (b) of the proposed protection method. In Stage 1 W_T is the number of edge pixels outside the white circle (outer iris boundary). In Stage 2 $P_R = P_{LF}/P_{HF}$ is the ratio between the power present in low-frequencies (within the white square boundary, P_{LF}), and in high-frequencies (outside the white square boundary, P_{HF}). A lighter shade denotes greater power.

4. THE SOLUTION: PROTECTION METHOD

From the results presented in Sect. 3 it is clear that efficient countermeasures must be developed and embedded in practical biometric applications in order to deal with the security flaws disclosed in the vulnerability assessment experiments.

The global solution proposed here classifies an input sample into the real or synthetic class following a two-stage approach as is shown in Fig. 3. Each of the two stages is developed as a specific countermeasure against a particular vulnerability of the system, using different discriminating properties of the iris samples.

- **Stage 1: Edge detection.** This is designed to protect the system against fraudulent access attempts carried out with very simple iris-like images such as the ones shown in the top row of Fig. 2, which should never be accepted by the system (Synth-Raw in Fig. 3). This protection mechanism is invoked right after the iris has been detected and before any preprocessing is performed. It operates on the whole captured ocular image in order to make sure that the sample being presented to the system is as close as possible to a real eye image.

To achieve this goal, edge detection is performed on the image using Sobel filters, and the number of edge pixels detected *outside* the iris boundaries is used as the discriminative feature between real and simple synthetic images (W_T). This ensures that no image with a homogeneous background (e.g., Fig. 2 top) is accepted into the system. An example of the computation of this protection method for a real and a synthetic image may be seen in Fig. 4 (a), where we can observe that, as expected, no edges are detected outside the iris boundaries for the reconstructed sample (i.e., $W_T = 0$).

Table 2. Performance of the proposed two-stage protection method.

	FGR (%)	FSR (%)	ACE (%)
Stage 1 (W_T)	0	0.05	0.025
Stage 2 (P_R)	0	0.24	0.12
Global	0	0.3	0.15

- **Stage 2: Power spectrum.** This is performed on the segmented and normalized iris image. The objective of this second module of the global protection scheme proposed, is to detect those reconstructed irises that have been embedded in a real eye (images shown in Fig. 2 bottom), bypassing this way the first protection stage (Synth-Emb in Fig. 3).

For this purpose, the power spectrum of the input images is used, computed according to the 2-D Discrete Fourier Transform (DFT2). Due to the characteristics of the reconstruction algorithm, the resulting synthetic images are formed by blocks with sharp edges. This configuration results in an abnormal amount of high-frequency energy compared to real irises (which have a much smoother surface).

The property mentioned above may be used to distinguish between real and synthetic irises by computing the ration between the total low-frequency power and that found in the high-frequencies, i.e., $P_R = P_{LF}/P_{HF}$. This ratio is expected to be lower in the synthetic images. An example of the computation of the power spectrum for real and synthetic images may be seen in Fig. 4 (b), where the square boundary which is used to compute P_{LF} (inside) and P_{HF} (outside) is highlighted in white. We can observe how the high frequency components are significantly higher (i.e., lighter shade) in the case of the reconstructed sample.

The proposed protection scheme was applied to the three experimental databases used in this work: Biosecure DB DS2 (1,680 real images), Biosecure Rec-Raw (2,100 synthetic images) and Biosecure Rec-Embedded (2,100 synthetic images). The discriminating parameters W_T and P_R were computed for all the samples. Then, the performance of each of the proposed individual stages and of the global protection method was estimated according to the Average Classification Error (ACE), defined as $ACE = (FGR+FSR)/2$, where the FGR (False Genuine Rate) represents the percentage of fake irides misclassified as real, and the FSR (False Synthetic Rate) computes the percentage of real irides assigned to the fake class.

The performance results of the protection scheme are shown in Table 2. We can observe that: *i*) the proposed approach is extremely effective to detect the attacks described in Section 3 (FGR=0%), *ii*) only very few real samples (mostly of bad quality, e.g., very bright with a white background) are rejected (FSR=0.3%), *iii*) the protection methodology proposed adds very little complexity or delay to the recognition system.

5. CONCLUSIONS

A two-stage protection method against attacks carried out with reconstructed iris images has been presented. The experiments have shown its ability to detect fraudulent access attempts using synthesized iris images, thereby solving important security flaws detected in the vulnerability evaluation of a state-of-the-art iris system.

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 and the development of effective countermeasures that minimize the effects of the detected threats, in order to increase the confidence of the final users in this thriving technology.

6. ACKNOWLEDGEMENTS

This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Challenge (TEC2009-11186) and Bio-Shield (TEC2012-34881) from Spanish MECD, TABULA RASA (FP7-ICT-257289) and BEAT (FP7-SEC-284989) from EU, and *Cátedra UAM-Telefónica*. Arun Ross was supported by the Center for Identification Technology Research (CITeR). Javier Galgally was awarded with a fellowship “José Castillejo” from the Spanish MECD in order to carry out this work.

REFERENCES

- [1] Jain, A. K., Ross, A., and Pankanti, S., "Biometrics: a tool for information security," *IEEE Trans. on Information Forensics and Security* **1**(2), 125–143 (2006).
- [2] Mansfield, A. and Wayman, J., "Best practices in testing and reporting performance of biometric devices," tech. rep., CESG Biometrics Working Group (August 2002).
- [3] Jain, A., Flynn, P., and Ross, A., eds., [*Handbook of Biometrics*], Springer (2008).
- [4] Daugman, J., "How iris recognition works," *IEEE Trans. on Circuits and Systems for Video Technology* **14**, 21–30 (2004).
- [5] International Biometric Group, "Generating images from templates." White paper (2002).
- [6] Venugopalan, S. and Savvides, M., "How to generate spoofed irises from an iris code template," *IEEE Trans. on Information Forensics and Security* **6**, 385–394 (2011).
- [7] Galbally, J., Ross, A., Gomez-Barrero, M., Fierrez, J., and Ortega-Garcia, J., "Iris image reconstruction from binary templates: An efficient probabilistic approach," *Journal of Computer Vision and Image Understanding* (2013). (Minor comments, under 3rd review).
- [8] Goldberg, D. E., [*Genetic Algorithms in Search Optimization and Machine Learning*], Addison Wesley (1989).
- [9] Masek, L. and Kovesi, P., *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).
- [10] Baker, J. E., "Reducing bias and inefficiency in the selection algorithm," in [*Proc. Int. Conf. on Genetic Algorithms and their Application (ICGAA)*], 14 – 21, L. Erlbaum Associates Inc. (1987).
- [11] Ortega-Garcia, J., Fierrez, J., F.Alonso-Fernandez, Galbally, J., Freire, M. R., Gonzalez-Rodriguez, J., Garcia-Mateo, C., Alba-Castro, J.-L., Gonzalez-Agulla, E., Otero-Muras, E., Garcia-Salicetti, S., Allano, L., B.Ly-Van, Dorizzi, B., Kittler, J., Bourlai, T., Poh, N., Deravi, F., Ng, M. W. R., Fairhurst, M., Hennebert, J., Humm, A., M.Tistarelli, Brodo, L., J.Richiardi, Drygajlo, A., Ganster, H., Sukno, F. M., Pavani, S.-K., Frangi, A., Akarun, L., and A.Savran, "The multi-scenario multi-environment BioSecure multimodal database (BMDB)," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **32**, 1097–1111 (2010).