# Face Verification Put to Test: A Hill-Climbing Attack Based on the Uphill-Simplex Algorithm

Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, Javier Ortega-Garcia
Biometric Recognition Group -ATVS, EPS, Universidad Autonoma de Madrid
C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain
marta.barrero@uam.es, javier.galbally@uam.es, julian.fierrez@uam.es, javier.ortega@uam.es

## Abstract

*The vulnerabilities of a PCA-based face verification system against a hill-climbing attack using the uphill-simplex algorithm are studied. Experiments are carried out on the face subcorpus of the publicly available BioSecure DB, where the attack has shown a remarkable performance proving the lack of robustness of the tested system to this type of threat. Furthermore, the proposed attacking scheme is not only able to bypass the security of the recognition system, but it is also capable of reconstructing the users face image, with the privacy concerns that this entails. As a possible countermeasure to minimize the effect of the attack, score quantization is applied. This protection method is able to reduce both the success rate and the efficiency of the attack, however it does not completely succeed in preventing a possible intruder from accessing the system. The study also highlights the high adaptation capabilities of the proposed attack which had already been used to break a signature-based verification system.*

## 1. Introduction

Due to the fact that biometrics [10], as an automatic means of human recognition, constitutes a relatively novel field of research, most efforts undertaken by the different parties involved in the development of this technology (researchers, industry, evaluators, etc.) have been mainly (but not exclusively) directed to the improvement of its performance [11]. This has left partially uncovered other important aspects involved in the complex biometric recognition problem.

In particular, it has not been until recently when biometric security assessment has emerged in the biometric community as a primary field of research, as a consequence of the concern arisen after the classification of the vulnerability points presented in [16] (shown in Fig. 1), and the different efficient attacking algorithms developed in order to compromise the security level given by biometric applications [6, 19].

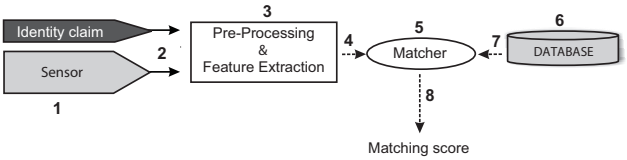


Figure 1. Architecture of an automated biometric verification system. Possible attack points given in [16] are numbered from 1 to 8.

These vulnerability studies have helped to improve the biometric technology by making public certain flaws and by encouraging the industry and researchers to look for solutions to the different threats [9, 22, 18]. This way, the level of security and the convenience offered to the final user are increased.

Most of the existing works studying the vulnerabilities of biometric systems to attacks against the inner modules of the system (those labeled from 2 to 8 in Fig. 1), use some type of variant of the hill-climbing algorithm presented in [17]. Some examples include an attack to a face-based system in [2], and to standard and Match-on-Card minutiae-based fingerprint verification systems in [21] and [12] respectively. These types of attacks take advantage of the score given by the matcher to iteratively change a synthetically generated template until the similarity score exceeds a fixed decision threshold and thereby access to the system is granted. Except for the algorithm proposed in [7], all of these hill-climbing approaches are highly dependent on the technology used, only being usable for very specific types of matchers.

However, recently, a general hill-climbing algorithm based on the uphill-simplex algorithm was presented and tested using a signature verification system [8]. In the present contribution this general method is successfully applied to attack an automatic face recognition system based on eigenfaces, proving this way its biometric independency

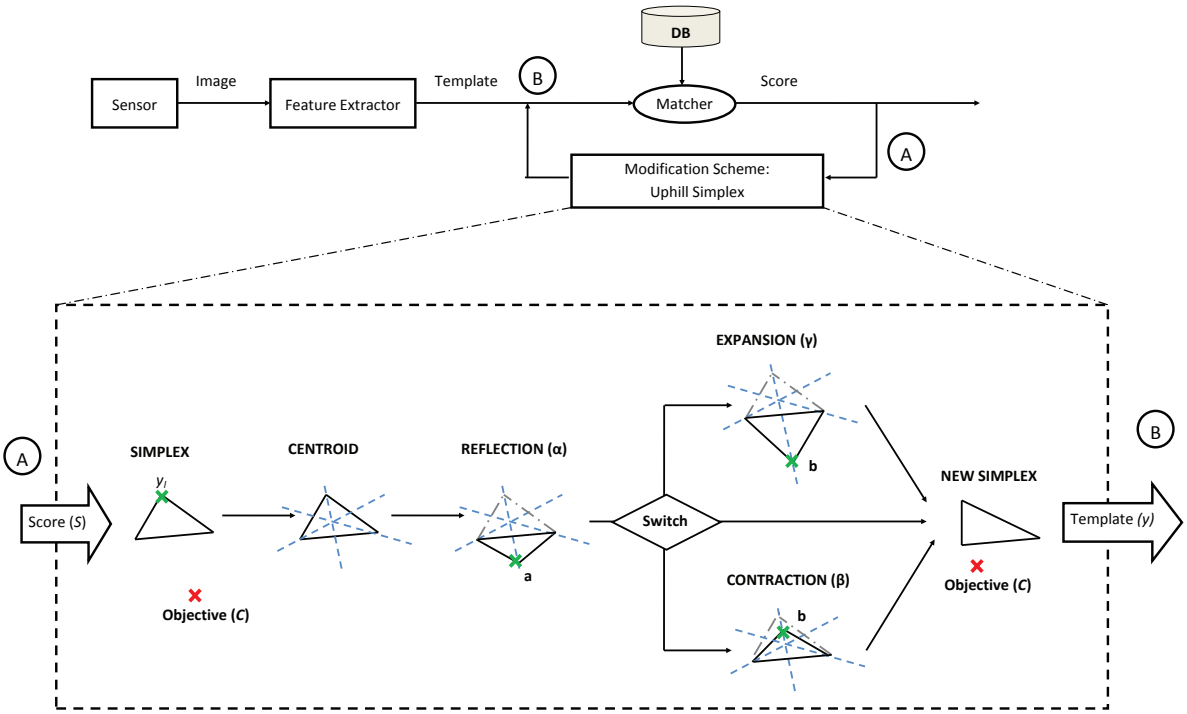


Figure 2. Diagram of the modification scheme for the uphill-simplex algorithm.

and its ability to adapt to different matchers that use fixed length feature vectors of real numbers.

The experiments are conducted on the BioSecure database [14], and clearly show the high attacking potential of the algorithm, pointing out the need to develop adequate countermeasures against it. Furthermore, the hill-climbing approach is shown to be faster than the previous general method proposed in [7] for all the operating points evaluated.

Score quantization is then analysed as a possible method to minimize the effect of the attack. Although it reduces both the success rate and the efficiency of the hill-climbing algorithm, it does not succeed in preventing a possible intruder from accessing the system.

The paper is structured as follows. The hill-climbing attack algorithm used in the experiments is outlined in Sect. 2, while the attacked system is presented in Sect. 3. The database and experimental protocol followed are described in Sect 4. The results of the attack and of the quantization scheme studied as countermeasure are detailed in Sect. 5. Conclusions are finally drawn in Sect. 6.

## 2. Hill Climbing Attack Based on the uphill-simplex Algorithm

In a generic hill-climbing attack, synthetic templates are generated and iteratively modified according to the similarity score given by a matcher, until the verification threshold $\delta$ is reached (as can be seen in Fig. 2 top).

In the present contribution we use the attack based on the uphill-simplex algorithm, first presented in [8]. The core idea behind the algorithm is to iteratively change a simplex (a polygon with $k+1$ vertices in the $k$-dimensional space) so that it approaches the objective (the user account being attacked, defined as $\mathcal{C}$). In each iteration, the similarity score ($s_i$) from each simplex vertex ($\mathbf{y}_i$) to the target ($\mathcal{C}$) is computed, according to a matching function ($\mathcal{J}$), $s_i = \mathcal{J}(\mathcal{C}, \mathbf{y}_i)$, with $i = 1, \ldots, k+1$. The vertex furthest to the objective, $\mathbf{y}_l$, is discarded and substituted by a new point, which, as can be seen in Fig. 2 (bottom), can be computed in three different ways, namely: $i$) *reflection*, according to a previously fixed $\alpha$ parameter; if reflection fails, either $ii$) *expansion* (with the $\gamma$ parameter) or $iii$) *contraction* (with the $\beta$ parameter) are used as a means to compute the new vertex. This process continues until the maximum score of the vertices exceeds the verification threshold or the maximum number of iterations allowed is reached.

## 3. Face Verification System Attacked

The described hill-climbing attack based on the uphill-simplex algorithm is used to evaluate the security of an Eigenface-based face verification system [20]. This technique uses Principal Component Analysis (PCA) to derive

Figure 3. Typical face images that can be found in the BioSecure DB.

a vector which represents the face images in a lower dimensional space, and it was used to present initial face verification results for the recent Face Recognition Grand Challenge [15].

The evaluated system uses cropped face images of size $64 \times 80$ to train a PCA vector space where $80\%$ of the variance is retained. This leads to a system where the original image space of 5120 dimensions is reduced to 100 dimensions (or eigenvectors). Similarity scores are computed in this PCA vector space using the Euclidean distance, as it showed a very competitive performance compared to the rest of the similarity measures tested.

## 4. Experimental Protocol

The experiments are carried out on the face subcorpus included in the Desktop Dataset of the BioSecure multimodal database [14]. BioSecure DB, which is publicly available through the BioSecure Foundation[1], was acquired thanks to the joint effort of 11 European institutions and has become one of the standard benchmarks for biometric performance and security evaluation [13].

The database comprises three datasets captured under different acquisition scenarios, namely: $i$) the Internet Dataset (DS1, captured through the Internet in an unsupervised setup), $ii$) the Desktop Dataset (DS2, captured in an office-like environment with human supervision), and $iii$) the Mobile Dataset (DS3, acquired on mobile devices and with uncontrolled conditions). The Desktop Dataset comprises voice, fingerprints, face, iris, signature and hand of 210 users, captured in two time-spaced acquisition sessions. The face subset used in this work includes four frontal images (two per session) with an homogeneous grey background, and captured with a reflex digital camera without flash. Typical examples of face images that can be found in BioSecure DS2 are shown in Fig. 3.

### 4.1. Performance Evaluation

The performance of the evaluated system is computed using the experimental protocol shown in Fig. 4. The database is divided into: $i$) a training set comprising the first three samples of 170 clients (used to compute both the PCA transformation matrix and the enrolment templates), and $ii$) an evaluation set formed by the fourth image of the previous 170 users (used to compute the genuine scores), and all the 4 images of the remaining 40 users with which the impostor scores are calculated. As a result of using the same subjects for PCA training and client enrolment, the system performance is optimistically biased, and therefore harder to attack than in a practical situation (in which the enrolled clients may not have been used for PCA training). This means that the results presented in this paper are a conservative estimate of the attack's success rate.

<table>
<tr><th></th><th></th><th colspan="2">BioSecure DB (210 Users)</th></tr>
<tr><th>Session</th><th>Sample</th><th>170 Users</th><th>40 Users</th></tr>
<tr><td rowspan="2">1</td><td>1</td><td rowspan="3">Training</td><td rowspan="3">Test (Impostors)</td></tr>
<tr><td>2</td></tr>
<tr><td rowspan="2">2</td><td>1</td></tr>
<tr><td>2</td><td>Test (Clients)</td><td></td></tr>
</table>

Figure 4. Diagram showing the partitioning of the BioSecure DB according to the performance evaluation protocol defined in the present work.

The final score given by the system is the average of the scores obtained after matching the input vector to the three templates of the attacked client model $\mathcal{C}$. In Fig. 5 we can see the False Acceptance Rate (FAR) and False Rejection Rate (FRR) curves of the Eigenface-based system using the described protocol. The system has an Equal Error Rate of $4.85\%$. The three operating points where the hill-climbing algorithm is evaluated (corresponding to FAR = $0.1\%$, FAR = $0.05\%$, and FAR = $0.01\%$) are also highlighted. These operating points correspond to a low, medium, and high security application according to [1].

### 4.2. Experimental Protocol for the Attacks

In order to generate the user accounts to be attacked with the hill-climbing algorithm, we used the train set defined in the performance evaluation protocol (i.e., three first samples of 170 users as shown in Fig. 4). The performance of the attack will be evaluated in terms of the success rate and efficiency, defined as [5]:

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

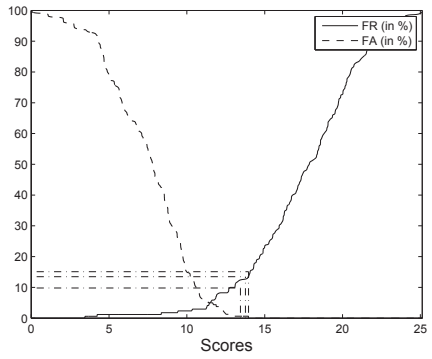


Figure 5. FAR and FRR curves for the system.

- **Success Rate (SR)**: it is the expected probability that the attack breaks a given account. It is computed as the ratio between the number of broken accounts ($A_B$) and the total number of accounts attacked ($A_T = 170$): $\mathrm{SR} = A_B/A_T$. This parameter indicates how dangerous the attack is: the higher the SR, the bigger the threat.

- **Efficiency (Eff)**: it indicates the average number of matchings needed by the attack to break an account. It is defined as $\mathrm{Eff} = 1/\left(\sum_{i=1}^{A_B} n_i/A_B\right)$, where $n_i$ is the number of matchings computed to bypass each of the broken accounts. This parameter gives an estimation of how easy it is for the attack to break into the system in terms of speed: the higher the Eff, the faster the attack.

A direct comparison between the attack performance results obtained in this work and those presented in the only previous work evaluating a general hill-climbing algorithm [7] will also be given in Sect. 5.

# 5. Results

In the first set of experiments, the performance of the attack at different operating points is studied and compared to the results obtained by other state of the art attacking algorithms. Then, we analyse score quantization as a possible protection method and we study its impact on the SR and Eff of the attacking scheme.

The goal of the security evaluation experiments is twofold, *i)* on the one hand, to study the vulnerability of an automatic face recognition system to the proposed hill-climbing algorithm, and *ii)* on the other hand, to test the efficiency of the attack against recognition systems working on different biometric traits (it was already successfully used to attack an on-line signature verification system in [8]).

| FAR | Uphill-simplex HC | | Bayesian HC [7] | |
|---|---|---|---|---|
| | SR | Eff ($\times 10^{-4}$) | SR | Eff ($\times 10^{-4}$) |
| 0.10% | 100% | 22.124 | 99.0% | 11.905 |
| 0.05% | 100% | 22.472 | 98.5% | 9.363 |
| 0.01% | 100% | 21.930 | 86.0% | 2.226 |

Table 1. Eff and SR at the operating points tested, compared to those obtained by the Bayesian hill-climbing attack in [7].

| FAR | Face Verif. System | | Signature Verif. Syst. [8] | |
|---|---|---|---|---|
| | SR | Eff ($\times 10^{-4}$) | SR | Eff ($\times 10^{-4}$) |
| 0.05% | 100% | 22.472 | 91.32% | 8.489 |
| 0.01% | 100% | 21.930 | 88.43% | 7.391 |

Table 2. Eff and SR at the operating points tested, compared to those obtained with the on-line signature verification system tested in [8].

## 5.1. Analysis of different operating points

The performance of the attack is tested at three different operating points, namely: *i)* FAR = 0.10%, *ii)* FAR = 0.05%, *iii)* FAR = 0.01%. These are the same operating points at which a very similar face verification system was evaluated in [7]. Therefore, the results obtained in both works may be compared in a fair fashion. The results of the experiments are detailed in Table 1.

As can be observed, the algorithm presented here successfully breaks all the attacked accounts, contrary to the Bayesian hill-climbing algorithm described in [7], existing a significant SR difference between the two approaches at the last operating point (i.e., 100% vs 86%). Moreover, while the efficiency decreases substantially along the three operating points for the previous algorithm, for the attack proposed in the present work it remains almost invariant, regardless of the operating point considered. This leads to an efficiency which is ten times faster at the last operating point between the two attacks (i.e., $21.930\times10^{-4}$ scores needed to break an account against almost $2.222\times10^{-4}$).

In Table 2 we show the Eff as well as the SR of the attack for the two common operating points (in terms of FAR) attacked in the present work and in [8], where this same algorithm was used to attack an on-line signature verification system. It has to be emphasized that the parameters of the algorithm used in the present work are the same ones which were optimized in [8]: $\alpha = 1.1, \beta = 0.8, \gamma = 1.1$. The fact that the SR and the Eff improve for the present case study proves the robustness of the algorithm: it is able to break totally heterogeneous systems working on different biometric traits and matchers, even improving its performance, without specifically adjusting its parameters.

It should also be noticed that in the present work the hill-climbing attack is initialized from a normal distribution of zero mean and unit variance, that is, the algorithm is able to

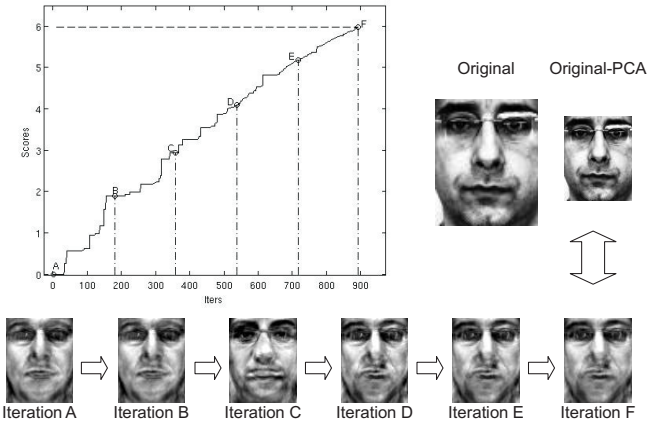


Figure 6. Example of the evolution of the score and the synthetic eigenfaces through the iterations of the attack for a broken account. The dashed line represents the objective threshold.

| QS | $10^{-6}$ | $10^{-2}$ | $10^{-1}$ | 1 | 2.5 | 5 |
|---|---|---|---|---|---|---|
| PI (%) | 9.83 | 7.62 | 2.22 | 0.25 | 0.01 | 0.01 |
| EER (%) | 4.85 | 4.85 | 4.87 | 4.90 | 5.85 | 9.12 |

Table 3. Percentage of the iterations of the hill-climbing attack with a positive score increase (PI), and EER of the system for different quantization steps (QS) of the matching score.

| QS | $10^{-6}$ | $10^{-1}$ | 2.5 |
|---|---|---|---|
| SR | 97.65% | 78.24% | 61.76% |
| Eff ($\times 10^{-4}$) | 20.284 | 20.202 | 13.158 |

Table 4. Performance (in terms of SR and Eff) of the hill-climbing attack against the system for different quantization steps (QS).

break the system without needing any real training faces to generate the first simplex.

In Fig. 6 an example of the execution of the attack is shown. The evolution of the score through the iterations of the algorithm is depicted together with six points (including the first and the last one) of the iterative process (marked with letters A to F). The dashed line represents the objective value to be reached (i.e, the threshold $\delta$). The two upper faces correspond to one of the original images of the attacked user and its representation in the PCA space. The sequence of the six faces below correspond to the feature vector that produced each of the six scores marked with A to F, starting from a random face (A) and finishing with an image (F) capable of breaking the system (i.e., similar enough to the target user labelled as "Original-PCA").

## 5.2. Countermeasuring the Attack: Score Quantization

The results achieved by the hill-climbing attack based on the uphill-simplex algorithm against the face recognition system considered in the experiments have shown its high vulnerability against this type of attacking approach and the need to incorporate some attack protection method that increases its robustness against this threat. In this section we analyse the performance of score quantization as a way to countermeasure the attack.

Score quantization has been proposed as an effective biometric-based approach to reduce the effects of hill-climbing attacks by quantizing the score so that the hill-climbing algorithm does not get the necessary positive feedback to iteratively increase the similarity measure. Although Adler presented a modified attacking algorithm for PCA-based face recognition systems robust to this countermeasure [3], the BioAPI consortium [4] recommends that biometric algorithms emit only quantized matching scores in order to prevent eventual hill-climbing attacks. Here, we will study the efficiency of this attack protection technique against the proposed hill-climbing algorithm.

We will consider the Eigenface-based system operating at medium security operating point (FAR = 0.05%). For the hill-climbing attack we will assume the same configuration used in the vulnerability assessment experiments, $[\alpha, \beta, \gamma] = [1.1, 0.8, 1.1]$ (which, as mentioned before, is taken from [8]).

In order to choose the quantization step we analyse the results obtained from the attack performed in Sect. 5 under the previously described conditions, and the findings are summarized in Table 3. QS stands for Quantization Step, PI is the percentage of iterations out of the total performed in the attack that produced a Positive Increase in the matching score (i.e., the score increase was higher than the quantization step), and EER is the Equal Error Rate of the system for the quantization step considered.

From the results shown in Table 3 we can see that for the last QS considered (5) the EER suffers a big increase (QS

is too big), while for the previous QS values the system performance is not significantly affected. Therefore, the hill-climbing attack is repeated considering the three QS values QS = $10^{-6}$, QS = $10^{-1}$, and QS = $2.5$ . Results are presented in Table 4, where we can see that score quantization reduces the success chances of the attack (for bigger QS, the SR decreases). However, it can also be noticed that the attacking algorithm is quite robust to this type of countermeasure, as even for the biggest value of QS (increasing it would imply a deterioration of the system EER as shown in Table 3), the SR of the attack is still over 60%.

## 6. Conclusions

The robustness of an Eigenface-based face verification system against a hill-climbing attack based on the uphill-simplex algorithm has been studied. Given the 100% success rate reached in all the experiments and the very few comparisons needed for breaking the accounts, its vulnerabilities against this kind of attacks have been clearly shown.

Furthermore, the experimental results reached in the present work added to those presented in [8], have proven that the hill-climbing attack based on the uphill-simplex algorithm can be successfully applied to break automatic recognition systems working on different biometric traits without even specifically adjusting its parameters.

As a possible way to minimize the effect of the attack, score quantization, with several quantization steps, was studied. Although it considerably reduced the SR of the attack and increased the number of comparisons needed, the proposed attacking scheme proved its robustness to this type of countermeasure.

## Acknowledgements

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

## References

[1] ANSI X9.84-2001, Biometric Information Management and Security.

[2] A. Adler. Sample images can be independently restored from face recognition templates. In *Proc. CCECE*, volume 2, pages 1163–1166, 2003.

[3] A. Adler. Images can be regenerated from quantized biometric match score data. In *Proc. CCECE*, pages 469–472, 2004.

[4] BioAPI Consortium. BioAPI specification (version 1.1), March 2001. www.bioapi.org/Downloads/BioAPI

[5] J. Galbally. *Vulnerabilities and Attack Protection in Security Systems Based on Biometric Recognition*. PhD thesis, Universidad Autonoma de Madrid, 2009.

[6] J. Galbally, R. Cappelli, et al. An evaluation of direct and indirect attacks using fake fingers generated from ISO templates. *Pattern Recognition Letters*, 31:725–732, 2010.

[7] J. Galbally, C. McCool, et al. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43:1027–1038, 2010.

[8] M. Gomez-Barrero, J. Galbally, et al. Hill-climbing attack based on the uphill simplex algorithm and its application to signature verification. In *Proc. BioID*, pages 83–94. LNCS-6583, 2011.

[9] A. K. Jain, K. Nandakumar, and A. Nagar. Biometric template security. *EURASIP Journal on Advances in Signal Processing, Special Issue on Biometrics*, 2008.

[10] A. K. Jain, A. Ross, and S. Pankanti. Biometrics: a tool for information security. *IEEE TIFS*, 1(2):125–143, 2006.

[11] A. Mansfield and J. Wayman. Best practices in testing and reporting performance of biometric devices. Technical report, CESG Biometrics Working Group, August 2002. (http://www.cesg.gov.uk/).

[12] M. Martinez-Diaz, J. Fierrez, et al. An evaluation of indirect attacks and countermeasures in fingerprint verification systems. *Patter Recognition Letters*, 32:1643–1651, 2011.

[13] A. Mayoue, B. Dorizzi, et al. *Guide to biometric reference systems and performance evaluation*, chapter BioSecure multimodal evaluation campaign 2007 (BMEC 2007), pages 327–372. Springer, 2009.

[14] J. Ortega-Garcia, J. Fierrez, et al. The multi-scenario multi-environment BioSecure multimodal database (BMDB). *IEEE TPAMI*, 32:1097–1111, 2010.

[15] J. Phillips, P. Flynn, et al. Overview of the face recognition grand challenge. In *Proc. IEEE CVPR*, pages 947–954, 2005.

[16] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication sytems. *IBM Systems Journal*, 40:614–634, 2001.

[17] C. Soutar. Biometric system security. http://www.bioscrypt.com/assets/security_soutar.pdf.

[18] B. Tan and S. Schuckers. A new approach for liveness detection in fingerprint scanners based on valley noise analysis. *Journal of Electronic Imaging*, 17:011009, 2008.

[19] L. Thalheim and J. Krissler. Body check: biometric access protection devices and their programs put to the test. *ct magazine*, pages 114–121, November 2002.

[20] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. IEEE CCVPR*, pages 586–591, 1991.

[21] U. Uludag and A. K. Jain. Attacks on biometric systems: a case study in fingerprints. In *Proc. SPIE-IE*, volume 5306, pages 622–633, 2004.

[22] U. C. von Seelen. Countermeasures against iris spoofing with contact lenses. Technical report, Iridian Technologies, 2005.