Targeted Impersonation As A Tool For The Detection Of Biometric System Vulnerabilities

John D. Bustard John N. Carter Mark S. Nixon
School of Electronics and Computer Science, University of Southampton
target@johndavidbustard.com

Abstract

This paper argues that biometric verification evaluations can obscure vulnerabilities that increase the chances that an attacker could be falsely accepted. This can occur because existing evaluations implicitly assume that an imposter claiming a false identity would claim a random identity rather than consciously selecting a target to impersonate. This paper shows how an attacker can select a target with a similar biometric signature in order to increase their chances of false acceptance. It demonstrates this effect using a publicly available iris recognition algorithm. The evaluation shows that the system can be vulnerable to attackers targeting subjects who are enrolled with a smaller section of iris due to occlusion. The evaluation shows how the traditional DET curve analysis conceals this vulnerability. As a result, traditional analysis underestimates the importance of an existing score normalisation method for addressing occlusion. The paper concludes by evaluating how the targeted false acceptance rate increases with the number of available targets. Consistent with a previous investigation of targeted face verification performance, the experiment shows that the false acceptance rate can be modelled using the traditional FAR measure with an additional term that is proportional to the logarithm of the number of available targets.

1. Introduction

This paper is concerned with the vulnerability of biometric verification. Verification occurs when a user claims an identity which is then validated by comparing a stored biometric signature against their presented biometric features. Whilst no verification process is infallible, significant progress has been made in improving verification accuracy and there are now many commercial biometric systems in regular use. However, recent research[18] has shown how these systems may be vulnerable to deliberate attempts to subvert them.

This paper describes 'targeted biometric impersonation'. Targeted attacks involve finding an existing person with a similar biometric signature and then fraudulently assuming that identity to spoof a verification check. Traditionally, the security of biometric verification has been measured using false acceptance rates. This measurement provides an estimate of the likelihood that an imposter would successfully be accepted by a biometric system if they randomly claimed a false identity. However, it does not accurately measure the vulnerability of such systems to more deliberate attacks. This paper focuses on targeted attacks applied to an iris verification algorithm to highlight how existing evaluation methods obscure system vulnerabilities. By deliberately selecting a legitimate user with similar biometric features, new weaknesses in the underlying biometric system can be revealed and used to increase the false acceptance rate of any imposter. Targeted attacks are a significant vulnerability as they have no artifical traits that can be recognised, either by an automated system or a human supervisor. They are also possible without control over the enrolment procedure or the need for a confederate whose true identity would be made known.

For the targeted attacks described in this paper, an imposter requires access to a copy of the biometric system being subverted and the enrollment information of users. This makes the attack of greatest concern for high risk applications where attackers are more likely to be sophisticated and well resourced. However, increases in the use of social networking, online dating and centralised biometric databases have made identity systems more vulnerable to targeted attacks. These large searchable collections of face and other biometric data increase the chance of finding a target who has a closely matching biometric signature. Such attacks are particularly dangerous as they can be effective both against automated biometrics and manual methods of identification, such as visual passport inspection. The assasination of Al-Mabhouh, a co-founder of the military wing of Hamas, in 2010 highlights this issues as it appears to have been an example of a sophisticated targeted attack. There are currently 27 suspects wanted by Interpol for this assassination. They

entered Dubai with stolen identity information and passed through multiple passport control processes without detection. If biometric systems are to prevent such attacks in future they need to be made robust to targeted impersonation.

The paper starts by surveying the existing literature on the measurement of biometric vulnerabilities. It then examines the effect of targeted impersonation on an iris verification system. The investigation uses a publicly available biometric algorithm and dataset to provide an estimate of how targeted attacks can reduce a biometric system's reliability. The paper then examines how the effectiveness of attacks increases with the number of potential targets. The paper concludes by proposing an improved false acceptance metric for verification performance. The paper's main contributions are the first investigation of the effects of targeted attacks on iris verification. This paper demonstrates the first use of targeted impersonation to identify specific system vulnerabilities that are concealed by traditional evaluations. Targeted attack analysis also provides a more accurate assesment of the importance of countermeasures to such vulnerabilities. The paper also verifies that a previous logarithmic model for the increase in false acceptance rate with number of subjects can also be applied to iris biometrics, further supporting its use as an improved metric for verification performance and as a possible model for the natural variation in biometric features.

2. Background

All biometric systems require some form of evaluation to assess their performance. Three approaches are identified in the literature:

- *Technology evaluation*, which tests computer algorithms against a database of previously obtained biometric data using an algorithm-independent sensor
- *Scenario evaluation*, which uses volunteers to test a system placed within a controlled environment modeled on a proposed application
- *Operational evaluation*, which attempts to analyse performance of biometric systems placed in real applications

Technology evaluations primarily measure verification performance using the false rejection and false acceptance rates of the system under test with different tradeoff priorities[4]. Technology evaluations are most commonly used because they are relatively straightforward to perform, particularly for the comparison of new and existing algorithms. Note, however, that evaluation in such an environment does not necessarily provide an accurate estimate of reliability when the biometric systems are deployed. For example, the relative distribution of subjects within the evaluation data may not be representative of real applications. In particular, most academic evaluation datasets consist of students and staff within engineering departments. Such datasets may also have been collected to maximise diversity and thus represent more dissimilar subjects than would occur in a typical deployment[10].

Many contextual factors can also have a significant effect on verification performance and, as the various biometrics have matured, these factors have been investigated. For example, within face recognition, the impact of image resolution, facial pose, lighting, focus, occlusion, facial expression and aging have all been studied[20]. Recent biometric evaluations have started to assess many of these factors.

More recently, deliberate attempts to attack biometric systems have been investigated. Ratha et al.[17] have identified eight different types of attack based on the part of the biometric system being subverted. Attacks from Type 1 are aimed at the sensor and are the focus of this paper. The remaining types are attacks on the electronic systems and enrolment procedures used to setup and perform verification.

In terms of sensor level attacks, three existing methods have been identified[8]:

- *Zero effort attacks*, in which a person claims a random identity and attempts to be incorrectly accepted by the system. Zero effort attacks are the attack type being measured in existing large scale performance evaluations that calculate false accept rates.
- *Brute force attacks*, which repeatedly attempt to access a system, adjusting a biometric feature until a sufficiently close match is obtained[13]. Such attacks generally require unrestricted access to the biometric system as is possible, for example, when picking a biometric lock on a stolen laptop. Secure access control scenarios, such as passport control at an airport, make such attacks less feasible as access failures can alert security.
- *Artifact attacks*, which use a synthetic biometric feature that has been produced from a genuine user. Such attacks would also cover the attempted use of a surgically removed biometric features and methods which exploit residual features on a sensor[16].

An additional consideration is that not all the users of a system will necesarily have the same level of security. This was highlighted by Doddington et al.[7], who measured the relative recognisability of different users of a speaker recognition system. Doddington et al. classified users into four different types: *sheep* who have normal performance, *goats* who are difficult to recognise, *lambs* who are easy to impersonate and *wolves* who can easily impersonate others. Attackers can exploit this variation to compromise a biometric system. For example, a lamb insertion attack[8] would

involve deliberately enrolling a person or synthetic feature that is known to have a similar signature to many subjects. The system containing the lamb subject would then be vulnerable to imposters claiming the lamb identity.

The effect of targeted attacks has been studied for both face[3] and gait verification systems[9]. In the case of gait verification the attack was used with another impersonation method to create the first successful spoofing attempt against that biometric. The second evaluation examined the effect of targeted attacks on forensic face verfication, this evaluation showed that the increase in false acceptance rate due to the increasing number of available targets could be modelled with a logathimic curve. This resulted in a new targeted false acceptance rate measure of biometric verification performance which combined the baseline false acceptance rate with a logarithmic term for the effect of targeting.

3. Baseline Targeting Performance

This section evaluates the effects of targeted attacks on a baseline iris biometric system. The evaluation is performed using a publicly available algorithm and database. The evaluations assume the attacker has complete access to the gallery of subjects and the verification algorithm used by the system. Half of the recordings of each subject are randomly selected and used as the gallery to which the attacker has access. Each subject in the gallery takes the role of an attacker. The gallery data is analysed to select a target that the attacker will impersonate. The non-gallery recordings of the target are then compared against the attacker to determine imposter scores. Score values are also calculated for all the true matching pairs of users of the system. These score values are used to produce DET curves showing the tradeoff of false accept and false reject rates for different verification thresholds. A traditional zero-effort DET curve is also produced to show the relative effect of targeted attacks. The curve is calculated by comparing each of the excluded recordings against each of the gallery recordings to produce a range of scores for both legitimate and zero-effort attacks. In all of the targeted attacks, a target was chosen based on the best score value of all of the possible combinations of attacker and target recordings within the gallery. For each experiment it is expected that real deployments may have more challenging input data and in turn may have more sophisticated verification systems; however, the experiments show that the relative effect of targeting is sufficient to warrant further investigation.

The iris baseline measurements are presented in figure 1. They have been obtained using the system created by Libor Masek[14], which is an open source implementation of an iris recognition system created by John Daugman[5]. The analysis was performed using the Casia-IrisV3-Interval subset of the CasiaV3 dataset. The dataset consists of a collection of controlled, high quality recordings of 249 subjects. The comparisons were performed using the Hamming distance measure with a maximum of 10 orientational shifts and 3 radial shifts, which are the same constants used by the open source VASIR iris recognition system[11].

The traditional DET iris evaluation has an equal error rate (EER) of 11%. However, at the same threshold the targeted attack scores have a much greater false accept rate of 75%. A close inspection of the target selection revealed that one subject was being targeted by all candidates. This subject had a much higher than average estimation of occlusion, reducing the discriminative power of the iris signature and thus acting in a manner similar to a lamb insertion attack. More recent work by John Daugman[6] has identified a score normalisation technique to address the problem of matching subjects with varying amounts of occlusion. This technique treats a signature matching operation as being equivalent to a set of Bernoulli trials. Under this assumption a normalisation scale of $HD_{norm} = 0.5 - (0.5 - HD_{raw})\sqrt{\frac{n}{911}}$ ensures that all subjects should have the same False Rejection Rate for a given verification threshold regardless of occlusion percentage. As can be seen from figure 1 when this technique is applied it has very little effect on the EER of the traditional evaluation. However, when the new system is subjected to targeted attacks there is a significant improvement from 75% to 31% FAR. This shows how important it is to evaluate systems using targeted attacks as traditional DET curve evaluations obscure vulnerabilities and underestimate the importance of countermeasures, such as score normalisation.

3.1. Number of targets

In the baseline experiments the number of targets available to the attacker is necessarily restricted by the size of the dataset. The size of the dataset is consistent with the number of subjects that might access a secure office environment but are much lower than many important identity scenarios such as passport control. To analyse the effect of increasing target numbers, experiments were performed using gallery subsets of increasing size. These subsets were used in the selection of targets for evaluation. To minimise any potential bias caused by subset selection, for a given size, all non-overlapping subsets within the first 249 subjects were combined to produce average false accept rates across the different subsets. This ensures that a subset size of 1 is virtually identical to the baseline performance. All gallery members took the role of attackers using the subset to generate the imposter scores. Figure 2 indicates how the false accept rate increases as the size of the target subset increases. The graph shows the false acceptance rate for a threshold that achieves the equal error rate of the baseline system under zero effort attacks. This is a plausible thresh-

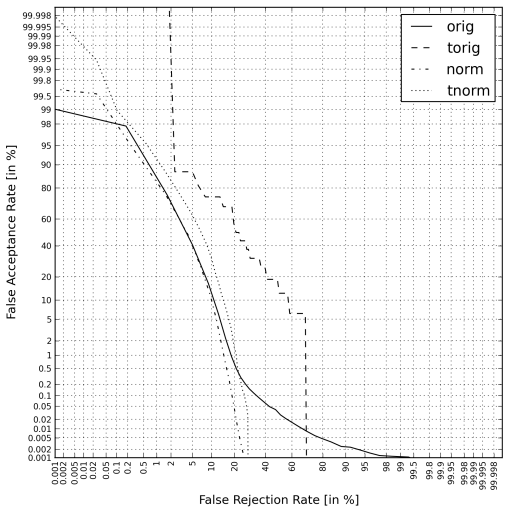

Figure 1. The effects of a targeted attack on Libor Masek's iris verification algorithm tested with the Casia-IrisV3-Interval dataset. The curve labelled *orig* shows the performance of the system when a zero effort attack is performed. The curve labelled *torig* shows how the false acceptance rates increase when targeted biometric attacks are performed. The curves labelled *norm* and *tnorm* show the performance of the baseline and targeted evaluations when the score normalisation has been applied.

old for systems that are unaware of the risks of targeted attacks. As the number of available targets are increased, the number of possible subsets decreases, raising the error in the measured false acceptance rate. Much of the curve, however, conforms reasonably well to a least squares fit of an $a.log(x) + b$ model, with $a = 3.8$ and $b = 10.5$. This logarithmic model is consistent with the pattern found in a previous investigation of face verification performance. It suggests that a logarithmic model captures the natural likelihood of similarity in human features. The fitted model can be used to provide estimates of the number of targets needed to achieve different false accept rates. For example, using this model, approximately 5.7 million targets are required for a 70% FAR, 78 million for 80% and 1 billion for 90%. Larger evaluations are needed to confirm these predictions.

Figure 3 compares the quality of the model fitting for both the baseline system and the system with score normalisation. Because of the presence of the lamb occluded subject in the baseline system, large step increases in FAR occur when the subject is included in the target list undermining the quality of the logarithmic fit. To evaluate the quality of the logarithmic model, the autocorrelation function of the residuals were calculated. These results can be seen in Figure 4. If the model is correctly capturing the behaviour of the function, the residuals of the fit should be uncorrelated, resulting in an autocorrelation function with a single large spike at 0. With only 250 samples it is still plausible that correlations will appear in a random noise function, however it does appear that the logarithmic model fits the face experiments slightly better than the iris ones. The logarithmic model applied to the 250 samples has an R^2 value of 0.983, a high value relative to linear or quadratic fittings which produce values of 0.757 and 0.892 respectively. However, it should be noted that such metrics have their limitations[15]. One explanation for the logarithmic effect comes from the central limit theorem. If biometric signatures are formed as a combination of many small random factors their underlying overall distribution amongst a population is likely to form a multivariate Gaussian. Under such circumstances a small percentage of attackers will have unlikely signatures and thus will require exponentially more targets to find a suitable subject to impersonate.

The precision of the target selection is also a factor in how effective an attack will be. Figure 5 shows how the false acceptance rate is reduced as noise is introduced into the match scores used for target selection. This is intended to simulate the effects of using a different verification algorithm or impaired gallery data for target selection purposes. The graph shows that the unnormalised baseline attacks become rapidly less effective as the target selection noise increases. This sensitivity is likely due to the single lamb subject contributing so heavily to the large baseline false acceptance rate. If this subject is not identified the FAR increase is lost. The normalisation function includes a constant intended to ensure that the average degree of iris occlusion results in no change to the threshold scores, thus enabling their comparison with the same level of noise. The normalised results have a much lower initial FAR, however, it is reduced much more slowly suggesting that each attacker has more potentially viable targets to select from and thus that there is a tradeoff in the benefits of using normalisation.

An additional consideration is how easily attackers can obtain information about the gallery subjects and the system being attacked. For small scale deployments, surveilance may be sufficient to establish possible targets. However, some biometrics will be much more vulnerable to targeted attacks than others. For example, face, voice and gait are relatively easy to record at a distance while fingerprint, iris and finger vein may require more elaborate social engineering to obtain. For identity applications with a large number of users, such as passports or driving licenses, public information may be sufficient. For example, a number of online dating websites have photographs of millions of users which can be anonymously searched using soft biometric constraints including, age, sex, race, hair colour and height [1]. Such information is primarily applicable for face verification attacks. However, as the resolution of digital photographs increases it may eventually become possible to extract iris patterns from public images. The red channel of

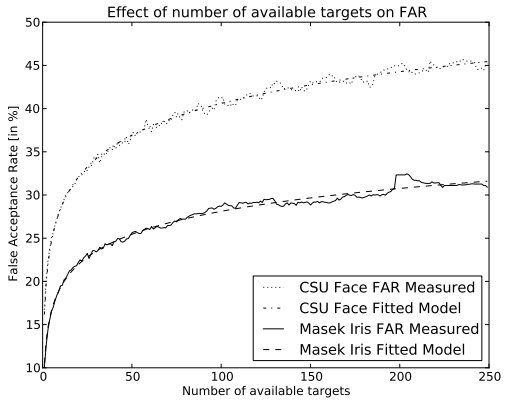

Figure 2. The effect of target numbers on the false accept rate of an iris verification system with a threshold set at the EER of the baseline system. The graph shows the results for the previously examined CSU face verification system[3] along with the Masek iris algorithm with score normalisation applied. Both systems appear to conform to a logarithmic model of system deterioration as the number of targets increase.

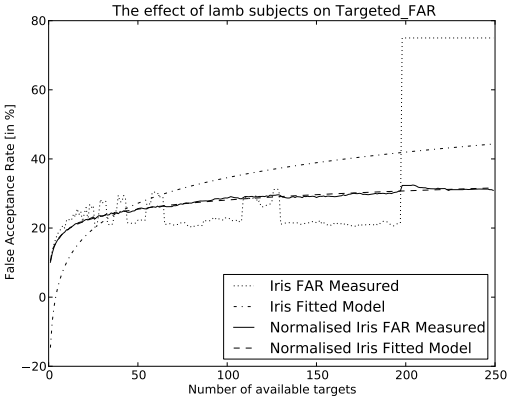

Figure 3. A comparison of the fitting quality of the normalised iris algorithm and the baseline. The presence of the lamb subject in the baseline system produces large step increases in the FAR undermining the logarithmic plot.

such images could then be used to estimate the similarity of the iris under the near infrared lighting used in commercial iris recognition systems[2]. Centralised databases of biometric information are of greater concern. For example, if the US Visit database were to be hacked into, or worse, publicly released by an activist group, then all of the recorded biometric information could be used to identify possible targets for face or fingerprint attacks[19].

4. Conclusions

This paper analyses the effect of targeted attacks, a new vulnerability that can reduce the effectiveness of automated and manual identity systems. The paper has evaluated a

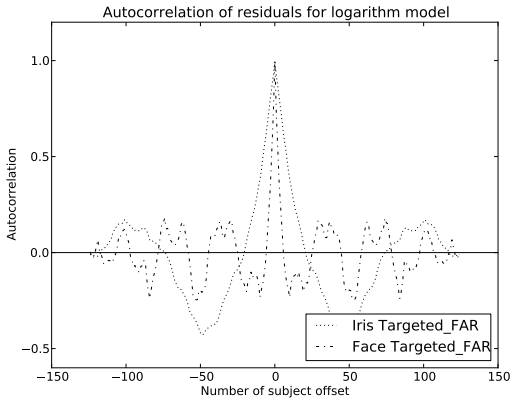

Figure 4. The autocorrelation of the residuals for the logarithmic model of the Targeted_FAR for face and for iris.

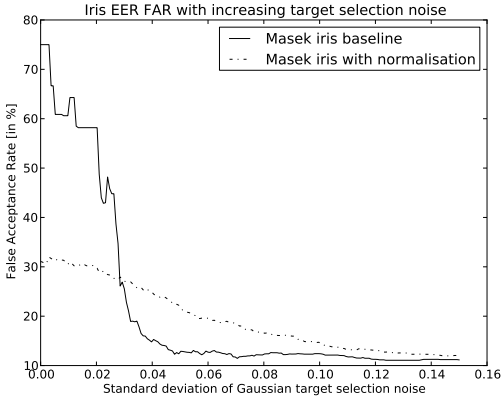

Figure 5. The reduction in FAR as the precision of the target selection is reduced. Imprecision is due to Gaussian noise being applied to the normalised Hamming distance scores during target selection.

baseline iris verification algorithm and revealed that targeted attacks can increase false acceptance rates at the EER of the baseline from 11% to 75% as a result of a single *lamb* user. By applying an existing score normalisation technique this vulnerability can be reduced to an FAR of 31%. The traditional DET curve evaluation obscures both the vulnerability and the importance of the score normalisation countermeasure. Further analysis suggests that the false acceptance rate can be estimated using a simple model that is proportional to the logarithm of the number of subjects. This model provides a means to estimate the vulnerability of systems with many users.

The analysis outlined in this paper indicates that if large biometric databases continue to be gathered and such information is not properly secured, all biometric systems might be at risk of targeted attacks. Although significant detail has been established in the best practice of performing accurate biometric evaluations, active spoofing attacks have yet to be

formalised[12][4]. As part of the formalisation process, this paper has highlighted the need for biometric evaluations to include a targeted false accept measurement as well as the traditional zero effort attack values in order to get a clearer indication of the true vulnerability of such systems in real deployments.

This paper has presented a method of evaluating targeted attacks by calculating the false acceptance rate for the various thresholds associated with the baseline DET curves of a biometric verification system. However, it should be noted that the effectiveness of targeted attacks are determined by both the verification threshold and the number of targets available to the attacker. The DET curve graphs displayed in this paper use the full gallery database as a source of target values. This paper shows how a more general, and more concise, description of system vulnerability can be obtained by fitting an $a.log(x) + b$ model to the increase in false acceptance rate associated with increasing target numbers. Within this model the value b corresponds to a traditional zero-effort false accept rate while the value a indicates how system performance deteriorates as the number of potential targets increase.

There are a number of areas for further investigation. For example, it would be valuable to determine the effect of targeted attacks on state of the art commercial algorithms, such as those used in the NIST IREX III Evaluation. Future work could also identify to what extent biometric systems are vulnerable to targeted attacks when the target is selected solely using human judgement, as this is the least sophisticated and most straightforward form of attack. It would also be informative to evaluate the effect of targeted attacks on other biometrics and other biometric algorithms to determine whether some features or systems are inherently more robust to such attacks. Finally, to facilitate further study into these forms of vulnerability it would be valuable for open implementations of state of the art biometric systems to be made available so that more detailed and realistic performance comparisons can be made.

5. Acknowledgements

References

[1] Plenty of fish. http://www.pof.com.

[2] C. Boyce, A. Ross, M. Monaco, L. Hornak, and X. Li. Multispectral iris analysis: A preliminary study. In *IEEE Computer Society Workshop on Biometrics at CVPR*.

[3] J. D. Bustard, J. N. Carter, and M. S. Nixon. Targeted biometric impersonation. *Int. Workshop on Biometrics and Forensics*, Under review.

[4] R. Cappelli, D. Maio, D. Maltoni, J. Wayman, and A. Jain. Performance evaluation of fingerprint verification systems. *IEEE Tran. PAMI*, 28(1):3–18, jan 2006.

[5] J. Daugman. How iris recognition works. *IEEE Trans. Circuits and Systems for Video Technology*, 14:21–30, 2002.

[6] J. Daugman. Score normalization rules in iris recognition. In *Encyclopedia of Biometrics*, pages 1135–1145. Springer US, 2009.

[7] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves a statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. In *Int. Conf. Spoken Language Processing*, 1998.

[8] T. Dunstone and G. Poulton. Vulnerability assessment. *Biometric Technology Today*, 2011(5):5 – 7, 2011.

[9] A. Hadid, M. Ghahramani, V. Kellokumpu, M. Pietikinen, J. Bustard, and M. Nixon. Can gait biometrics be spoofed? In *Proc. 21st International Conference on Pattern Recognition*, 2012.

[10] R. Jenkins and A. M. Burton. Stable face representations. *Trans. Royal Society B Biological Sciences*, 366(1571):1671–1683, 2011.

[11] Y. Lee, R. Micheals, and P. Phillips. Improvements in video-based automated system for iris recognition (vasir). In *Workshop on Motion and Video Computing*, pages 1–8, dec 2009.

[12] A. J. Mansfield and J. L. Wayman. Best practices in testing and reporting performance of biometric devices. Technical report, Center for Mathematics and Scientic Computing, National Physical Laboratory, 2002.

[13] M. Martinez-Diaz, J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Siguenza. Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification. In *Proc. 40th IEEE Int. Carnahan Conferences Security Technology*, pages 151–159, oct 2006.

[14] L. Masek. Recognition of human iris patterns for biometric identification. Master's thesis, University of Western Australia, 2003.

[15] NIST/SEMATECH. e-handbook of statistical methods, 2013.

[16] K. Y. T. Matsumoto, H. Matsumoto and R. L. v. R. S. Hoshino. Impact of articial "gummy" fingers on fingerprint systems. In *Proc. SPIE Optical Security and Counterfeit Deterrence Techniques IV*, volume 4677, pages 275–289, 2002.

[17] U. Uludag and A. K. Jain. Attacks on biometric systems: a case study in fingerprints. In *Security, Steganography, and Watermarking of Multimedia Contents'04*, pages 622–633, 2004.

[18] T. van der Putte and J. Keuning. Biometrical fingerprint recognition: don't get your fingers burned. In *Proc. 4th Conf. Smart card research and advanced applications*, pages 289–303, 2001.

[19] G. C. Wilshusen and K. A. Rhodes. Gao-07-870:homeland security needs to immediately address significant weaknesses in systems supporting the us-visit program. Technical report, United States Government Accountability Office, 2002.

[20] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35:399–458, December 2003.