Multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters

P. A. Johnson [1], B. Tan [2], S. Schuckers [3]

ECE Department, Clarkson University
Potsdam, NY 13699, USA

[1] johnsopa@clarkson.edu
[2] tanb@clarkson.edu
[3] sschuckers@clarkson.edu

Abstract—**In biometric systems, the threat of "spoofing", where an imposter will fake a biometric trait, has lead to the increased use of multimodal biometric systems. It is assumed that an imposter must spoof all modalities in the system to be accepted. This paper looks at the cases where some but not all modalities are spoofed. The contribution of this paper is to outline a method for assessment of multimodal systems and underlying fusion algorithms. The framework for this method is described and experiments are conducted on a multimodal database of face, iris, and fingerprint match scores.**

I. Introduction

Biometric systems have been found to be useful tools for person identification and verification. A biometric characteristic is any physiological of behavioral trait of a person that can be used to distinguish that person from other people. A few key aspects of a human physiological or behavioral trait that make for a strong biometric for recognition are universality, distinctiveness, permanence, and collectability. These ensure that the trait is available from all people, is adequately variable among all people, does not change significantly over time, and is reasonably able to be measured. The problem with any human trait that meets these criteria is in the performance, acceptability, and circumvention of the biometric trait [1]. Performance is an issue resulting mainly from the combination of lack of variability in the biometric trait, noise in the sensor data due to environmental factors, and robustness of the matching algorithm. Acceptability indicates how willing the client pool will be to use the biometric identifier regularly. Circumvention is the possibility of a non-client (impostor) getting past the system using deceptive methods. Typically these methods involve the forgery of the biometric trait, an act commonly termed "spoofing". The security threat of spoofing will be the main focus of this paper.

To better deal with the performance of a biometric system as well as hopefully diminishing the possibility of circumvention, the idea of using multi-biometric systems has gained popularity [2]. A multi-biometric system is one that combines information from multiple sources in an attempt to reduce the effect of poor performance in any one source. Multi-biometric systems have commonly taken three forms; single biometric trait multiple representation, single biometric trait multiple matcher, and multiple biometric trait [3]. These three methods seek to reduce errors due to noisy sensor data, poor matcher performance, and poor performance in a biometric trait in general. Implementing multiple modalities (i.e., biometric traits) in a system, for instance, face, iris, and fingerprint, requires an imposter to spoof more than one biometric trait, making it much more difficult to fool the system. This gives multimodal systems a leading edge over the other two classes of multi-biometric systems in terms of security.

The key to creating a secure multimodal biometric system is in how the information from the different modalities is fused to make a final decision. There are two different categories of fusion schemes for multiple classifiers; rule based and supervised based. The rule based strategies consist of but are not limited to sum rule, product rule, max rule, min rule, median rule, and majority voting, of which sum rule was found to be the best performer as described by Kittler [4]. Supervised methods, on the other hand, require training but can often provide better results than the rule based methods. Fierrez et al. [5], for example, have shown that a fusion strategy using a support vector machine (SVM) was able to out-perform a fusion algorithm using the sum rule. Introducing a quality measure into a fusion algorithm is one method that has been used to boost performance in multi-biometric systems, such as the multiple matcher approach for fingerprint recognition of Fierrez et al. [6], as well as the bimodal approaches presented in [7] and [8].

The main focus of this paper is the security risk in multimodal biometric systems due to spoof attacks. It is commonly believed that in order for an imposter to fool a multimodal system, every modality would need to be spoofed. The threat being addressed here is the case where a subset of biometric modalities of a multi-modal system is spoofed (e.g. one of three, or two of three modalities).

Rogrigues et al. [9] proposed the idea of using quality measures in a fuzzy logic based fusion algorithm to protect against spoof attacks. The methods proposed in [9] implement auxiliary information such as a sample quality measure and a measure of security (i.e., how easy it is to spoof each biometric trait), to weight the contribution of each biometric modality to the system. If for instance, a more secure biometric of high quality gives a low match score and

WIFS'2010, December, 12-15, 2010, Seattle, WA, USA.
978-1-4244-9080-6/10/*$26.00 ©2010 IEEE.*

a less secure biometric gives a high match score, then there is a high likelihood of a spoof attack. We further extend the work of Rogrigues in exploring the multimodal vulnerability and strategies for fusion in the case where partial spoofing has occurred.

It is commonly understood that one of the strengths of a multimodal system is in its ability to accommodate for noisy sensor data in an individual modality. In contrast, a more secure algorithm, in order to address the issue of a spoof attack on a partial subset of the biometric modalities, must require adequate performance in all modalities. This type of algorithm would invariably negate, to some extent, the contribution of a multimodal system to performance in the presence of noisy sensor data. Therefore this creates a tradeoff between performance and security in multimodal systems as is presented in Table I. This simplified table shows that a unimodal system suffers from low performance in terms of noisy data and low security in terms of spoof attacks. A multimodal system improves the performance aspect but increases the security only slightly since it is still vulnerable to partial spoof attacks. Enhanced fusion methods, which utilize approaches to improve security, will again suffer decreased performance when presented with noisy data. A framework in order to study the tradeoffs will be presented, as well as quantitative results to support the conjecture of Table I.

The contribution of this paper is the development of a framework for the assessment of the performance of multimodal biometric systems under non-zero effort (spoof) attacks. It will be shown that there is a significant security risk where only a subset of the modalities used in the system are spoofed and the remainder are zero effort attempts, where the imposter merely presents their own biometric traits to the system. This framework will be a basis for analyzing systems as well as the underlying fusion algorithms. In addition, this assessment will provide information to quantify various performance metrics to allow for a more informed choice on how to qualitatively set the system decision threshold to ensure adequate performance and security under the threat of a spoof attack.

II. Partial Spoof Assessment Framework

The typical method for assessing a multimodal biometric system is first to make genuine comparisons, where all modalities for enrollment and authentication are from the same subjects. Then, for imposter comparisons, pairs of two different subjects are chosen, one to represent all modalities for enrollment and the other to represent all modalities for authentication. In multimodal fusion, the relative tradeoff of performance for genuine versus imposters is given using a DET curve, as described below. A DET curve can be created for each modality, as well as the fusion of the multiple modalities. The process of fusion includes score normalization. Details of the fusion rules and score normalization are given later in this section. In this paper, we assess security of the fusion rule given a multimodal system using three modalities; we assume that one or two of the modalities are spoofed. To simulate a successful spoofing attempt, we assume that the genuine score is a successful spoof attempt for the spoofed modality and the imposter scores are used for the other modalities (i.e., assume the user supplied his/her own biometric). This structure is outlined in Table II. Additional performance metrics for this alternative structure are given below.

TABLE I
Trade-off Between Performance and Security in Multimodal Fusion

Number of Modalities	Performance	Security
Single Modality	Low	Low
Multiple Modalities	High	Medium
Multiple Modalities with Enhanced Fusion	Medium	High

A. *Score Normalization*

In a multimodal biometric system it is important to normalize the scores from each source so as to avoid an unwanted bias towards any of the modalities. There are many methods for normalizing scores, as described by Jain et al. [10]. Here, z-score normalization, which scales the scores based on the mean and standard deviation of the score distribution, is shown to give the best results. The min-max method, which scales the scores to the range of zero to one using the lowest and highest score, is described as one of the simplest method and is shown to provide near optimal results. In the following experiments, the scores will be normalized using the min-max technique, as having the scores bound to a range of zero to one will allow for a more convenient assessment.

B. *Fusion*

Once the scores have been normalized so that they are within the same range, they are combined using fusion to make a final score. There are a multitude of fusion strategies to choose from. A simple rule based strategy is sufficient to present the proposed framework. Given that the sum rule has been found to be the top performer of the rule based strategies, it will be implemented in the following experiments.

C. *Assessment*

After the scores from each modality have been fused, giving a combined score in the range of zero to one, a threshold is implemented to make a final accept or reject decision. By varying this threshold, a performance curve known as a Detection Error Tradeoff (DET), [11], can show the relationship between the false reject rate (FRR), the percentage of genuine users that are rejected and false accept rate (FAR), the percentage of imposters that are accepted.

In our proposed security framework, to distinguish from FAR, a new metric is introduced called the spoof false accept rate (SFAR), which is the percentage of false accepts given that one or more of the modalities have been successfully

TABLE III
OUTLINE FOR PARTIAL SPOOF ASSESSMENT FRAMEWORK. G INDICATES A GENUINE MATCHING AND I INDICATES AN IMPOSTER MATCHING. GENUINE MATCHINGS ARE USED TO SIMULATE SUCCESSFUL SPOOF ATTEMPTS.

Categories	Face	Iris	Fingerprint	Performance Metric
Clients	G	G	G	FRR
Imposters	I	I	I	FAR
1 Modality Spoofed	G I I	I G I	I I G	SFAR
2 Modalities Spoofed	G G I	G I G	I G G	SFAR

spoofed. As has been outlined in [12], spoof false accept rate (SFAR) should be distinguished from traditional FAR, as it involves a non-zero effort to spoof the system. A multimodal system will be analyzed by the comparison of the traditional FRR and FAR with the new SFAR. Table II shows the outline for this method of assessment and related performance metrics.

III. EXPERIMENTS

For the following experiments, a multimodal database consisting of face, iris, and fingerprint match scores from genuine and imposter pairs is used. This database was created by West Virginia University and is available on the CITeR website [13]. A spoof attempt is simulated using a genuine match score in place of an imposter match score. Given the availability of three modalities, a three modality system, where one or two modalities are spoofed as well as a two modality system where one modality is spoofed will be illustrated.

Fusing the match scores from face, iris, and fingerprint should allow for a high performing biometric recognition system. Common practice in the designing of such a system is to obtain a large number of match scores from genuine and imposter comparisons and calculate the FRR and FAR values at varying threshold levels. A DET curve can then be made by plotting FRR vs. FAR on a log scale. The point at which FRR equals FAR is known as the equal error rate (EER) and is often chosen as the operating point of the system by setting the corresponding threshold level. Here we choose the EER as the operating point for our analysis, but similar analysis can be performed for other points on the curve.

With a set threshold (here at the equal error between FRR and FAR), the additional curves based on the SFAR can be analyzed based on the same threshold. At a specific operating point, for the rule-based fusion, FRR will be the same for both FAR and SFAR, as the change in score fusion as shown in Table II only impacts SFAR. As shown in Figure 1, there are two groups of DET curves based on SFAR, for one modality spoofed and two modalities spoofed.

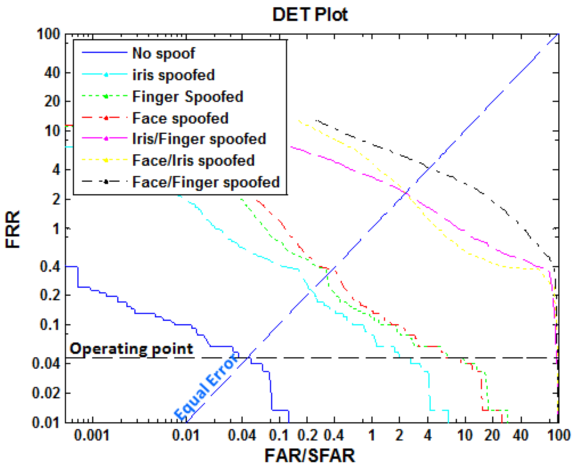

Fig. 1. DET plot for 3 modality multimodal system. The dark blue line is the traditional fusion with tradeoff between FRR and FAR. For a fixed FRR, the operating point line can be followed to the right in order to observe the intersection of the DET curves with the remaining lines which indicate the tradeoff between the same FRR (for a set threshold) and SFAR. An improved tradeoff between SFAR and FRR can be seen by the intersection of the curves with the equal error line.

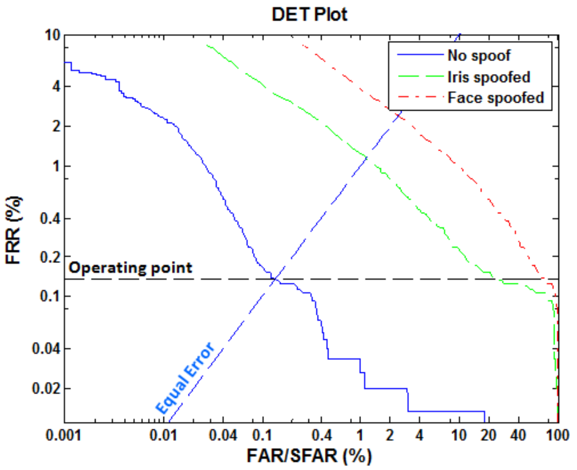

Fig. 2. DET plot for 2 modality multimodal system. For this system, like the 3 modality system, a high SFAR is shown at an operating point selected to be at the EER between FRR and FAR, which can be corrected to achieve a better tradeoff at EER where FRR equals SFAR.

In addition to the three modality system, a two modality system is conceptualized using the face and iris match scores. The DET curve for this type of system is shown in Figure 2.

IV. DISCUSSION

The effect of a spoof attack on the system under study is represented by the SFAR, which can be seen as a shifting to the right of the FAR in Figures 1 and 2. By following the operating point line to the right, the performance during a spoof attack, in terms of SFAR, in shown by the intersection of the DET curve with the operating point line. These results are pulled out and displayed in Figures 3 and 4 for the three

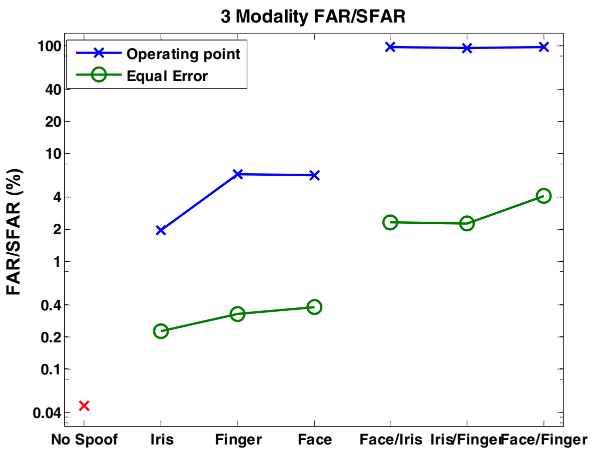

Fig. 3. Performance of 3 modality system comparing FAR and SFAR. The data points from the original operating point (-x-) as well as after adjustment to spoof equal error (-o-) for the 3 modality system are presented. An improved SFAR is shown.

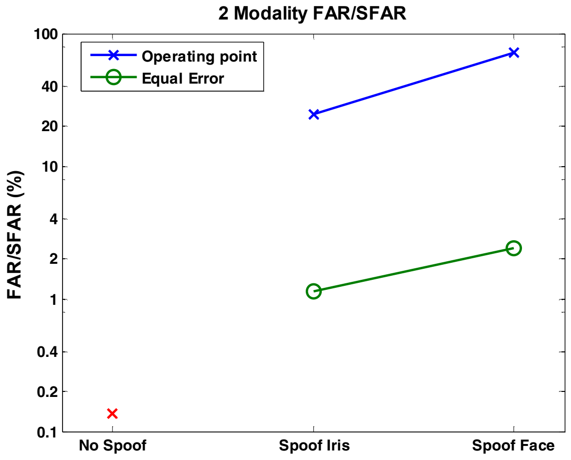

Fig. 4. Performance of 2 modality system comparing FAR and SFAR. The data points from the original operating point (-x-) as well as after adjustment to spoof equal error (-o-) for the 2 modality system are presented. An improved SFAR is again shown after adjusting to the equal error.

modality and two modality cases respectively.

The results of Figure 3 show an EER (FRR/FAR) of 0.05%. For this operating point, when one of three modalities is spoofed, the average SFAR is 4.9%, with an associated FRR of 0.05%. When two of three modalities are spoofed the SFAR jumps up to an average of 97.4%, that is, over 97% of the time, a person will be able to spoof the system by spoofing two of three modalities.

To improve security of the fusion algorithm, we consider a new threshold where FRR equals SFAR; we will call this the Spoof EER (EER_{spoof}) line. At this new threshold of the system, a better tradeoff between FRR and SFAR can be achieved at 0.31% EER_{spoof} for the case where one modality is spoofed and 2.89% EER_{spoof} for the case where two modalities are spoofed. It is evident that the tradeoff in these adjusted error rates may be preferred given the threat of a spoof attack.

TABLE IIIII

Error Rate Trade-off Between Performance and Security in Multimodal Fusion. F and I Indicate Face and Iris Respectively.

Number of Modalities	Performance		Security
	FRR	FAR	SFAR
Single Modality	F: 0.35% I: 2.06%	F: 0.35% I: 2.06%	~100%
Multiple Modalities	0.14%	0.14%	48.37%
Multiple Modalities with Enhanced Fusion	1.76%	0.014%	1.76%

It should be noted that the operating points selected are used as an example. It is recognized that other operating points could be selected depending on the application requirements for FRR, FAR, and SFAR.

These results are comparable to the two modality system, where a SFAR given one modality is spoofed of 48.37%, averaged over the two modalities, with a FRR of 0.14% can be equalized to an EER of 1.76%. The two modality system results are now used to recreate Table I in terms of actual error rates from the above presented results to demonstrate the high, medium, and low performance and security measures. These error rates are tabulated in Table III. This table shows that a bimodal system is able to improve performance under normal operating conditions over a single modality system but is not secure in regards to a spoof attack on one modality. An enhanced fusion algorithm with increased security is much more able to protect against a spoof attacks but looses performance with respect to the FRR.

One limitation to the fusion framework to perform the tradeoff between performance and security is that it rests on the assumption that spoof match scores would be similarly distributed as live match scores. More study needs to be performed which will compare the score distributions of matched live-live and live-spoof image pairs. If there is a difference, the distribution of genuine scores can be adjusted to reflect a typical live-spoof distribution.

In future work, it is proposed that this same procedure be implemented for a number of different fusion strategies as a means of comparing performance between algorithms.

The sum rule for fusion at the score level is among the simplest of such algorithms and is used here merely to demonstrate the structure of the method of assessment. Enhanced fusion strategies such as proposed here and in Rodrigues, et al [9], need to fully consider the tradeoff in FRR, FAR, and newly introduced SFAR, to utilize the power of multimodal fusion for increased performance and security. Our future efforts will be the use of this framework to assess fusion algorithms across an expanded set of databases consisting of a broad range of modalities.

V. Summary/Conclusions

Given the increasing concern about security in biometric systems, the threat of non-zero effort attacks on these systems must be addressed. Traditional FAR accounts only for zero effort imposter attempts on a system. Real spoof attempts may be rare but are becoming more feasible and need to be taken into account. This paper has shown that a multimodal biometric system is not impervious to spoof attacks and is even vulnerable to partial spoof attacks. If a system is designed based only on traditional FRR and FAR, a partial spoof attack has been shown to be possible with a high probability.

This paper proposes a method for determining the best practices for using multimodal fusion to minimize spoof attacks. A new performance measure, SFAR, is introduced to represent conditions of a partial spoof attack. It is shown that after a system assessment based on SFAR is conducted, a calculated adjustment of the operating point can ensure for a more secure system, at a cost of decreased FRR performance. While intuitively expected, we quantitatively demonstrate how to assess the tradeoff.

As a future step in this research, additional multimodal databases will be sought out to be implemented in this study. A number of fusion algorithms will also be collected and compared using the methods outlined in this paper. Through this assessment of multimodal biometric systems, it is likely that insight will be gained as to how a liveness detection method might best be integrated into an overall system.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," 2004.

[2] A. Ross, A. K. Jain, "Multimodal biometrics: An overview," *12th European Signal Processing Conference (EUSIPCO),* pp. 1221-1224, Sep. 2004.

[3] A. Ross, A. Jain, and J. Qian, "Information Fusion in Biometrics," *Audio- and Video-Based Biometric Person Authentication*, pp. 354-359, 2001.

[4] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Analysis & Applications*, vol. 1, pp. 18-27, Mar. 1998.

[5] J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Fusion strategies in multimodal biometric verification," *IEEE International Conference on Multimedia and Expo*, Los Alamitos, CA, USA: IEEE Computer Society, pp. 5-8, 2003.

[6] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain, "Incorporating Image Quality in Multi-algorithm Fingerprint Verification," *Advances in Biometrics*, pp. 213-220, 2005.

[7] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun, "Discriminative multimodal biometric authentication based on quality measures," *Pattern Recognition*, vol. 38, pp. 777-779, May 2005.

[8] K. Nandakumar, Y. Chen, and A. K. Jain, "Quality-based Score Level Fusion in Multibiometric Systems," *International Conference on Pattern Recognition,* pp. 1059-1070, 2006.

[9] R.N. Rodrigues, L.L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," *Journal of Visual Languages & Computing*, vol. 20, pp. 169-179, Jun. 2009.

[10] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition*, vol. 38, pp. 2270-2285, Dec. 2005.

[11] A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," *Fifth European Conference on Speech, Communication, and Technology,* 1997.

[12] A. Adler and S. Schuckers, "Security and Liveness, Overview," *Encyclopedia of Biometrics*, pp. 1146-1152, 2009.

[13] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, *A Protocol for Multibiometric Data Acquisition, Storage and Dissemination*, WVU, Lane Department of Computer Science and Electrical Engineering, 2007.