Journal of Visual Languages and Computing 20 (2009) 169–179

Contents lists available at ScienceDirect

Journal of Visual Languages and Computing

journal homepage: www.elsevier.com/locate/jvlc

# Robustness of multimodal biometric fusion methods against spoof attacks

Ricardo N. Rodrigues [a,*], Lee Luan Ling [b], Venu Govindaraju [a]

[a] University at Buffalo, Center for Unified Biometrics and Sensors, 216C Southlake Vlg., Buffalo, NY 14261, USA
[b] State University of Campinas (Unicamp), Communications Department, Campinas SP, Brazil

ARTICLE INFO

*Keywords:*
Multimodal biometrics
Secure biometrics
Face recognition
Fingerprint

ABSTRACT

In this paper, we address the security of multimodal biometric systems when one of the modes is successfully spoofed. We propose two novel fusion schemes that can increase the security of multimodal biometric systems. The first is an extension of the likelihood ratio based fusion scheme and the other uses fuzzy logic. Besides the matching score and sample quality score, our proposed fusion schemes also take into account the intrinsic security of each biometric system being fused. Experimental results have shown that the proposed methods are more robust against spoof attacks when compared with traditional fusion methods.



## 1. Introduction

Several different types of biometric traits can be used to perform automatic personal authentication (e.g. fingerprint, face, voice, hand geometry, etc.). The biometric characteristics used in an authentication process needs to meet some basic requirements like universality, distinctiveness, permanence, among others [1]. However, in practical applications, no biometric characteristic fully meets these requisites, consequently no single biometric mode is free of errors. Some of these limitations can be overcame or minimized if multiple biometric modalities are used.

In order to take full advantage of the multimodal approach, it is essential to implement a good method for fusing different sources of biometric information. Many fusion methods have being recently proposed [2–7], being that all them show that multimodal biometric systems can significantly increase the recognition rates when compared to unimodal biometric systems. However, none of these have explored the security issues, which is the focus of this paper.

Intuitively, a multimodal system is intrinsically more secure than unimodal systems since it is more difficult to spoof two or more biometric traits than a single one. However, is it really necessary to spoof all the fused biometric traits to crack a multimodal system? This question is specially important when a very secure biometric (e.g retina scan) is combined with another that is easily spoofed (e.g. face). In this scenario, the benefits of adding the face information may be negated by reduction in overall security.

The likelihood ratio (LLR) between the genuine and impostor distribution is known to be the optimal fusion method, in the sense that it minimizes the probability of error [8,9]. This assumes a representative estimation for both distributions is available, which usually is learnt from training data. For example, in [8], a Gaussian mixture model is used to model these distributions. However, these distributions are learnt without considering the hypothesis that an impostor user may have spoofed a biometric mode since only "non-spoofed" impostor samples are used in the training process. The weaknesses of a fusion scheme trained this way will only show up in a scenario when a biometric mode was successfully

* Corresponding author.
*E-mail addresses:* rnr4@cse.buffalo.edu (R.N. Rodrigues), lee@decom.fee.unicamp.br (L.L. Ling), venu@cubs.buffalo.edu (V. Govindaraju).

1045-926X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jvlc.2009.01.010

spoofed, which has never been investigated before. In fact, the hypothesis that an impostor may spoof a biometric mode is not modeled in any proposed fusion method so far.

Some recent works have proposed the introduction of auxiliary information, like biometric sample quality [2–4] and user specific parameters [5], in the fusion method to build more adaptable and reliable biometric systems. The general idea in these methods is to "weight" (directly or indirectly) the contributionof each unimodal biometric based on the auxiliary information. We explore a similar concept by incorporation in the fusion process an auxiliary information that indicates how secure each biometric mode is. Our main objective is to develop biometric fusion strategies that are robust against spoof attacks and that are capable of combining biometric systems with different levels of security without compromising the overall security of the multimodal system.

We propose two novel multimodal biometric fusion schemes that consider the spoofing hypothesis and take into account the security of each biometric system being fused. The first scheme is an extension of the LLR and the second is modeled using fuzzy logic. Both models share the same basic ideas, but differ in details and implementation.

The proposed fusion schemes are described in Section 2. In Section 3 we describe the experiments conducted to analyze the fusion schemes robustness against bad quality samples and spoof attack. The results of these experiments are shown in Section 4. The conclusions are presented in Section 5.

## 2. Proposed fusion schemes

### 2.1. General concepts

In this work we consider only the verification task (i.e. the user has claimed an identity and the system needs to decide if the user is genuine or impostor). Let $M$ be the number of biometric systems to be fused. Fig. 1 illustrates the general multimodal architecture when $M = 2$. The biometric information is fused at the matching score level. This means that each biometric system $i = 1, \ldots, M$ individually performs a matching between the enrolled sample and the test sample, and computes a similarity score $s_i$ between the two samples. We consider that for each biometric system $i$, there exists an expert that, given a test sample, can provide a score $q_i$ that measures the quality of the biometric sample. The set $X = \{s_1, \ldots, s_M, q_1, \ldots, q_M\}$ forms the input for the fusion scheme that process these inputs and produces a single scalar output $z$ such that higher values of $z$ indicate that the user is genuine (or impostor). A threshold operation is applied to the output $z$ for final classification between impostor or genuine.

The security of each biometric system $i$ is modeled by the parameter $c_i$, which represents how difficult it is to spoof the biometric system $i$. It should be noted that it is very hard (if not impossible) to measure the security of a biometric system [10]. In this work, we manually set the parameter $c_i$ based on general knowledge about the security of each biometric. A qualitative assessment of the security for some biometrics can be found in [1].

### 2.2. Extended LLR

Let $I$ be a binary random variable that indicates if a user is impostor ($I = 1$) or genuine ($I = 0$). Our final objective is to evaluate the LLR between the genuine distribution $p(X|I = 0)$ and the impostor distribution $p(X|I = 1)$ as follows:

$$z = \frac{p(X|I = 0)}{p(X|I = 1)} \tag{1}$$

Since we are considering the fusion of different types of biometrics, we can consider that $s_i$ and $q_i$ are independent

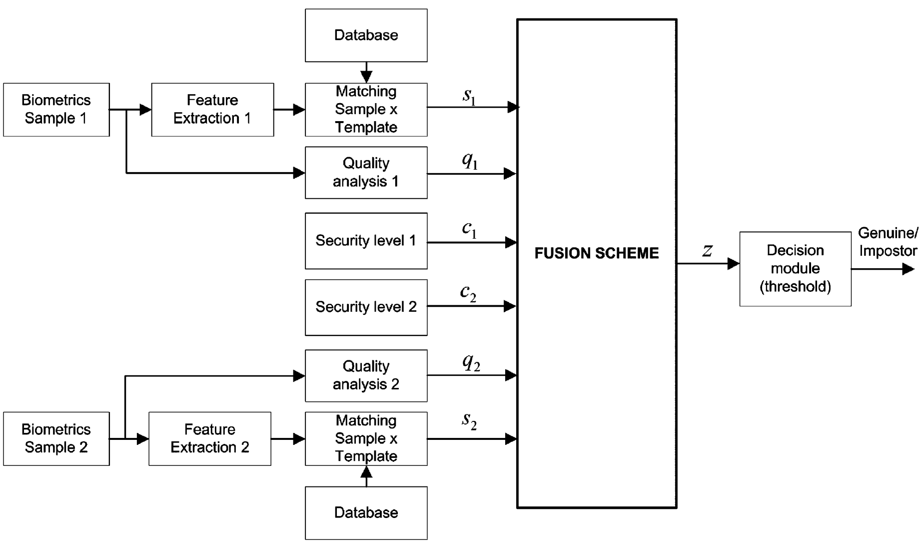


**Fig. 1.** Multimodal system architecture overview.

of $s_j$ and $q_j$ for $i \neq j$, thus we can write

$$p(X|I) = \prod_{i=1}^{M} p(s_i, q_i|I) = \prod_{i=1}^{M} p(q_i)p(s_i|I, q_i) \tag{2}$$

Usually, the conditional distributions in Eq. (1) are learnt using a training dataset where $I$ is know for a set of given $X$'s. However, this approach does not consider the fact that, in practice, an impostor may have spoofed one or more biometric modes, and therefore the estimated impostor distribution may not be representative. To solve this problem, we could try to create a multimodal dataset where some samples (for example, 5% of the total) are spoofed samples and then use standard techniques to learn the impostor distribution. However, this approach would require a significant effort since it is not always easy to spoof a biometric sample. We propose a model that can estimate the true impostor distribution without the need of training spoofed samples.

Our model relies on the assumption that if a biometric system was successfully spoofed by an impostor, the similarity score (in this case) will follow a genuine probability distribution. Based on this assumption, we first introduce hidden variables that model the probability of a spoof attack and the probability of a spoof attack to be successful; then, these variables are marginalized to obtain the true impostor distribution. The marginalization over the hidden variables can be interpreted as if we were considering every possible situation in which the multimodal system can be spoofed.

We start by defining $T_i$ as a binary random variable that indicates if there was an attempt to spoof biometric system $i$:

$$T_i = \begin{cases} 1 & \text{if there was an attempt of spoofing biometric system } i \\ 0 & \text{if there was no attempt of spoofing biometric system } i \end{cases} \tag{3}$$

The conditional joint distribution $p(T_1, \ldots, T_M|I = 1)$ indicates "how often" impostors try to spoof the biometric systems. This distribution depends on the application since some applications may be more frequently targeted by forgers than others. Obviously, there is no sense in an authentic user try to spoof the system. Therefore, we can define $p(T_1 = 0, \ldots, T_M = 0|I = 0) = 1$. For impostor users we define

$$p(T_1, \ldots, T_M|I = 1) = \begin{cases} 1 - \alpha & \text{if } T_i = 0, \; i = 1, \ldots, M \\ \dfrac{\alpha}{2^M - 1} & \text{otherwise} \end{cases} \tag{4}$$

where $\alpha$ is a parameter that indicates the probability that some spoof attack has been attempted. Note that there are $2^M - 1$ possible combinations for spoofing the multimodal biometric systems, being that we assign an equal probability for each of these combinations.

We define another binary random variable $F_i$ that indicates if a given spoof attack was successful or not. The probability distribution for this variable is conditioned only by the existence (or not) of a spoof attempt. Obviously, if there was no spoof attack attempt ($T_i = 0$), the probability of a successful spoof attack ($F_i = 1$) is zero: $p(F_i = 1|T_i = 0) = 0$; this implies that $p(F_i = 0|T_i = 0) = 1$. The probability of a spoof attack being successful is directly related to how secure a biometric system is. Therefore, we define the probability of a spoof attack being successful as $p(F_i = 1|T_i = 1) = 1 - c_i$. This implies that the probability of a spoof attack be unsuccessful is $p(F_i = 0|T_i = 1) = c_i$.

Fig. 2 illustrate the proposed model for the case when $M = 2$. We can marginalize the introduced hidden variables to estimate $p(X|I)$ as follows:

$$\begin{aligned} p(X|I) &= \sum_{T_1} \cdots \sum_{T_M} \sum_{F_1} \cdots \sum_{F_M} p(X, T_1, \ldots, T_M, F_1, \ldots, F_M|I) \\ &= \sum_{T_1} \cdots \sum_{T_M} \sum_{F_1} \cdots \sum_{F_M} p(T_1, \ldots, T_M|I) \\ &\quad \times \prod_{i=1}^{M} [p(F_i|T_i)p(s_i, q_i|F_i, I)] \end{aligned} \tag{5}$$

In order to evaluate Eq. (5) we still need to know the distributions $p(s_i, q_i|F_i, I)$, $i = \{1, \ldots, M\}$. In the case when the biometric system $i$ was not spoofed (i.e. $F_i = 0$), this distribution can be learnt directly from the training data. We note that there is no need to know $p(s_i, q_i|F_i = 1, I = 0)$ since we have assumed that a genuine user will never try to spoof the system. Using the underlying assumption that $s_i$ follows a genuine distribution when a biometric system was successfully spoofed, we consider $p(s_i, q_i|F_i = 1, I = 1) = p(s_i, q_i|F_i = 0, I = 0)$. Now, given a new test input $X$, we can use Eq. (5) to evaluate the LLR in Eq. (1).

We use a gamma distribution to model the genuine and impostor distribution when $F_i = 0$. The choice for the

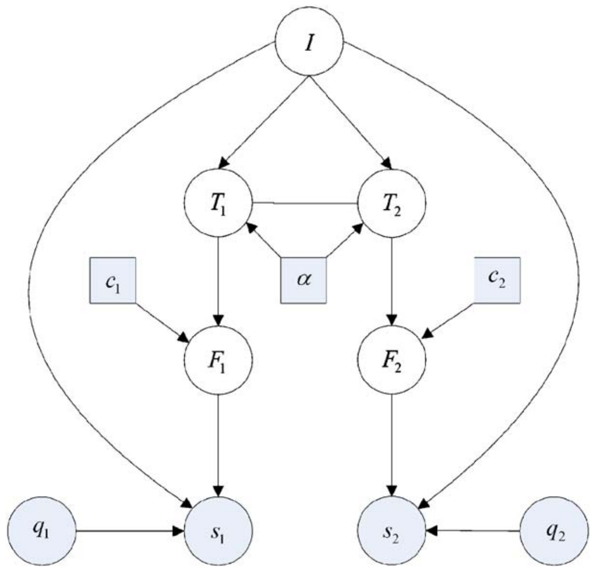


**Fig. 2.** Graphical model showing the relation between the variables and parameters. The shaded circles are observed random variables, while the white circles are latent random variables. The shaded squares represent fixed parameters. Our objective is to infer if an user is impostor ($I = 1$) or authentic ($I = 0$) based on the observed variables.

gamma distribution is due to the empirical evidence that biometric similarity scores tend to have long tails [8]. Note, however, that this distribution depends on the matcher that is being used in the fusion and can be easily be changed for applications that use different matchers. Let $Gamma(x;\theta,\beta)$ represent a gamma distribution where $\theta$ and $\beta$ are the shape and inverse scale parameters, respectively. Then we have

$$p(s_i,q_i|F_i=0,I=1)=Gamma(X_i;\theta_i^{imp},\beta_i^{imp}) \tag{6}$$

$$p(s_i,q_i|F_i=0,I=0)=Gamma(X_i;\theta_i^{gen},\beta_i^{gen}) \tag{7}$$

where the parameters $\theta_i^{imp}$, $\beta_i^{imp}$, $\theta_i^{gen}$, $\beta_i^{gen}$ are learnt via maximum likelihood based on training data.

### 2.3. Fuzzy logic fusion scheme

The probabilistic fusion scheme described in Section 2.2 uses many heuristics in the modeling. Usually, one tries to use training data to learn all probability distributions. However, in our case this is not possible, thus we are forced to heuristically define some probability distributions. In this section we present a fusion method based on fuzzy logic that allows us to explicitly describe the heuristics using linguistic expressions. For sake of simplicity, we describe the fusion schemes for combining two biometric systems ($M=2$) and indicate what should be modified for the case of an arbitrary number of biometric systems.

The development of a fuzzy logic system involves three main steps: (i) defining fuzzy variables and their membership functions (fuzzification process); (ii) creating the fuzzy rules that describe relations between the fuzzy variables; (iii) establishing an appropriated defuzzification method [11]. The fuzzy logic system is based on a first order Takagi–Sugeno–Kang scheme [12].

Fig. 3 shows a block diagram for the fuzzy fusion scheme. In the fuzzification step, each one of the six inputs ($s_1$, $s_2$, $q_1$, $q_2$, $c_1$, $c_2$) is modeled as a fuzzy variable. A membership function maps each fuzzy variable into a real number on the $[0,1]$ range that describes the linguistic expression *high parameter*.

The similarity score ranges may be very distinct for different biometric modes as well as for different matching algorithms that use the same biometric modality. Therefore, choosing an appropriate membership function is crucial for keeping the linguistic expression meaningful. For the *high similarity* linguistic expression we propose a membership function where the "fuzzy" region is the region where the false acceptance rate (FAR) and false rejection rate (FRR) are non zero:

$$s_i^{high}=\min\left[1,\max\left(0,\frac{s_i-s_{i.ZeroFRR}}{s_{i.ZeroFAR}-s_{i.ZeroFRR}}\right)\right] \tag{8}$$

where $s_{i.ZeroFRR}$ and $s_{i.ZeroFAR}$ are the points in the $s_i$ domain where the FRR and FAR are zero, respectively. Fig. 4 illustrates this function in relation to an hypothetical FAR and FRR graphs. The values for $s_{i.ZeroFRR}$ and $s_{i.ZeroFAR}$ are learnt using a training dataset for each biometric system *i*. In fact, any monotonically increasing function bounded between zero and one could be used as a membership function. However, it should be capable of interpreting the *high similarity* linguistic expression in a meaningful way.

For the *high quality* linguistic expression, we choose a min–max function:

$$q_i^{high}=\frac{q_i-\min(Q_i)}{\max(Q_i)-\min(Q_i)} \tag{9}$$

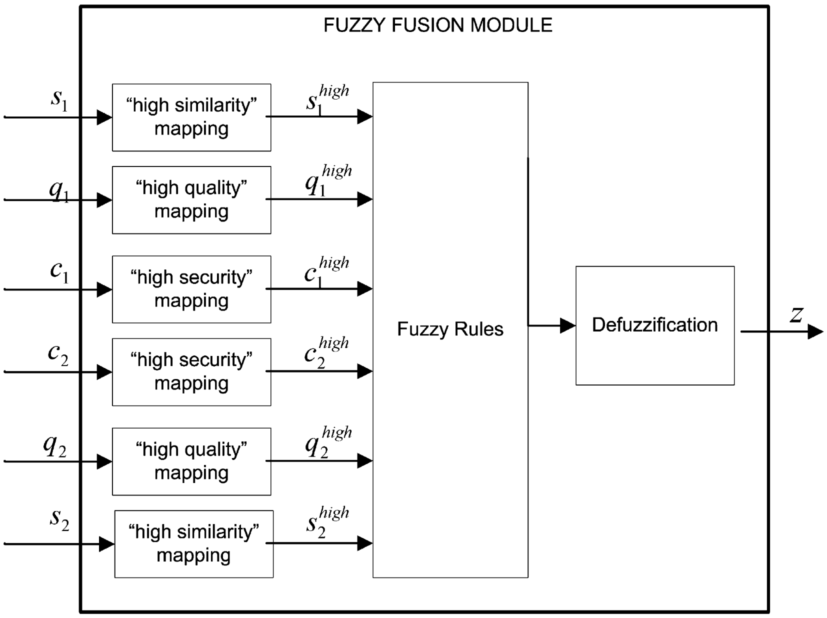


**Fig. 3.** Fuzzy fusion.

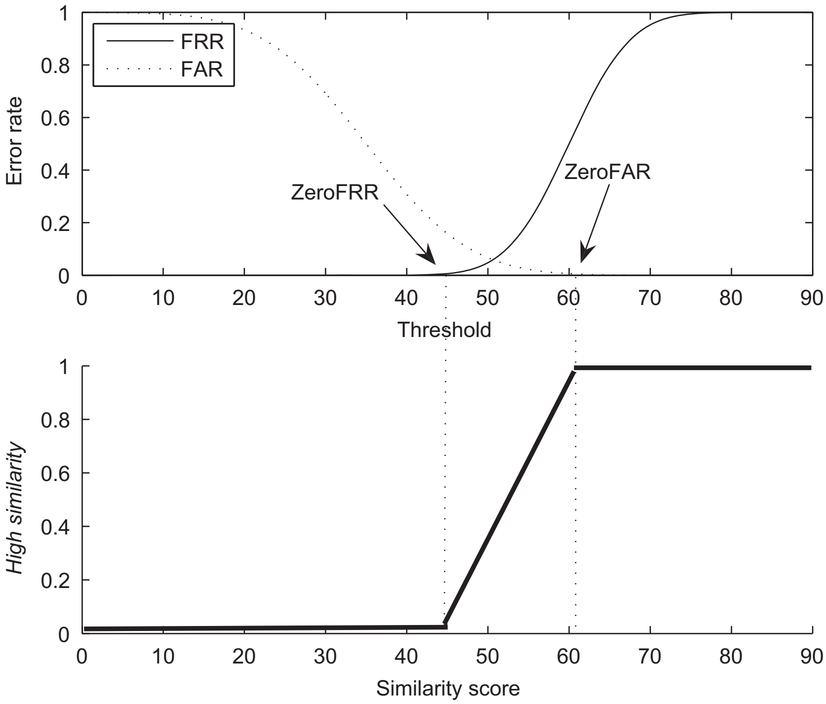


**Fig. 4.** High similarity membership function.

| | | | $s_2$ is high | | | | $s_2$ is low | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $c_2$ is high | | $c_2$ is low | | $c_2$ is high | | $c_2$ is low | |
| | | | $q_2$ is low | $q_2$ is high | $q_2$ is low | $q_2$ is high | $q_2$ is low | $q_2$ is high | $q_2$ is low | $q_2$ is high |
| $s_1$ is high | $c_1$ is high | $q_1$ is low | $z$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is medium | $z$ is low | $z$ is medium | $z$ is low |
| | | $q_1$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is medium | $z$ is high | $z$ is medium |
| | $c_1$ is low | $q_1$ is low | $z$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is low | $z$ is low | $z$ is low | $z$ is low |
| | | $q_1$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is high | $z$ is low | $z$ is low | $z$ is low | $z$ is low |
| $s_1$ is low | $c_1$ is high | $q_1$ is low | $z$ is medium | $z$ is high | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low |
| | | $q_1$ is high | $z$ is low | $z$ is medium | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low |
| | $c_1$ is low | $q_1$ is low | $z$ is medium | $z$ is high | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low |
| | | $q_1$ is high | $z$ is low | $z$ is medium | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low | $z$ is low |

**Fig. 5.** Fuzzy rules.

where $\min(Q_i)$ and $\max(Q_i)$ specify the minimum and maximum values for the $q_i$ domain.

Since the security parameters are already constrained to the $[0,1]$ range we define

$$c_i^{high} = c_i \tag{10}$$

The output of the fusion module, denoted by $z$, is also modeled as a fuzzy variable and can assume one of the three linguistic values: *low*, *medium* and *high* with the corresponding numeric values as $z_{low} = 0$, $z_{medium} = 0.5$ and $z_{high} = 1$; where larger the numerical value, greater is the indication that the user is genuine.

Once all linguistic variables are adequately mapped into membership functions, they are processed by a set of fuzzy rules. These rules are elaborated based on human expertise. Due to space constrains, we will not justify each rule individually. The main ideas followed in the compilation of these fuzzy rules are simple: low security biometric system cannot faithfully perform the recognition task alone; and similarity scores with low quality should have low weights in the final output. The fuzzy rules implemented for the proposed fuzzy bimodal biometric system are listed below. The same rules are illustrated in Fig. 5.

(1) If $s_1$ is high and $s_2$ is high then $z$ is high.
(2) If $s_1$ is not high and $s_2$ is not high then $z$ is low.
(3) If $c_1$ is high and $q_1$ is not high and $q_2$ is not high and $s_1$ is high and $s_2$ is not high then $z$ is medium.
(4) If $c_1$ is high and $q_1$ is not high and $q_2$ is high and $s_1$ is high and $s_2$ is not high then $z$ is low.
(5) If $c_1$ is high and $q_1$ is high and $q_2$ is not high and $s_1$ is high and $s_2$ is not high then $z$ is high.
(6) If $c_1$ is high and $q_1$ is high and $q_2$ is high and $s_1$ is high and $s_2$ is not high then$z$ is medium.
(7) If $c_1$ is not high and $s_1$ is high and $s_2$ is not high then $z$ is low.
(8) If $c_2$ is high and $q_1$ is not high and $q_2$ is not high and $s_1$ is not high and $s_2$ is high then $z$ is medium.

(9) If $c_2$ is high and $q_1$ is high and $q_2$ is not high and $s_1$ is not high and $s_2$ is high then $z$ is low.
(10) If $c_2$ is high and $q_1$ is not high and $q_2$ is high and $s_1$ is not high and $s_2$ is high then $z$ is high.
(11) If $c_2$ is high and $q_1$ is high and $q_2$ is high and $s_1$ is not high and $s_2$ is high then $z$ is medium.
(12) If $c_2$ is not high and $s_1$ is not high and $s_2$ is high then $z$ is low.

The *and* operation in the fuzzy rules is implemented through the multiplication between the two involved membership values. The *not* operation is defined as $1 - x^{high}$, where $x^{high}$ is the corresponding membership function.

The defuzzification process is responsible for combining the results produced by all fuzzy rules and producing one single scalar output suitable for the final classification. In our method, the defuzzification process is carried out by a weighted average as follows:

$$z = \frac{\sum_{j=1}^{12} m_j z_j}{\sum_{j=1}^{12} m_j}, \tag{11}$$

where $m_j$ is the activation value for rule $j$ and $z_j$ its correspondent output value. For example, if we consider the second rule, we have $m_2 = (1 - s_1^{high})(1 - s_2^{high})$ and $z_2 = 0$. After the defuzzification step, the output $z$ will be in the $[0, 1]$ range, being that the bigger its value the bigger is the indication that the user is genuine.

The main challenge in extending this fusion scheme for an arbitrary number of biometric systems ($M > 2$) is in the fuzzy rules definition. This is because the number of rules grow exponentially with $M$. In [13] we present a preliminary analysis for $M = 3$.

## 3. Experiments

The performance of the proposed fusion schemes were evaluated using two biometric systems: a face recognition system implemented using *eigenfaces* [14] and a publicly available fingerprint system developed by NIST [15]. The quality of a fingerprint sample was computed using the NFIQ software [16], also developed by NIST. The quality for a face image was manually assigned based on the face rotation, illumination and facial expression. In a practical application, the face image quality could be automatically calculated using methods described in [17,18]. We use the subscript *face* and *fing* to refer to the face and fingerprint systems, respectively (e.g. $s_{fing}$ refers to the similarity score for fingerprint). We set the security parameters as $c_{face} = 0.3$ and $c_{fing} = 0.7$. For the extended LLR we set $\alpha = 0.01$ (i.e. the probability of a spoof attack is 1%).

A multimodal dataset was created by randomly combining users from the FVC2004-DB1 dataset [19] with users from the FERET-*b series* face dataset [20], creating a *virtual* multimodal user [21]. The biometric samples for each virtual user are randomly and uniquely combined to create multimodal samples. The FVC2004-DB1 dataset contains 1000 fingerprints from 100 different users (10 fingerprints per user), while the FERET-*b series* dataset contains 2200 faces from 200 users (11 face images per user). We run the experiments 10 times, being that a new multimodal dataset is randomly created at each time. The results presented here are based on the average from these 10 runs. We randomly choose 40 users from the multimodal dataset to train the fusion models, and use the other 60 users to run the tests. Three different experiments were performed using the multimodal test dataset:

*Experiment* I: The objective of this experiment is to test the fusion methods under normal operation conditions, where no poor quality samples are presented. Only those samples with $q_{face}^{hiqh} \geqslant 0.6$ and $q_{fing}^{hiqh} \geqslant 0.6$ are used in the tests.

*Experiment* II: In this experiment, all samples are used, independently of its quality. Comparing the results from this experiment with the results from Experiment I, we can evaluate which fusion method is more robust in noisy environments.

*Experiment* III: What happens when one of the unimodal biometrics in a multimodal system is successfully spoofed? To answer this question, this experiment simulates a scenario where the face system is spoofed. In this experiment, the multimodal impostor comparisons were performed using an impostor fingerprint sample (as usual) together with a genuine face sample. We assume that there is no difference between a genuine face sample and a successfully spoofed one. As in Experiment I, we only use samples that have $q_{face}^{hiqh} \geqslant 0.6$ and $q_{fing}^{hiqh} \geqslant 0.6$.

Beside the fusion schemes proposed in Section 2, we also run the experiments on the following fusion schemes:

*LLR* [8]: Eq. (1) describe the LLR fusion. Here, we use the traditional approach to estimate the genuine and impostor distributions (i.e. we do not consider the possibility of a spoof attack). Note that this is equivalent to the extended LLR fusion scheme described in Section 2.2 where the probability of a spoof attack is zero (i.e. $\alpha = 0$).

*Weighted sum*: Some articles [6,7] have reported that the sum rule fusion have presented very good results, even when compared with sophisticated methods like neural networks [7] and decision trees [6]. The weighted sum fusion performs a linear combination between the similarity scores as follows:

$$z = \sum_{i=1}^{M} w_i s_i \tag{12}$$

where $w_i$ is the weight of biometric system $i$. In our case, since we are combining only two biometric systems, we can write

$$\frac{z}{w_{face}} = z' = k s_{fing} + s_{face} \tag{13}$$

where $k = w_{fing}/w_{face}$. In our experiments, the parameter $k$ was found through an extensive search for an optimal value using the training data.

## 4. Results and analysis

The experiment results are analyzed with the use of the receiving operating curve (ROC) [1]. This curve is obtained by varying the decision threshold value and plotting the genuine acceptance rate (GAR) *versus* the FAR for the same (implicit) threshold. The GAR is the probability of a genuine user being correctly accepted as genuine and the FAR is the probability that an impostor user being mistakenly accepted as genuine. Note that $GAR = 1 - FRR$, where FRR is the false rejection rate.

Fig. 6 shows the ROC curves obtained in Experiment I. The results show that in this scenario (no bad quality samples), all fusion methods have improved compared to the unimodal systems. The LLR and weighted sum give the best results. This is expected and can be justified. In summary, this happens because traditional methods do not consider the hypothesis of spoof attack. For the LLR, we can use Neyman–Pearson lemma to prove that any fusion method that consider the possibility of spoof attack will not be better (on average) than the LLR in Experiment 1: Let $H_1$ and $H_0$ be two hypotheses that represent the presence or absence of spoof attack, respectively. Then we can write $p(error|s) = p(error|s, H_1).p(H_1) + p(error|s, H_0).p(H_0)$, where $s$ represent the similarity scores and *error* is a linear combination between FAR and FRR that depends on the chosen threshold. In the traditional LLR fusion method only hypothesis $H_0$ is considered during training (i.e. $p(error|s, H_0)$ is minimized), so it is guaranteed that when $H_0$ is true it will have the lowest error rate possible. In contrast, our method minimizes $p(error|s)$. Considering that the fusion scheme that minimizes $p(error|s, H_1)$ is different than the one that minimizes $p(error|s, H_0)$ and that $p(H_1) > 0$, we have (on average): $p_*(error|s, H_0) < p_{LLR}(error|s, H_0)$, where $p_*(\cdot)$ represent the error rate for any method that consider the possibility of spoof attack and $p_{LLR}(\cdot)$ the error for the traditional LLR fusion method.

Fig. 7 shows the ROC curves for the Experiment II. In this scenario (where bad quality samples were introduced), we note that the fusion methods that use sample quality score are not so affected as the weighted sum. These results suggest that the use of sample quality indeed increase the robustness of fusion methods as expected. The more surprising results were observed in Experiment III, shown in Fig. 8. In this scenario (where the face trait was successfully spoofed), all fusion strategies had worse result than fingerprint alone. However, the introduction of security parameter in the probabilistic and fuzzy fusion methods resulted in a more secure multimodal system when compared with the LLR and weighted sum.

An important parameter for the proposed extended LLR is $\alpha$, which represents the prior probability of spoof attack. This parameter is application dependent and may even vary for the same application (e.g. for some applications, the probability of a spoof attack may be higher at night). We can understand the effect of $\alpha$ by considering $p(error|s)$, where $\alpha = p(H_1)$. For lower values of $\alpha$, the term $p(error|s, H_1)$ will have a lower weight during the minimization of $p(error|s)$. Therefore, the average test error rates will be higher when $H_1$ is true (Experiment III) but lower when $H_0$ is true (Experiment I).

In practice, a biometric system has to operate with a unique threshold for all situations. For example, we do not have *a priori* knowledge about the spoof attack on face system, so we cannot set a specific threshold for this situation. However, when plotting the ROC curve for each experiment separately, the threshold remains implicit, causing the loss of reference when comparing the same biometric system in different experiments. Therefore, we use Experiment I ROC curve to fix some reference thresholds where FAR is 1%, 0.1% and 0.01% (referred as $trs_1$, $trs_{0.1}$ and $trs_{0.01}$, respectively). Then, we evaluate the FAR and FRR for the other experiments using these thresholds. Ideally, the FAR should be the same as the reference values for all experiments. Table 1 shows the error rates for the fixed thresholds.

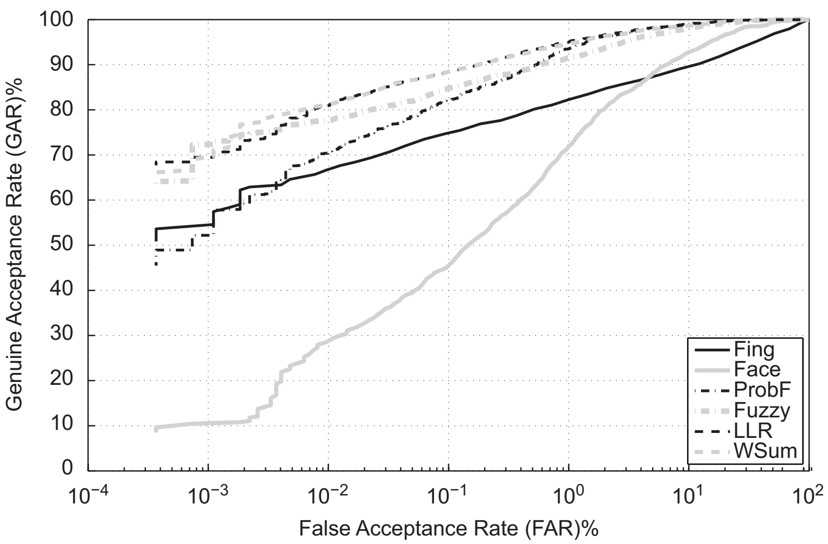


**Fig. 6.** Experiment I ROC curves.

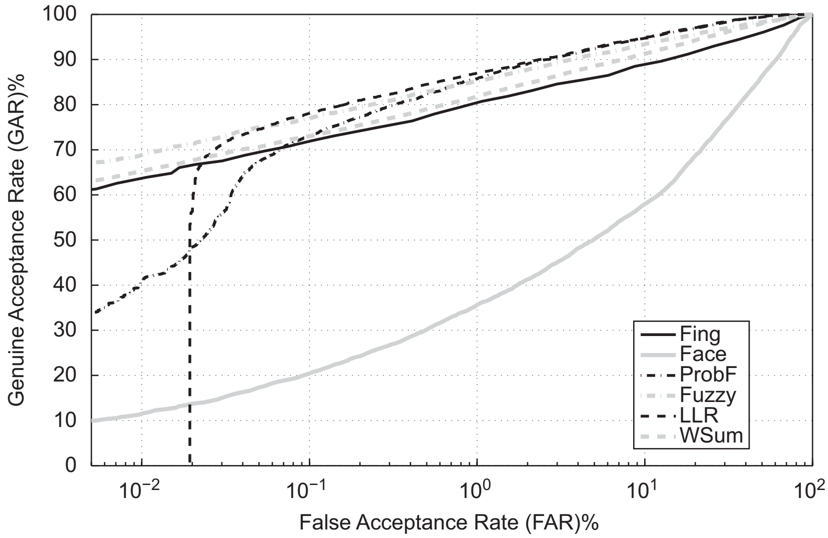


**Fig. 7.** Experiment II ROC curves.

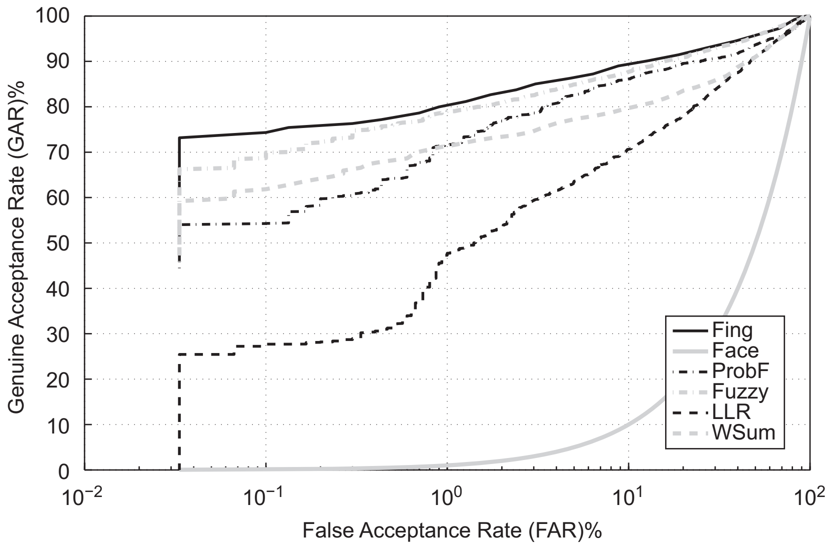


**Fig. 8.** Experiment III ROC curves.

Analyzing this table, the following observations can be made:

- In Experiment I, the weighted sum and LLR fusion methods have the overall best results.
- In Experiment II, the LLR fusion had the overall best result. The methods that use the sample quality score had a smaller increase in the FRR when compared with Experiment I. This suggests that the introduction of the sample quality score in the fusion scheme can indeed increase the multimodal robustness against noisy samples. However, note that some methods had a slight increase in the FAR as well.
- In Experiment III, both, weighted sum and LLR fusion methods presented a dramatic increase in the FAR when compared with the reference FAR (Experiment I). For example, when the LLR fusion operates with threshold $trs_{0.1}$, a forger that has successfully spoofed the face biometric system has a chance of 42.09% of being accepted using his own fingerprint; while for the fuzzy logic fusion, his chances are 4.71% and for the probabilistic fusion, 4.33% (bolded numbers in $trs_{0.1}$ row). This suggests that the traditional fusion methods which do not use the security parameter can be cracked by spoofing only one biometric trait more easily than the proposed methods.

We extend the analysis between Experiments I and III as follows: for each FAR value in Experiment I (referred as $FAR_1$) we fix the threshold and evaluate the respective FAR in Experiment III (referred as $FAR_3$). The results are shown in Fig. 9. It can be seen that the LLR fusion, which had the

**Table 1**
Error rates for some fixed reference thresholds.

| Reference | System | Thresholds | Experiment I | | Experiment II | | Experiment III | |
|---|---|---|---|---|---|---|---|---|
| | | | FAR (%) | FRR (%) | FAR (%) | FRR (%) | FAR (%) | FRR (%) |
| *trs*$_1$ | Fing | 17.34 | 1.01 | 17.70 | 1.08 | 19.43 | 1.27 | 18.83 |
| | Face | 0.55 | 1.00 | 28.01 | 0.54 | 68.73 | 72.38 | 27.62 |
| | ProbF | 2.68 | 1.00 | 6.96 | 1.57 | 11.95 | 52.26 | 6.52 |
| | Fuzzy | 0.21 | 1.00 | 8.52 | 1.25 | 14.02 | 24.40 | 8.02 |
| | LLR | 1.85 | 1.00 | 4.88 | 1.14 | 12.59 | 69.36 | 4.77 |
| | WSum | 0.71 | 1.00 | 5.36 | 0.61 | 20.29 | 68.02 | 5.01 |
| *trs*$_{0.1}$ | Fing | 24.07 | 0.10 | 25.16 | 0.12 | 27.63 | 0.06 | 26.41 |
| | Face | 0.70 | 0.10 | 54.49 | 0.05 | 82.53 | 45.83 | 54.17 |
| | ProbF | 31.42 | 0.10 | 18.04 | 0.22 | 22.31 | **4.33** | 18.83 |
| | Fuzzy | 0.30 | 0.10 | 15.21 | 0.16 | 21.46 | **4.71** | 15.36 |
| | LLR | 6.30 | 0.10 | 11.63 | 0.11 | 21.54 | **42.09** | 11.69 |
| | WSum | 0.88 | 0.10 | 11.60 | 0.06 | 28.85 | **38.51** | 11.85 |
| *trs*$_{0.01}$ | Fing | 30.94 | 0.01 | 33.23 | 0.02 | 34.63 | 0.03 | 34.09 |
| | Face | 0.78 | 0.01 | 71.22 | 0.00 | 90.20 | 28.51 | 71.49 |
| | ProbF | 185.94 | 0.01 | 29.59 | 0.00 | 63.12 | 0.80 | 31.70 |
| | Fuzzy | 0.39 | 0.01 | 22.33 | 0.02 | 28.91 | 0.80 | 22.24 |
| | LLR | 10.57 | 0.01 | 18.48 | 0.03 | 29.12 | 26.30 | 18.27 |
| | WSum | 1.01 | 0.01 | 18.87 | 0.01 | 36.13 | 12.25 | 19.49 |

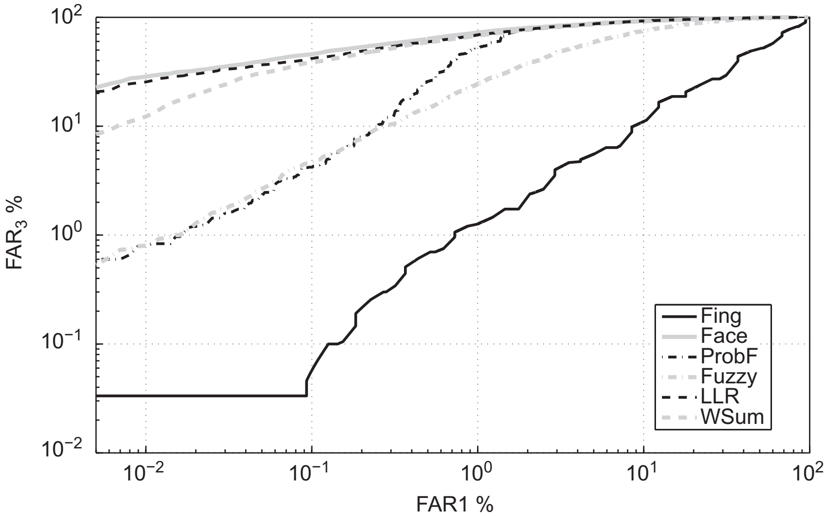


**Fig. 9.** Comparison between the FAR for Experiment I ($FAR_1$) and FAR for Experiment III ($FAR_3$) using the same threshold.

best performance in Experiments I and II, was the most affected by the spoofing of the face biometric. The weighted sum was also significantly affected. The fuzzy logic fusion presented the best results, which for low values of $FAR_1$, were similar to the probabilistic fusion. This result indicates that the introduction of the security parameter can indeed result in a more secure fusion scheme.

### 4.1. Validation

We validate the results shown in the previous section by using a different dataset: the NIST Biometric Scores Set—Release 1 [22], which is a public multimodal score dataset that includes the similarity scores for a face biometric system and a fingerprint biometric system. We use the left index fingerprint set with the "C" face recognition set. Since there are no sample quality scores in this dataset, we fix the quality of all samples as one.

Fig. 10 show the comparison between the FAR of Experiments I and III for this dataset in the scenario where the face biometric was successfully spoofed. Fig. 11 show the same graph but in the scenario where the fingerprint biometric was spoofed. The results support the conclusion that the proposed fusion schemes are more robust against spoof attack when compared with the LLR and weighted sum.

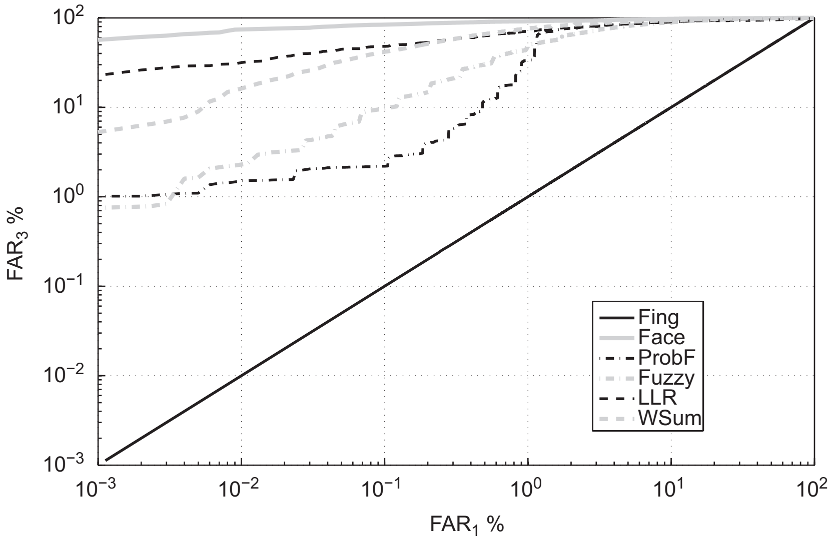


**Fig. 10.** Comparison between the FAR for Experiment I ($FAR_1$) and FAR for Experiment III ($FAR_3$) when the face trait was spoofed.

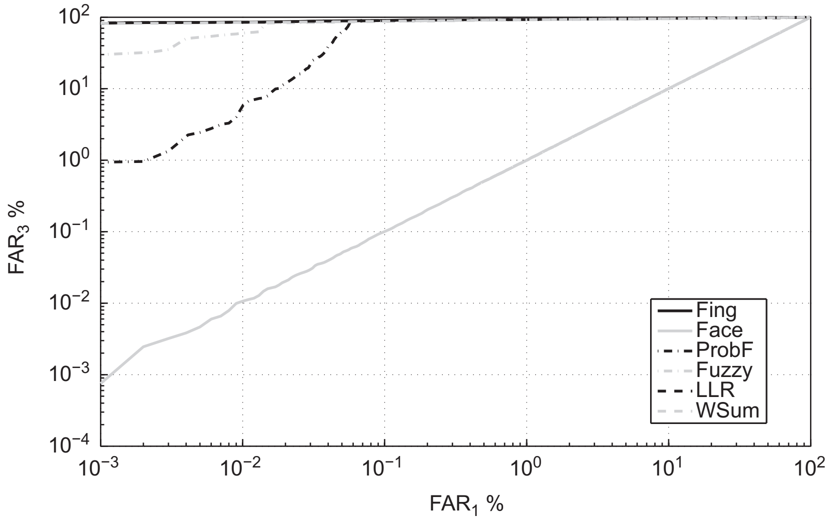


**Fig. 11.** Comparison between the FAR for Experiment I ($FAR_1$) and FAR for Experiment III ($FAR_3$) when the fingerprint was spoofed.

## 5. Conclusion

In this work, we have analyzed the impact of a spoof attack in multimodal biometric systems. Our experiments show that when using traditional fusion schemes (i.e. LLR or weighted sum), a forger can dramatically increase the chances of cracking a multimodal system by spoofing only one of the biometrics. To reduce this weakness, we proposed two new fusion schemes that take into account the security of each unimodal biometric system. The experiments indicate the existence of a tradeoff between recognition accuracy and robustness against spoof attacks.

The experiments also indicate that the fuzzy fusion scheme had a better overall performance when compared with the probabilistic fusion scheme. In the future, we will implement a training process to automatically optimize the membership functions in the fuzzy logic fusion, and test both fusion schemes with a broader range of parameters.

## References

[1] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, IEEE Transactions on Circuits and Systems for Video Technology 14 (1) (2004).
[2] C. Sanderson, K.K. Paliwal, Identity verification using speech and face information, Digital Signal Processing 14 (2003) 449–480.
[3] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, Multimodal biometric authentication using quality signals in

mobile communications, in: Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP03), August 2003.

[4] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, J. Bigun, Discriminative multimodal biometric authentication based on quality measures, Pattern Recognition 38 (5) (2005) 777–779.

[5] K.-A. Toh, X. Jiang, W.-Y. Yau, Exploiting global and local decisions for multimodal biometrics verification, IEEE Transactions on Signal Processing 52 (10) (2004) 3059–3072.

[6] A. Ross, A.K. Jain, Information fusion in biometrics, Pattern Recognition Letters 24 (2003).

[7] Y. Wang, T. Tan, A.K. Jain, Combining face and iris biometrics for identity verification, in: Lecture Notes in Computer Science, vol. 2688, 2003, pp. 805–813.

[8] K. Nandakumar, Y. Chen, S.C. Dass, A.K. Jain, Likelihood ratio-based biometric score fusion, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (2) (2008) 342–348.

[9] S. Tulyakov, V. Govindaraju, Classifier combination types for biometric applications, in: CVPRW '06: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, 2006, IEEE Computer Society, Washington, DC, USA, pp. 58, ISBN: 0-7695-2646-2, doi: 10.1109/CVPRW.2006.54.

[10] C. Soutar, Biometric system security, 2004 ⟨http://www.bioscrypt.com/assets/security_soutar.pdf⟩.

[11] W. Pedrycz, F.A.C. Gomide, An Introduction to Fuzzy Sets: Analysis and Design, Complex Adaptive Systems, MIT Press, Cambridge, MA, 1998.

[12] M. Sugeno, K.T. Kang, Structure identification of fuzzy model, Fuzzy Sets and Systems 28 (1991) 191–212.

[13] A.A.F. Ynui, L.L. Ling, Estudo e implementação de um sistema biométrico multimodal com lógica nebulosa, UNICAMP, Technical Report, March 2008.

[14] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of Computer Vision and Pattern Recognition, 1991, pp. 586–591.

[15] Nist fingerprint software, National Institute of Standards and Technology (NIST), Internet page, March 2006 ⟨http://fingerprint.nist.gov/NFIS/⟩.

[16] E. Tabassi, C.L. Wilson, C.I. Watson, Fingerprint image quality, National Institute of Standards and Technology (NIST), Technical Report NISTIR 7151, August 2004.

[17] Y. Yao, B. Abidi, M. Abidi, Quality Assessment and Restoration of Face Images in Long Range/High Zoom Video, Springer, Berlin, 2007, pp. 43–60, (Chapter 4).

[18] Q. Xiong, C. Jaynes, Mugshot database acquisition in video surveillance networks using incremental auto-clustering quality measures, in: AVSS '03: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance, IEEE Computer Society, Washington, DC, USA, 2003, p. 191.

[19] Fvc2004: Third international fingerprint verification competition, University of Bologna, Internet page, 2006 ⟨http://bias.csr.unibo.it/fvc2004/⟩.

[20] NIST, The facial recognition technology (feret) database, National Institute of Standards and Technology (NIST), Technical Report NISTIR 7151, August ⟨http://www.itl.nist.gov/iad/humanid/feret/feret_master.html⟩.

[21] R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain, Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (3) (2005) 450–455.

[22] NIST, Biometric scores set, National Institute of Standards and Technology (NIST), Technical Report BSSR1, October 2004 ⟨http://www.itl.nist.gov/iad/894.03/biometricscores/⟩.