# Combining Match Scores with Liveness Values in a Fingerprint Verification System

Emanuela Marasco, Yaohui Ding, Arun Ross
Lane Department of Computer Science and Electrical Engineering
West Virginia University, PO Box 6109 Morgantown, WV 26506
emanuela.marasco@mail.wvu.edu, yding@mix.wvu.edu, arun.ross@mail.wvu.edu

## Abstract

*We discuss the problem of combining biometric match scores with liveness measure values in the context of fingerprint verification. Recent literature has focused on the development of methods to assess if an input fingerprint sample is a "live" entity or a "spoof" artefact. This is commonly done by generating a single-valued numerical entity referred to as the liveness measure value. However, the problem of combining this liveness value with match scores has not been rigorously investigated. The goal of this work is to design a framework in which a liveness detector is incorporated with a fingerprint matcher. We first design and analyze three different methods to combine match scores with liveness values. Next, we introduce a Bayesian Belief Network (BBN) scheme that models the relationship between match scores and liveness values. Experiments carried out on a publicly available database of the Fingerprint Liveness Detection Competition 2009 (LivDet09) show the effectiveness of assuming a certain degree of influence of liveness values on match scores.*

## 1. Introduction

Recent research has highlighted the vulnerability of biometric systems to spoof attacks, commonly realized by presenting duplicated biometric traits to the sensor [14]. In the case of fingerprints, spoofs are usually made of materials that can be imaged by the fingerprint sensor, such as playdoh, gelatin and silicone [3]. Fig. 1 shows examples of live and spoof fingerprints. An efficient countermeasure that is being studied to handle this problem is liveness detection [24]. Liveness detection refers to the ability of a system to correctly distinguish between a live human biometric presentation and spoof artefacts [16]. This is commonly done by generating a single-valued numerical entity referred to as the liveness measure value. Fingerprint samples that are assigned higher liveness values are less likely to be a spoof, and vice-versa.

While previous research has considered the problem of designing spoof-resilient fusion schemes for multibiometric systems [20], [19], [7], [10], the specific problem of combining liveness values with match scores has not been investigated.

In this work, we assume that spoofs may be presented during both enrollment and verification stages. Two potentially dangerous cases are possible: (a) a person creates a new identity using a spoof during enrollment and uses another sample of the spoof during verification; and (b) a person enrolls using a live finger but an intruder uses a spoof artefact of the true finger during verification.

The central theme of the work is based on the observation that match scores are impacted by the presentation of a spoof artefact to a fingerprint system. While this observation has been made in the literature, the nature of the impact has not been modeled. Further, there is no systematic evaluation of methods to incorporate liveness values with match scores. In this work, we first discuss three different methods to combine match scores with liveness values, which do not explicitly model the interaction between match scores and liveness values. Then, we design a method for combining match scores and the corresponding liveness value based on a Bayesian Belief Network (BBN) model that assumes a certain influence of the liveness value on match scores.

The contribution of this paper is two-fold: a) designing four methods for combining match scores with liveness measure values; b) designing an evaluation scheme for assessing the robustness of a biometric system in the presence of spoof attacks. We investigate if combining liveness values with match scores can improve the verification performance in addition to improving the robustness to spoof attacks.

The paper is organized as follows. In Section 2, we describe four possible configurations for combining match scores with liveness measure. Section 3 presents the proposed evaluation framework. Section 4 discusses the evaluation procedure and the experimental results. Section 5



978-1-4673-1228-8/12/$31.00 ©2012 IEEE

(a) (b) (c)

Figure 1. Fingerprint samples taken from the CrossMatch database of the Fingerprint Liveness Detection Competition 2009 [13]: a) live; b) silicone spoof and c) gelatin spoof.

compares the four approaches and draws conclusions.

## 2. The Proposed Combination Methods

In this section we present four different methods for consolidating the output of a biometric modality matcher with that of a liveness detector. We focus on two main goals: *i)* Goal 1: determine the probability that the two samples being compared belong to the same identity (it does not matter if either is a spoof or a live sample); *ii)* Goal 2: determine the probability that the two samples being compared are from the same identity *and* are both live samples.

### 2.1. Sequential Methods

In this work, we assume that the matcher and liveness detector are "classifiers". The inputs to the matcher are two fingerprint samples (e.g., gallery and probe images). The output is a match score that indicates the proximity of the two samples. A threshold is applied to this match score in order to determine if the samples correspond to the same identity ("Genuine (G)") or different identities ("Impostor (I)"). Thus, the verification stage has two output classes: G and I. The input to the liveness detector is a fingerprint sample (e.g., gallery or probe image). The output is a liveness value indicating the degree of liveness of the sample. A threshold is applied to this liveness value in order to determine if the sample is "Live (L)" or "Spoof (S)". Since there are two samples, liveness detection stage has four output classes: LL, LS, SL, SS (see Table A). We consider various arrangements of the matcher and the liveness detector modules. Some configurations may not be operationally tenable - however, these have been considered only for completeness sake.

- In **Method 1**, the classifier is invoked before the liveness detector as seen in Fig. 2. The matcher in the first stage is used to distinguish genuine from impostor based only on match scores. In the liveness detection stage there are two pairs of classifiers: one pair that is invoked if the input samples are deemed to belong

**Table A: Notation**

**Inputs:**

Let $m$ be the match score between the gallery and probe samples as computed by the matcher.

Let $l_g$ be the liveness measure value assigned by the liveness detector to the gallery sample.

Let $l_p$ be the liveness measure value assigned by the liveness detector to the probe sample.

**Events:**

Let $I = 0$ $(1)$ denote a genuine (impostor) user.

Let $S_g = 0$ $(1)$ denote the presence of a live (spoof) biometric presentation at enrollment time.

Let $S_p = 0$ $(1)$ denote the presence of a live (spoof) biometric presentation at verification time.

**Output classes:**

Live-Live-Genuine (LLG): the gallery and the probe are both live and they have the same identity.

Live-Spoof-Genuine (LSG): the gallery is live, the probe is spoofed but they correspond to the same identity.

Spoof-Live-Genuine (SLG): the gallery is spoofed, the probe is live but they correspond to the same identity.

Spoof-Spoof-Genuine (SSG): the gallery and the probe are both spoofed and they are of the same identity.

Live-Live-Impostor (LLI): the gallery and the probe are both live but they correspond to different identities.

Live-Spoof-Impostor (LSI): the gallery is live, the probe is spoofed and they correspond to different identities.

Spoof-Live-Impostor (SLI): the gallery is spoofed, the probe is live and they correspond to different identities.

Spoof-Spoof-Impostor (SSI): the gallery and the probe are both spoofed and they correspond to different identities.

**Evaluation:**

Goal 1: compute the probability that the presentation characteristic belongs to a genuine user (determine $P(I = 0|m, l_g, l_p)$).

Goal 2: compute the probability that the presentation characteristic belongs to a genuine user and both gallery and probe are live (determine $P(I = 0, S_g = 0, S_p = 0|m, l_g, l_p)$).

to the Genuine (G) class and another that is invoked if they are deemed to belong to the Impostor (I) class. This arrangement may be redundant (i.e., the use of four different liveness detectors may not be necessary).

- In **Method 2**, the liveness detector is invoked before the matcher as seen in Fig. 3. Depending upon the output of the two liveness classifiers in the first stage (LL, LS, SL or SS), one of four matchers in the verification stage is invoked. For example, the first matcher (Classifier 3) operates only on gallery and probe samples that are both classified as Live, while the fourth matcher (Classifier 6) operates only on gallery and probe samples that are both classified as Spoof.

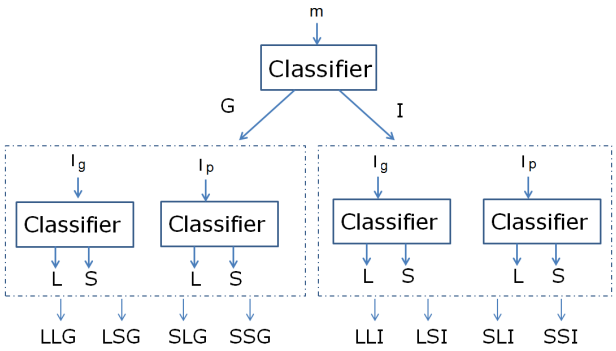



Figure 2. Architecture of Method 1. Here, the matcher is invoked before the liveness detector. The classifier in the first stage (classifier 1) is used to distinguish genuine from impostor based only on match scores. In the spoof detection stage there are two pairs of classifiers: one pair (classifier 2 and 3) that is invoked if the input samples are deemed by the matcher to belong to the Genuine (G) class and another pair (classifier 4 and 5) that is invoked if they are deemed to belong to the Impostor (I) class. This arrangement may be redundant (i.e., the use of four different liveness classifiers may not be necessary). See Table A for notations.

## 2.2. Classifier-based Fusion

In **Method 3** (see Fig. 4), the match score and the liveness values are provided as inputs to a single classifier [21], [23], [22]. This classifier has one of eight possible outputs: LLG, LSG, SLG, SSG, LLI, LSI, SLI, SSI. This is an example of a multi label problem [1], [18]. For each class label, the first two letters denote the liveness state of the samples, while the third letter denotes whether the samples correspond to the Genuine or Impostor class (see Table A). In this method, no explicit assumption is made regarding a possible relationship between liveness values and match scores.

## 2.3. Bayesian Belief Network Framework

The three methods described above do not explicitly model the relationship between liveness values and match

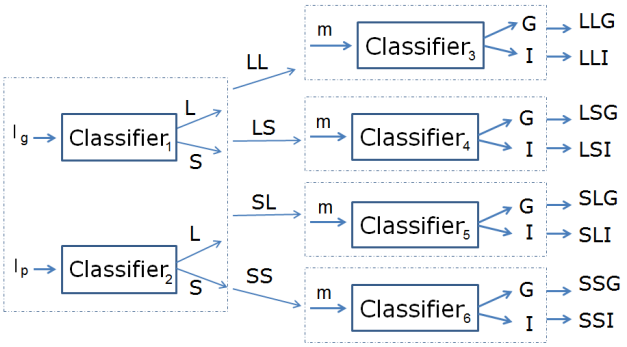



Figure 3. Architecture of Method 2. Here, the liveness detector is invoked before the matcher. Depending upon the output of classifier 1 and 2 (LL, LS, SL or SS), one of four classifiers in the verification stage is invoked. For example, classifier 3 operates only on input scores between gallery and probe samples that are both classified as Live, while classifier 6 operates only on scores between gallery and probe samples that are both classified as Spoof. See Table A for notations.

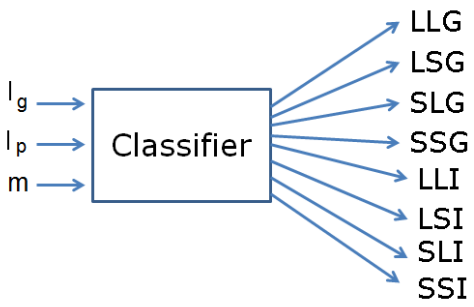



Figure 4. Architecture of Method 3. Here, the classifier has three inputs: match score, liveness value of gallery sample and liveness value of probe sample. All 3 inputs are used simultaneously in order to determine the output class. See Table A for notation.

scores. A powerful framework for modeling causal relationships among a set of variables $X$ is offered by graphical models such as Bayesian Belief Networks [6]. A graph is able to capture the way in which the joint distribution over all of the random variables can be decomposed into a product of factors each depending only on a subset of the involved variables.

Fig. 5 shows a BBN-based representation for our domain of interest, referred to as **Method 4**. The variable $I$ denotes the event related to the presence or absence of a genuine user. It assumes value equal to '0' when the samples belong to the Genuine class and '1' when the samples belong to the Impostor class. The variable $m$ denotes the match score between the two samples (e.g., gallery and probe) whose value is affected by the state of the variable $I$ [15], [4]. For example, a match score between two samples of different individuals (I=1) is likely to be lower than that of samples com-

ing from the same individual (I=0). The variables $S_g$ and $S_p$ represent the events related to the presence of a spoof biometric presentation at enrollment and verification times, respectively. Each assumes the value '1' when the presentation characteristic is a spoof and the value '0' when it is live. The variables $l_g$ and $l_p$ denote the liveness values of the gallery and probe samples, respectively.

In the proposed method, we assume that the liveness values $l_g$ and $l_p$ influence the corresponding match score, $m$. The interactions among the involved variables are based on the idea that the events $S_g$, $S_p$ and $I$ influence a common effect, i.e., the decision made by the biometric system, through variables $l_g$, $l_p$ and $m$. We study how the impact of the event $I$ on the final decision depends on the other events $S_g$ and $S_p$ [9], [17]. This approach has one of eight possi-

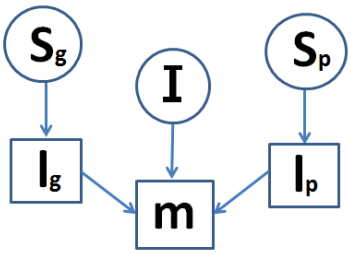



Figure 5. Architecture of Method 4. The Bayesian Network combines match scores and the corresponding liveness measure values. In this configuration, the liveness measure is assumed to influence match scores.

ble outputs: LLG, LSG, SLG, SSG, LLI, LSI, SLI, SSI (see Table A).

#### 2.3.1 Probabilistic Representation

The computational paradigm of Bayesian Networks is based on probabilistic evidence where new evidence has to be propagated to other parts of the network. When performing Bayesian inference, a combination of observed data with prior knowledge is required. In our study, we seek to integrate the biometric matcher, the liveness detector, and prior of the three distributions P($I$), P($S_g$) and P($S_p$). In the Bayesian Network model, all the conditional probabilities are given and the goal is to determine the maximum posterior value of the unknown variables in the network, through careful application of the Bayes rule [2], [26]. The joint probability distribution, represented as $P(I, S_g, S_p, m, l_g, l_p)$, is factorized according to the structure of the network, as follows:

$$P(I, S_g, S_p, m, l_g, l_p) = P(I)P(S_g)P(S_p)P(l_g|S_g)P(l_p|S_p)P(m|l_g, l_p, I) \quad (1)$$

For goal 2, the probability can be updated as follows:

$$\begin{aligned}
P(I, S_g, S_p|m, l_g, l_p) &= \frac{P(I, S_g, S_p, m, l_g, l_p)}{P(m, l_g, l_p)} \\
&= \frac{P(I)P(S_g)P(S_p)P(l_g|S_g)P(l_p|S_p)P(m|l_g, l_p, I)}{P(m|l_g, l_p)P(l_g, l_p)} \\
&\quad \text{(from Eqn. (1))} \\
&= \frac{P(I)P(S_g)P(S_p)P(l_g|S_g)P(l_p|S_p)P(m|l_g, l_p, I)}{P(m|l_g, l_p)P(l_g)P(l_p)} \\
&\quad \text{(from Fig. 5, } l_g \text{ and } l_p \text{ are independent)} \\
&= \frac{P(S_g)P(l_g|S_g)}{P(l_g)} \frac{P(S_p)P(l_p|S_p)}{P(l_p)} \frac{P(I)P(m|l_g, l_p, I)}{P(m|l_g, l_p)} \\
&= P(S_g|l_g)P(S_p|l_p)\frac{P(I|l_g, l_p)P(m|l_g, l_p, I)}{P(m|l_g, l_p)} \\
&\quad \text{(as shown in Fig. 5, } P(I) = P(I|l_g, l_p)) \\
&= P(S_g|l_g)P(S_p|l_p)\frac{P(m|l_g, l_p)P(I|m, l_g, l_p)}{P(m|l_g, l_p)} \\
&= P(S_g|l_g)P(S_p|l_p)P(I|m, l_g, l_p)
\end{aligned} \quad (2)$$

The above equation shows that the proposed BBN can be considered as being composed of three independent components: the first two terms indicate that both the gallery and probe samples are classified as being live or spoof based only on their liveness measure values, while the third term indicates that, the input biometric presentation is classified as being genuine or impostor based on both match scores and liveness measure values.

## 3. Performance Metrics

The four proposed methods are evaluated considering two main categories of errors (i.e., spoof detection and identity verification) and using a set of performance metrics related to the two goals mentioned in Section 2.

For each method the evaluation is conducted at two different levels:

- *Verification level*: When distinguishing genuine from impostor pairs, the performance can be measured based on the errors made by a typical biometric system: False Match Rate (FMR) and False Non-Match Rate (FNMR) [5].

- *Spoof detection level*: When distinguishing spoof from live samples, the robustness of the system can be measured based on the errors related to the False Live Rejection Rate (FLRR) defined as the proportion of $L$ samples that are incorrectly classified as being $S$, and False Spoof Acceptance Rate (FSAR) defined as the proportion of $S$ samples that are incorrectly classified as being $L$.

Below we define the metrics to measure the errors that may occur for Method 3 and 4.

- False Match Rate (FMR): Proportion of samples belonging to classes $LLI$, $LSI$, $SLI$, $SSI$ that are incorrectly classified as belonging to one of the classes $LLG$, $LSG$, $SLG$, $SSG$;

- False Spoof Acceptance Rate (FSAR): Proportion of samples belonging to classes $LSI$, $SLI$, $SSI$, $LSG$, $SLG$, $SSG$ that are incorrectly classified as belonging to one of the classes $LLG$, $LLI$;

- False Non-Match Rate (FNMR): Proportion of samples belonging to classes $LLG$, $LSG$, $SLG$, $SSG$ that are incorrectly classified as belonging to one of the classes $LLI$, $LSI$, $SLI$, $SSI$;

- False Live Rejection Rate (FLRR): Proportion of samples belonging to classes $LLG$, $LLI$ that are incorrectly classified as belonging to one of the classes $LSI$, $SLI$, $SSI$, $LSG$, $SLG$, $SSG$.

The aforementioned evaluation scheme specifically focuses on errors pertaining to a) the matching errors between probe and gallery images, b) the spoof detection errors of the probe and gallery images. However, when the system operates in real-world applications, it should only accept samples that belong to the class $LLG$ (i.e., both the probe and gallery samples are live and pertain to the same identity); it should reject all other samples. To capture this, the notion of *global* errors in introduced below.

- Global False Acceptance Rate (GFAR): Proportion of samples belonging to classes $LLI$, $LSI$, $SLI$, $SSI$, $LSG$, $SLG$, $SSG$ that are incorrectly classified as belonging to the class $LLG$;

- Global False Rejection Rate (GFRR): Proportion of samples belonging to classes $LLG$ that are incorrectly classified as belonging to one of the classes $LLI$, $LSI$, $SLI$, $SSI$, $LSG$, $SLG$, $SSG$.

# 4. Experimental Results

## 4.1. Dataset

The performance of the proposed methods was evaluated on a subset of the CrossMatch database taken from the Fingerprint Liveness Detection Competition 2009 [13]. It is made up of live and spoof fingerprint samples imaged using a CrossMatch optical scanner with a resolution factor of 500 dpi and an image size of 480x640 pixels. Two spoof materials were considered in our experiments: gelatin and silicone. Table 1 provides details about the database. Match scores were extracted using the VeriFinger software by matching all pairs of images across all subjects. The scores, therefore, correspond to four different matching scenarios: Live vs Live, Live vs Spoof, Spoof vs Live, and Spoof vs Spoof. For each image, the liveness measure was extracted by using an algorithm which combines morphological and perspiration-based characteristics [11], [12].

Table 1. Details about the dataset adopted for the experiments.

| Material | Subjects | Samples[a] | Scores |
|---|---|---|---|
| Gelatin | 41 | 10 live<br>10 spoof<br>per subject | Gen: 41*20*19 = 15,580<br>Imp: 41*40*20*19 = 623,200<br>Liveness values: 41*20 = 820 |
| Silicone | 23 | 10 live<br>10 spoof<br>per subject | Gen: 23*20*19 = 8,740<br>Imp: 23*22*20*19 = 192,280<br>Liveness values: 23*20 = 460 |

[a] 5 of these samples are acquired at 0 seconds while the remaining 5 are acquired after 2 seconds

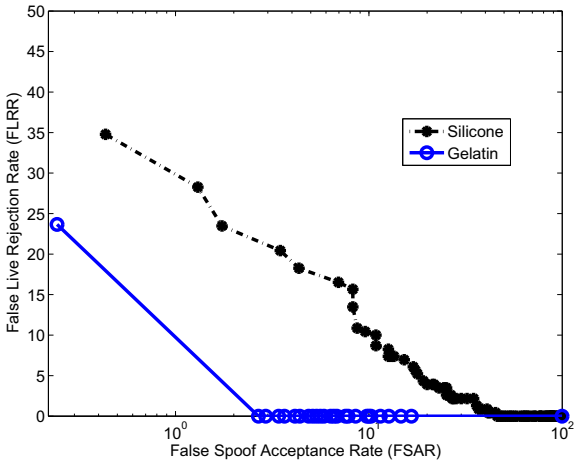
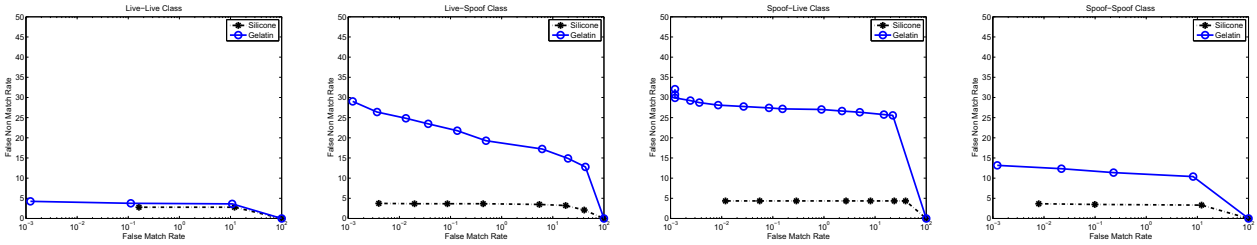



Figure 6. ROC curves for gelatin and silicone.

The verification performance of the fingerprint recognition system is analyzed using the Receiving Operating Characteristic (ROC) curve as shown in Fig. 7. ROC curves are obtained for both spoof materials under the four different matching scenarios (live-live, live-spoof, spoof-live and spoof-spoof). On the CrossMatch database, the liveness measure seems to be better in detecting spoof samples made with gelatin and poor in detecting spoof samples made with silicone. So the liveness detector has higher reliability in the case of gelatin and lower reliability in the case of silicone. The ROC curves of the liveness detector for the two spoof materials are shown in Fig. 6.

## 4.2. Evaluation Procedure

The sequential methods (Method 1 and Method 2) require a threshold, i.e., the classifiers seen in Fig. 2 and Fig. 3 are threshold-based [25]. In order to determine a suitable threshold, a training set is needed for each classifier (matchers and liveness detectors). The threshold selected for each classifier minimizes the Total Error Rate (TER) on the train-




(a) (b) (c) (d)

Figure 7. Verification performance of the fingerprint recognition system for different spoof materials. Four different scenarios are considered: a) Live-Live case: the gallery and the probe samples are both live; b) Live-Spoof case: the gallery sample is live and the probe sample is a spoof; c) Spoof-Live case: the gallery sample is a spoof and the probe sample is live; d) Spoof-Spoof case: the gallery and the probe samples are both spoofs.

ing set. TER is defined as follows:

$$TER_V = FMR + FNMR \tag{3}$$

for the verification performance and

$$TER_{SD} = FSAR + FLRR \tag{4}$$

for the spoof detection performance.

The training set is composed by randomly sampling the data set of subjects at 3 different rates: 25%, 50%, 75%. In order to avoid overfitting the training sets, a 10-fold cross validation is used. In each fold, the threshold that yields the minimum TER is determined [25]. In some cases, two or more thresholds may have the same minimum value. To resolve such a tie, the threshold corresponding to the lowest Type-I error (FMR for matchers and FSAR for liveness detectors) is selected. Once the threshold is determined for every training fold, the average threshold of all 10 folds is used as the final threshold. The performance is then evaluated on all the test folds using this average threshold.

The evaluation of Method 3 was carried out by implementing four different classifiers and choosing the one that resulted in the best performance. Classifiers were trained at different rates (25%, 50% and 75%) as well. The Neural Network (NN) presented the lowest FMR, compared to the Decision Tree (DT), the Naive Bayes (NB) and the K-Nearest Neighbor (KNN). For Method 3, we report results obtained by using the NN since it was the best classifier. NN was then employed in Method 4 as well, as an estimator to compute the conditional probability obtained by the mathematical deviation expressed in Eqn. (2) [8].

The classifiers were implemented by using the Matlab Version 7.6.0.324 (R2008a) software.

### 4.3. Results

The performance of the proposed sequential schemes, i.e., Method 1 and Method 2, can be analyzed via ROC curves. The performance of Methods 3 and 4, on the other hand, can be analyzed using an 8x8 class-confusion matrix. Therefore, in order to compare the performance of Method 1 and 2 with that of Methods 3 and 4, we report the error rates only at a specific operating point where all the four proposed methods have comparable Type-I error rates. In the case of Method 3, since the FNMR obtained by the Neural Network was not comparable with the other three proposed methods, we also report the error rates of the Full Bayesian classifier which showed a comparable FNMR. The results are summarized in Tables 2 and 3.

- Tables 2 and 3 indicate that the best *verification* performance is achieved by Method 4. This outcome suggests that combining liveness information with match scores leads to a verification performance improvement compared to the case where the liveness measure is not used (see the error rates of stage 1 of Method 1). For example, at a training rate of 25%, FMR is 0.11% for Method 4 and 0.18% when liveness measure is not used.

- In the presence of a reliable liveness measure (see Table 3 which corresponds to gelatin spoof), the best *spoof detection* performance is achieved by Method 3, while when dealing with a less reliable liveness measure (see Table 2 which corresponds to silicone spoof) it is achieved by Method 1.

- The best *global* performance is achieved by Method 4. This result demonstrates that the configuration of the Bayesian Network is effective and the assumption that liveness values influence match scores works well. Lowest global error rates are observed in the presence of a reliable liveness measure (see Table 3).

## 5. Conclusions

In this paper, we propose four different configurations for combining biometric match scores with liveness measure values. Methods 1 and 2 are sequential, Method 3 is classifier-based, and Method 4 is based on the Bayesian Belief Network (BBN) model. In Method 4, the problem is for-

Table 2. Verification, Spoof Detection and Global performance of the proposed combination methods. The spoof material employed here is silicone. 3 $(NN)$ denotes Method 3 based on the Neural Network classifier; 3 $(FB)$ denotes Method 3 based on the Full Bayesian classifier; 4 $(NN)$ denotes Method 4 based on the Neural Network as estimator and 4 $(FB)$ denotes Method 4 based on the Full Bayesian as estimator.

| | | **Verification** | | **Spoof Detection** | | **Global Error** | |
|---|---|---|---|---|---|---|---|
| Rates | Method | FMR % | FNMR % | FSAR % | FLRR % | GFA % | GFR % |
| 25% | 1 | 0.175 | 5.261 | **1.619** | **12.487** | 0.543 | 13.739 |
| | 2 | 0.323 | 5.493 | 10.468 | 12.403 | 0.701 | 14.899 |
| | 3 (NN) | 0.336 | 10.101 | 3.909 | 12.744 | 0.253 | 25.652 |
| | 3 (FB) | 0.625 | 5.000 | 12.066 | 12.476 | 0.505 | 13.391 |
| | 4 (NN) | **0.002** | **5.528** | 5.688 | 12.544 | **0.248** | **15.290** |
| | 4 (FB) | 0.109 | 5.058 | 5.272 | 12.514 | 0.326 | 14.725 |
| 50% | 1 | 0.234 | 5.022 | **1.467** | **12.492** | 0.503 | 14.435 |
| | 2 | 0.547 | 5.044 | 10.224 | 12.330 | 0.862 | 14.348 |
| | 3 (NN) | 0.345 | 9.500 | 3.769 | 12.770 | 0.217 | 26.348 |
| | 3 (FB) | 0.637 | 4.957 | 12.013 | 12.471 | 0.478 | 13.130 |
| | 4 (NN) | **0.002** | **5.187** | 5.297 | 12.558 | **0.241** | **15.696** |
| | 4 (FB) | 0.103 | 5.044 | 5.244 | 12.516 | 0.308 | 14.217 |
| 75% | 1 | 0.144 | 4.783 | **0.660** | **12.517** | 0.317 | 18.783 |
| | 2 | 0.482 | 4.783 | 7.314 | 12.346 | 0.690 | 17.046 |
| | 3 (NN) | 0.360 | 11.435 | 4.000 | 12.994 | 0.192 | 30.078 |
| | 3 (FB) | 0.593 | 4.478 | 11.877 | 12.482 | 0.535 | 14.435 |
| | 4 (NN) | **0.002** | **5.065** | 5.338 | 12.561 | **0.222** | **19.304** |
| | 4 (FB) | 0.112 | 4.565 | 5.351 | 12.501 | 0.313 | 18.044 |

Table 3. Verification, Spoof Detection and Global performance of the proposed combination methods. The spoof material employed here is gelatin. 3 $(NN)$ denotes Method 3 based on the Neural Network classifier; 3 $(FB)$ denotes Method 3 based on the Full Bayesian classifier; 4 $(NN)$ denotes Method 4 based on the Neural Network as estimator and 4 $(FB)$ denotes Method 4 based on the Full Bayesian as estimator.

| | | **Verification** | | **Spoof Detection** | | **Global Error** | |
|---|---|---|---|---|---|---|---|
| Rates | Method | FMR % | FNMR % | FSAR % | FLRR % | GFA % | GFR % |
| 25% | 1 | 0.191 | 15.707 | 0.194 | 12.458 | 0.045 | 4.423 |
| | 2 | 0.218 | 16.008 | 0.569 | 12.458 | 0.047 | 4.585 |
| | 3 (NN) | 0.039 | 26.894 | 0.648 | 12.502 | 0.074 | 13.171 |
| | 3 (FB) | 0.265 | 15.285 | **0.008** | **12.437** | 0.086 | 5.309 |
| | 4 (NN) | **0.011** | **16.634** | 0.333 | 12.479 | **0.006** | **5.724** |
| | 4 (FB) | 0.142 | 15.602 | 0.485 | 12.434 | 0.040 | 4.618 |
| 50% | 1 | 0.251 | 15.915 | 0.184 | 12.455 | 0.051 | 4.488 |
| | 2 | 0.176 | 16.463 | 0.652 | 12.460 | 0.047 | 4.488 |
| | 3 (NN) | 0.402 | 26.939 | 1.336 | 12.773 | 0.051 | 19.385 |
| | 3 (FB) | 0.271 | 15.537 | **0.002** | **12.451** | 0.056 | 5.854 |
| | 4 (NN) | **0.011** | **16.415** | 0.320 | 12.475 | **0.008** | **4.634** |
| | 4 (FB) | 0.089 | 15.866 | 0.316 | 12.434 | 0.046 | 4.195 |
| 75% | 1 | 0.188 | 15.634 | 0.114 | 12.461 | 0.042 | 4.488 |
| | 2 | 0.223 | 15.537 | 0.572 | 12.461 | 0.042 | 4.488 |
| | 3 (NN) | 0.402 | 28.171 | 1.650 | 12.773 | 0.039 | 21.453 |
| | 3 (FB) | 0.284 | 15.756 | $\sim$ **0** | **12.449** | 0.062 | 6.049 |
| | 4 (NN) | **0.010** | **16.951** | 0.302 | 12.474 | **0.005** | **4.293** |
| | 4 (FB) | 0.117 | 15.146 | 0.306 | 12.432 | 0.029 | 4.293 |

mulated such that an influence of liveness values on match scores is assumed.

Results show that liveness values do impact verification performance. Our experiments also show that the best

global accuracy is achieved when using a BBN-based combination scheme. However, the performance is affected by the reliability of the liveness detector used. We note that the primary purpose of this paper was to indicate the multiple architectures that are possible when combining liveness values with match scores. The precise performance numbers (e.g., Tables 2 and 3) will rely on the matcher and liveness detector used.

We will extend this work to additional databases in which spoof samples are realized with other spoof materials. Further, while our framework has been proposed for unimodal scenarios, it can be extended to multimodal systems as well.

## 6. Acknowledgments

This work was partially supported by U.S. NSF CAREER Award IIS 0642554. The authors are grateful for valuable discussions with researchers from Clarkson University, CSC and NIST.

## References

[1] F. Brucker, F. Benited, and E. Sapozhnikova. Multi-label classification and extracting predicted class hierachies. *Pattern Recognition*, 44:724–738, 2011. 3

[2] R. Duda, P. Hart, and D. Stork. *Pattern Classification*. Wiley, New York, 2. edition, 2001. 4

[3] J. Galbally-Herrero, J. Fierrez-Aguilar, J. D. Rodriguez-Gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador. On the vulnerability of fingerprint verification systems to fake fingerprint attacks. pages 130–136, October 2006. 1

[4] D. Heckerman. A tutorial on learning with bayesian networks. *Technical Report MSR-TR-95-06*, November 1996. 3

[5] A. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transaction on Circuits and Systems for Video*, 14(1):4–20, January 2004. 4

[6] F. Jensen and T. Nielsen. *Bayesian networks and decision graphs*. Springer, 2007. 3

[7] P. Johnson, B. Tan, and S. Schuckers. Multimodal fusion vulnerability to non-zero effort (spoof) imposters. *IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–5, 2010. 1

[8] A. Likas. Probability density estimation using artificial neural networks. *Computer Physics Communications*, 135:167–175, 2001. 6

[9] P. Lucas. Bayesian network modelling through qualitative patterns. *Artificial Intelligence*, pages 163–233, 2005. 4

[10] E. Marasco, P. Johnson, C. Sansone, and S. Schuckers. Increase the security of multibiometric systems by incorporating a spoofing detection algorithm in the fusion mechanism. *The 10th International Workshop on Multiple Classifer Systems (MCS)*, June 2011. 1

[11] E. Marasco and C. Sansone. An anti-spoofing technique using multiple textural features in fingerprint scanners. *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMs)*, pages 8–14, 2010. 5

[12] E. Marasco and C. Sansone. Combining perspiration- and morphology-based static features for fingerprint liveness detection. *Pattern Recognition Letters*, 33:1148–1156, 2012. 5

[13] G. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers. First international fingerprint liveness detection competition (LivDet09). *The 15th International Conference on Image Analysis and Processing (ICIAP)*, pages 12–23, September 2009. 2, 5

[14] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. *Optical Security and Counterfeit Deterrence Techniques IV*, 4677:275–289, January 2002. 1

[15] D. Maurer and J. Baker. Fusing multimodal biometrics with quality estimates via a bayesian belief network. *Pattern Recognition*, 41(3):821–832, March 2008. 3

[16] K. Nixon, V. Aimale, and R. Rowe. Spoof detection schemes. *Handbook of Biometrics*, 2008. 1

[17] N. Poh and J. Kittler. A unified framework for biometric expert fusion incorporating quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(1):3–18, 2012. 4

[18] J. Quevedo, O. Luaces, and A. Bahamonde. Multilabel classifiers with a probabilistic thresholding strategy. *Pattern Recognition*, 44:724–738, 2011. 3

[19] R. Rodrigues, N. Kamat, and V. Govindaraju. Evaluation of biometric spoofing in a multimodal system. *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2010. 1

[20] R. Rodrigues, L. Ling, and V. Govindaraju. Robustness of multimodal biometric fusion methods against spoof attacks. *Journal of Visual Languages and Computing*, 2009. 1

[21] A. Ross and A. Jain. Information fusion in biometrics. *Pattern Recognition Letters 24*, pages 2115–2125, 2003. 3

[22] A. Ross, A. Jain, and K. Nandakumar. *Introduction to Biometrics: A Textbook*. Springer, 2011. 3

[23] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multi-Biometrics*. Springer, 2006. 3

[24] S. Schuckers. Spoofing and anti-spoofing measures. *Information Security Technical Report*, 7(4):56–62, 2002. 1

[25] V. Sheng and C. Ling. Thresholding for making classifiers cost-sensitive. *the 21st National Conference on Artificial Intelligence*, pages 476–481, July 2006. 5, 6

[26] S. Yanushkevich. Belief network design for biometric systems. *Computational Intelligence in Biometrics and Identity Management (CIBIM)*, pages 1–10, April 2011. 4