Bayesian Belief Models for Integrating Match Scores with Liveness and Quality Measures in a Fingerprint Verification System

Yaohui Ding
Department of Computer Science and Engineering, Michigan State University
East Lansing, MI 48824, USA
dingyaoh@msu.edu

Ajita Rattani
Department of Computer Science and Electrical Engineering, University of Missouri-Kansas City
Kansas City, MO 64110, USA
rattania@umkc.edu

Arun Ross
Department of Computer Science and Engineering, Michigan State University
East Lansing, MI 48824, USA
rossarun@cse.msu.edu

Abstract

Recent research has sought to improve the resilience of fingerprint verification systems to spoof attacks by combining match scores with both liveness measures and image quality in a learning-based fusion framework. Designing such a fusion framework is challenging because quality and liveness measures can impact the match scores and, therefore, the influence of these variables on the match score has to be modelled. Further, these measures themselves are influenced by many latent factors, such as the fabrication material used to generate fake fingerprints. We advance the state-of-the-art by proposing two Bayesian Belief Network (BBN) models that can utilize these measures effectively, by appropriately modelling the relationship between quality, liveness measure and match scores with the consideration of latent variables. We demonstrate the efficacy of the proposed models on the LivDet 2011 fingerprint spoof dataset.

1. Introduction

Biometrics is the science of recognizing individuals based on their physical (such as face, fingerprint, iris) and behavioral (such as speech and gait) traits [5]. Among the various biometric traits used, fingerprint is one of the more popular traits that has been adopted in several applications such as access control systems, border crossing programs, time and attendance systems, and self-service kiosks such as ATMs. The popularity of the fingerprint stems from its low error rates and ease of use [6]. Recently, a growing number of mobile applications are being equipped with fingerprint sensors for access control.

Despite its popularity, the fingerprint acquisition process is inherently affected by factors such as noise in the acquisition sensor, non-linear skin deformations and partial prints arising during the physical interaction between the human and the device, and environmental/seasonal factors (*e.g.*, dry fingerprints during the winter season). Poor quality prints may occur both during the enrollment stage and the verification stage that can lead to errors in the biometric verification process.

Additionally, several studies have established that fingerprint verification systems are quite vulnerable to spoof attacks [15, 17]. A spoof attack occurs when an adversary mimics the biometric trait of another individual to circumvent the system[1]. Fake fingers can also be used during the enrollment stage, especially in mobile applications where the enrollment process is not monitored. Further, replicates of a fake finger can be delegated across multiple impostors for illegal access.

In order to realize a fingerprint system capable of handling variations in the image quality as well as robustness against spoof attacks, three major components are required: (a) image quality estimator yielding *quality measure* to in-

[1] https://www.tabularasa-euproject.org/

dicate how good the image quality is [4, 18], (b) spoof detector yielding *liveness measure* to indicate how likely the fingerprint is from a live finger [8, 16], and (c) an effective *fusion framework* capable of incorporating quality measures and liveness measures with the fingerprint match scores to make an optimal accept/reject decision. Figure 1 shows a block diagram where image quality, liveness measures and match scores extracted from a pair of fingerprint images are integrated together in a fusion framework to render the final accept/reject decision.

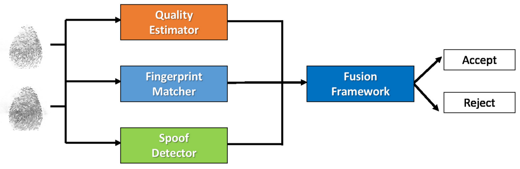


Figure 1: Illustration of the fusion framework integrating match scores with quality measures and liveness measures from two fingerprint samples, and rendering a final accept/reject decision.

The aim of this work is to design fusion frameworks that can *mitigate the limitations of existing frameworks* by simultaneously incorporating liveness and quality measures in a fingerprint verification system. This paper offers two main contributions:

- Categorizing existing fusion frameworks incorporating liveness measures into two categories: (a) direct modelling, and (b) graphical modelling, based on the relationship assumed between the variables (*i.e.*, match scores, liveness measure and quality measures).
- Development of two Bayesian Belief Network (BBN) based fusion frameworks, referred to as BBN-MLQ and BBN-MLQc. Experimental results show that the proposed models consistently obtain lower error rates over existing frameworks from two perspectives: (i) spoof detection capability, and (ii) verification of an identity.

2. Related Literature

Existing fusion frameworks that combine match scores with liveness measures and image quality can be categorized as (i) *Direct modelling* or (ii) *Graphical modelling* (as shown in Figure 2). This taxonomy is based on whether the dependence between the variables involved is purely learned from the data or assumed via causal understandings.

(i) Direct modelling: Direct modelling based schemes attempt to favor an equivalent impact from each involved variable, and the relationship among variables are purely learned from the data.

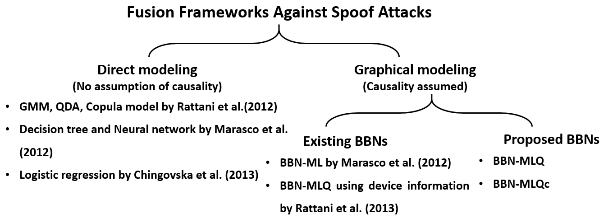

Fusion Frameworks Against Spoof Attacks
Direct modeling
(No assumption of causality)
GMM, QDA, Copula model by Rattani et al.(2012)
Decision tree and Neural network by Marasco et al. (2012)
Logistic regression by Chingovska et al. (2013)
Graphical modeling
(Causality assumed)
Existing BBNs
BBN-ML by Marasco et al. (2012)
BBN-MLQ using device information by Rattani et al. (2013)
Proposed BBNs
BBN-MLQ
BBN-MLQc

Figure 2: Taxonomy of existing fusion frameworks incorporating match scores, liveness measures and image quality.

Marasco *et al.* [7] proposed and compared different schemes for combining liveness measures with match scores. Compared to sequential schemes that invoked the spoof detector and the fingerprint matcher sequentially, parallel schemes that combined liveness measures and match scores as a two-dimensional input variable to classifiers such as Decision Trees, Naive Bayes and Neural Networks, were observed to result in a consistently higher accuracy. The authors remark that existing spoof detectors cannot be used for automated rejection of biometric samples until their detection accuracies are substantially improved.

Rattani and Poh [12] proposed a fusion framework that combined biometric sample quality and liveness measures with fingerprint match scores. The framework was implemented using three generative classifiers based on Gaussian Mixture Model (GMM), Gaussian Copula and Quadratic Discriminant Analysis (QDA). The results indicated that the GMM classifier provided the lowest overall error rate. The authors also established the benefit of fusing *both* quality and liveness measure in a fingerprint verification system. Chingovska *et al.* [3] proposed a fusion framework that incorporated LBP-based liveness measures with face match scores using logistic regression.

(ii) Graphical modelling: Graphical modelling based schemes assume a causal relationship between the variables. These schemes are often more accurate than direct modelling based schemes because the estimation of conditional probabilities is often simplified by such assumptions. Based on the assumptions about the relationship between the involved variables, different configurations of graphical models may be designed.

Marasco *et al.* [7] proposed a Bayesian Belief Network (BBN) model that combined match scores with liveness measures. This BBN (referred to as BBN-ML in this work) assumed a one-directional influence of match scores on liveness measure. Based on this configuration, the conditional probability of an input fingerprint sample being from a genuine user, given its liveness measure and match score, was inferred. *The authors also demonstrated the effectiveness of the proposed BBN over direct modelling schemes that did not explicitly assume any relationship between match scores* *and liveness measures.*

However, the image quality was not incorporated by Marasco *et al.* in their proposed framework. *Thus, the variation in the match score and liveness measure as a function of the change in the sample quality was not taken into account.* Further, the framework also did not take into account the influence of latent factors - such as the type of sensor and fake fabrication material (*i.e.*, material-specific characteristics) - on the liveness measures. Note that the fabrication materials can influence the quality of the fabricated spoofs and the liveness measure as pointed in [14].

Rattani *et al.* [13] proposed a fusion framework that fused the match scores, quality and liveness measures, while also accounting for the sensor influence, using a Bayesian framework. Although the model was not further generalized to consider the influence of other latent variables, it provided a good insight into the advantage of graphical modelling. The results indicated that the performance of the proposed model in a multi-sensor scenario, was comparable to a fusion framework that was trained and tested using fingerprint images from the same sensor. As Rattani *et al.*'s model is based on modelling a specific factor (*i.e.* the *sensor* influence) on match scores, it is not further discussed in this manuscript.

3. Bayesian Belief Networks in Biometrics

Notation: Let the observation be $\mathbf{x} = [m, l_1, l_2, q_1, q_2]$ where $m \in \mathbb{R}$ is a fingerprint match score, $l_1 \in \mathbb{R}$ ($l_2 \in \mathbb{R}$) denotes liveness measure of the gallery sample (probe sample), and $q_1 \in \mathbb{R}$ ($q_2 \in \mathbb{R}$) is the quality value of the gallery sample (probe sample). Let $\mathbf{K} = \{\mathbf{G}, \mathbf{I}\}$ denote two possible outputs: genuine (two fingerprint samples are from the same finger) and impostor (two fingerprint samples are from different fingers). Note that **K** does not include any assumptions about whether the pair of matched samples are live or fake. Further, let S_1 and S_2 denote the liveness states of the gallery and probe samples, which can be either **L**ive or **S**poof, *i.e.*, $\mathbf{S}_i = \{\mathbf{L}, \mathbf{S}\}$ for $i \in \{1, 2\}$. Thus, the output of a fingerprint matcher working in conjunction with a spoof detector can result in 8 possible events $\{\mathbf{S}_1, \mathbf{S}_2, \mathbf{K}\}$: **LLG**, **LLI**, **LSG**, **LSI**, **SLG**, **SLI**, **SSG**, **SSI**.

The Bayesian Belief Network (BBN) is a useful graphical model to express the causal relationship between variables. The relationship is often depicted as a directional graph with nodes representing variables and arrows representing conditional probabilities [1]. In the context of biometrics, a conventional generative classifier attempts to model the match scores (m) conditioned on the ground truth of the image pair being compared (**K**), *i.e.*, $p(m|\mathbf{K})$. The BBN model representing this conventional classifier is denoted as $\mathbf{K} \rightarrow m$. This conventional classifier can be extended to include all eight events and can be effectively realized using likelihood ratio-based test statistics as in Eqn (1). This conventional classifier, referred to as BBN-M, is considered as one of the baseline classifiers in this work.

$$f^{llr} = \frac{p(\mathbf{LLG}|m)}{p(\sim \mathbf{LLG}|m)}. \quad (1)$$

As mentioned before, for spoof detection, the variables $\mathbf{S}_1$ and $\mathbf{S}_2$ represent the ground truth of the states of liveness of the two fingerprint samples. If $\mathbf{S}_i$ is a spoof, then it will likely result in a lower liveness value l_i. The BBN model representing the relationship between liveness state $\mathbf{S}_i$ and the liveness measure l_i is denoted as $\mathbf{K} \rightarrow m$ and $\mathbf{S}_i \rightarrow l_i$.

One advantage of BBN is to explicitly depict the dependence between predictor variables, such as the match scores and the quality measures from two fingerprint samples, by the prior knowledge or the causal understanding from a human perspective rather than just the data. Take the quality values of the gallery and probe sample (*i.e.*, q_1 and q_2) as an example. Firstly, the variable q_1 and q_2 are supposed to be independent (denoted as $q_1 \perp\!\!\!\perp q_2$), because two fingerprint samples can be arbitrary. Moreover, they can be assumed to influence the match score m (denoted as $q_1 \rightarrow m$ and $q_2 \rightarrow m$) from a causal understanding, but they are expected to be independent with the ground truth **K**. This is because the ground truth of two fingerprint samples being from the same finger or from two different fingers cannot be influenced by the quality measures of these samples [2] This advantage is further discussed regarding the calculation of likelihood ratio-based test statistics below.

3.1. Existing Bayesian Belief Networks

(a) **BBN-MQ**: Figure 3 (a) show the BBN model proposed in [9] that combined fingerprint match score (m) with the image quality (q_1 and q_2). The model is based on the following assumption:

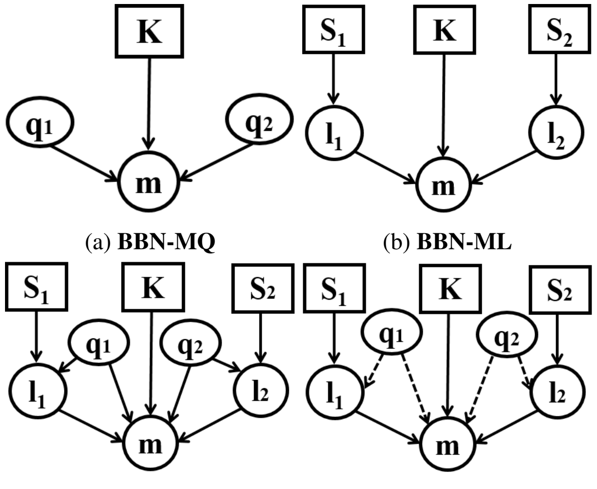
Assumption: *Quality measure of a sample influences the corresponding match score.*

The assumption is shown as $q_i \rightarrow m$ for $i \in \{1, 2\}$ in Figure 3 (a), and the joint density represented by the **BBN-MQ** model can be directly calculated as,

$$p(\mathbf{K}, q_1, q_2, m) = p(m|\mathbf{K}, q_1, q_2)p(q_1)p(q_2)p(\mathbf{K}). \quad (2)$$

Since this model does not consider spoof attacks, the conditional probability of $\mathbf{K} = \{\mathbf{G}, \mathbf{I}\}$ does not include the liveness states ($\mathbf{S}_1$ and $\mathbf{S}_2$). The final decision $\mathbf{K} = \{\mathbf{G}, \mathbf{I}\}$ is made based on the likelihood ratio-based test statistic (f^{llr})

[2] Of course, there could be cases where a person's fingerprint is consistently poor due to implicit skin issues.



(a) **BBN-MQ** (b) **BBN-ML**

(c) **BBN-MLQ** (d) **BBN-MLQc**

Figure 3: Several possible BBNs for fusing fingerprint match scores with liveness and quality measures. BBN-MQ and BBN-ML are based on previous literature, while BBN-MLQ and BBN-MLQc are the proposed ones.

as follows:

$$\begin{aligned} f^{llr} &= \frac{p(\mathbf{K}{=}\boldsymbol{G}|m,q_1,q_2)}{p(\mathbf{K}{=}\boldsymbol{I}|m,q_1,q_2)} = \frac{p(\mathbf{K}{=}\boldsymbol{G},m,q_1,q_2)}{p(\mathbf{K}{=}\boldsymbol{I},m,q_1,q_2)} \\ &\quad \textit{(based on Eqn (2) and since } \boldsymbol{K} \perp\!\!\!\perp q_1 \perp\!\!\!\perp q_2\textit{)} \\ &= \frac{p(\mathbf{K}{=}\boldsymbol{G})p(m,q_1,q_2|\mathbf{K}{=}\boldsymbol{G})}{p(\mathbf{K}{=}\boldsymbol{I})p(m,q_1,q_2|\mathbf{K}{=}\boldsymbol{I})}. \end{aligned} \tag{3}$$

Assuming the prior probability of $p(\mathbf{K} = \mathbf{G})$ and $p(\mathbf{K} = \mathbf{I})$ are equal, the above f^{llr} can be obtained by estimating the joint probability of $\{m, q_1, q_2\}$ given the target class **K**.

(b) **BBN-ML**: Figure 3 (b) show the BBN model proposed by Marasco *et al.* [7] for combining match scores (m) with the corresponding liveness measures (l_1 and l_2). BBN-ML is based on the following assumption:
Assumption: *Liveness measure of a sample influences the corresponding match score.*

This assumption is shown as two directional arrows (*i.e.*, $l_i \to m$) in Figure 3 (b). The joint densities represented by the **BBN-ML** model can be written as:

$$\begin{aligned} &p(\mathbf{K},\mathbf{S}_1,\mathbf{S}_2,m,l_1,l_2) \\ &= p(m|\mathbf{K},l_1,l_2)p(l_1|\mathbf{S}_1)p(l_2|\mathbf{S}_2)p(\mathbf{K})p(\mathbf{S}_1)p(\mathbf{S}_2). \end{aligned} \tag{4}$$

The final decision is made using the likelihood ratio based test statistic (f^{llr}) of the conditional probability of eight possible events (classes) given the match score (m) and liveness scores (l_1 and l_2). Taking the only acceptance case[3](**LLG**) as an example, the conditional probability given (m, l_1, l_2) can be estimated by:

$$\begin{aligned} p(\boldsymbol{LLG}|m,l_1,l_2) &\to \frac{p(\mathbf{K}{=}\boldsymbol{G},\mathbf{S}_1{=}\boldsymbol{L},\mathbf{S}_2{=}\boldsymbol{L},m,l_1,l_2)}{p(m,l_1,l_2)} \\ &\quad \textit{(from Eqn (4))} \\ &= \frac{p(m|\mathbf{K},l_1,l_2)p(l_1|\mathbf{S}_1)p(l_2|\mathbf{S}_2)p(\mathbf{K})p(\mathbf{S}_1)p(\mathbf{S}_2)}{p(m|l_1,l_2)p(l_1,l_2)} \\ &\quad \textit{(since } l_1 \perp\!\!\!\perp l_2\textit{)} \\ &= \frac{p(\mathbf{S}_1)p(l_1|\mathbf{S}_1)}{p(l_1)}\,\frac{p(\mathbf{S}_2)p(l_2|\mathbf{S}_2)}{p(l_2)}\,\frac{p(\mathbf{K})p(m|\mathbf{K},l_1,l_2)}{p(m|l_1,l_2)} \\ &\quad \textit{(since } \boldsymbol{K} \perp\!\!\!\perp l_1 \textit{ and } \boldsymbol{K} \perp\!\!\!\perp l_2\textit{)} \\ &\to p(\mathbf{S}_1{=}\boldsymbol{L}|l_1)\;p(\mathbf{S}_2{=}\boldsymbol{L}|l_2)\;p(\mathbf{K}{=}\boldsymbol{G}|m,l_1,l_2). \end{aligned} \tag{5}$$

It must be noted that the above mathematical derivation can simplify the calculation of the likelihood ratio (f^{llr}) between the classes **LLG** and $\sim$ **LLG**.

3.2. Proposed Bayesian Belief Networks

(a) **BBN-MLQ**: Figure 3 (c) shows one of the proposed BBN model that combines match scores with quality and liveness measure. This model is based on the following three assumptions:
Assumption 1: *Quality measure of a sample influences the corresponding match score,* i.e., $q_i \to m$
Assumption 2: *Liveness measure of a sample influences the corresponding match score,* i.e., $l_i \to m$
Assumption 3: *Quality measure of a sample influences the corresponding liveness measure,* i.e., $q_i \to l_i$.

The joint probabilities represented by **BBN-MLQ** are factorized as:

$$\begin{aligned} &p(\mathbf{K},\mathbf{S}_1,\mathbf{S}_2,m,l_1,l_2,q_1,q_2) \\ &\quad = p(m|\mathbf{K},l_1,l_2,q_1,q_2)p(l_1|\mathbf{S}_1,q_1)p(l_2|\mathbf{S}_2,q_2) \\ &\quad p(\mathbf{K})p(\mathbf{S}_1)p(\mathbf{S}_2)p(q_1)p(q_2). \end{aligned} \tag{6}$$

BBN-MLQ can be realized using the likelihood ratio based test statistic (f^{llr}) as follows:

$$\begin{aligned} f^{llr} &= \frac{p(\mathbf{LLG}|m,l_1,l_2,q_1,q_2)}{p(\sim \mathbf{LLG}|m,l_1,l_2,q_1,q_2)} \qquad (7) \\ &\quad \textit{(from Eqn (6) and since } \boldsymbol{K} \perp\!\!\!\perp \boldsymbol{S}_1,\boldsymbol{S}_2\textit{)} \\ &= \frac{p(m|\mathbf{K}{=}\boldsymbol{G},l_1,l_2,q_1,q_2)p(l_1|\mathbf{S}_1{=}\boldsymbol{L},q_1)p(l_2|\mathbf{S}_2{=}\boldsymbol{L},q_2)p(\boldsymbol{LLG})}{\sum_{\sim \boldsymbol{LLG}} p(m|\mathbf{K},l_1,l_2,q_1,q_2)p(l_1|\mathbf{S}_1,q_1)p(l_2|\mathbf{S}_2,q_2)p(\sim \boldsymbol{LLG})} \\ &\quad \textit{(since } \boldsymbol{K} \perp\!\!\!\perp l_1,l_2,q_1,q_2\textit{)} \\ &= \frac{p(m,l_1,l_2,q_1,q_2|\mathbf{K}{=}\boldsymbol{G})p(l_1,q_1|\mathbf{S}_1{=}\boldsymbol{L})p(l_2,q_2|\mathbf{S}_2{=}\boldsymbol{L})p(\boldsymbol{LLG})}{p(m,l_1,l_2,q_1,q_2|\mathbf{K})p(l_1,q_1|\mathbf{S}_1)p(l_2,q_2|\mathbf{S}_2)p(\sim \boldsymbol{LLG})}. \end{aligned}$$

The configuration of this BBN model can be considered as a direct extension of BBN-ML by adding quality measures as new predictor variables. Although the inference of the model is straightforward, the influence of latent factors has not been considered. As a result, we propose another

[3] The *acceptance case* indicates the event where the two samples are live and they originate from the same finger.

configuration of the BBN model to utilize the quality measures in a more effective way.

(b) **BBN-MLQc**: Figure 3 (d) shows another configuration of the BBN model. This model is based on the fact that a simpler BBN configuration with fewer assumptions is more likely to generalize over unseen data. This is because additional assumptions of causal relationships can lead to a more complex joint probability function (such as in Eqn (6)) which may be difficult to estimate and interpret. Therefore, this model incorporates quality measures into the existing **BBN-ML** model without making any additional assumptions, while the match scores and liveness measures are calibrated/normalized based on the quality measure. The model is referred to as **BBN-MLQc** in this work, and the assumption made in this model is as same as the one made in the **BBN-ML** model:
Assumption: *Liveness measure of a sample influences the corresponding match score.*

The conditional probability can be estimated in a manner similar to the **BBN-ML** model:

$$\begin{aligned}&p(\mathbf{K}{=}\boldsymbol{G}, \mathbf{S}_1{=}\boldsymbol{L}, \mathbf{S}_2{=}\boldsymbol{L}\ |m^{norm}, l_1^{norm}, l_2^{norm}) \qquad (8)\\ &= p(\mathbf{S}_1{=}\boldsymbol{L}|l_1^{norm})\ p(\mathbf{S}_2{=}\boldsymbol{L}|l_2^{norm})\ p(\mathbf{K}{=}\boldsymbol{G}|m^{norm}, l_1^{norm}, l_2^{norm})\end{aligned}$$

where m^{norm} and l_i^{norm} for $i \in \{1, 2\}$ are the quality-normalized match scores and liveness measures, respectively. The proposed quality-based calibration is based on the following observations:

1. Similar quality measures are likely to share a similar combination of factors, such as image resolution, noise level, clarity of ridges/valley structures, or fabrication materials used. Quality categorization can, therefore, capture these latent factors.

2. Certain liveness values may result in higher spoof detection accuracy than others. In such cases, the quality measure of the biometric samples can be ignored by the spoof detector. This suggests the use of a piecewise function to calibrate liveness values by the quality measure only over certain ranges.

The rationale behind the proposed **BBN-MLQc** model is to categorize the quality measure into discrete states, and then apply different calibration functions for each quality based on the liveness detection accuracy.

The categorization (or discretization) of continuous quality measures is achieved using the Minimum Optimal Description Length (MODL) algorithm based on the minimal description length (MDL) principle [2]. The class entropy of a set of quality measures q is defined as:

$$Ent(q) = -\sum_{i=1}^{Z} p(c_i, q) log(p(c_i, q))), \qquad (9)$$

where $p(c_i, q)$ is the proportion of samples lying in category c_i, and Z is the total number of categories. Suppose the first bin B_1 is added as a cut-off point and the set q is partitioned into subsets q_{c_1} and q_{c_2}, then the entropy of the partition is:

$$Ent(q, B_1) = \frac{|q_{c_1}|}{|q|} Ent(q_{c_1}) + \frac{|q_{c_2}|}{|q|} Ent(q_{c_2}), \qquad (10)$$

where $|q|$ denotes the number of samples in the set q. There could be $Z - 1$ bins. The original MODL algorithm in [2] scores all possible categorization possibilities and selects the one with the lowest entropy, and is also employed to decide the number of categories Z in this work.

The quality categorization is followed by an exploration of optimal calibration functions for liveness values. There are multiple ways to transform the liveness values using quality. In this work, the basic Fisher's linear discriminant analysis (LDA) is employed. The calibrated liveness values can be considered as a linear combination of variables (l, q).

$$l_i^{norm} = \begin{cases} l_i & i \in c_1, \\ f_{LDA}^{c_i}(l_i, q_i) & i \in c_{\{2,...Z\}}. \end{cases} \qquad (11)$$

Basically, Eqn (11) indicates that if the samples lie in the quality state c_1 - corresponding to the quality state obtaining the highest liveness detection accuracy - the liveness measures do not need to be calibrated by any image quality. However, if the samples lie in other quality states, the liveness values are calibrated using $f_{LDA}^{c_i}$, where c_i denotes the corresponding quality state. The output classes used for training the LDA functions are **L**ive or **S**poof, *i.e.*, $\mathbf{S}_i = \{\mathbf{L}, \mathbf{S}\}$ for $i \in \{1, 2\}$.

It should be noted that the above quality-based calibration is non-linear with respect to the liveness values, and the estimation of the joint probability function represented by the proposed BBN is greatly simplified by the calibration process.

4. Experiments and Results

Experimental analysis is conducted on the LivDet 2011 [19] database. It consists of 1,000 live and 1,000 fake fingerprint samples in the training set, and the same number of samples from different subjects in the test set.

The spoof artifacts in the LivDet 2011 database are fabricated using five materials, viz., gelatine, silicone, woodglue, ecoflex, and latex. For each material, 200 fingerprints were fabricated from 20 fingers using the consensual method (*i.e.*, with the consent and collaboration of the user). Both live fingers and spoof artifacts were obtained using *four* different sensors, *i.e.*, Biometrika, Italdata, Sagem and DigitalPersona. Because of space constraints, only the results from the Biometrika and Italdata sensor are presented in this paper.

Table 1: Spoof detection performance of the various BBN frameworks on the LivDet 2011 database.

Various Frameworks	Biometrika		Italdata		Digital		Sagem	
	1-FerrLive at 1% FerrFake	1-FerrLive at 10% FerrFake	1-FerrLive at 1% FerrFake	1-FerrLive at 10% FerrFake	1-FerrLive at 1% FerrFake	1-FerrLive at 10% FerrFake	1-FerrLive at 1% FerrFake	1-FerrLive at 10% FerrFake
BBN-MLQc	70.1	91.1	52.6	84.8	81.2	95.8	85.6	97.2
BBN-MLQ	62.3	91.1	49.8	83.2	77.1	95.8	84.1	97.2
BBN-ML	45.3	80.3	34.1	67.8	67.1	91.4	76.6	92.5
DM-GMM	61.7	91.0	46.0	82.1	75.3	93.3	83.0	95.6
Spoof Detector	42.0	80.0	22.9	66.9	61.9	88.0	72.1	92.5

The VeriFinger SDK[4] is used to generate match scores by matching all pairs of images within and across all subjects for live and spoof impressions. The quality of live and spoof impressions was obtained using the IQF freeware developed by MITRE[5]. The quality measure ranges between 0 and 100, with 0 being the lowest and 100 being the highest quality. Finally, fingerprint liveness was assessed using the recently proposed spoof detection algorithm based on local binary patterns (LBP) [11]. A two class Support Vector Machine (SVM) (implemented using LIBSVM package) was trained using LBP features extracted from live and fake images in the training set. The output score (probability estimate) of the SVM was then used as a liveness measure. The LBP-SVM spoof detector provides a better spoof detection accuracy over existing techniques as reported in [10].

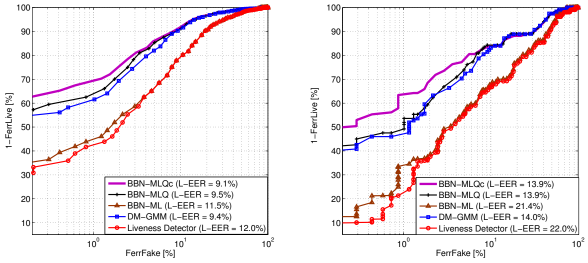
The evaluation of the various BBN frameworks is conducted in terms of the spoof detection accuracy and overall performance.

Experiment 1. Performance Against Spoof Attacks

This experiment evaluates the performance of the proposed frameworks against spoof attacks. The **FerrLive** denotes the proportion of live samples that are incorrectly classified as being spoof, while the **FerrFake** denotes the proportion of spoof samples that are incorrectly classified as being live. Further, EER of the spoof detection (indicated as L-EER) is the rate at which FerrLive is equal to FerrFake. Table 1 shows the spoof detection performance of the BBN-MLQc, BBN-MLQ, BBN-ML, BBN-MQ and GMM-based direct modelling scheme (DM-GMM) on four sensors.

It can be seen that the proposed BBN-MLQ and BBN-MLQc obtained better spoof detection performance in comparison to the existing frameworks and the baseline LBP-based spoof detector. This is due to appropriate modelling of quality with the liveness measure.

- The FerrLive error (at 1% FerrFake rate) of the BBN-MLQc reduced by 28.0%, 41.0%, 19.3% and 13.5% over the baseline LBP-based liveness detection scheme for the Biometrika, Italdata, DigitalPersona and Sagem sensors, respectively. It demonstrates the advantage of incorporating the quality of images for spoof detection.



(a) **Biometrika** (b) **Italdata**

Figure 4: Liveness detection performance of the various BBN frameworks on the LivDet 2011 database. Note that the liveness detection accuracy of these frameworks is *not* the same as that of the LBP-based liveness detection algorithm used. This is because the interaction of liveness measures with match score and quality is taken into account when rendering the final decision.

- The spoof detection accuracy of the BBN-MLQc is significantly better than the DM-GMM direct modelling scheme. However, the BBN-MLQ is only slightly better than DM-GMM. It indicates that the benefits of the graphical modelling algorithm depends on the *configuration* of networks.

- Furthermore, the FerrLive error (at 1% FerrFake rate) of the BBN-MLQc model reduced by 24.8%, 19.8%, 14.1% and 9.0% over the existing BBN-ML model, although both of them share the same causal assumptions. It demonstrates the benefits of utilizing the proposed quality-based calibration scheme.

The BBN-M and BBN-MQ do not incorporate the liveness measure, hence, they are not evaluated in this experiment. The ROC curves of spoof detection performance on data from the Biometrika and Italdata sensors are shown in Figure 4. The above observations are consistent across all the four sensors.

Experiment 2. Overall Performance

This experiment evaluates the overall performance of the proposed BBN-MLQc and BBN-MLQ frameworks under

[4] http://www.neurotechnology.com/vf_sdk.html

[5] http://www.mitre.org/tech/mtf/

Table 2: Performance of the various frameworks when all eight events are considered for the Biometrika and Italdata sensors. BBN-MLQc is seen to outperform all other frameworks.

Various	Biometrika		Italdata		Digital		Sagem	
Frameworks	GAR [%] at OFAR = 1%	GAR [%] at OFAR = 5%	GAR [%] at OFAR = 1%	GAR [%] at OFAR = 5%	GAR [%] at OFAR = 1%	GAR [%] at OFAR = 5%	GAR [%] at OFAR = 1%	GAR [%] at OFAR = 5%
BBN-MLQc	80.5	88.3	72.5	89.0	84.8	88.6	75.3	88.3
BBN-MLQ	75.6	85.2	72.1	88.6	83.0	87.6	72.3	87.2
BBN-ML	74.2	86.7	72.5	88.7	83.0	87.7	73.3	87.4
BBN-MQ	77.9	85.5	72.0	88.2	77.5	87.1	68.1	84.4
BBN-M	76.7	85.1	71.8	88.1	77.5	86.9	68.1	83.5
DM-GMM	79.8	87.1	72.5	89.0	82.9	87.5	72.5	87.4

all possible spoof attack scenarios. Comparative assessment is made with the existing Bayesian Networks and GMM-based direct modelling scheme (DM-GMM).

As the fingerprint verification system operates under both zero-effort impostor and spoof attacks, the overall performance rates can be defined as follows:

- **Genuine Acceptance Rate (GAR)**: Proportion of the **LLG** class that are incorrectly classified as genuine and accepted by the system.

- **Overall False Acceptance Rate (OFAR)**: Proportion of zero-effort impostor and spoof samples that are incorrectly classified as the **LLG** class.

- **Overall Equal Error Rate (O-EER)**: The rate at which OFAR equals 1 minus the Genuine Acceptance Rate (GAR). The **O-EER** of each fusion scheme is shown in the ROC curves.

Table 2 demonstrates that BBN-MLQc performs much better than all the existing frameworks and the baseline BBN-M. This is due to its high spoof detection capability and better performance under spoof attacks (see Experiment 1). The ROC curves of each fusion scheme are shown in Figure 5.

- At a fixed 1% OFAR, the GAR of the BBN-MLQc increased by 17.0%, 5.77%, 9.49% and 6.66% (range [5.77%, 17.0%]) over the BBN-M for the Biometrika, Italdata, Sagem and DigitalPersona sensors, respectively. For instance, the GAR of the BBN-MLQc is 95.5% whereas GAR of the BBN-M is 81.7% at a 1% OFAR for the Biometrika sensor.

- At a fixed 1% OFAR, the GAR of the BBN-MLQ increased by 16.5%, 5.13%, 9.03% and 6.02% (range [16.5%,5.13%]) over the BBN-M for all the four sensors, respectively. The GARs of the BBN-ML increased in the range [13.7%,5.17%], and are similar to the GARs of the DM-GMM that increased in the range [13.7%,5.14%]). Further, BBN-MQ performed just a little better than BBN-M by 1.47%, 0.22%, 0.14% and 0.13% (range [1.47%,0.13%]), respectively.

5. Summary

In this work, we proposed two Bayesian Belief Network (BBN) models that can effectively integrate liveness measures with quality measures and match score. The proposed BBN models have two different configurations distinguished on the basis of how the quality measures are incorporated. This study also compares the proposed BBN models with existing fusion frameworks against spoof attacks. Comprehensive experiments are conducted on the LivDet 2011 dataset. Results indicate that the proposed BBN-MLQ and BBN-MLQc methods consistently outperform existing fusion frameworks. Based on the experiments, the following conclusions can be drawn:

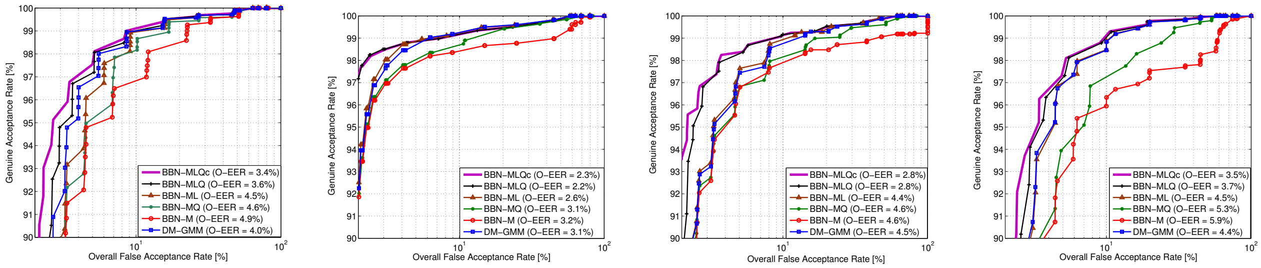
- **Causal relationship:** Fusion frameworks that model the appropriate relationship between the considered variables, such as the influence of the quality on liveness measure, obtain better performance.

- **Benefits of quality:** Incorporating image quality is beneficial in the fusion framework (BBN-MLQ and BBN-MLQc). This is because quality measures can take into account the material-specific characteristics of spoof fabrication materials. Further, the models incorporating quality also have benefits (better performance) when evaluated on novel spoof fabrication materials [14].

- **The role of quality:** These quality measures can be incorporated as features (as in BBN-MLQ) or used as a normalization parameter (as in BBN-MLQc). Experimental results suggest the efficacy of quality when used as a normalization parameter rather than a feature, since the latter makes the Bayesian Belief Network more complicated to be interpreted and calculated.

References

[1] C. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.



(a) **Biometrika** (b) **Italdata** (c) **Digital** (d) **Sagem**

Figure 5: Performance of the various frameworks when all eight events are considered for all four sensors. It can be seen that BBN-MLQc outperforms all other frameworks.

[2] M. Boulle. MODL: A Bayes optimal discretization method for continuous attributes. *Machine Learning*, 65(1):131–165, May 2006.

[3] I. Chingovska, A. Anjos, and S. Marcel. Anti-spoofing in action: joint operation with a verification system. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–8, USA, 2013.

[4] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):531–543, Apr. 2007.

[5] A. K. Jain, P. Flynn, and A. Ross. *Handbook of Biometrics*. Springer, 2007.

[6] D. Maltoni, D. Maio, A. Jain, and S. Prabhakar. *Handbook of Fingerprint Recognition*. Springer, 2009.

[7] E. Marasco, Y. Ding, and A. Ross. Combining match scores with liveness values in a fingerprint verification system. *Proc. of IEEE Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012.

[8] E. Marasco and A. Ross. A survey on anti-spoofing schemes for fingerprint recognition systems. *ACM Computing Surveys*, 47(2):1–36, 2014.

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

[10] V. Mura, L. Ghiani, G. L. Marcialis, F. Roli, D. A. Yambay, and S. A. Schuckers. LivDet 2015 - fingerprint liveness detection competition. In *Proc. of IEEE International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–6, 2015.

[11] S. Nikam and S. Aggarwal. Wavelet energy signature and glcm features-based fingerprint anti-spoofing. In *Proc. of IEEE Int. Conf. On Wavelet Analysis and Pattern Recognition*, pages 717–723, Hong Kong, China, 2008.

[12] A. Rattani and N. Poh. Biometric system design under zero and non-zero effort attacks. In *Proc. of IEEE Intl. Conf. on Biometrics*, pages 1–8, Madrid, Spain, 2013.

[13] A. Rattani, N. Poh, and A. Ross. A bayesian approach for modeling sensor influence on quality, liveness and match score values in fingerprint verification. In *Proc. of IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 37–42. IEEE, 2013.

[14] A. Rattani, W. J. Scheirer, and A. Ross. Open set fingerprint spoof detection across novel fabrication materials. *IEEE Transactions on Information Forensics and Security*, 10(11):2447–2460, 2015.

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

[16] C. Sousedik and C. Busch. Presentation attack detection methods for fingerprint recognition systems: a survey. *IET Biometrics*, 2014.

[17] B. Toth. Introduction to Biometric Liveness Detection. *Information Security*, 10:291–298, October 2005.

[18] L. Wein and M. Baveja. Using fingerprint image quality to improve the identification performance of the U.S. VISIT program. In *Proc. of the National Academy of Sciences*, volume 102, pages 7772–7775, 2005.

[19] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. LivDet 2011 - fingerprint liveness detection competition. In *Proc. of IEEE Intl. Conf. on Biometrics*, pages 208–215, Delhi, India, 2012.