Appeared in Proc. of IEEE International Workshop on Information Forensics and Security (WIFS), (Guangzhou, China), November 2013

# A Bayesian Approach for Modeling Sensor Influence on Quality, Liveness and Match Score Values in Fingerprint Verification

Ajita Rattani [#1], Norman Poh [*2], Arun Ross [#3]

[#] *Dept. of Computer Science and Engineering, Michigan State University, USA*
[1] ajita@msu.edu
[3] rossarun@msu.edu

[*] *Dept. of Computing, University of Surrey, UK*
[2] normanpoh@ieee.org

***Abstract*—Recently a number of studies in fingerprint verification have combined match scores with quality and liveness measures in order to thwart spoof attacks. However, these approaches do not explicitly account for the influence of the sensor on these variables. In this work, we propose a graphical model that accounts for the impact of the sensor on match scores, quality and liveness measures. The proposed graphical model is implemented using a Gaussian Mixture Model based Bayesian classifier. Effectiveness of the proposed model has been assessed on the LivDet11 fingerprint database using Biometrika and Italdata sensors.**

## I. Introduction

Recent research has highlighted the vulnerability of biometric systems to spoof attacks. A spoof attack occurs when an adversary mimics the biometric trait of another individual in order to circumvent the system. For instance, it has been shown that a person can fool a fingerprint system by using a finger-like object made of gelatin or play-doh that has the fingerprint ridges of another individual impressed on it [1].

In the context of fingerprints, liveness detection algorithms have been proposed as a counter-measure against spoof attacks. These algorithms attempt to discriminate live biometric samples from spoof (fake) artefacts by examining the textural, anatomical and/or physiological attributes of the finger [2], [3]. The output of these liveness detection algorithms is a single-valued numerical entity referred to as *liveness measure*.

Liveness detection algorithms are not designed to operate in isolation; rather, they have to be integrated with the overall fingerprint recognition system. Accordingly, recent studies have combined match scores generated by a fingerprint matcher with liveness values [4], [5] as well as image quality value [6], in order to render a decision on the recognition process. Typically, a learning-based scheme is used in such a fusion framework [7]. For example, in [4] the authors combine fingerprint match scores with liveness measures using a Bayesian Belief Network. In [6], fingerprint match scores are combined with quality and liveness measures using a density-based fusion framework. The work in [6] also established the benefits of incorporating *both* image quality and liveness measures in the fusion framework. In [5], face match scores are combined with liveness measures using logistic regression. However, in the aforementioned schemes, the influence of the sensor on the 3 variables - match scores, liveness values and quality - has not been considered. Such a consideration is essential for several reasons: (a) the quality of an image is impacted by the sensor used; (b) most liveness measures are learning-based and are impacted by the sensor that was used to collect live and spoof training data; (c) understanding sensor influence, can help in facilitating sensor interoperability [8] for fingerprint matchers and liveness detectors.

In order to address this issue, we propose a fusion framework based on graphical models where the influence of the sensor on match scores, liveness measures and quality values is accounted for. The proposed graphical model is based on the assumption that data from a set of fingerprint sensors are available during the training stage. However, the actual sensor identity is not known during the testing stage[1].

The contributions of this work are as follows: (1) development of a graphical model for fusing match scores, liveness measures and quality values while accounting for sensor influence; (2) implementation of the proposed model using a Gaussian Mixture Model (GMM) based Bayesian classifier in designing a fingerprint verification system that is robust to zero-effort impostors as well as non-zero-effort spoof attacks; and (3) evaluation of the proposed model using fingerprint data from two different sensors in the LivDet 2011 fingerprint database.

This paper is organized as follows: Section 2 presents the proposed graphical model. Section 3 discusses the database and experimental protocol used in this work. Experimental results are reported and discussed in section 4. Conclusions are drawn in section 5.

[1]In principal, data from the sensor used during testing does not have to be available during training.

*WIFS'2013, November 18-21, 2013, Guangzhou, China.*
*ISBN 978-1-4673-5593-3* 

Appeared in Proc. of IEEE International Workshop on Information Forensics and Security (WIFS), (Guangzhou, China), November 2013

## II. The Proposed Graphical Model and its Implementation using GMM

A graphical model is an effective tool to express the relationship between different variables [9], [10]. This is often depicted as a directional graph with nodes representing probabilities of a variable and arrows representing conditional probabilities (e.g., an edge from $A$ to $B$ defines $P(B|A)$).

We formulate the spoof-resilient fingerprint verification problem as follows. An input fingerprint sample has to be compared against a live [2] template fingerprint sample. Liveness measures and quality values are extracted from both samples. Further, a match score is computed between the two samples. Let $y \in \mathbb{R}$ be the match score, $q_t \in \mathbb{R}$ ($q_i \in \mathbb{R}$) represent the quality of the template (input) fingerprint sample ($q = [q_t, q_i]$) and $l_t \in \mathbb{R}$ ($l_i \in \mathbb{R}$) denote liveness measure of the template (input) fingerprint sample ($l = [l_t, l_i]$). The output is one of three classes: the genuine class, $G$, when the input is deemed to be a live sample with the same identity as that of the template; the impostor class, $I$, when the input is deemed to be a live sample whose identity is *not* the same as that of the template; the spoof class, $S$, when the input is not deemed to be a live sample. Let $k \in \{G, I, S\}$. Further, let $d$ signify one of $N$ sensors used for training dataset acquisition. While the set of sensors used for training dataset acquisition is known, the actual sensor used to acquire biometric data during system deployment (or testing) does not have to be known.

Table I shows three graphical models and the conditional probabilities between match score $(y)$, quality $(q)$, liveness measure $(l)$ and sensor information $(d)$, conditioned on the class label $(k)$ to represent a) a conventional classifier (Model $A$); b) a fusion framework based on [6] that combines $l$, $q$ and $y$ (Model $B$); and c) the proposed classifier that incorporates the influence of the sensor on $l$, $q$ and $y$ (Model $C$). Next, we will explain these graphical models and how they can be realized through a GMM-based Bayesian classifier [7] .

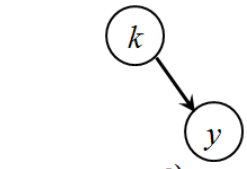
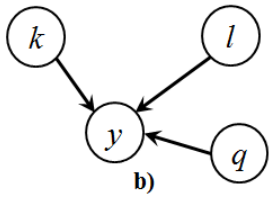
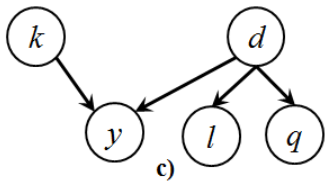
| Graphical Model | Conditional Probabilities |
|---|---|
|  a) | • $k \rightarrow y$ ; $p(y\|k)$ |
|  b) | • $q \rightarrow y$ ; $p(y\|q)$<br>• $l \rightarrow y$ ; $p(y\|l)$<br>• $k \rightarrow y$ ; $p(y\|k)$<br>$= p(y\|k, q, l)$ |
|  c) | • $d \rightarrow y$; $p(y\|d)$<br>• $k \rightarrow y$; $p(y\|k)$<br>$= p(y\|k, d)$<br>• $d \rightarrow q$; $p(q\|d)$<br>• $d \rightarrow l$; $p(l\|d)$<br>$= p(q, l\|d)$ |

TABLE I: Three graphical models that describe the relationship between match scores $(y)$, quality $(q)$, liveness measures $(l)$ and sensor information $(d)$, conditioned on the class label $(k)$.

**a) Conventional classifier:** Model $A$ in Table I represents the conventional generative classifier that attempts to model the score $(y)$ conditioned on class label ($k \in \{G, I\}$), i.e., $p(y|k)$. A conventional classifier assumes the attacker is simply a zero-effort impostor ($k = I$) and does not consider the possibility of concerted spoof attacks. It can be implemented using the log-likelihood ratio based test statistic as follows:

$$y_a^{llr} = \log \frac{p(y|k=G)}{p(y|k=I)}. \tag{1}$$

**b) Fusion framework against spoof attacks:** Model $B$ in Table I models the joint density of match scores, quality and liveness measures conditioned on class $k$. This model is based on the framework mentioned in [6] for fingerprint verification against zero-effort impostor and spoof attacks. The joint distribution represented by model $B$ is in the following form:

$$p(y, k, q, l) = p(y|k, q, l)p(q)p(l)P(k) \tag{2}$$

Model $B$ can be realized using the log-likelihood ratio based test statistic as follows:

$$\begin{aligned} y_b^{llr} &= \log \frac{p(y, q, l|k=G)}{p(y, q, l|k \neq G)}, \\ &= \log \tfrac{p(y|k=G,q,l)}{p(y|k\neq G,q,l)} + \underbrace{\log \frac{p(q|k=G)}{p(q|k \neq G)}} + \underbrace{\log \frac{p(l|k=G)}{p(l|k \neq G)}}. \end{aligned} \tag{3}$$

where $(k \neq G) \in \{I, S\}$. The underbraced terms in (3) will be zero as quality $(q)$ and liveness $(l)$ measures are assumed to have no discriminatory information for distinguishing between the genuine and impostor classes. Therefore, the log-ratio for these terms will be zero. Further, $p(y|k, q, l)$ cannot be directly estimated using off-the-shelf algorithms. This is because the conditioning variables $q$ and $l$ in (3) are continuous. Alternatively, model $B$ can be effectively realized by the joint density estimate of $y$, $q$, $l$ for class $k$ i.e., $p(y, q, l|k)$. This method was reported to be more effective than model $A$ (the baseline)

[2] In this work, the template is assumed to be that of a live fingerprint sample. Its liveness value is nevertheless computed.

under zero-effort impostors as well as spoof attacks in [6], as will also be confirmed by our experiments here.

**c) Proposed model:** Model $C$ in Table I is an extension of model $B$ and models the influence of the sensor. This model attempts to model the dependency of score $(y)$ on class label $(k \in \{G, I, S\})$ as well as the sensor $(d)$. Further, the quality $(q)$ and liveness measures $(l)$ are categorized according to the sensor $(d)$ used in the training dataset.

The joint densities represented by model $C$ are as follows:

$$p(y, k, q, l, d) \quad = \quad p(y|k, d)p(q, l|d)P(d)P(k) \qquad (4)$$

Model $C$ can be effectively realized by extending (3) as:

$$y_c^{llr} \quad = \quad \log \frac{\sum_d p(y, q, l|k = G, d)P(d|q)}{\sum_d p(y, q, l|k \neq G, d)P(d|q)}. \qquad (5)$$

$P(d|q)$ in (5) can be estimated using the Bayes rules as:

$$P(d|q) = \frac{p(q|d)P(d)}{p(q)}. \qquad (6)$$

There is an integration over the sensor $(d)$ in (5) because sensor $(d)$ to be used during the testing (or system deployment) is not known in advance and its probability is inferred from the quality $(q)$ of the operational data i.e., $P(d|q)$ in (5).

Model $C$ is based on a conjecture that match scores, quality and liveness measures are sensor dependent ($d \rightarrow \{y, q, l\}$ in model $C$, Table I(c)). Further, their exists no significant correlation between quality $(q)$ and liveness $(l)$ measures (absence of an arrow between $q$ and $l$ in model $C$). These conjectures will be validated by empirical evidence (see Section IV).

The densities $p(y, q, l|k)$ in (3), $p(y, q, l|k, d)$ in (5) and $p(q|d)$ in (6) to estimate $P(d|q)$ in (5) are themselves estimated using the Gaussian Mixture Model (GMM). GMM has been successfully used to estimate joint densities [11]. Let $\phi^N(\mathbf{x}, \mu, \mathbf{\Sigma})$ be the $N$-variate gaussian density with mean vector $\mu$ and covariance matrix $\mathbf{\Sigma}$, i.e.,

$$\phi^N(\mathbf{x}, \mu, \mathbf{\Sigma}) = (2\pi)^{-N/2}|\mathbf{\Sigma}|^{-1/2}\exp(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)) \qquad (7)$$

The estimates of $p(\mathbf{x}|k)$ (where $\mathbf{x}$ is an observation vector which is $(y, q, l)$ in our case) for class $k$ is obtained as a mixture of Gaussians as:

$$p(\mathbf{x}|k) = \sum_{j=1}^{M_k} w_{k,j}\phi^N(\mathbf{x}, \mu_{\mathbf{k,j}}, \mathbf{\Sigma}_{\mathbf{k,j}}) \qquad (8)$$

where $M_k$ is the number of mixture components used to model the densities of class $k$. $w_{k,j}$ is the weight assigned to the $j^{th}$ mixture component in $p(\mathbf{x}|k)$, $\sum_{j=1}^{M_k} w_{k,j} = 1$. Selection of the appropriate number of components is one of the most challenging issues in mixture density estimation. The GMM fitting algorithm proposed in [11] automatically estimates the appropriate number of components and the component parameters using an EM algorithm and the minimum message length criterion. Hence, the GMM fitting algorithm proposed in [11] was used in this study[3].

## III. Database, Protocol and Performance Metrics

The LivDet11 dataset was used to evaluate fingerprint liveness detection algorithms submitted to the Second International Competition on Fingerprint Liveness Detection (LivDet11) [12]. It consists of 1000 live and 1000 fake fingerprint images in the training set and the same number of images in the test set. All images collected using the Biometrika and Italdata sensors were used in this study[4]. The live images were obtained from 200 different fingers with 5 samples per finger for each set. The fake fingerprints were fabricated using the following five materials: gelatine, silicone, woodglue, ecoflex and latex. 200 fake fingerprints were fabricated per material ($200 \times 5 = 1000$) from 20 fingers with ten samples per finger for each set (training and testing).

The NIST Bozorth3[5] software was used for obtaining a match score between a pair of fingerprint images. The quality of live as well as fake fingerprint impressions was measured using the IQF freeware developed by MITRE[6]. The quality factor ranges between 0 and 100, with 0 being the lowest and 100 being the highest quality. Finally, fingerprint liveness was assessed using the liveness measure proposed by Nikam and Aggarwal [3], which is based on Local Binary Pattern (LBP) features. A two class Support Vector Machine (SVM) (implemented using LIBSVM package[7]) was trained using LBP features extracted from live and fake images in the training set. The output score (probability estimate) of SVM was then used as a liveness measure. Equal Error Rate (EER) using this liveness measure was evaluated to be 10.95% and 18.95% on the test partition of the LivDet11 database for Biometrika and Italdata sensors, respectively.

**Protocol and Performance metrics:** Following the LivDet2011 protocol described in [12], we used 1000 live and 1000 fake images to train the models and the remaining 1000 live and 1000 fake images were used to evaluate the performance of the models.

*a) Learning-based fusion framework*: The observation vector consists of a match score $(y)$ and a pair of quality values $(q)$ as well as liveness measures $(l)$ extracted from a pair of training images - the input and the template. This observation vector is mapped to one of three output classes: $G$, $I$, or $S$ (4000 observation vectors were used for each class). This information, i.e., $(y, l, q, k)$ is used to train the GMM-based Bayesian classifiers in model $B$ (3) and model $C$ (5).

*b) Performance assessment metric:* In a spoof-resilient fingerprint verification system, as defined in this work, the $G$

[3]We used the MATLAB code available at http://www.lx.it.pt/~mtf/mixturecode.zip

[4]Data from other sensors in the LivDet11 dataset could not be used due to few correspondences in the subject identity between live and fake images.

[5]http://www.nist.gov/itl/iad/ig/nbis.cfm

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

[7]http://www.csie.ntu.edu.tw/~cjlin/libsvm/