Improving Biometric Identification Through Quality-based Face and Fingerprint Biometric Fusion*

Yan Tong Frederick W. Wheeler Xiaoming Liu
Visualization and Computer Vision Lab
GE Global Research
Niskayuna, NY
tongyan,wheeler,liux@research.ge.com

Abstract

Multi-modal biometric fusion is more accurate and reliable compared to recognition using a single biometric modality. However, most existing fusion approaches neglect the influence of the qualities of the biometric samples in information fusion. Our goal is to advance the state-of-the-art in biometric fusion technology by providing a more universal and more accurate solution for personal identification and verification with predictive quality metrics.

In this work, we developed score-level multi-modal fusion algorithms based on predictive quality metrics and employed them for the task of face and fingerprint biometric fusion. The causal relationships in the context of the fusion scenario are modeled by Bayesian Networks. The recognition/verification decision is then made through probabilistic inference. Our experiments demonstrated that the proposed score-level fusion algorithms significantly improve the verification performance over the methods based on the raw match score of a single modality (face or fingerprint). Furthermore, the fusion framework with both face and fingerprint image qualities achieves the best verification performance and outperforms all other baseline fusion algorithms tested including other straightforward quality-based fusion methods.

1. Introduction

Biometric identification has the potential of becoming an increasingly powerful tool for public safety. However it is challenged by the imperfect nature of the image data, intra-subject variation, inter-subject similarities, and subject-dependent characteristics [14], especially when the biometric data is from a single source. These issues can be overcome by biometric fusion, where multiple biometric modalities and/or multiple biometric samples are combined to improve performance by making the best use of all available biometric technologies and devices.

Generally speaking, there are three major approaches to biometric fusion: *multi-modal*, *multi-sample*, and *multi-algorithm* [2]. *Multi-modal* biometric fusion, where the recognition is performed on multiple biometric samples acquired from different biometric sources of a subject (e.g., face and fingerprint) or from different sensor types (e.g., optical sensor and thermal sensor) [2], has received increasing interest and is demonstrated to be more accurate and reliable compared to recognition performed on a single biometric modality. In addition, it will "make 'spoofing' more difficult" [2]. Furthermore, a biometric system often consists of three levels: *data/feature extraction*, *match score*, and *decision making*; and fusion can occur at any level [14].

A number of projects have looked at the fusion of multiple biometric modalities. The modalities studied include fingerprint, face, iris, palmprint, hand geometry, ear, voice, and 3D face. A detailed overview on multi-modal biometric fusion approaches can be found in [14] and [2]. However, current biometric fusion technologies often suffer in operational environments where the quality of biometric samples varies significantly. For example, the acquisition condition affects the fingerprint image quality [16], while intrapersonal variations and imaging conditions (e.g., face pose and facial expression) affect the facial image quality significantly. Recently, biometric quality metrics have been developed and utilized to improve the performance of biometric verification [12] and influence the multi-modal biometric fusion [4, 5, 11, 9, 13].

Our approach is differentiated from most existing work

*This project was supported by award #2007-MU-CX-K001 awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the Department of Justice.

The research in this paper made use of the CAS-PEAL-R1 face database collected under the sponsorship of the Chinese National Hi-Tech Program and ISVISION Tech. Co. Ltd.

by the *explicit* incorporation of predictive quality metrics in the fusion process. To do this, we develop a unified probabilistic framework to introduce and model the relationships among multiple modalities along with their relationships with their quality assessments. The quality-based biometric fusion is performed by probabilistic inference through the framework. Specifically, we propose and investigate two quality-based *multi-modal* fusion algorithms at the *match score* level and demonstrate them with an example of fusing face and fingerprint image samples: 1) fusion with facial image quality and 2) fusion with both face and fingerprint image qualities. Although Maurer and Baker [9] also employed a probabilistic network to perform quality-based multi-modal biometric fusion, they only consider the quality of probe samples by assuming that the gallery samples have uniformly high quality. In contrast, we assume that both the probe and gallery sample qualities may impact matcher performance, so the proposed fusion frameworks account for both quality values. Furthermore, quality-related facial image features are explicitly modeled in our approach, allowing the inference of facial image quality as a hidden node, utilizing both the quality-related features and the match scores.

Experimental results demonstrate that the proposed score-level face and fingerprint biometric fusion algorithms improve the biometric verification performance significantly compared to methods based solely on raw match scores of face/fingerprint. Furthermore, we also compare the proposed algorithms with several other fusion algorithms including other straightforward quality-dependent multi-modal fusion algorithms and show that the proposed fusion algorithm with both face and fingerprint image qualities outperforms all the other algorithms for comparison.

2. Score-level Face and Fingerprint Biometric Fusion with Quality Metrics

2.1. Quality Metrics

Biometric quality assessment has been increasingly used in various aspects of biometric systems such as rejecting poor quality samples in the image enrollment process [1], incorporating quality assessment in the verification process [12], predicting algorithm performance [18, 7, 15], and multi-modal biometric fusion [4, 5, 11, 9, 13]. As described in [16, 7], the quality of a biometric sample can be defined as a scalar quantity; and an effective quality metric should be predictive of the performance of a biometric system such that high quality biometric samples would result in high recognition performance given a matching algorithm.

In this work, we employ the NIST Fingerprint Image Quality (NFIQ) derived by Tabassi et al. [16], where 5 discrete quality levels are defined for the fingerprint images based on a normalized match score. To assess the quality values for facial image samples, we define the facial image quality metric as a symmetric normalized match score following the work by Ozay et al. [12] and further discretize the continuous quality values into 3 discrete levels. These quality metrics are designed to be predictive of the performance of matching algorithms.

Table 1. Nomenclature

$\mathbf{f}_{g,a}$	A quality-related image feature vector of a gallery facial image
$\mathbf{f}_{p,a}$	A quality-related image feature vector of a probe facial image
$q_{g,a}$	Image quality for a gallery facial image.
$q_{p,a}$	Image quality for a probe facial image.
$q_{g,i}$	Image quality for a gallery fingerprint image.
$q_{p,i}$	Image quality for a probe fingerprint image.
s_a	Match score for a gallery-probe facial image pair
s_i	Match score for a gallery-probe fingerprint image pair
$match$	Whether the gallery and probe samples belong to same subject for both fingerprint and face

2.2. Causal Relationships in Score-level Quality-based Face and Fingerprint Fusion

There are several key elements involved in a quality-based face and fingerprint fusion process including *quality-related image features* of facial images, *image quality* for both probe and gallery facial images, *image quality* for both probe and gallery fingerprint images, the *match scores* for facial and fingerprint images, and the fact of *match/no-match* of the probe and gallery images[1]. The nomenclature and terminology that will be used hereafter are listed in Table 1. We use subscript *a* for face and *i* for fingerprint.

From a Bayesian inference viewpoint, for quality-based score-level face and fingerprint biometric fusion, there are three causal relationships among the nine elements defined in Table 1. First, by assuming that the facial image quality is affected by the quality-related facial image features, and thus can be directly derived from the image features, $\mathbf{f}_{g,a}$ and $\mathbf{f}_{p,a}$ can be regarded as the causes to generate $q_{g,a}$ and $q_{p,a}$, respectively. Second, given a face recognition engine, the match score of a gallery-probe facial image pair s_a is affected by the image qualities of the gallery and probe facial images ($q_{g,a}$ and $q_{p,a}$) and the state of $match$ (*match/no-match*). Finally, given a fingerprint recognition algorithm, the match score of a gallery-probe fingerprint image pair s_i is affected by the image qualities of the gallery and probe fingerprint images ($q_{g,i}$ and $q_{p,i}$) and the state of $match$ (*match/no-match*).

These causal relationships can be represented by a graphical model, as shown in Fig. 1. Specifically, we propose to use a Bayesian Network (BN) to model and learn such relationships. A BN is a Directed Acyclic Graph that represents a joint probability distribution among a set of variables. In

[1] In this work, we do not directly utilize quality-related image features for fingerprint images since we are using the NFIQ metric.

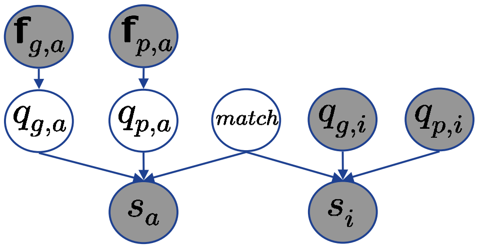


Figure 1. A graphical model for score-level quality-based face and fingerprint biometric fusion, where the shaded nodes and the unshaded nodes represent measurement nodes and hidden nodes, respectively.

a BN, nodes denote variables and the links among nodes denote the conditional dependencies among the variables. The conditional dependency is characterized by the conditional probability associated with each node. As shown in Fig. 1, the direct links between the nodes represent the causal relationships described above. The shaded nodes are measurement nodes ($\mathbf{f}_{g,a}$, $\mathbf{f}_{p,a}$, s_a, s_i, $q_{g,i}$, and $q_{p,i}$), whose states can be obtained through computer vision techniques; and the unshaded nodes ($q_{g,a}$, $q_{p,a}$, and $match$) are hidden nodes, whose states are what we will estimate via the model.

2.3. Model Parameterization and Learning

Given the model structure shown in Fig. 1, we need to define the states for each node, and then parameterize the model parameter associated with each node. A node $\mathbf{X}$ is parameterized by its conditional probability $p(\mathbf{X}|pa(\mathbf{X}))$ given its parents $pa(\mathbf{X})$ or its prior probability $p(\mathbf{X})$ if it does not have a parent.

Node $match$ has binary states ($match \in 0, 1$) representing *no-match* or *match* of the face and fingerprint gallery-probe pairs and is parameterized by its prior probability $p(match)$. If $match = 1$, the gallery and probe samples belong to the same subject for both fingerprint and face.

The quality $q_{g,i}$ for the gallery fingerprint image has $K_i = 5$ discrete states as in [16], where "1" represents the highest quality level and "5" represents the lowest quality level. It is parameterized by its prior probability $p(q_{g,i})$. The states of $q_{p,i}$ are defined and $p(q_{p,i})$ is parameterized likewise for the quality of the probe fingerprint image.

The continuous vector $\mathbf{f}_{g,a}$ contains the quality-related image features for the gallery facial image and parameterized by its prior probability $p(\mathbf{f}_{g,a})$. The quality-related image features can be coordinates of a set of facial landmarks, PCA shape coefficients of a facial shape, and/or PCA appearance coefficients of a face region. The quality-related image feature vector $\mathbf{f}_{g,a}$ is assumed to satisfy a multivariate Gaussian distribution, which is required by the BN implementation (Bayes Net Toolbox for Matlab [10]) used in this work, such that:

$$p(\mathbf{f}_{g,a}) = (2\pi)^{-\frac{d_{f_{g,a}}}{2}} |\Sigma_{f_{g,a}}|^{-\frac{1}{2}} exp(-\frac{\gamma^2_{f_{g,a}}}{2}) \quad (1)$$

where $d_{f_{g,a}}$ is the dimension of vector $\mathbf{f}_{g,a}$, and parameter $\gamma^2_{f_{g,a}}$ is defined as a Mahalanobis distance

$$\gamma^2_{f_{g,a}} = (\mathbf{f}_{g,a} - \bar{\mathbf{f}}_{g,a})^T \Sigma^{-1}_{f_{g,a}} (\mathbf{f}_{g,a} - \bar{\mathbf{f}}_{g,a}) \quad (2)$$

with the corresponding mean vector $\bar{\mathbf{f}}_{g,a}$ and covariance matrix $\Sigma_{f_{g,a}}$. The states of $\mathbf{f}_{p,a}$ are defined and $p(\mathbf{f}_{p,a})$ is parameterized similarly for the probe facial image.

The image quality $q_{g,a}$ for the gallery facial image can be defined as a continuous variable or be discretized into several discrete states. In this work, we employ discrete image qualities for facial image samples since discrete fingerprint image qualities are used. Given its parent $\mathbf{f}_{g,a}$, $q_{g,a}$ is parameterized by $p(q_{g,a}|\mathbf{f}_{g,a})$. For $q_{g,a}$ with K_a possible discrete states[2], $p(q_{g,a}|\mathbf{f}_{g,a})$ is defined as a multinomial logit function defined as follows:

$$p(q_{g,a} = k|\mathbf{f}_{g,a}) = \frac{exp(W_{gk} \times \mathbf{f}_{g,a} + b_{gk})}{\sum_{k=1}^{K_a} exp(W_{gk} \times \mathbf{f}_{g,a} + b_{g,k})} \quad (3)$$

where $q_{g,a} = k$ means $q_{g,a}$ is at its k^{th} state with $k \in \{1, ..., K_a\}$; W_{gk} and b_{gk} are model parameters that should be learned. Similarly, we define the states of $q_{p,a}$ and parameterize $p(q_{p,a}|\mathbf{f}_{p,a})$ for the probe facial image quality.

Match scores s_a and s_i are continuous variables and can be parameterized by $p(s_a|q_{g,a}, q_{p,a}, match)$ and $p(s_i|q_{g,i}, q_{p,i}, match)$, respectively. However, $p(s_a|q_{g,a}, q_{p,a}, match)$ and $p(s_i|q_{g,i}, q_{p,i}, match)$ are generally not well-modeled by parametric distributions apparently in Figs. 6 and 7. Hence, we discretize the match scores into discrete numbers such that s_a and s_i have N_a and N_i discrete states, respectively. Note that in the choice of N_a or N_i there is a tradeoff between distribution modeling accuracy and the amount of required training data.

Given the model structure and the definitions of the model parameters, we will learn the model parameters associated with each node given a set of training data. Since we can obtain training data for all nodes, learning the model parameters can be performed by Maximum Likelihood Estimation (MLE).

2.4. Fusion through Probabilistic Inference

Once the measurement nodes ($\mathbf{f}_{g,a}$, $\mathbf{f}_{p,a}$, $q_{g,i}$, $q_{p,i}$, s_a, and s_i) are observed, we can perform the quality-based face and fingerprint score-level fusion through probabilistic inference via the model as shown in Fig. 1. The decision for

[2]In this work, $K_a = 3$ representing three facial image quality levels, where "3" represents the highest quality level and "1" represents the lowest quality level.

match or *no-match* can be made by maximizing the probability of $match$ given the states of the measurement nodes $p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$.

$$match* = \underset{match}{\operatorname{argmax}}\, p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \quad (4)$$

Based on the conditional independence encoded in the BN, $p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$ can be factored as follows:

$$\begin{aligned} & p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \quad (5) \\ &= \sum_{k_g=1}^{K_a} \sum_{k_p=1}^{K_a} p(match, q_{g,a}, q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \\ &= \sum_{k_g=1}^{K_a} \sum_{k_p=1}^{K_a} c \times p(\mathbf{f}_{g,a}) \times p(q_{g,a}|\mathbf{f}_{g,a}) \times p(\mathbf{f}_{p,a}) \times p(q_{p,a}|\mathbf{f}_{p,a}) \times p(match) \\ & \times p(q_{g,i}) \times p(q_{p,i}) \times p(s_a|q_{g,a}, q_{p,a}, match) \times p(s_i|q_{g,i}, q_{p,i}, match) \end{aligned}$$

where c is a normalization factor. The factorized probabilities in Eq. 5 are the conditional probabilities as discussed in the previous section.

In addition to estimating the state of $match$, we can also assess the image qualities of gallery and probe facial images by maximizing $p(q_{g,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$ and $p(q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$ as follows:

$$\begin{aligned} & p(q_{g,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \quad (6) \\ &= \sum_{match=0,1} \sum_{k_p=1}^{K_a} p(match, q_{g,a}, q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \end{aligned}$$

$$\begin{aligned} & p(q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \quad (7) \\ &= \sum_{match=0,1} \sum_{k_g=1}^{K_a} p(match, q_{g,a}, q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i) \end{aligned}$$

where $p(match, q_{g,a}, q_{p,a}|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$ is factored as in Eq. 5. Specifically, we use the Bayes Net Toolbox for Matlab [10] to perform the BN inference.

3. Score-level Face and Fingerprint Biometric Fusion with Facial Image Quality

Our related approach is readily adapted to other fusion situations. Assuming that we do not have any knowledge of the fingerprint sample quality or that the fingerprint matching algorithm already takes account of the sample quality, we can perform score-level face and fingerprint biometric fusion with only facial image qualities. In this case, we model the causal relationships among $\mathbf{f}_{g,a}$, $\mathbf{f}_{p,a}$, s_a, s_i, $match$, $q_{g,a}$, and $q_{p,a}$. To do this, we use a BN model as shown in Fig. 2, which is similar to that in Fig. 1 except that we do not model $q_{g,i}$ and $q_{p,i}$ and that s_i has one parent, $match$.

We define the states for the nodes in Fig. 2 exactly the same as in Fig. 1. The model parameterization for the nodes in Fig. 2 is similar to those described in Section 2.3. The only difference is that s_i is parameterized by $p(s_i|match)$ given a single parent $match$. Learning the model parameters is performed by MLE given the training data. The *match/no-match* decision can be made by maximizing $p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, s_a, s_i)$, given the observations of the measurement nodes.

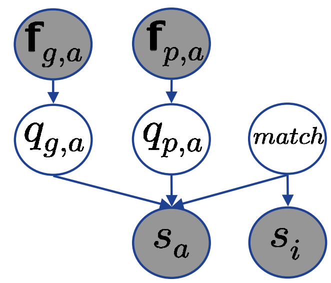


Figure 2. A graphical model for score-level face and fingerprint biometric fusion with facial image qualities.

4. Experimental Results

4.1. Face and Fingerprint Databases

Since face and fingerprint can reasonably be regarded as two independent resources of biometric information, we create a face and fingerprint image database of 325 "virtual subjects" [17], each of which has 7 chimeras[3] consisting of a facial image sample and a fingerprint image sample. Specifically, we use facial images from CAS-PEAL face database [6] and fingerprint images from the Fingerprint Verification Competition (FVC2004) database [3]. No real subject is paired with more than one other subject in forming these virtual subjects.

The CAS-PEAL face database [6] contains facial images with various facial appearance variations caused by face pose, facial expression, accessories, and lighting (PEAL) for a large population. It has been widely used for evaluating performance of face recognition algorithms. Since the face pose variation has shown to be the most significant factor affecting the face recognition performance [12], we randomly select 325 subject with 7 facial images, each of which represents one of the facial appearance variations caused by the face pose (frontal, $\pm 15°$, $\pm 30°$, and $\pm 45°$ in the horizontal direction). Fig. 3 shows a set of exemplar facial images of a subject we used.

FVC2004 database [3] consists of 4 subsets: DB1 and DB2 captured from different optical sensors, DB3 captured from thermal sweeping sensor, and DB4 containing synthetic fingerprints. We select 325 subjects from the three subsets containing real fingerprint images (DB1, DB2, and DB3). Each subject has 7 fingerprint images representing different sources of variations including vertical positions,

[3]Chimera is a type of data that is generated from multiple biometric modalities of different subjects and forms a synthetic biometric set for a subject.

Figure 3. Exemplar facial images of a subject in the CAS-PEAL face database [6] representing the facial appearance variations caused by different face poses: frontal, $\pm 15^{\circ}$, $\pm 30^{\circ}$, and $\pm 45^{\circ}$ in the horizontal direction.

Figure 4. Sample fingerprint images in FVC2004 database [3]. Each row contains fingerprint images from one of the three subsets (DB1, DB2, and DB3 top-to-bottom) respectively; and each column consists of fingerprint images at different quality levels (1–5 left-to-right) defined by NIST NFIQ [16].

pressure against the sensor surface, skin distortion, rotation, and skin humidity. Fig. 4 shows some example fingerprint images: each row containing fingerprint images from one of the three subsets (DB1, DB2, and DB3) respectively, and each column consisting of fingerprint images at different quality levels (1–5) determined by NIST NFIQ [16].

In the experiments, we divided the face and fingerprint chimera database into ten subsets such that the subjects in each subset are mutually exclusive. In a 10-fold testing framework, we use one subset for testing, while the remaining nine subsets are used for training the proposed fusion frameworks.

4.2. Experiment Setup

4.2.1 Match Score Acquisition and Quantization

We use the commercial face matching engine FaceIt® from Identix[4] to obtain match scores of facial images (s_a) and a NIST MINEX certified fingerprint matching algorithm VeriFinger[5] to get match scores of fingerprint images (s_i). The continuous face match score is then quantized into $N_a = 40$ levels and the fingerprint match score is discretized into $N_i = 50$ levels in our experiments.

[4]http://www.l1id.com/

[5]http://www.neurotechnology.com/verifinger.html

4.2.2 Facial Image Features

For the quality-related facial image features, we use the PCA appearance coefficients of the image intensities enclosed in a warped face region. For the i^{th} facial image in the database, the image feature vector $\mathbf{f}_i$ is computed as follows:

$$\mathbf{f}_i = \mathbf{P}_I^T(\mathbf{I}_i - \bar{\mathbf{I}}) \tag{8}$$

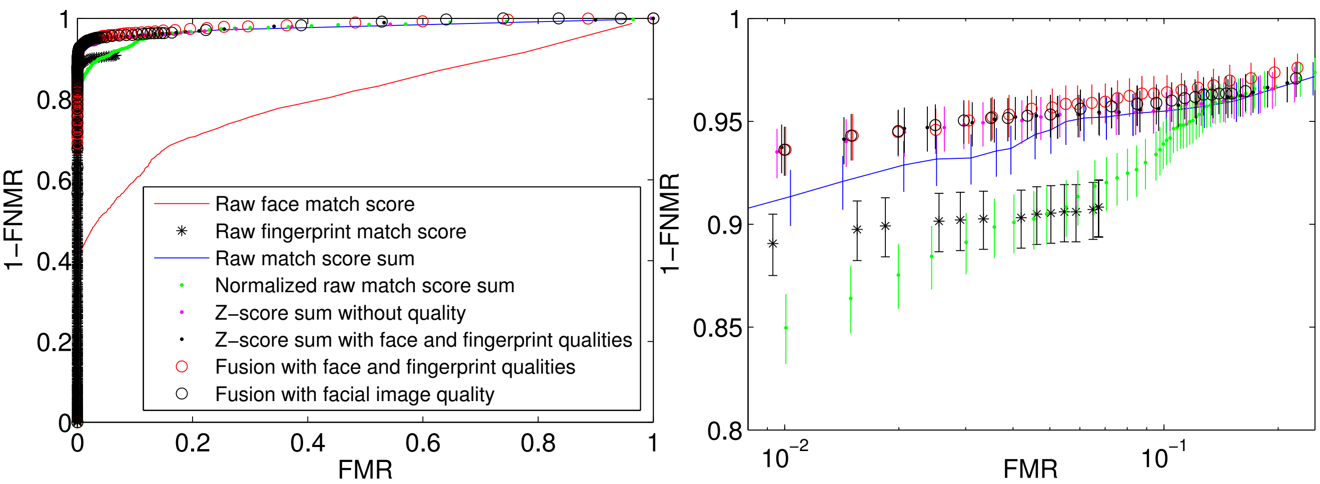
where $\mathbf{I}_i$ is the image intensity vector obtained by warping the i^{th} image to a common face region through a global affine transform[6]. $\bar{\mathbf{I}}$ and $\mathbf{P}_I$ are the mean face and eigenfaces representing the major facial appearance variation modes in the dataset. $\bar{\mathbf{I}}$ and $\mathbf{P}_I$ are trained from over 3000 facial images of other subjects in the CAS-PEAL face database [6] that are not selected to form our chimera database. The dimension of the eigenfaces is selected such that 50% of the energy of the appearance variations is preserved.

4.2.3 Facial Image Quality and Fingerprint Image Quality

In this work, we employ the NIST Fingerprint Image Quality (NFIQ) derived by Tabassi et al. [16] to obtain the qualities for fingerprint samples for training and testing the proposed biometric fusion models. NFIQ defines the fingerprint quality at 5 quality level (1–5 from "high" to "low"). Since there are a few fingerprint samples that have the worst qualities (levels 4 and 5), we combine the quality levels 4 and 5 into a single quality level. As a result, the fingerprint quality has 4 levels in our experiments.

We use the facial quality metrics developed by Ozay et al. [12] to obtain the image qualities for all facial images in the training set. The continuous quality values obtained in

[6]In this work, the global affine transformation matrix is obtained based on the eye positions for each facial image by using a face and eye detector. Advanced face alignment techniques such as Boosted Appearance Model [8] can be employed to extract more sophisticated image features accounting for variations caused by face pose, facial expression, and illumination.



(a) (b)

Figure 5. (a) Verification ROC curves for each algorithm. (b) Zoomed version of (a). On (b) error bars indicate 95% confidence intervals in the FNMR (False Non-Match Rate). For both (a) and (b), the x-axis represents the False Match Rate (FMR), and the y-axis represents the true matching rate (1-FNMR).

this way are further quantized into three discrete quality levels (1–3 from “low” to “high”). During the testing process, the facial image qualities for both probe and gallery images are unknown; and we use the proposed fusion algorithms to estimate them.

4.3. Evaluation of the Face and Fingerprint Biometric Fusion Algorithms

In this experiment, we will compare the biometric verification performance of the proposed algorithms with several baseline algorithms including:

- *Raw face match score*
- *Raw fingerprint match score*
- *Raw match score sum*:

  A summation of raw match scores of face and fingerprint is used for biometric verification.
- *Normalized raw match score sum*:

  The raw match scores of face and fingerprint are scaled into a range of $[0, 1]$, respectively. Their summation is used for biometric verification.
- *Z-score sum without quality assessment*:

  The z-score of a pair of gallery-probe biometric samples performs score normalization such that the reliable samples have higher scores, and is computed as $z = \frac{s-\mu}{\sigma}$, where s is the raw match score of a pair of biometric samples; μ and σ are the mean and standard deviation of the *no-match* match score distribution in the training set. The summation of the z-scores of the face and fingerprint is used for biometric verification.
- *Z-score sum with face and fingerprint quality assessments:*

  The z-score of a pair of gallery-probe fingerprint samples with the fingerprint quality assessment is computed by

  $$z_{q_{g,i},q_{p,i}} = \frac{s_i - \mu_{q_{g,i},q_{p,i}}}{\sigma_{q_{g,i},q_{p,i}}} \tag{9}$$

  where s_i is the raw match score of a pair of fingerprint samples; $q_{g,i}$ and $q_{p,i}$ are the corresponding quality assessments of gallery and probe fingerprint images; $\mu_{q_{g,i},q_{p,i}}$ and $\sigma_{q_{g,i},q_{p,i}}$ are the mean and standard deviation of the *no-match* match score distribution corresponding to the gallery-probe fingerprint quality combination in the training set. $z_{q_{g,a},q_{p,a}}$ is computed likewise, where $q_{g,a}$ and $q_{p,a}$ are estimated by the proposed fusion algorithm with both face and fingerprint quality metrics as described in Section 2. Then, the summation of $z_{q_{g,i},q_{p,i}}$ and $z_{q_{g,a},q_{p,a}}$ is used for biometric verification, where the higher quality and thus more reliable sample has a larger effective weight.

In the experiments, we will evaluate our proposed biometric fusion algorithms including fusion with facial image quality metric as discussed in Section 3 and with both face and fingerprint image qualities as described in Section 2. For the biometric fusion methods, the scores (e.g., z-scores, $p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, q_{g,i}, q_{p,i}, s_a, s_i)$, and $p(match|\mathbf{f}_{g,a}, \mathbf{f}_{p,a}, s_a, s_i)$) for all 10 subsets are collected together, and then a single threshold is applied for each point on the ROC curve, for each method.

Fig. 5 shows the overlaid ROC curves for the proposed algorithms and the baseline algorithms. The Equal Error Rate (EER) for each algorithm is given in Table 2. From

Method	EER
Raw face match score	0.2619
Raw fingerprint match score	0.0795
Raw match score sum	0.0522
Normalized raw match score sum	0.0766
Z-score sum without quality assessment	0.0476
Z-score sum with face and fingerprint quality assessment	0.0467
Face and fingerprint fusion with face quality	0.0484
Face and fingerprint fusion with face and fingerprint qualities	0.0443

Table 2. Performance comparison of the algorithms in terms of Equal Error Rates (ERR).

Fig. 5 and Table 2, we can see that the proposed score-level face and fingerprint fusion algorithms perform much better than the methods based solely on the raw match score of a single modality (face or fingerprint). That demonstrates the effectiveness of the proposed biometric fusion models. Furthermore, the method using score-level face and fingerprint fusion with both face and fingerprint qualities achieves the best verification performance.

From Fig. 6, we can see that the *match* and *no-match* match score distributions are well separated when the gallery and probe facial images have the best qualities (Fig. 6c) and are not separated when they have the worst qualities (Fig. 6a). This observation demonstrates that the computed facial image quality is indicative of the performance of the face matching algorithm. On the contrary, even when both the gallery and probe fingerprint images have the best qualities (Fig. 7a), the *match* and *no-match* match score distributions are not well separated as we expected and are similar to those with lower qualities (Fig. 7b and Fig. 7c). The reason is likely that the NFIQ metric is built on the *match* and *no-match* distributions of the match scores, and thus is matching algorithm dependent as described in Section 2.1. Since we use a matching algorithm that is different from those used for training the NFIQ, the NFIQ cannot accurately predict the performance of the fingerprint matching algorithm we used. Although the fingerprint image quality is not as informative as facial image quality, using both face and fingerprint image qualities still helps to improve the verification performance significantly through the proposed biometric fusion framework.

5. Conclusion and Future Work

The purpose of this work is to investigate how multiple biometric technologies can together be made a more effective tool for public safety. Individual biometrics often lack the accuracy that is needed to be effective in operational environments. Our goal is to advance the state-of-the-art in biometric fusion technology in order to provide a more universal and more accurate solution for personal identification and verification.

To do this, we developed score-level multi-modal fusion algorithms based on predictive quality metrics and employed them for the task of face and fingerprint biometric fusion. The causal relationships in the context of each fusion scenario are modeled in a principled way by a probabilistic framework. The recognition/verification decision is made through probabilistic inference. The experiments demonstrated that the proposed score-level fusion algorithms significantly improve the verification performance over the methods based on the raw match score of a single modality (face or fingerprint). Furthermore, the fusion framework with both face and fingerprint image qualities achieves the best verification performance and outperforms the other baseline fusion algorithms.

We should note that the proposed quality-based fusion framework is not restricted to face and fingerprint and can be generalized to include other biometric modalities such as iris, with or without quality metrics. Also, model parameter learning is a data-driven process and it is to be determined the degree to which additional training data will improve MLE performance or reduced data will hurt performance.

References

[1] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A comparative study of fingerprint image-quality estimation methods. *IEEE Transactions on Information Forensics and Security*, 2(4):734–743, 2007.

[2] K. W. Bowyer, K. I. Chang, P. Yan, P. J. Flynn, E. Hansley, and S. Sarkar. Multi-modal biometrics: An overview. *Second Workshop on Multi-Modal User Authentication*, 2006.

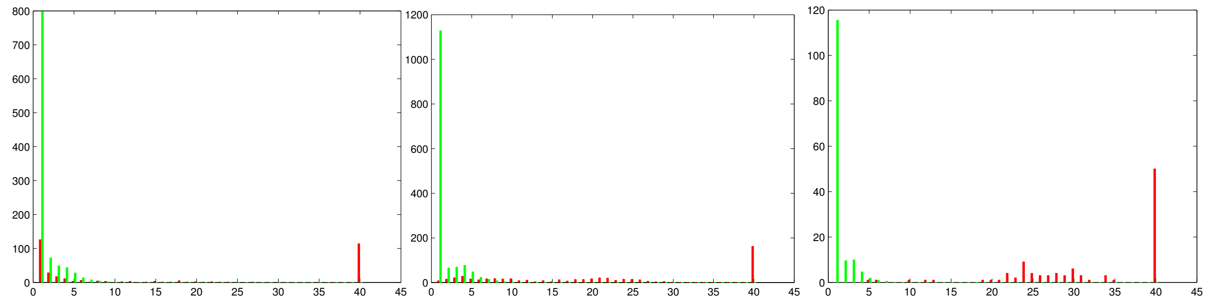
[3] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance evaluation of fingerprint verification systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):3–18, 2006.

[4] M. T. Chan, R. Brown, and W. Turner. Incorporating quality metrics in multimodal biometric fusion. In *IEEE Computer Society Workshop on Biometrics (in association with CVPR 2006)*, 2006.

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

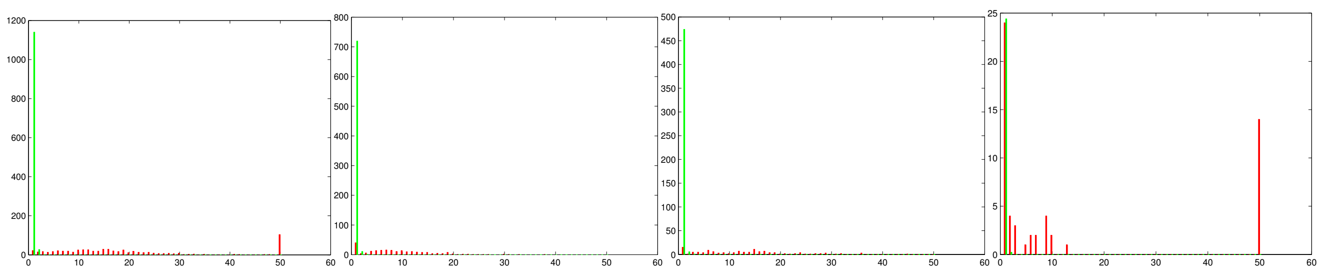
[6] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, and D. Zhao. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 38(1):149–161, 2008.

[7] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):1–13, 2007.



(a) $q_{g,a} = 1, q_{p,a} = 1$ (b) $q_{g,a} = 2, q_{p,a} = 2$ (c) $q_{g,a} = 3, q_{p,a} = 3$

Figure 6. Histograms of the *match* and *no-match* distributions of the discretized face match scores for a subset of the gallery-probe facial quality combinations. Red and green bars represent the *match* and *no-match* match score distributions, respectively. For clarity, the histograms of *no-match* match scores are divided by 100. The quality levels range from $q = 1$ (lowest) to $q = 3$ (highest).



(a) $q_{g,i} = 1, q_{p,i} = 1$ (b) $q_{g,i} = 2, q_{p,i} = 3$ (c) $q_{g,i} = 3, q_{p,i} = 2$ (d) $q_{g,i} = 4, q_{p,i} = 4$

Figure 7. Histograms as in Fig. 6, but for fingerprints using modified NFIQ quality metrics [16] where the quality levels range from $q = 1$ (highest) to $q = 4$ (lowest).

[8] X. Liu. Discriminative face alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1941–1954, 2009.

[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] K. P. Murphy. The Bayes Net Toolbox for Matlab. *Computing Science and Statistics*, 33:331–350, 2001.

[11] K. Nandakumar, Y. Chen, A. K. Jain, and S. C. Dass. Quality-based score level fusion in multibiometric systems. In *International Conference on Pattern Recognition*, 2006.

[12] N. Ozay, Y. Tong, F. W. Wheeler, and X. Liu. Improving face recognition with a quality-based probabilistic framework. In *IEEE Computer Society Workshop on Biometrics (in association with CVPR 2009)*, 2009.

[13] N. Poh, T. Bourlai, J. Kittler, L. Allano, F. Alonso-Fernandez, O. Ambekar, J. Baker, B. Dorizzi, O. Fatukasi, J. Fierrez, H. Ganster, J. Ortega-Garcia, D. Maurer, A. A. Salah, T. Scheidat, and C. Vielhauer. Benchmarking quality-dependent and cost-sensitive multimodal biometric fusion algorithms. *IEEE Trans. on Information Forensics and Security*, 4(4):849–866, 2009.

[14] A. Ross and A. K. Jain. Multimodal biometrics: An overview. In *Proc. of the Twelfth European Signal Processing Conference*, pages 1221–1224, 2004.

[15] W. J. Scheirer, A. Bendale, and T. E. Boult. Predicting biometric facial recognition failure with similarity surfaces and support vector machines. In *IEEE Computer Society Workshop on Biometrics (in association with CVPR 2008)*, 2008.

[16] E. Tabassi, C. L. Wilson, and C. I. Watson. Fingerprint image quality. Technical report, NIST, 2004.

[17] B. Ulery, W. Fellner, P. Hallinan, A. Hicklin, and C. Watson. Studies of biometric fusion appendix c evaluation of selected biometric fusion techniques. *Technical Report IR 7346 NIST*, 2006.

[18] P. Wang, Q. Ji, and J. L. Wayman. Modeling and predicting face recognition system performance based on analysis of similarity scores. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):665–670, 2007.