Improving Face Recognition with a Quality-based Probabilistic Framework[*]

Necmiye Ozay[†] Yan Tong Frederick W. Wheeler Xiaoming Liu

ECE Department Visualization and Computer Vision Lab

Northeastern University GE Global Research

Boston, MA Niskayuna, NY

`ozay.n@neu.edu` `tongyan,wheeler,liux@research.ge.com`

Abstract

This paper addresses the problem of developing facial image quality metrics that are predictive of the performance of existing biometric matching algorithms and incorporating the quality estimates into the recognition decision process to improve overall performance. The first task we consider is the separation of probe/gallery qualities since the match score depends on both. Given a set of training images of the same individual, we find the match scores between all possible probe/gallery image pairs. Then, we define symmetric normalized match score for any pair, model it as the average of the qualities of probe/gallery corrupted by additive noise, and estimate the quality values such that the noise is minimized. To utilize quality in the decision process, we employ a Bayesian network to model the relationships among qualities, predefined quality related image features and recognition. The recognition decision is made by probabilistic inference via this model. We illustrate with various face verification experiments that incorporating quality into the decision process can improve the performance significantly.

1. Introduction

Biometric sample quality values can be used in many different stages of biometric operations (e.g., enrollment phase quality assessment, verification/identification quality assessment, prediction of algorithm failure, quality-based adaptation of the processing phase and multimodal biometric fusion [5, 1, 12, 2, 7, 10]). Although principled quality measures have been developed for fingerprint samples, the

facial image quality problem still remains open [5]. This is partially due to the fact that how face recognition actually works is less certain since classifiers are learned using abstract features which hide the actual matching method. Also, there is a large number of factors, including intra-personal variations and imaging conditions, affecting the matching performance. Different face matching algorithms are designed to be robust to particular subsets of these factors. Hence, a high-quality image for one algorithm is not necessarily of the same quality for another one. Therefore, quality should be *learned* for a specific face matching algorithm. In this paper, we take a black box identification approach. That is, we do not make any assumption about the matching algorithm. Each probe/gallery image pair in a training set produces a match score. The input pairs together with the match scores are used to *learn* what kind of images are of high-quality for the underlying algorithm.

Following the work of Tabassi *et. al* in [11], we define the quality of a biometric sample as a scalar quantity that is predictive of the performance of a biometric system. Considering the match score as a similarity measure, a predictive quality measurement algorithm should satisfy the following property: a biometric sample of a subject should be assigned a high quality value if it is similar to the samples from the same subject while it is different from the samples from other subjects. Hence, the match scores of high-quality samples of the same subject should be well separated from the non-match score distribution of the same subject. However, it is important to consider that both the quality of the probe and the quality of the gallery play a role. Although a common assumption in building quality metrics (see for instance [11]) is that the gallery is of good quality so that the normalized match score reflects the quality of the probe, in many applications, one should not assume that the gallery is of high-quality without having a measure of quality. Instead, we assume that multiple samples for each subject are available in the training dataset; and present a probe/gallery quality separation scheme to estimate the quality value for each facial image sample using

a symmetric normalized match score as a measure of the match quality. After assigning facial image qualities to the images in the training set, it is possible to learn the relationship between some predefined image features and the facial image quality. Therefore, we can predict the quality value for each test sample using this relationship and the image features extracted from the sample.

Once the quality of a facial image is estimated, there is still the issue of how the estimated quality can be employed to improve the face recognition performance. Face recognition from a high-quality probe and high-quality gallery image pair would produce the highest recognition performance among all the possible probe/gallery quality combinations. As a result, one solution is to use the estimated quality value as a guidance to select *good* image samples for recognition. However, this is only useful in applications where it is possible to collect samples until a high-quality sample is acquired. A more general and desirable approach is to combine the quality assessment with the recognition and to improve the recognition performance across the whole quality scale.

There are several statistical relationships in a quality-based face recognition problem, including the relationships between the image features and quality assessment of an image and the relationships between the quality assessments of probe and gallery images and the match score obtained through a particular matching algorithm. In this work, we propose a unified probabilistic framework to simultaneously predict the quality of the facial image samples and perform quality-based face recognition by exploiting these relationships. Specifically, we use a Bayesian Network (BN) to model and learn such relationships. Then, the quality-based face recognition is performed by probabilistic inference through the proposed framework. Our experimental results show that the quality-based face recognition improves the face verification performance significantly compared to the methods based solely on raw match score.

Our ability to improve overall recognition performance by utilizing quality metrics derived from the facial image samples themselves implies some inefficiency on the part of the core matching algorithm. After all, the matching algorithm has access to the images, and could compute and use the quality metrics just as we do. In effect, the matching algorithm could internalize our entire additional quality analysis mechanism. Still, the improvement we achieve should not be too surprising, as face recognition systems are imperfect, especially when used on data not similar to that used to develop and train the algorithms.

2. Separation of Probe/Gallery Qualities

A matching algorithm $\mathcal{A}$ produces a score for a given pair of images:

$$s_{i_k j_l} = \mathcal{A}(i_k, j_l), \qquad (1)$$

where i_k denotes the kth image of the ith individual. In a verification task, a probe image i_p is compared against the gallery image i_g of the claimed identity i using algorithm $\mathcal{A}$. If the match score $s_{i_p i_g}$ is above a predefined threshold, the claim is accepted.

As mentioned earlier, an authentic high-quality probe/gallery pair should produce a match score that is well separated from the non-match scores. In [11], Tabassi *et. al* proposed the normalized match score as a measure of this separation. The normalized match score between the kth and l^{th} images of ith individual, when i_k is the probe and i_l is the gallery image, is defined as:

$$NMS(i_k, i_l) = \frac{s_{i_k i_l} - \mu_{i_k}(s_{non-match})}{\sigma_{i_k}(s_{non-match})}, \qquad (2)$$

where $s_{i_k i_l}$ is the match score between i_k and i_l as in Eq. (1), $\mu_{i_k}(s_{non-match})$ and $\sigma_{i_k}(s_{non-match})$ are respectively the mean and standard deviation of the non-match scores between the image i_k and the images from other individuals $j \neq i$.

The *NMS* provides some information about the quality of the probe sample. However, it is also sensitive to the quality of the gallery since both probe and gallery affect $s_{i_k i_l}$. Hence, it should only be used as a measure of the quality of the probe when the gallery samples are of uniformly high-quality.

The other problem with *NMS* is that it is not symmetric in its arguments as the non-match score distribution will vary when probe and gallery images are interchanged. This variation could be severe especially when the probe and gallery images are of different quality. Hence, it is not an ideal measure of the quality of the match either. A better measure of the quality of a match should be symmetric with respect to probe and gallery images since verification decisions are based on thresholding the match score $s_{i_k i_l}$ which is approximately symmetric for most matching algorithms. Before proceeding with a separation method, we define *symmetric normalized match score* as:

$$SNMS(i_k, i_l) = \frac{1}{2}(NMS(i_k, i_l) + NMS(i_l, i_k)). \qquad (3)$$

SNMS will be high for high-quality probe/gallery pairs and will be low for low-quality probe/gallery pairs. Hence, *SNMS* can be used as a measure of the quality of a match. Once we have a way to measure the quality of the match, the next step is to find a way to estimate the quality of each image sample by separating probe and gallery qualities.

Although separation of probe/gallery qualities has not received much attention in the literature, there has been some work on how to combine the sample qualities to obtain the quality of the match [5]. Similar functions can be used for formulating a quality separation scheme. In particular, we model the quality of the match as the average of

probe/gallery qualities; that is:

$$Q(i_k, i_l) = \frac{q(i_k) + q(i_l)}{2}, \qquad (4)$$

where $Q(i_k, i_l)$ is the quality of the match and $q(i_k)$, $q(i_l)$ are qualities of the samples.

A predictive quality measure should estimate the quality of the match which we measure using *SNMS*. Formally, we would like to find the values of the scalar function q such that

$$\frac{q(i_k) + q(i_l)}{2} \approx SNMS(i_k, i_l). \qquad (5)$$

If we have at least three images for each individual in the training dataset, the separation problem can be solved. In particular, by combining all the equations for an individual i, we obtain the following least squares problem:

$$\mathbf{A}\mathbf{q}_i + \boldsymbol{\epsilon} = \mathbf{y}_i, \qquad (6)$$

where

$$\mathbf{A} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & \dots & 0 \\ 0.5 & 0 & 0.5 & 0 & \dots & 0 \\ \vdots & & \ddots & & & \vdots \\ 0 & \dots & 0 & 0 & 0.5 & 0.5 \end{bmatrix}, \mathbf{q}_i = \begin{bmatrix} q(i_1) \\ \vdots \\ q(i_N) \end{bmatrix}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \vdots \\ \epsilon_T \end{bmatrix}, \mathbf{y}_i = \begin{bmatrix} SNMS(i_1, i_2) \\ SNMS(i_1, i_3) \\ \vdots \\ SNMS(i_{N-1}, i_N) \end{bmatrix}.$$

Here, $\mathbf{A}$ is a $T \times N$ matrix with two non-zero elements in each row where N is the number of samples from individual i, and T is the size of an index set containing the pairs (i.e., the index set has $T = N(N-1)/2$ elements and every row of $\mathbf{A}$ corresponds to a different pair of images from the ith individual). When $N \geq 3$, $\mathbf{A}$ is full column rank; hence the solution with minimum squared error (i.e., minimum $\|\boldsymbol{\epsilon}\|_2^2$) is given by $\mathbf{q}_i = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}_i$.

The sample qualities for all the images in the training set can be obtained using this separation scheme. The quality values obtained with this method are continuous variables. If desired, it is possible to quantize these values into several bins to obtain discrete quality values (e.g., high, medium, low). Once we have a quality value assigned to each image in the training dataset, the next step is to develop a methodology that can predict the quality values for the test samples and use these estimated quality values to boost recognition performance.

3. Quality-based Face Recognition Model

In this section, we introduce our probabilistic framework that simultaneously estimates the quality values from predefined image features in the facial images and utilizes the quality values to improve the face recognition decisions.
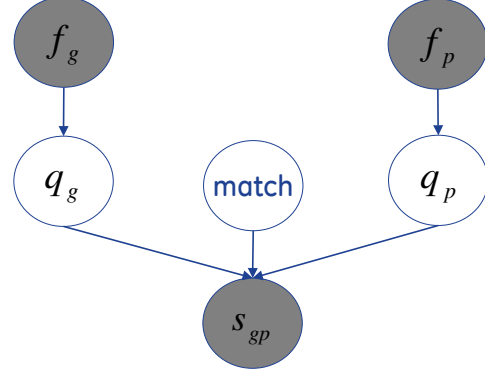

Figure 1. A graphical model for quality-based face recognition. The shaded nodes are measurement nodes, whose states can be obtained; the unshaded nodes are hidden nodes, whose states are inferred via the model.

3.1. Causal Relationships in Quality-based Face Recognition

There are several key elements involved in the quality-based face recognition problem including *image features*, *image quality* for both probe and gallery images, the matching algorithm, and the *match* or *no-match* status of the probe/gallery image pair. Let $\mathbf{f}_g$ be a feature vector containing some predefined *image features* such as coordinates of a set of facial landmarks, shape coefficients of a statistical facial shape model, and/or appearance coefficients of a statistical appearance model. Let q_g be an assessment of *image quality*, for a gallery image. Similarly, $\mathbf{f}_p$ and q_p are the corresponding feature vector and the quality assessment for a probe image, respectively. Let s_{gp} be the match score for a probe/gallery image pair obtained by some face matching algorithm, and $match$ denote whether the gallery and probe images belong to same individual.

For quality-based face recognition, we model the causal relationships among the six elements defined above (i.e., $\mathbf{f}_g$, $\mathbf{f}_p$, q_g, q_p, s_{gp}, and $match$). First, by assuming that the image quality is affected by the image features, and thus can be directly derived from the image features, $\mathbf{f}_g$ and $\mathbf{f}_p$ can be regarded as the sole causes to generate q_g and q_p, respectively. Second, given a face matching algorithm, the match score s_{gp} is affected by the image qualities of the gallery and probe images (q_g and q_p) and the state of $match$ (*match/no-match*). These 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 (DAG) that represents a joint probability distribution among a set of variables. In a BN, nodes denote variables and the links among nodes denote the conditional dependency among variables. The 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 like $\mathbf{f}_g$ and $\mathbf{f}_p$, whose states can be obtained directly; and the unshaded nodes like *match* are hidden nodes, whose states will be estimated.

3.2. Model Parameterizations

Given the model structure shown in Fig. 1, we need to define the states for each node and, then, parameterize the model parameter (the conditional probability) associated with each node.

$\mathbf{f}_g$ is a continuous vector that contains the image features for the gallery image and is parameterized by its prior probability. Here, we assume that $\mathbf{f}_g$ satisfies a multivariate Gaussian distribution with mean vector $\bar{\mathbf{f}}_g$, covariance matrix Σ_{fg}. $\mathbf{f}_p$ is defined and parameterized similarly for the probe image.

match has binary states ($match \in \{0, 1\}$) representing *no-match* or *match* status of a probe/gallery pair. It is parameterized by its prior probability $p(match)$.

The image quality q_g for the gallery image can be defined as a continuous variable or can have discrete states. Given its parent $\mathbf{f}_g$, q_g is parameterized by its conditional probability $p(q_g|\mathbf{f}_g)$. If q_g has continuous states, its Conditional Probability Distribution (CPD) is defined as a Gaussian distribution as follows:

$$p(q_g|\mathbf{f}_g) = \frac{1}{\sqrt{2\pi}\sigma_g} exp(-\frac{(q_g - \bar{q}_g - W_g\mathbf{f}_g)^2}{2\sigma_g^2}), \quad (7)$$

where W_g is a regression matrix that maps $\mathbf{f}_g$ to q_g and $\bar{q}_g + W_g\mathbf{f}_g$ is the mean quality value, and σ_g is the variance.

For q_g with K possible discrete states, its CPD is defined as a multinomial logit function as follows:

$$p(q_g = k|\mathbf{f}_g) = \frac{exp(W_{gk}\mathbf{f}_g + b_k)}{\sum_{k=1}^{K} exp(W_{gk}\mathbf{f}_g + b_k)}, \quad (8)$$

where $q_g = k$ means q_g is at its kth state with $k \in \{1, ..., K\}$; W_{gk} and b_k are model parameters that will be learned. Similarly, q_p is defined and parameterized.

The match score s_{gp} is a continuous variable. Its CPD $p(s_{gp}|match, q_g, q_p)$ is assumed to be a Gaussian distribution and defined according to the states of its parents.

3.3. Model Learning

Given the model structure and the definition of the model parameters, we learn the model parameters associated with each node given a set of training data. We can get the image features for each gallery or probe image, and have the groundtruth labels of match/no-match for each probe/gallery image pair in a training image set. Hence, learning the model parameters for $\mathbf{f}_g$, $\mathbf{f}_p$, and *match* can be performed by Maximum Likelihood Estimation (MLE).

There are two situations with respect to learning the parameters of q_g and q_p. On the one hand, suppose we do not have any knowledge of the image quality, and thus cannot obtain the labels of the q_g and q_p for training. The model parameters $p(q_g|\mathbf{f}_g)$ and $p(q_p|\mathbf{f}_p)$ can be estimated by an Expectation-Maximization (EM) algorithm. On the other hand, assuming that we can get the labels for q_g and q_p through the quality assessment algorithm presented in Section 2, $p(q_g|\mathbf{f}_g)$ and $p(q_p|\mathbf{f}_p)$ can be learned by MLE given the training data of $\mathbf{f}_g$, $\mathbf{f}_p$, q_g, and q_p.

Thus, learning the parameters for s_{gp} is conducted by MLE if q_g and q_p have labels in the training data or by EM otherwise.

4. Quality-based Face Recognition through Inference

Once the measurement nodes ($\mathbf{f}_g$, $\mathbf{f}_p$, and s_{gp}) are observed, we can perform the quality-based recognition through probabilistic inference via the model as shown in Fig. 1. The match or no-match decision can be made by maximizing the joint probability of *match*, q_g, and q_p given the measurements of image features of gallery and probe images ($\mathbf{f}_g$ and $\mathbf{f}_p$), and the match score (s_{gp}) as follows:

$$match, q_g, q_p = \underset{match, q_g, q_p}{\operatorname{argmax}} \; p(match, q_g, q_p|\mathbf{f}_g, \mathbf{f}_p, s_{gp})$$
$$(9)$$

Based on the conditional independency encoded in the BN, $p(match, q_g, q_p|\mathbf{f}_g, \mathbf{f}_p, s_{gp})$ can be factorized as follows:

$$p(match, q_g, q_p|\mathbf{f}_g, \mathbf{f}_p, s_{gp}) = c \times p(\mathbf{f}_g) \times p(q_g|\mathbf{f}_g) \times p(\mathbf{f}_p)$$
$$\times p(q_p|\mathbf{f}_p) \times p(match) \times p(s_{gp}|q_g, q_p, match),$$
$$(10)$$

where c is a normalization factor. The factorized probabilities in Eq. (10) are the conditional probabilities that are learned as discussed in the previous section. In this work, we use the Bayes Net Toolbox for Matlab [6] by Murphy to implement the BN inference.

5. Experimental Results

5.1. Experiment Setup

We use the IMM Face Database [8] to test the proposed quality-based face recognition algorithm. This database consists of 40 different subjects each of which has 6 images. The image categories are: (i) neutral, (ii) smile, (iii) 30 degree left, (iv) 30 degree right, (v) spotlight on left, and vi) arbitrary expression. Fig. 2 shows some example images of a subject in the IMM Face Database. We use a commercial face recognition product, FaceIt®, from Identix[2] to obtain the measurements of match scores for facial image pairs.

---

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

Figure 2. Sample images for a subject in the IMM Face Database with the mesh connecting facial landmarks superimposed.

In our experiments, unless stated otherwise, the shape coefficient of a Point Distribution Model (PDM) [3] is used as the feature representation of an image. To be specific, for the ith image in the dataset, its shape coefficient $\mathbf{f}_i$ is computed as follows:

$$\mathbf{f}_i = \mathbf{P}^T(\mathbf{x}_i - \bar{\mathbf{x}}), \qquad (11)$$

where the 66-dimensional vector $\mathbf{x}_i$ is a concatenation of the x and y coordinates of 33 facial landmarks around the facial features, such as eyes, eyebrows, nose, mouth, and face contour shown as vertices of facial mesh in Fig. 2. $\bar{\mathbf{x}}$ and $\mathbf{P}$ are the mean and shape basis of PDM representing the major facial shape variations. They are trained using multi-view facial images of the FERET database [9] where the number of the shape basis is determined such as 90% energy of shape deformations is preserved. For 2D images, the shape coefficient $\mathbf{f}_i$ encodes the shape deformations due to identity, facial pose and expression variations. Since our focus is on evaluating the quality-based recognition model, we provide manually labeled facial landmarks ($\mathbf{x}_i$) for each image in the experiments.

5.2. Quality Assessment via Probe/Gallery Quality Separation

In this section, we demonstrate how our quality assessment scheme works. Each image in the IMM Face Database is assigned a quality value by using our separation scheme. Fig. 3 shows the histogram of the obtained quality values together with estimated Gaussian distributions of the quality values within different image categories. These results show that the separation scheme is capable of capturing the quality variations among the training data.

5.3. Evaluation of the Quality-based Face Recognition Algorithm

In the experiments, we compare four methods for making face recognition decisions. The first one is simply face recognition based on the raw match score. The other three are based on the quality-based face recognition decision
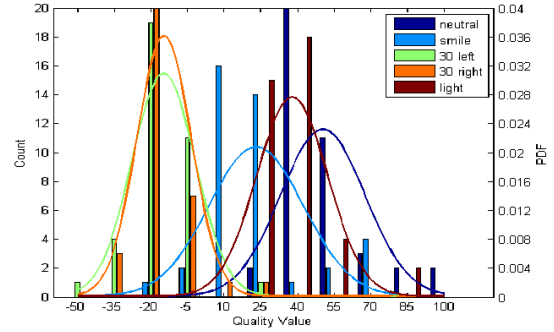

Figure 3. Results of quality separation. Histogram shows distribution of quality values among different image types. For each image category a Gaussian distribution with mean and standard deviation equal to those of the quality distribution of that category is superimposed on the histogram.

Method	EER
Raw Match Score	0.139
Unsupervised Discrete Quality	0.118
Supervised Continuous Quality	0.107
Supervised Discrete Quality	0.103

Table 1. Equal Error Rates (ERR) for the four face recognition methods.

algorithm proposed in this work. In the second method, q_g and q_p have three discrete states and are learned without training data of quality values (i.e., q_g and q_p). In the third method, q_g and q_p have continuous states and are learned using training data obtained by our quality separation scheme. For the fourth method, q_g and q_p have discrete states and are learned using training data obtained by discretizing the results of our quality separation scheme into three bins. The leave-one-subject-out cross-validation methodology is used to evaluate the proposed quality-based face recognition algorithm. Each time, we use all the probe/gallery image pairs in the database to train the model except that we leave out the probe/gallery image pairs containing the images of one subject for testing, such that the subjects in the training and testing sets are exclusive.

Fig. 4 shows the overlaid ROC curves for all four methods. The Equal Error Rate (EER) for each method is given in Table 1. From Fig. 4 and Table 1, we can see that all the quality-based face recognition methods perform better than the method based solely on the raw match score. That demonstrates the effectiveness of the proposed quality-based face recognition model. The fourth method achieves the best verification performance, which implies that the quality metrics based on the symmetric normalized match score is effective for image quality assessment.
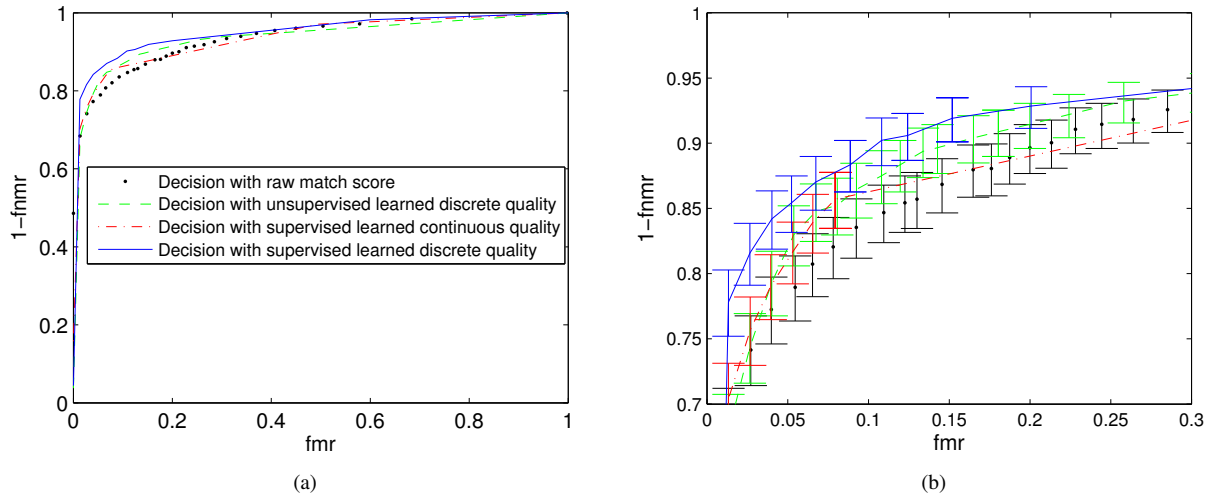
Figure 4. (a) Verification ROC curves for each method. The x-axis represents the False Matching Rate (FMR), and the y-axis represents the true matching rate (1-FNMR). (b) Zoomed version of (a), showing statistically significant performance improvement of method four over method one. On (b) error bars indicate 95% confidence intervals in the FNMR (False Non-Match Rate) estimate.

5.4. Study on the Relationship between Image Quality and Verification Performance

We also study the relationship between the estimated image quality and verification performance. For this purpose, the results of the fourth method are used. Besides estimating the state of $match$, we obtain the states of q_g and q_p through the probabilistic inference as described in Section 4. Based on the estimated q_g and q_p, we determine the verification ROC curves according to each combination of probe/gallery image quality levels as shown in Fig. 5.

It is clear that high-quality probe/gallery image pairs achieve superior performance, which indicates that our face quality assessment method is predictive of the recognition performance as desired. Therefore, the estimated facial image quality can also be used as a guidance to automatically select *good* samples for face recognition.

5.5. Quality-based Face Recognition Using Appearance-based Image Features

We also evaluate the quality-based face recognition algorithm using different types of quality related features. Specifically, we use the appearance coefficient of an Active Appearance Model (AAM) [4], which models the variations of image intensities enclosed in a warped face region. For the ith image in the dataset, its appearance coefficient $\mathbf{f}_i$ is computed as follows:

$$\mathbf{f}_i = \mathbf{P}_I^T (\mathbf{I}_i - \bar{\mathbf{I}}), \qquad (12)$$

where $\mathbf{I}_i$ is the image intensity vector obtained by warping the ith image to a common face region through a global

affine transform[3]. $\bar{\mathbf{I}}$ and $\mathbf{P}_I$ are the mean and appearance basis of AAM representing the major appearance variations modes. $\bar{\mathbf{I}}$ and $\mathbf{P}_I$ are trained from facial images in the FERET database [9] where the basis dimensionality is selected such that 50% of the energy of the appearance variations is preserved.

For model training, we use the discretized quality values based on symmetric normalized match scores as discrete image qualities for probe and gallery images and again perform leave-one-subject-out cross-validation. Fig. 6 shows the verification ROC of quality-based face recognition using the appearance-based image features. We can see that a similar verification performance (EER is 0.102) is achieved by using appearance-based image features compared to that of shape-based image features. That implies that the face pose and expression variations, which cause the changes in both facial appearance and facial shape, are the major factors to affect image quality and verification performance.

6. Conclusions and Future Work

In this work, we have developed a unified probabilistic framework for quality-based face recognition decisions, where the quality assessments of facial images are integrated into face recognition in a principled way. The experiments demonstrated that the proposed algorithm significantly improves face recognition performance over a wide range of facial image quality.

We should note that the proposed quality-based recognition framework is not restricted to face recognition and can

---

[3]In this work, the global affine transformation matrix is obtained based on the eye positions and the lowest point of chin on each facial image by using a face and eye detector.
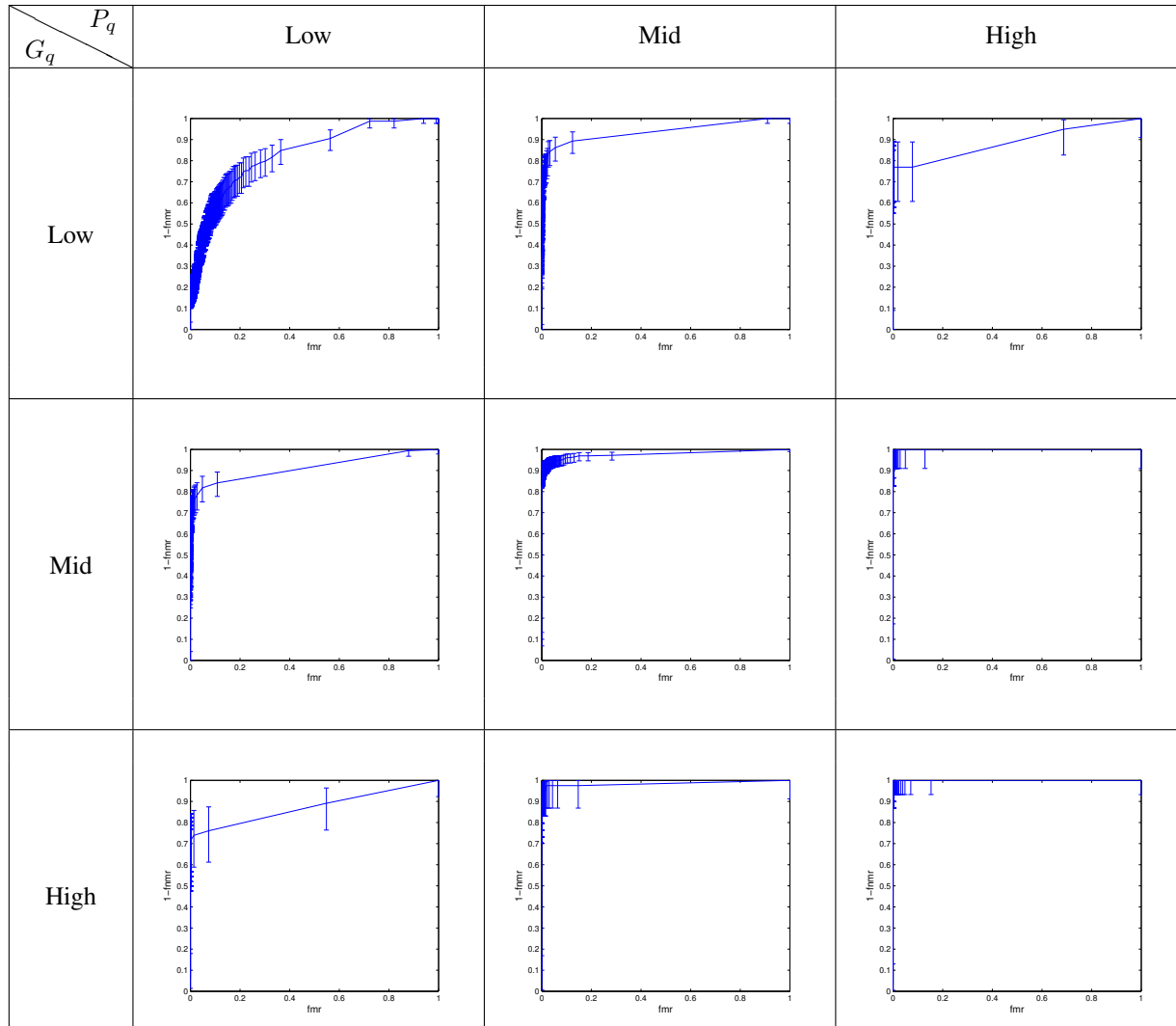
Figure 5. ROCs of quality-based face recognition results according to different probe/gallery image quality combinations. The fourth method is used in this study. For each figure, error bars indicate 95% confidence intervals; FMR represents false match rate; and FNMR represents false non-match rate.

be generalized to other biometric modalities by employing the appropriate quality-related features. Next, we intend to extend the framework to quality-based multi-modal biometrics fusion by introducing and modeling the relationships among multiple modalities.

One of the open questions is the robustness of quality separation and MLE (EM) model learning schemes to the size of the training set. For example, although continuous quality values used in the third method in Section 5.3 are expected to carry more information than the discrete values used in the fourth method, the performance of the fourth method is better. This might be due to the fact that the discrete case is more robust to the error in the quality separation process. Also, limited number of training data might have caused overfitting and poor estimates in MLE for the

continuous case. A sensitivity analysis in this regard is subject to further research.

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] 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.

[3] T. Cootes, D. Cooper, C. Taylor, and J. Graham. A trainable method of parametric shape description. In *2nd British*
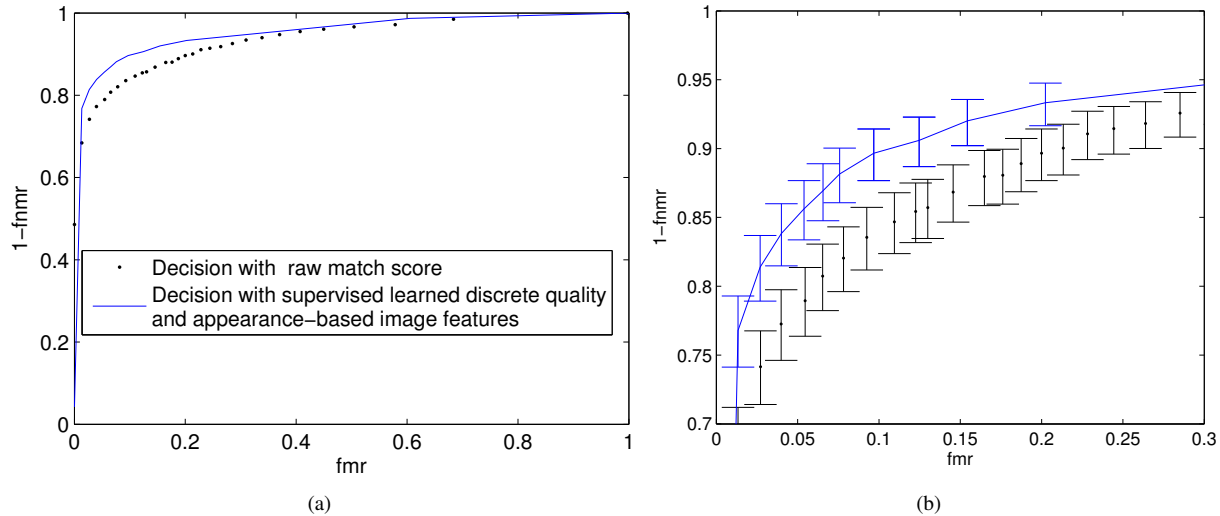
(a)

(b)

Figure 6. (a) Verification ROC curves for the quality-based face recognition using appearance-based image features and method based on raw match score. The x-axis represents the False Matching Rate (FMR), and the y-axis represents the true matching rate (1-FNMR). (b) Zoomed version of (a), showing statistically significant performance improvement over method based on raw match score. On (b) error bars indicate 95% confidence intervals in the FNMR (False Non-Match Rate) estimate.

Machine Vision Conference, 1991.

[4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001.

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

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

[7] 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.

[8] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathe-matical Modelling, Technical University of Denmark, DTU, May 2004.

[9] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[10] 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.

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

[12] 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.