

# A NOVEL APPROACH TO FINGERPRINT IMAGE QUALITY

*Elham Tabassi and Charles L. Wilson*

National Institute of Standards and Technology, Gaithersburg, MD, USA  
{tabassi,cwilson}@nist.gov

## ABSTRACT

We present a novel measure of fingerprint image quality, which can be used to estimate fingerprint match performance. This means presenting the matcher with *good quality* fingerprint images will result in high matcher performance, and vice versa, the matcher will perform poorly for poor quality fingerprints. We discuss the implementation of our fingerprint image quality metric and we present the results of testing it on 280 different combinations of fingerprint image data and fingerprint matcher system. We found that the metric predicts matcher performance for all systems and datasets. Our definition of quality can be applied to other biometric modalities and upon proper feature extraction can be used to assess quality of any mode of biometric samples.

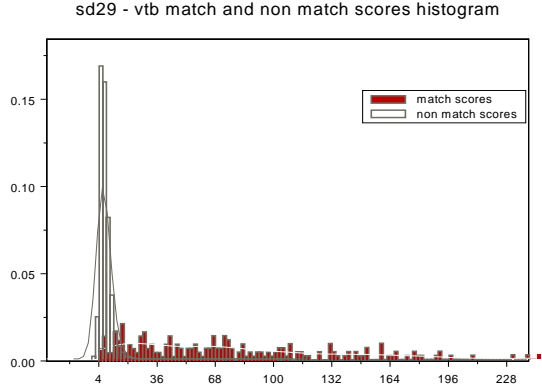
## 1. INTRODUCTION

Automatically and consistently determining the quality of a given biometric sample for identification and/or verification is a problem with far reaching ramifications. If one can identify low quality biometric samples, this information can be used to improve the acquisition of new data. This same quality measure can be used to selectively improve an archival biometric gallery by replacing poor quality biometric samples with better quality samples. The weights for multimodal biometric fusion can be selected to allow better quality biometric samples to dominate the fusion. All of these applications require that the quality of the biometric sample be determined prior to identification or verification. Most of these applications also require that quality of the biometric sample be computed in real-time during data acquisition. The image quality measure presented here meets all the above requirements: it assesses quality of a fingerprint before any matching process, and is fast enough to meet the speed requirement.

A fingerprint is a pattern of friction ridges on the surface of a fingertip. A good quality fingerprint has distinguishable patterns and features that allow the extraction of features, which are useful for subsequent matching of fingerprint pairs. A minutia-based automatic

fingerprint matching algorithm uses features that compare local ridge characteristics (minutia) of two fingerprints  $x_{g(i)}$  and  $x_{p(j)}$  and produces a real-valued similarity score  $s_{ij}$ , where subscript  $g(i)$  denotes  $i$ -th gallery (file) and  $p(j)$  denoted  $j$ -th probe (search) fingerprint images and  $s_{ij}$  is the similarity score of the  $i$ -th gallery and the  $j$ -th probe samples. We call similarity scores of a genuine (i.e. same person) comparisons *match scores*, and similarity scores of imposter (i.e. different person) comparisons *non-match scores*. A higher similarity score is construed to indicate a higher likelihood that the samples come from the same individual. Let  $s_m(x_i)$  denote the match score for sample  $x_{p(i)}$  and  $s_n(x_{ji})$  the non-match scores of  $x_{p(i)}$  and  $x_{g(j)}$ , with  $i \neq j$ . Let  $M(s_m)$  denote the cumulative distribution function (CDF) of the match scores, and  $N(s_n)$  the CDF of non-match scores. The Detection Error Tradeoff characteristic (DET) is a plot of the false non-match rate,  $FNMR=M(s_m)$  against the false match rate,  $FMR=1-N(s_n)$  for all values of  $s_m$  and  $s_n$ . The DET, and the equivalent ROC, are the commonest statement of performance of a verification system. We define *fingerprint image quality as a predictor of matcher performance before a matcher algorithm is applied*. This means presenting the matcher with good quality fingerprint images will result in high matcher performance, and vice versa, the matcher will perform poorly for poor quality fingerprints. Despite some ongoing and past efforts in the investigation of fingerprint image quality [5,6,7], to our knowledge, nobody has publicly defined fingerprint image quality as a scalar predictor of matcher performance. Before proceeding any further, we need to quantify matcher performance.

The similarity score is the ultimate statement of expected performance: in conjunction with the underlying match and non-match distributions it yields likelihood for the samples coming from the same person or different people. Figure 1 shows the histogram of NIST VTB[2] match and non-match scores of fingerprint impressions of 216 subjects in data set NIST SD29. It is typical for the match distribution to be wider than non-match distribution and it is quite typical for the two distributions to overlap. The overlap of match and non-match distributions means a given sample  $x_i$  will match falsely if its match score  $s_m(x_i)$  is less than some non-match scores  $s_n(x_{ji})$ . If the quality measure  $q$  is to be predictive of a matcher,



**Figure 1 - SD29 VTB match and non-match scores histogram**  
 performance, *good* quality fingerprints must be those with high match scores and well separated from the non-match distribution. Similarly, *poor* quality fingerprints are those with lower match scores, in particular those whose match scores are in the region of overlap with non-match scores. Therefore, the quality measure  $q$  should be indicative of the degree to which the match distribution  $M(s_m)$  is separated from the non-match distribution  $N(s_n)$ . Specifically, we define the quality  $q_i$  of biometric sample  $x_i$  to predict

$$o(x_i) = [\sigma(s(x_{ji}))]^{-1} (s_m(x_i) - E[s_n(x_{ji})]) \quad \forall x_i \in \mathcal{G} \text{ or } \mathcal{P} \quad (1)$$

where  $E[.]$  is the mathematical expectation, and  $\sigma(.)$  is standard deviation. Comparing a probe sample  $x_i$  with an internal gallery of  $G$  samples, which includes one and only one sample from the same subject (person), results in a vector of  $G$  scores,  $s$ . Only one element of vector  $s$  is  $x_i$ 's match score, and the other  $G-1$  are its non-match scores.  $E[.]$  is evaluated by computing mean of all non-match scores of probe sample  $x_i$  to all  $G-1$  non-matching gallery entries. Likewise,  $\sigma(.)$  is standard deviation estimated solely from the non-matching elements of  $s$ . We call  $o(x_i)$  *normalized match score* of sample  $x_i$ . Basically, we are comparing the subject's biometric sample to the claimed match sample and to other non-matching samples, and adjusting the raw score on the basis of the extra scores.

## 2. FINGERPRINT IMAGE QUALITY

For each fingerprint, we define its image quality as the prediction of its normalized match score. Similarity scores, and therefore normalized match scores (1) are functions of both probe and gallery samples, but quality as defined here is a scalar value which is measured for each sample separately. Therefore, pairwise quality  $q=H(q_{gallery}, q_{probe})$  should be predictive of recognition performance of pair  $(x_{gallery}, x_{probe})$ . Extensive testing at NIST [1,2] has shown that recognition errors are triggered

by low quality samples. That is,  $H(.)$  is simply the minimum of the individual numbers  $q_{probe}$ , and  $q_{gallery}$ , and so pairwise quality is defined as  $q=\min(q_{gallery}, q_{probe})$ . In an operational setting, if the enrolled samples are assured to have high quality, then a measurement of quality of a subject's biometric sample (probe) can be sufficient for predicting its normalized match score. We measure (scalar value) quality  $q_i$  for biometric sample  $x_i$  by first computing a feature vector  $v_i$  which contains appropriate signal or image fidelity characteristics of  $x_i$ , and then finding some (nonlinear) mapping from  $v_i$  to  $o(x_i)$ . Mathematically speaking

$$v_i=L(x_i) \text{ and } q_i=o^*(x_i)=I(v_i) \quad (2)$$

The function  $L(.)$  will be realized by computing characteristics and features of  $x_i$  that convey information useful to a matching algorithm. Applying  $L(.)$  to a sample  $x_i$  results in an  $n$ -dimensional feature vector  $v_i$ . For fingerprints, this includes measured clarity of ridges and valleys, size of the image, and measures of number and quality of minutiae. The function  $I(.)$  is a mapping from the space of feature vectors  $v$  to normalized match scores  $o(.)$ .  $o^*(x_i)$  is the predicted value for  $o(x_i)$ . Equation 2 suggests use of various regression methods to estimate the response function  $o(.)$  from a vector of variables  $v$ . We tried various regression methods and failed to find a good fit primarily because sample-specific measures (components of vector  $v$ ) are not linearly related to the response variable  $o(.)$  and hence nonlinear functions have to be found, outliers heavily influence data, and residual errors are not Gaussian for any of the regression methods tried. In addition, it is sufficient to know the level of quality (e.g. high, good, medium, low) and, since quality is defined as the prediction of  $o(.)$  (i.e. normalized match scores), it is sufficient to know the range of  $o(x_i)$  (e.g. high, good, medium, low) rather than its exact value for each sample  $x_i$ . These facts lead us to restate the problem in terms of classification. This means we define sample quality as a measure that predicts the *bin*  $o(x_i)$  falls in. Now function  $I(.)$  from Equation 2 is basically a classifier that maps feature vector  $v_i$  of sample  $x_i$  to a quality number  $Q$  such that  $o(x_i)$  falls in the  $k$ -th bin where, without loss of generality,  $Q = k$ . That is, the quality number  $q_i$  of sample  $x_i$  is the bin in which  $o(x_i)$  will fall.

The number  $K$  of allowed bins constitutes a coarseness parameter, against which the quality number can be traded off. For example, a high / low level ( $K = 2$ ) is easier to achieve than the continuous case, where  $K$  tends to infinity (regression). In this paper we chose  $K = 5$ , that is we have five levels of quality: poor ( $q=5$ ), fair ( $q=4$ ), good ( $q=3$ ), very good ( $q=2$ ), and excellent ( $q=1$ ). The choice of 5 levels of image quality is a compromise. Studies at NIST [2] show that eight to ten levels would be needed to fully characterize a matcher that is very sensitive to image

Q	QUALITY	RANGE
5	POOR	$[0, W^{-1}(0.75)]$
4	FAIR	$(W^{-1}(0.75)], C^{-1}(0.05)]$
3	GOOD	$(C^{-1}(0.05), C^{-1}(0.2)]$
2	VERY GOOD	$(C^{-1}(0.2), C^{-1}(0.6)]$
1	EXCELLENT	$(C^{-1}(0.6), C^{-1}(1)]$

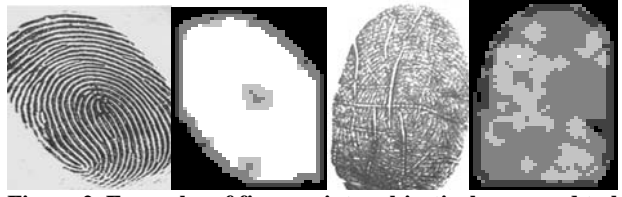
**Table 1. Bin boundary for normalized match scores  $o(\cdot)$ . The boundaries were set by inspection to give useful categorization of the normalized match scores statistic.**

quality. Three levels (good, bad, reject) of quality are sufficient to characterize matchers that are not very sensitive to image quality. We have selected five levels, using irregularly spaced quantiles of  $o(\cdot)$  as  $o(\cdot)$ 's bin boundaries. Table 1 shows the  $o(\cdot)$  bin boundaries, where  $W(\cdot)$  and  $C(\cdot)$  denote the CDF of falsely (wrongly) and correctly matched samples, respectively. Our bin boundary selection agrees with our definition of quality; we are labeling samples with poorest recognition rate, and hence matched falsely, as poor. Samples with fair quality are those that are borderline; some of these samples are matched falsely. Most of the good quality samples are matched correctly, and very good and excellent are those samples that are almost entirely matched correctly. In other words, excellent quality samples are those with very high FNMR and poor quality samples are those with high FMR.

### 3. IMPLEMENTATION

In this section we discuss our implementation of  $L(\cdot)$  and  $I(\cdot)$  in Equation 2 for fingerprint images. We first apply  $L(\cdot)$  to a biometric sample  $x_i$  to get feature vector  $\mathbf{v}_i$  and then use  $\mathbf{v}_i$  as input to a neural network  $I(\cdot)$ .  $L(\cdot)$  is realized by computing the characteristics and features of biometric sample  $x_i$  that convey information useful to a matching algorithm. It is known that fingerprint matcher algorithms commonly in use are sensitive to clarity of ridges and valleys, measures of number and quality of minutiae, and size of the image. We have used NIST Fingerprint Image Software (NFIS2) [3] package to extract features, thereby implementing  $L(\cdot)$ . The MINDTCT package of NFIS2 has a fingerprint minutia detector algorithm that reads an ANSI/NIST formatted file and searches the file structure for a grayscale fingerprint record, automatically detects minutia, assesses minutia quality, and generates an image quality map.

MINDTCT generates the quality map by measuring the quality of localized regions in the image including determining the directional flow of ridges and detecting regions of low contrast, low ridge flow, and high curvature. These last three conditions represent unstable areas in the image where minutiae detection is unreliable. The information in these regions is integrated into one



**Figure 2. Examples of fingerprints subjectively assessed to be of good and poor quality and their grayscale quality maps**

general map that contains 5 levels of quality (4 being the highest quality and 0 being the lowest). The quality assigned to a specific block is determined based on its proximity to the blocks flagged in the above mentioned regions. The background has a score of 0, and a score of 4 means a very good region of fingerprint. Figure 2 shows quality maps grayscale image with black, dark gray, medium gray, light gray, and white corresponding to scores of 0 to 4, respectively, for two fingerprints of different quality. It is notable that the gray scale quality map image is mostly dark gray or black for the poor quality print, and mostly white for the good quality print. For each fingerprint we used MINDTCT to generate its quality map. Blocks with quality 0 are regarded as background. We compute the total number of blocks with quality 1 or better as the effective size of the image or foreground. Then percentages of foreground blocks with qualities 1, 2, 3, or 4 are computed. We call them quality zones 1, 2, 3, and 4, respectively. Fingerprint images with higher number of quality zone 4 (equivalently smaller number of quality zone 1 and/or 2) are more desirable.

NFIS2 also computes a quality/reliability associated with each detected minutia point. Two factors are combined to produce a quality measure for each detected minutia point. The first is taken directly from the location of the minutia point within the quality map described above. The second factor is based on simple pixel intensity statistics (mean and standard deviation) within the immediate neighborhood of the minutia point. Based on these two factors, NFIS2 assigns a quality value in the range 0.01 to 0.99 to each minutia.

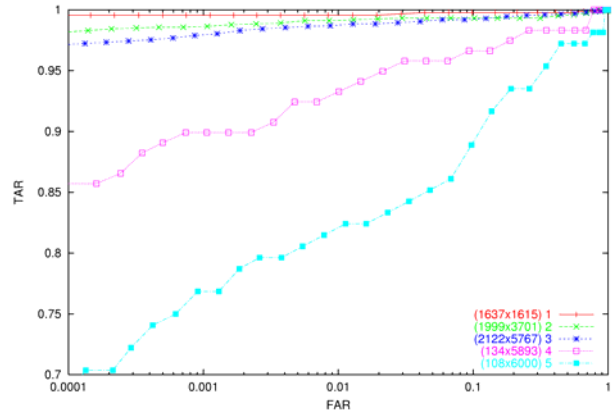
We used the quality map and minutia quality assessment of NFIS2 to define our feature vectors as shown in Table 2. Therefore, for each fingerprint image, an 11-dimensional feature vector is computed using MINDTCT of NFIS2. We performed exploratory data analysis and found that distribution of none of the components of our feature vector is normal or even nearly normal and the correlation factors of none of the features are strong enough to predict the normalized match score all by itself. However, there is strong evidence that some linear or nonlinear combination of these features can predict normalized match scores defined in Equation 1. We chose the artificial neural network as the nonlinear classification method. The neural network has the capability of acting as an approximation function for any

	NAME	DESCRIPTION
1	foreground	#of blocks that have quality 1 or better
2	total #of minutia	# of total minutiae found in the fingerprint
3	min05	# of minutiae that have quality 0.5 or better
4	min06	# of minutiae that have quality 0.6 or better
5	min075	# of minutiae that have quality 0.75 or better
6	min08	# of minutiae that have quality 0.8 or better
7	min09	# of minutiae that have quality 0.9 or better
8	quality zone 1	percentage of the foreground blocks of quality map with quality =1
9	quality zone 2	percentage of the foreground blocks of quality map with quality =2
10	quality zone 3	percentage of the foreground blocks of quality map with quality =3
11	quality zone 4	percentage of the foreground blocks of quality map with quality =4

**Table 2. Feature vector description**

arbitrary nonlinear function and is not dependent on model-based distribution functions in either the feature or the classification space. We used the neural network implementation in NFIS2 package. The theory behind the machine learning techniques used in this program is discussed in[4]. It trains a 3-layer feed-forward nonlinear perceptron model. The five classes of quality explained in Table 1 are the neural network outputs. The input to neural network is the 11-dimension feature vector discussed in Table 2. We chose 22 hidden nodes. Boltzmann pruning, i.e. dynamic removal of connections, is performed during training. The activation functions used for hidden nodes as well as output nodes are sinusoids. NIST has acquired a collection of live-scan and scanned paper fingerprints datasets collected at different operational settings from different government agencies [2]. We used a subset of these datasets (5244 images) for training and the rest (234700) for testing our system. Care was taken to design a training set balanced in terms of numbers of different finger positions of different quality chosen from different datasets.

Our evaluation criterion is ranked ROC as a function of image quality. Given a quality number taking on  $K$  integral values,  $K$  ROC characteristics are generated. Similarity scores of probe images of quality  $q_k$  and gallery images of quality  $q_k$  or better are used in the computation of the  $k$ -th ROC. As it is shown in Figure 3, the highest recognition performance is achieved for samples with quality 1 (excellent), and samples with quality 5 (poor) have the lowest performance. We have used similarity scores of 14 different fingerprint systems supplied by 8 commercial fingerprint vendors[2] and fingerprint images of 20 different fingerprint data sets to test our fingerprint image quality metric. Our fingerprint quality measure successfully predicts matcher performance in all 280 cases.



**Figure 3. ROC as a function of image quality for right index fingerprints of 6000 subjects with 2 prints per subject, therefore a similarity matrix of size 6kx6k.**

#### 4. CONCLUSION

We presented a novel measure of fingerprint image quality, which can be used to estimate fingerprint match performance. Results of extensive test of our system showed that it generalizes very well to other matchers and other datasets. Our implementation is publicly, but export controlled, available as part of NFIS2.

#### 5. REFERENCES

- [1] R. A. Hicklin, H. Korves, B. Ulery, M. Zoepfl, M. Bone, P. Grother, R. Micheals, S. Otto, C. Watson, *Fingerprint Vendor Technology Evaluation (FpVTE) NISTIR 7123*, NIST, Gaithersburg MD, 2003.
- [2] C. Watson, A. Wilson, K. Marshall, M. Indovina, R. Snelick, *Studies of one-to-one fingerprint matching with vendor SDK matchers NISTIR 7119*, NIST, Gaithersburg MD, 2003.
- [3] M.D. Garris, C.I. Watson, R.M. McCabe, C.L. Wilson, *User's guide to NIST fingerprint image software NISTIR 6813*, NIST, Gaithersburg MD, 2004.
- [4] C.L. Wilson, J.L. Blue, O.M. Omidvar, "Training Dynamics and Neural Network Performance", *Neural Networks*, vol. 10, No. 5, pp. 907-923, 1997.
- [5] E. Lim, X. Jiang, W. Yau, "Fingerprint quality and validity analysis", *IEEE ICIP*, 2002.
- [6] N.K. Ratha and R.M. Bolle, "Fingerprint image quality estimation", *IBM computer science research report RC21622*, 1999.
- [7] L.L. Shen, A. Kot and W.M. Koo, "Quality measures of fingerprint images", *3rd international conference AVBPA 2001*, p182-271, June 2001.