Quantifying Quality: A Case Study in Fingerprints

Michael Yi-Sheng Yao, Sharath Pankanti, Norman Haas, Nalini Ratha, Ruud M. Bolle
IBM T J Watson Research Center
Yorktown Heights, NY 10598

Abstract

For a particular biometric to be effective, it should be universal: every individual in the target population possesses the biometrics and every acquisition from each individual provides useful information for personal identity verification or recognition. In other words, everybody should have the biometrics and it should be easy to sample or acquire. In practice, adverse signal acquisition conditions and inconsistent presentations of the signal often results in unusable or nearly unusable biometrics signals (biometrics samples). This is confounded by the problem that the underlying individual biometrics signal can vary over time due, for example, to aging. Hence, poor quality of the actual machine sample of a biometrics constitutes the single most cause of poor accuracy performance of a biometrics system. Therefore, it is important to quantify the quality of the signal for either seeking a better re-presentation of the signal or for subjecting the poor signal to alternative methods of processing (e.g., enhancement). In this paper, we explore a definition of quality of fingerprint impressions and present detailed algorithms to measure image quality. The proposed quality measure has been developed with the use of human annotated images and tested on a large number of fingerprints of different modes of fingerprint acquisition methods.

1 Introduction

Because of adverse and/or hostile signal and image acquisition situations, a biometrics system performance suffers from random false rejects/accepts. In a deployed system, poor acquisition of samples perhaps constitutes the single most important reason for high false reject/accept rates. There are two solutions to this. One can either probabilistically model and weigh all the adverse situations into the feature extraction/matching system or one can try to dynamically and interactively obtain a desirable input sample.[1]

Automatic implementation of either strategy entails algorithmically characterizing what one means by *desirable* pattern samples and by quality of pattern samples. The term "quality" is then somehow related to "process-ability" of the sample. That is, the system faces difficulty in analyzing poor quality samples and performs well when presented with good quality samples. It is important to quantify the concept of quality so that consistent actions can be applied to samples having different degrees of undesirability. Additionally recognizing specific reasons by which a sample is undesirable (e.g., partial face view) so that an appropriate corrective action can be suggested or taken by the system (e.g., apply enhancement) or by the subject (e.g., present the biometrics in a different, "better" way). Finally, it is desired that assessment of quality be fast, and not require actually matching the given sample either with its mates or with its non-mates. Often, one single best sample is desired.

Gracefully and conveniently handling biometrics samples of diverse quality is of great significance to any practical biometrics identity system. However, in theory, for almost all applications it is possible to compromise some convenience (ease of use) by accepting only a certain quality of input. Therefore, it comes as no surprise that almost all operational biometrics systems have an implicit or explicit strategy for handling samples of poor quality. Some of the simplest measures for quality control include provision for presenting multiple samples to the system. In such systems, it is hoped that multiple opportunities to present a sample may alleviate the problem of overcoming occasional poor quality sample. In other schemes, the system provides a user with live visual display of the biometrics that has been sampled (which is of course not practical for all biometrics). It is expected that this feedback to the user will provide an opportunity to self-correct some of the simple mistakes such as improper placement of a finger or coughing during a speech sample. This feedback is of course especially important during the enrollment process. Enrolling undesirable samples here can cause lasting, more serious impact on system accuracy.

The rest of this paper concentrates on fingerprints. Sections 2 presents quality problems specific to fingerprint domain. Section 3 describes in detail our algo-

[1] This is only possible in interactive overt online systems involving cooperative users.

rithms for quality assessment. Section 4 presents experimental results. Finally, Section 5 provides conclusions and ideas for further research.

2 Assessing Fingerprint Quality

Precise fingerprint image acquisition up to the minutiae has some peculiar and challenging aspects [1], many because of contact[2] problems. Specifically, (i) *Inconsistent contact:* The act of sensing distorts the finger. Determined by the pressure and contact of the finger on the imaging surface (e.g., 2D glass platen), the 3D shape of the finger is mapped onto the 2D surface. Typically, this mapping is uncontrolled and results in different inconsistently mapped regions across impressions. (ii) *Nonuniform contact:* The ridge structure of a finger would be completely captured if the ridges of all imaged parts of the finger were in complete optical contact with the glass platen. In practice, due to various reasons, this is not the case. (iii) *Irreproducible contact:* Manual work, accidents etc. inflict injuries to the finger, thereby, changing the ridge structure of the finger either permanently or semi-permanently. This may introduce additional spurious minutiae or "minutiae-like" features. (iv) The act of sensing itself adds noise to the image. For example, residues leftover from the previous fingerprint capture. Additionally, (v) a typical imaging system distorts the image of the sensed finger due to imperfect imaging conditions. All these factors contribute to poor samples and feature extraction artifacts during image processing and hence increase false accept/reject rates.

Most of the poor quality prints are due to non-uniform and inconsistent contact. Among the prints with inconsistent contact, the undesirability of the differently distorted impressions (due to application of different pressure on the finger) is very difficult to assess without actually matching the prints and hence is not within the scope of quality assessment.[3] Many quality assessment systems however do detect the amount of the print area and the relative placement/orientation of the finger; these systems can provide simple feedback to the user about proper placement of the finger to the image acquisition device but cannot quantify the quality of the fingerprint image itself. More sophisticated systems have an explicit method of quantifying the quality of the fingerprint being captured. In this work, we only address issues related to poor quality fingerprints due to non-uniform contact. Issues related to inconsistent contact area from print to print are deferred to later.

3 Non-uniform Contact

The dryness of the finger skin, skin disease, sweat, dirt, and humidity in the air all contribute to a non-uniform and non-ideal contact situation: some parts of the ridges may not come in complete contact with the platen and regions representing some valleys may come in contact with the glass platen. Non-uniform contact manifests itself in dry prints (too little ridge contact) and smudgy prints (neighboring ridges touching each other obliterating the intervening valleys) or in prints with combinations. Non-uniform contact may result in "noisy" low contrast images and could lead to many feature extraction artifacts, e.g., spurious minutiae or missing minutiae. For instance, in a dry fingerprint (see Figure 1(b)), the ridge is in intermittent contact with the platen; hence dryness of the fingerprint manifests itself in significant variation of pixel intensities along a dry finger ridge. In extreme situations, there is no particularly dominant direction of a very dry ridge because too small a fraction is in contact with the platen.

On the other hand, in a smudgy portion of the fingerprint (see Figure 1(d)), the neighboring ridges touch each other; thus completely obliterating the intervening valley. As a result, variation in pixel intensities across the ridge direction is significantly lower than the typical expected variation across an ideal ridge. In extreme situations, the directionality of the ridges is obliterated due to a large number of ridges touching each other (this is analogous to image saturation).

For the pattern recognition task of fingerprint representations, prints have been (implicitly or explicitly) modeled as smoothly flowing directional textures (ridges) that can be extracted by typical fingerprint feature extraction algorithms [1, 7]. Since directionality of the finger ridges is an essential attribute of its image texture, we propose that this anisotropy constitute a basis for assessing the overall quality of the fingerprint. Below, we summarize an algorithm for fingerprint quality assessment based on the directionality of its texture.

Sub-sampling and blocking: For efficiency reasons, the quality analysis uses a sub-sampled image. The analysis samples the image at rate s in x and y-directions. The sub-sampled image is further divided into square blocks of size B.

Direction and foreground estimation: This step determines if a given block depicts a portion of a fingerprint and extracts a nominal direction from a foreground block. Any number of existing strategies can be adopted [1, 7, 9]. For efficiency reasons, we use the method proposed by Mehtre [10]. At each pixel in a given block, a number of pixels are selected along

[2] The term contact does not necessarily mean physical contact and needs to be appropriately interpreted according to sensing modality.

[3] If the sensor is explicitly equipped with force sensors, the distorted fingerprints can be recognized [6].

a line segment of an orientation (d) and pre-specified length (l) centered around that pixel; variation in the intensities of the selected pixels is then determined by computing the sum of intensity differences $D_d(i,j)$ between the given pixel and the selected pixels,

$$D_d(i,j) = \sum_{(i',j')} |f(i,j) - f_d(i',j')|, \qquad (1)$$

with $d = 0, \pi/n, ..., \pi$ and where $f(i,j)$ is the intensity of pixel (i,j) and $f_d(i',j')$ are the intensities of the neighbors of pixel (i,j) along direction d. This indicates the summation of differences between the given pixel of interest, pixel (i,j), and a number l (say 6) neighboring pixels along each of the directions. The variation in intensities is computed for n discrete orientations; the orientation at a pixel $\hat{d}$ is the orientation of the line segment for which the intensity variation thus computed is minimal.

Regions of background and portions of impressions having faint residual leftover of earlier captured prints on a dirty input device usually exhibit small intensity variation around their neighborhoods. To determine if an image pixel belongs to the background, the intensity variation $D(i,j)$ at the pixel (i,j) of interest is subsequently obtained by summing up the differences in the n directions with $D(i,j) = \sum_d D_d(i,j)$ and when D is less than a background threshold τ for each d, the pixel is classified as a background pixel. When more than a fraction of pixels in a block are background pixels, the block is regarded as background block.

Using connected component analysis, foreground components that are smaller than a certain threshold fraction of the total image area τ_3 are considered spurious. A print with no legitimate foreground area is of poorest quality.

Dominant direction: After the foreground blocks are marked, it is determined if the resulting direction for each block is prominent. The idea is that a block with a prominent direction should exhibit a clear ridge/valley direction that is consistent with most of the pixel directions in the block. Existence of a dominant direction can be assessed by computing a histogram of directions D_d (Expression 1) at each pixel in a given block. If the maximum value of the histogram is greater than a prominent threshold T_1, the block is said to have a dominant direction, and is labeled as prominent. Bifurcations of ridges may often result in two dominant directions in a block. Therefore, if two or more directions of the direction histogram are greater than a bifurcation threshold, $T_2 < T_1$, the corresponding block is labeled as such. A post-processing step removes blocks that are inconsistent with their neighbors. If a "directional" block is surrounded by "non-directional" blocks, it is relabeled as a non-directional block. Similarly, a non-directional block surrounded by neighboring directional blocks is changed to a directional block. Using connected component analysis, finally, regions of dominant blocks with area smaller than a threshold number of blocks β are discarded. The result is that the fingerprint foreground image is partitioned into (i) regions of contiguous blocks with direction and (ii) regions of blocks without direction or non-contiguous blocks with direction.

Quality computation: Since regions (or accordingly minutiae) near the centroid are likely to provide more information for biometrics authentication, the overall quality of the fingerprint image is computed from the directional blocks by assigning relative weight w_i for foreground block i at location x_i given by

$$w_i = \exp\{-||x_i - x_c||^2/(2q^2)\} \qquad (2)$$

where x_c is the centroid of foreground, and q is a normalization constant.

The overall quality Q of a fingerprint image is obtained by computing the ratio of total weights of directional blocks to the total weights for each of the blocks in the foreground, $Q = \sum_{\mathcal{D}} w_i / \sum_{\mathcal{F}} w_i$. Here $\mathcal{D}$ is the set of directional blocks and $\mathcal{F}$ the set of foreground blocks. The quality Q is used as a measure of how much reliable directional information is available in a fingerprint image. If the computed Q is less than the quality threshold, T, the image is considered to be of poor quality.

Dryness and smudginess: Once it is determined that the fingerprint is of a certain poor quality, it is desirable to be able to identify a more specific cause of the low quality. We describe a method of distinguishing smudged poor quality prints from dry poor quality prints based on simple statistical pixel intensity based features. The idea is that for a smudged impression, there are a relatively large number of blocks whose contrast is very small. Similarly, for a dry impression, there are a relatively large number of blocks where the contrasts of their neighbors vary significantly.

First, the mean intensity of pixels within each foreground block is computed. The pixels whose intensities are smaller than the mean intensity of all the pixels in the block are considered pixels on a ridge, i.e., ridge pixels.[4] Let μ be the true mean intensity of ridge pixels. For each block, the mean intensity (μ) is estimated using pixels whose intensities are smaller than the mean intensity of all pixels within the block. Further, the standard deviation (σ) of intensities of all pixels within the same block is determined. For a block with good contrast, μ is small and σ is large; for a block

[4] Assume that the finger imaging depicts ridges as darker pixels.

with low contrast due to smudginess, μ is small and σ is small. The contrast within a block is measured using the product ($c_s = \mu\,\sigma$). If the contrast measure c_s is smaller than a threshold ρ_1, the block is classified as a smudged block. Lastly, the smudginess measure is determined as the ratio of the number of smudged blocks to total number of foreground blocks. If the resulting ratio is larger than a threshold, ρ_2, a smudged impression is reported.

For a block with good contrast, on the other hand, μ is small and σ is large. For a block with low contrast due to dryness, however, μ is large and σ is small. Consequently, to measure the contrast within a block, we compute the ratio of corresponding μ to corresponding σ, i.e., $c_d = \mu/\sigma$ where c_d is the contrast measure. A block is considered to be dry if c_d is greater than a dryness threshold, δ_1. Alternatively, the fingerprint is considered dry if the contrasts of its neighboring blocks vary significantly. Specifically, let the c_{max} (c_{min}) be maximum (minimum) value of contrast difference between the contrast of the given block and those of its neighboring blocks. If the difference between c_{max} and c_{min} is larger than a dryness threshold δ_2, then the block is a dry block. The dryness measure of the image is computed as the ratio of the total number of dry blocks to the total number of foreground blocks. If the resulting measure is greater than a threshold, δ_3, it is reported that a dry impression causes the quality problems.

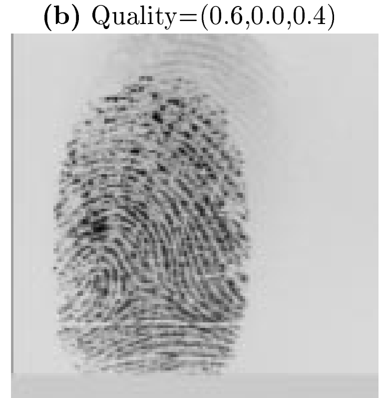
4 Experiments

The algorithm has been implemented and tested on a large number of the fingerprints captured in the laboratory and in the real world using inked, optical live scans, as well as solid-state live scan fingerprint scanners. The algorithms is extremely fast (e.g., less than 100 ms on a 933 MHz Pentium III processor on a 512×512 image). Typical operational parameters for the proposed quality assessment algorithm for and 8-bit, 512 *dpi* 512×512 gray scale images are: $s = 4$ and $B = 8$. Figure 1 illustrates results of fingerprint quality assessment for a few representative optically scanned fingerprint images. The results of the quality assessment typically appear consistent with visual human assessment.

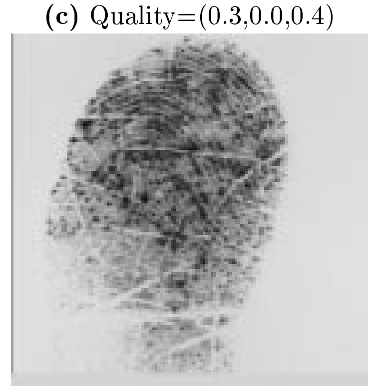
We use the fingerprint image quality assessment to compare fingerprints in three data sets: NIST-9, FVC-2000, and IBM-99. NIST-9 [12] is a public database of 8-bit gray scale inked fingerprint images scanned from mated fingerprint card pairs. The database contains 90 mated card pairs of segmented 8-bit gray scale fingerprint images (900 fingerprint image pairs per CD-ROM). Each segmented image is 832 by 768 pixels and classified using the National Crime Information Center

(a) Quality=(0.9,0.0,0.0)

(b) Quality=(0.6,0.0,0.4)



(c) Quality=(0.3,0.0,0.4)



(d) Quality=(0.1,0.2,0.5)

Figure 1: A fingerprint quality assessment measure. Quality (x, y, z) indicates print of overall quality x, smudginess y, and dryness z.

(NCIC) classes given by the FBI. Only the first 600 images have been arbitrarily sampled from the entire data set for quality assessment. FVC-2000 is another public database. It is one of the databases used in a recent fingerprint verification contest [11] and was collected using optical sensing techniques. This database contains impressions of 110 fingers with eight impressions per finger. This set was acquired from a set of 19 volunteers in the 5-73 age range (55% male). Without interleaving, two images of six fingers were taken from each volunteer per session. Each volunteer attended four sessions, with no more than two sessions per day, and depending on the volunteer, over a span of three days to three months. Much care was given to cleaning the sensor and the fingers between acquisitions. Some visual inspection of the data set was performed to ensure minimal overlap and not too much rotation between impressions. Database IBM-99 is a private database. This database was acquired from a group of 57 subjects in two sessions five weeks apart. There were approximately the same number of adult males and females and the age group is 22-65 year old. During each session, five prints of both the left and right index finger were obtained for each subject. Hence, the database contains a total of 1,140 impressions, 10 prints of 114 fingers. All prints in data set IBM-99 were collected using the same optical fingerprint scanner with 512 dpi image resolution.

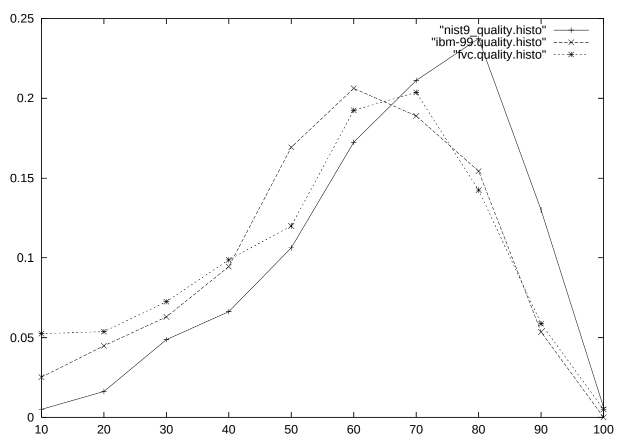


Figure 2: Image quality histograms for three fingerprint data sets: NIST-9, FVC-2000, and IBM-99.

Histograms of fingerprint qualities for the fingerprints in the data sets are presented in Figure 2. It can be clearly seen that the inked prints taken under careful supervision of trained personnel (e.g., NIST-9) appear to be of significantly better quality than the live scan prints acquired under less supervised conditions. It is also interesting to note that the optical live scan fingerprints in FVC-2000 and IBM-99 data sets appear

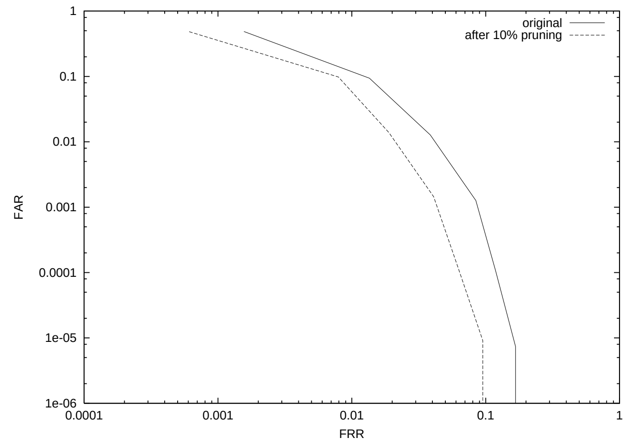


Figure 3: Effect of quality pruning on fingerprint verification accuracy performance.

to be of approximately comparable quality. However, FVC-2000 may contain slightly poorer overall quality prints than IBM-99 (note the higher frequency of lower quality prints in FVC-2000) possibly due to visual preselection used by the FVC-2000 designers ("three bears rule"). Even though the absolute values of quality measurements are somewhat arbitrary, a consistent evaluation can be used to assess/compare biometric data sets [14].

Does the quality as computed by the quality assessment algorithm really reflect on the desirability of the print to be processed by an automatic fingerprint verification system? To investigate this issue, we plot ROC curves showing accuracy performance of a fingerprint matcher [13] for the IBM-99 data set. Subsequently, using the proposed quality assessment algorithm, some of the worst quality prints (10%) from IBM-99 data set were culled from the data set; the ROC curve was recomputed using the remaining prints of IBM-99. Figure 3 presents effect of culling fingerprints of poor quality from IBM-99. As expected, the accuracy of the matcher did indeed significantly improve.

5 Conclusions and Future Research

Automatically and consistently determining suitability of a given input measurement (biometrics sample) for automatic identification is a challenging problem. This paper proposes a method of assessment of quality of a fingerprint based on a simple ideal fingerprint image model. We present an algorithm for overall quality assessment. We further propose a basis for categorizing a poor quality fingerprint into either "dry" or "smudged" prints using simple local statistical measures of the pixel intensities. Algorithms for assessment of dryness and smudginess indices are also presented.

The qualities of images in a few databases have been computed. The quality assessment results appear to be mostly consistent with visual quality assessment. A more objective evaluation of the quality assessment is a topic of our future research.

There are a number of limitations to the impression quality measure. Most significantly, the proposed quality assessment algorithms are based on an intuitive domain-specific strategy, and it appears that a more generic framework for quality assessment would be a more useful tool for realizing quality assessment strategies for different representations of fingerprints and for different biometric identifiers. Secondly, the proposed algorithms are meant to present an overall judgment of the quality of the fingerprint and are incapable of providing pointers to the most problematic portions of the finger. Finally, it is to be noted that the test for fingerprint quality does not imply existence of a fingerprint. The present model is sufficiently naive (for efficiency reasons) and many non-fingerprint images may successfully pass the fingerprint quality test.

Matcher/representation independent description of complexities of biometric data sets has not received much attention in the literature. We hope that objective assessment of the quality assessment through careful ground-truth marking, development of generic frameworks for data-modeling and of the quality assessment will be not only open new approaches to comparing/assessing biometric data sets but also to characterizing the different categories of patterns [8]. For any specific applications, regulating the sample quality in effect results in artificially increasing the failure to enroll (FTE) rate. For applications where convenience is the motivating factor, this may not be a problem because increasing the FTE rate only decreases the pool of subjects that can "smoothly" use the installation. Enforcing quality control while maintaining a low FTE rate, for security specific applications, on the other hand, invariably will imply that the application becomes less convenient for many subjects. This is compounded by the fact that enrollment of poor samples will result in higher false accept and false reject rates caused by these samples.

References

[1] A. Jain, L. Hong, S. Pankanti, R. Bolle, "Identity Authentication Using Fingerprints", Proceedings of the IEEE, Vol. 85, No. 9, pp. 1365–1388. 1997.

[2] R. M. Bolle, S. Pankanti, Y-S. Yao, System and Method for Determining if a Fingerprint Image Contains an Image Portion Representing a Partial Fingerprint Impression, U. S. Patent No. 6,005,963, Dec. 21, 1999.

[3] R. M. Bolle, S. Pankanti, Y-S. Yao, System and Method for Determining the quality of Fingerprint Images, U. S. Patent No. 5,963,656, Oct. 5, 1999.

[4] R. M. Bolle, S. Pankanti, Y-S. Yao, System and Method for Determining if a Fingerprint Image Contains an Image Portion Representing a Dry Fingerprint Impression, U. S. Patent No. 5,995,640, Nov. 30, 1999.

[5] R. M. Bolle, S. Pankanti, Y-S. Yao, System and Method for Determining if a Fingerprint Image Contains an Image Portion Representing a Smudged Fingerprint Impression, U. S. Patent No. 5,883,971, Mar. 16, 1999.

[6] J. L. Levine, M. A. Schappert, N. K. Ratha, R. M. Bolle, S. Pankanti, R. S. Germain, and R. L. Garwin, System and Method for Distortion Control in Live-Scan Inkless Fingerprint Images, U. S. Patent No. US06064753, May 16, 2000.

[7] N. K. Ratha and S. Chen and A. K. Jain, Adaptive flow orientation based texture extraction in fingerprint images, Pattern Recognition, vol. 28, no. 11, pp. 1657-1672, November, 1995.

[8] S. Pankanti, N. K. Ratha, and R. Bolle, Pattern Specific Inhomogeneity, Submitted to ICPR2002, Quebec City, Canada.

[9] D. Maio and D. Maltoni, Direct gray-scale minutiae detection in fingerprints, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(1), Jan. 1997, pp. 27-40.

[10] B.M. Mehtre, Fingerprint Image Analysis for Automatic Identification, Machine Vision and Applications, Springer-Verlag, vol. 6, pp. 124–139, 1993.

[11] D. Maio, D. Maltoni, R. Capelli, J.L. Wayman, and A.K. Jain. FVC2000: Fingerprint verification competition. Technical report, Univ. of Bologna, September 2000.

[12] C. I. Watson, NIST Special Database 9, Mated Fingerprint Card Pairs, National Institute of Standards and Technology, 1993.

[13] N. Ratha, K. Karu, S. Chen, and A. K. Jain, A Real-time Matching System for Large Fingerprint Database, IEEE Trans. on PAMI, Vol. 18, No. 8, pp. 799–813, 1996.

[14] R. Bolle, N. Ratha, and S. Pankanti, Evaluating Authentication Systems Using Bootstrap Confidence Intervals, *Proc. 1999 IEEE Workshop on Automatic Identification Advanced Technologies* (Morristown, NJ), October 28-29, 1999, pp. 9–13.