

Assessing Fingerprint Individuality In Presence Of Noisy Minutiae

Sarat C. Dass

Abstract

Fingerprint image quality is an important source of intra-class variability. When the underlying image quality is poor, human experts as well as automatic systems are more likely to make errors in minutiae detection and matching by either missing true features or detecting spurious ones. As a consequence, fingerprint individuality estimates change depending on the quality of the underlying images. The goal of this paper is to quantitatively study the effect of noise in minutiae detection and localization, resulting from varying image quality, on fingerprint individuality. The measure of fingerprint individuality is modeled as a function of image quality via a random effects model and methodology for the estimation of unknown parameters is developed in a Bayesian framework. Empirical results on two databases, one in-house and another publicly available, demonstrate how the measure of fingerprint individuality increases as image quality becomes poor. The measure corresponding to the “12-point match” with 26 observed minutiae in the query and template fingerprints increases by several

Sarat C. Dass is with the Department of Statistics & Probability, A430 Wells Hall, Michigan State University, East Lansing, MI 48824. Phone: 517-355-9589. Fax: 517-432-1405. Email: sdass@msu.edu. This research is partially supported by the NSF DMS grant 0706385.

orders of magnitude when the fingerprint quality degrades from “best” to “poor”.

Index Terms

Biometric authentication, fingerprint-based authentication, fingerprint individuality, fingerprint image quality.

I. Introduction

Fingerprint individuality is the study of the extent of uniqueness of fingerprints in a target population, and is the central premise for expert testimony in court. Fingerprint experts compare the salient features, typically minutiae, of a latent print lifted from a crime scene with those taken from the defendant. A reasonably high degree of match between minutiae leads the experts to testify irrefutably that the owner of the latent print and the defendant are one and the same person. Central to establishing an identity based on fingerprint evidence is the assumption of discernible uniqueness; fingerprint minutiae of different individuals are observably different, and therefore, when two prints share many common minutiae, the experts conclude that the owners of the two different prints are one and the same person.

For decades, the testimony of fingerprint experts was almost never excluded from these cases, and on cross-examination, the foundations and basis of this testimony were rarely questioned. A significant event that questioned this trend occurred in 1993 in the case of *Daubert vs. Merrell Dow Pharmaceuticals* [1], where the U.S. Supreme Court ruled that in order for expert forensic testimony to be allowed in courts, it had to be subject to the criteria of scientific validation (see [2] for details). Following the *Daubert* ruling, forensic evidence based on fingerprints was first challenged in the 1999 case of *U.S. v. Byron C. Mitchell* [3], and subsequently, in 20 other cases involving fingerprint evidence. The main concern with the admissibility of fingerprint evidence is the problem of individualization, namely, that the fundamental premise for asserting the uniqueness of fingerprints has not been scientifically validated and matching error rates are unknown [2], [4].

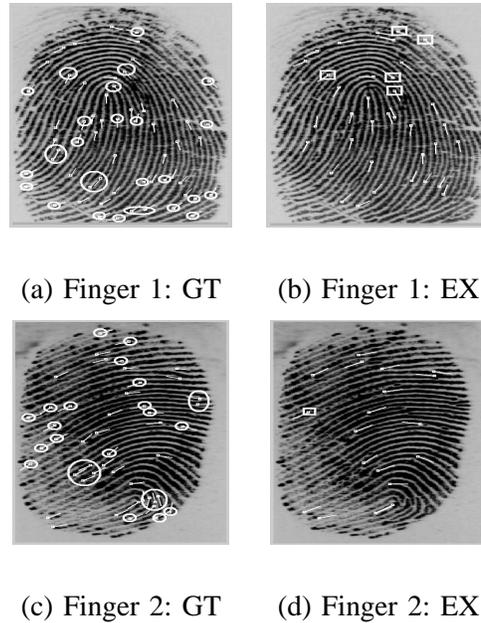


Fig. 1. Spurious and missed minutiae in two fingerprint images in the IBM database [2].

More recently, in the Bryan K. Rose case in 2008 [5], a Baltimore County Circuit judge ruled that fingerprint evidence was not reliable enough to be used in a capital murder case, and she denied the prosecutors' request to reconsider that decision.

The central question in a court of law is: What is the uncertainty associated with the experts' judgement when matches are decided by fingerprint evidence? How likely can an erroneous decision be made for the given latent print? For example, if the latent print is of poor quality, then it is reasonable to believe that the match made by the experts is more prone to error. The main issue with expert testimony is the lack of quantification of the uncertainty in their decision. It is not known how fingerprint individuality estimates change as a function of the image quality, given the uncertainty present in recovering the true minutiae.

A measure of fingerprint individuality is given by the probability of a random correspondence (PRC), which is the probability that two minutiae, one from the query and the other from the template fingerprint, randomly corresponds with each other. Spurious features have the effect of increasing random correspondences, whereas missed minutiae decreases the number of true correspondences. These two types of errors, in turn, affect the

PRC as a function of image quality. Figure 1 shows two fingerprint images from the IBM database [2]. The left⁴ and right panels of Figure 1 illustrate two different types of ‘extractors’ used for these images. The notations ‘GT’ (for panels (a) and (c)) and ‘EX’ (for panels (b) and (d)), respectively, indicates minutiae extraction by manual inspection and by an automatic algorithm reported in [6]. Minutiae in panels (a) and (c) (respectively, (b) and (d)) enclosed in a circle (respectively, square) are missed (respectively, spurious) minutiae. Figure 2 shows the random correspondences generated when noisy minutiae is used (i.e., EX panels) compared to ground truth minutiae (GT panels). Panel (a) (respectively, (b)) matches the ground truth (extracted) minutiae of Fingers 1 and 2 in Figure 1. Clearly, the information on the number of matches is different. Nine true correspondences are obtained for panel (a). Only one true correspondence (number 4) is recovered in panel (b) while the remaining true correspondences are not recovered at all due to missed minutiae. Further, three false correspondences are generated due to spurious minutiae in panel (b). Finally, Figure 3 shows the classification of minutiae (either missed or spurious) as a function of the underlying image quality. Panels (a) and (c) (respectively, (b) and (d)) show GT (respectively, EX) minutiae for two fingerprint images with the image in the top row having a better quality measure. As image quality degrades, more missed and spurious minutiae are generated resulting in a greater number of missed and spurious correspondences.

The goal of this paper is to extend the framework for assessing fingerprint individuality of [4] by modeling errors in minutiae detection and localization. The aim is to analyze the probability of obtaining random correspondences in the presence of noisy minutiae (Figure 2(b)) as opposed to ground truth minutiae (Figure 2(a)). The methodology involves two main steps: First, a flexible class of statistical models is elicited for noisy minutiae, and used to derive PRCs. Second, the estimated PRCs are modeled as a function of the quality of the fingerprint pair via a log-linear random effects model. Table I gives an outline of our work in comparison to two previous models of fingerprint individuality, [2] and [4].

The rest of this paper is organized as follows: Section II gives a brief introduction to the modelling of fingerprint minutiae in terms of statistical mixture models. Section III discusses the log-linear random effects

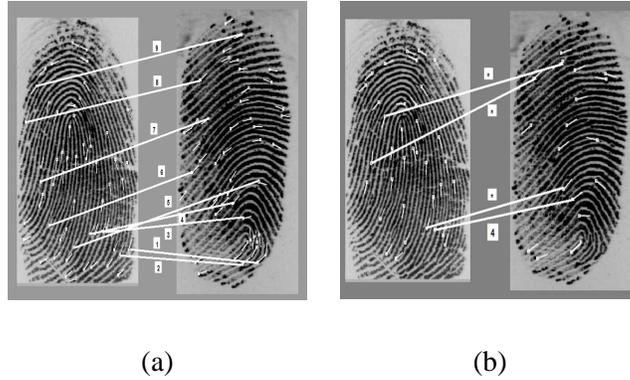


Fig. 2. Effect of spurious and missed minutiae on random correspondences. Panels (a): GT and (b):

EX minutiae

Source of fingerprint variability	Pankanti et al. [2]	Zhu et al. [4]	Proposed approach
Minutiae variability within a finger	Y	Y	Y
Minutiae variability in different fingers	N	Y	Y
Spurious and missed minutiae	N	N	Y

TABLE I. Summary of sources of variability considered in three studies on fingerprint individuality.

Y=Yes, N=No.

model that is used to study how the probability of random correspondences change as a function of the underlying image quality. Section IV presents the methodology for fitting the log-linear models in a Bayesian framework. Section V presents the fitting of the models to two databases, namely, the IBM ground truth and NIST fingerprint databases. Several tests for the goodness of fit of the log-linear random effects models to the two databases is presented. Section VI presents the summary and conclusion.

II. Mixture Models on Minutiae

The first stage of our methodology involves fitting mixture models to the distribution of minutiae in each fingerprint image. The motivation for using mixture models is that they are a flexible family of distributions capable of capturing salient characteristics of minutiae occurrence; for example, clustering tendencies and

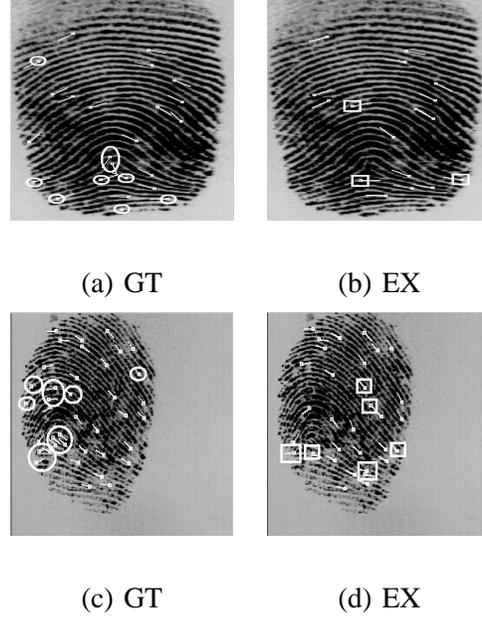


Fig. 3. Spurious and missed minutiae in two fingerprint images of different quality in the IBM database.

spatial dependence between the location and orientation of minutiae. The current discussion closely follows the framework developed in [4], but is presented here for completeness. For each fingerprint, let k denote the total number of minutiae which are represented by their location and direction information $\{(S_j, D_j), j = 1, 2, \dots, k\}$, where $S_j \in R^2$ and $D_j \in [0, 2\pi)$. Minutiae referred to here is assumed to be made available by some feature extractor; later on, the ‘GT’, ‘EX’ and ‘MINDTCT’ (which corresponds to minutiae extracted using the NIST minutiae extractor) type minutiae will be considered in Section V. The distribution proposed in [4] for the minutiae is given by the G -component mixture

$$f(s, \theta | \Theta_G) = \sum_{g=1}^G \tau_g f_g^X(s | \mu_g, \Sigma_g) \cdot f_g^D(\theta | \nu_g, \kappa_g), \quad (1)$$

where $s \equiv (x, y) \in R^2$, $\theta \in [0, 2\pi)$, and

$$f_g^X(s | \mu_g, \Sigma_g) = \phi_2(s | \mu_g, \Sigma_g) \quad (2)$$

is the bivariate Gaussian density with mean μ_g and covariance matrix Σ_g that represents the distribution of minutiae location from the g -th component. The second density in the mixture of equation (1)

$$f_g^D(\theta | \nu_g, \kappa_g, p_g) = \begin{cases} p_g v(\theta) & \text{if } \theta \in [0, \pi), \\ (1 - p_g) v(\theta - \pi) & \text{if } \theta \in [\pi, 2\pi), \end{cases} \quad (3)$$

is a two-component mixture of Von-Mises distributions that represents the distribution of minutiae direction from the g -th component; here, $v(\theta)$ is the Von-Mises distribution defined as

$$v(\theta) \equiv v(\theta | \nu_g, \kappa_g) = \frac{2}{I_0(\kappa_g)} \exp\{\kappa_g \cos 2(\theta - \nu_g)\}, \quad (4)$$

with $I_0(\kappa_g)$ given by

$$I_0(\kappa_g) = \int_0^{2\pi} \exp\{\kappa_g \cos(\theta - \nu_g)\} d\theta. \quad (5)$$

In (4), ν_g and κ_g represent the mean angle and the precision (inverse of the variance) of the Von-Mises distribution, respectively. The proposed mixture models were shown to be a flexible class of distributions [4] that satisfactorily represent all characteristics of minutiae variability, including clustering tendencies (see, for example, [7] and [8]), and dependence between the minutiae location and direction. Estimation of the unknown parameters $\Theta_G = \{G, ((\tau_g, \mu_g, \Sigma_g, \nu_g, \kappa_g, p_g), g = 1, 2, \dots, G)\}$ is carried out using the EM algorithm [9], and the optimal number of components, G^* , is selected using the Bayes Information Criteria (BIC) (see [4] for more details).

Subsequently, the mixture models are used to obtain the probability of a random correspondence as a measure of fingerprint individuality. For an impostor query and template pair of fingerprints (Q, T) , Zhu et al. [4] obtained a closed form expression for the probability of a random correspondence (PRC) corresponding to w minutiae matches. The probability of obtaining exactly w matches given that there are m and n minutiae in Q and T , respectively, is derived to have the Poisson probability mass function

$$p^*(w; Q, T) = \frac{e^{-\lambda(Q,T)} \lambda(Q, T)^w}{w!} \quad (6)$$

with mean $\lambda(Q, T)$. The expression of $\lambda(Q, T)$ was obtained in [4] as

$$\lambda(Q, T) = m n p(Q, T), \quad (7)$$

where

$$p(Q, T) = P(|S^Q - S^T|_s \leq r_0 \text{ and } |D^Q - D^T|_a \leq d_0) \quad (8)$$

denotes the probability of a match when (S^Q, D^Q) (respectively, (S^T, D^T)) is a random minutiae from the mixture distribution fitted to Q (respectively, T). In (8), $|\cdot|_s$ (respectively, $|\cdot|_a$) is the Euclidean distance (respectively, angular distance) with the following expressions for points $s_1 \equiv (x_1, y_1)$ and $s_2 \equiv (x_2, y_2) \in R^2$, and angles θ_1, θ_2 in $[0, 2\pi)$: $|s_1 - s_2|_s = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ and $|\theta_1 - \theta_2|_a = \min(|\theta_1 - \theta_2|, 2\pi - |\theta_1 - \theta_2|)$. In terms of the fitted mixture models to Q and T , f_Q and f_T , respectively, $p(Q, T)$ is calculated from the formula

$$p(Q, T) = \int \int_{(s_2, \theta_2) \in B(s_1, \theta_1)} f_Q(s_1, \theta_1) f_T(s_2, \theta_2) ds_2 d\theta_2 ds_1 d\theta_1, \quad (9)$$

where $B(x, \theta_1)$ is the (r_0, d_0) -ball of (s_1, θ_1) (that is, the set of all (s_2, θ_2) with spatial and angular distances less than r_0 and d_0 , respectively, from (s_1, θ_1)).

The mean parameter $\lambda(Q, T)$ in equation (6) can be interpreted to be the expected number of matches; there are a total of $m n$ possible match pairings with the probability of each match being $p(Q, T)$. The PRC corresponding to w matches is the probability of observing w or more matches between Q and T , and has the formula

$$PRC(w | m, n) = \sum_{u \geq w} p^*(u; Q, T), \quad (10)$$

where $p^*(u; Q, T)$ is as given in (6).

III. Incorporating Fingerprint Image Quality

The probability of random correspondence (PRC) measures the likelihood of a random match between two different fingerprints, and therefore, reflects the amount of uncertainty in the observed match. When poor quality

images are used for determining a match, the associated uncertainty is much larger. It follows that the PRC is an increasing function of quality degradation; lower (respectively, higher) quality fingerprint images results in higher (respectively, lower) values of the PRC. The Poisson distribution is an example of a monotone likelihood ratio (MLR) family in its parameter λ ; see [10]. The MLR property implies that the PRC in equation (10) is an increasing function of λ . From (7), it follows that $PRC(w | m, n)$ is an increasing function of the probability of a match $p(Q, T)$ for fixed m and n . Thus, it is sufficient to investigate how $p(Q, T)$ (as opposed to the PRC in equation (10)) can be modelled as a function of the quality degradation of Q and T .

The image quality of Q and T is obtained using a quality extractor that quantifies numerically the clarity of ridge valley structures. In this study, two quality extractors are used: (i) the quality extractor, ‘QUAL’, reported in [11] which uses a wavelet transform to measure image clarity, and (ii) the NIST quality extractor, ‘NFIQ’, as reported in [12]. The quality extractor ‘QUAL’ outputs a number Q^* between 0 and 1 with higher values indicating better quality images. The transformation $Q = 1 - Q^*$ is then used to reverse the direction of image degradation. Subsequently, a total of $B = 3$ equal proportion quality bins are constructed to group the quality measure with $\mathcal{G}_1 = [0, 0.4)$, $\mathcal{G}_2 = [0.4, 0.7)$ and $\mathcal{G}_3 = [0.7, 1]$ (see Figure 5). The ‘QUAL’ quality extractor eventually outputs a quality label 1, 2 or 3, with 1 and 3, respectively, corresponding to the best and worst quality images. The NFIQ algorithm is an implementation of the “NIST Image Quality” algorithm based on neural networks as described in [12]. For the NFIQ quality extractor, the output is a quality label numbered 1, 2, \dots , 5 with 1 and 5 corresponding to the best and worst quality images, respectively. The final quality bins are arranged according to increasing quality degradation, namely, 1, 2, \dots , B where $B = 3$ and $B = 5$ in the cases of ‘QUAL’ and ‘NFIQ’, respectively.

The following model is a first step towards investigating how the PRCs change as a function of image quality degradation: For the query and template pair (Q, T) ,

$$\log p(Q, T) = \mu + \gamma_{(q(Q), q(T))} + \epsilon(Q, T) \quad (11)$$

where $q(S)$ is the quality measure of fingerprint S based on either ‘QUAL’ or ‘NFIQ’ taking values in the set

$\{1, 2, \dots, B\}$, μ is an overall mean of the logarithm of the probability of a match $p(Q, T)$, calculated based on formula (9), and $\epsilon(Q, T)$ are independent and identically distributed random errors distributed as normal with mean 0 and unknown variance σ^2 . This will be called Model I from now on. The motivation for Model I in (11) is that the probability of a match (and subsequently the PRC) can be expressed as a linear combination of an overall mean μ and fixed effects $\gamma_{(1,1)}, \gamma_{(1,2)}, \dots, \gamma_{(B,B)}$ that depend on the quality of the underlying fingerprint images. Since the bins are ordered in terms of decreasing quality, the probability of a match will increase as the index of the quality bin increases (as discussed earlier). Thus, it is natural to have a restriction on the parameters $\gamma_{(r,s)}$ given by

$$\begin{aligned} \gamma_{(r,s_1)} &\leq \gamma_{(r,s_2)} \quad \text{for } s_1 \leq s_2 \text{ and every fixed } r = 1, 2, \dots, B, \text{ and} \\ \gamma_{(r_1,s)} &\leq \gamma_{(r_2,s)} \quad \text{for } r_1 \leq r_2 \text{ and every fixed } s = 1, 2, \dots, B. \end{aligned} \quad (12)$$

Parameters of Model I cannot be estimated uniquely due to model non-identifiability, meaning that different combinations of the model parameters give rise to the same distribution for $\log p(Q, T)$. To avoid the non-identifiability problem, it is necessary to impose some restrictions on the $\gamma_{(r,s)}$ parameters. We assume that $\gamma_{(1,1)} = 0$ with the interpretation that μ is the the baseline log probability of a match corresponding to the best quality label pair $(q(Q), q(T)) = (1, 1)$ and all other quality combinations represent increases in $p(Q, T)$ over this baseline.

Model I is a first step towards building a framework that incorporates image quality effects into fingerprint individuality estimates. The model may be too simple to adequately fit to large databases as demonstrated in Section V. The reason for this under-fitting is that other main as well as correlation effects have not been accounted for in the model formulation of (11). For example, there is significant correlation between the estimates of $p(Q, T)$ and $p(Q', T')$ when Q and Q' (and T and T') are different impressions of the same fingers. One way of incorporating correlation into the framework of (11) is to consider the following mixed

effects model, Model II, as follows:

$$\log p(Q, T) = \mu + \gamma(q(Q), q(T)) + \beta_{ch(Q)} + \beta_{ch(T)} + \alpha_{Finger(Q)} + \alpha_{Finger(T)} + \epsilon(Q, T), \quad (13)$$

where $Finger(S)$ represents the finger label for the impression S and $\alpha_{Finger(S)}$ are independent and identically distributed random variables distributed as normal with mean zero and unknown variance σ_α^2 . For the model in (13), the correlation between $\log p(Q, T)$ and $\log p(Q', T')$ can be computed to be

$$Corr(\log p(Q, T), \log p(Q', T')) = \frac{2\sigma_\alpha^2}{2\sigma_\alpha^2 + \sigma^2} = \frac{2}{2 + \nu} \quad (14)$$

where $\nu \equiv \sigma^2/\sigma_\alpha^2 > 0$ is the ratio of the two variance components. Further, in (13), extra fixed effects terms can be incorporated to represent additional information or characteristics of the input image pair (Q, T) . These fixed effects are represented in (13) by the $\beta_{ch(S)}$ terms where $ch(S)$ represents some additional characteristic of impression S taking values in the label set $\{1, 2, \dots, C\}$. For the NIST database, for example, $ch(S)$ will represent the fingerprint class information ‘L’, ‘R’, ‘W’ and ‘A’ with $C = 4$. Then, $\beta_{ch(S)}$ represents the effect on $\log p(Q, T)$ corresponding to class $ch(S)$.

IV. A Bayesian Framework for Inference

The estimation of model parameters in (11) and (13) is carried out in a Bayesian framework which is briefly discussed here. The set of all parameters corresponding to the two models is denoted by Θ with the understanding that $\Theta = (\mu, \gamma_{(r,s)}, r, s = 1, 2, \dots, B, \sigma^2)$ and $\Theta = (\mu, \gamma_{(r,s)}, r, s = 1, 2, \dots, B, \beta_1, \beta_2, \dots, \beta_C, \sigma^2, \nu)$ for Models I and II, respectively. The Bayesian inferential framework needs two inputs: (1) the likelihood of the data, $\log p(Q, T)$, given Θ , which we denote by $\ell(data | \Theta)$, and (2) a prior on Θ denoted by $\pi_0(\Theta)$. Given the likelihood and prior, the posterior of Θ is computed via Bayes Theorem:

$$\pi_0(\Theta | data) = \frac{\ell(data | \Theta) \cdot \pi_0(\Theta)}{\int_{\Theta} \ell(data | \Theta) \cdot \pi_0(\Theta) d\Theta}. \quad (15)$$

In (15), the integral in the denominator is taken over the entire space of Θ . The posterior $\pi_0(\Theta | data)$ represents the best distribution of uncertainty associated with the unknown parameters once the data has been observed.

A point estimate of Θ can be obtained by computing the posterior mean from (15) while the variability can be ascertained by computing the posterior variance. The best range of most likely values of Θ can be obtained from the highest posterior density sets (HPD sets) of (15).

For the present application, the prior on Θ is taken to be “flat” or non-informative. For Model I, we have

$$\pi_0(\Theta) = \left(\frac{1}{\sigma^2} d\sigma^2 \right) (d\mu) \left(\prod_{r,s=1}^B d\gamma_{(r,s)} \right) \quad (16)$$

on the set $\gamma_{(r,s)}$ where restriction (12) is satisfied. For Model II, the prior is

$$\pi_0(\Theta) = \left(\frac{1}{\sigma^2} d\sigma^2 \right) (G(\nu; \alpha, \beta) d\nu) (d\mu) \left(\prod_{u=1}^C d\beta_u \right) \left(\prod_{r,s=1}^B d\gamma_{(r,s)} \right) \quad (17)$$

on the set where (12) is satisfied; in (17), $G(\alpha, \beta)$ is the Gamma probability density function given by

$$G(\nu; \alpha_0, \beta_0) = \frac{1}{\Gamma(\alpha_0) \beta_0^{\alpha_0}} \nu^{\alpha_0-1} \exp\{-\nu/\beta_0\} \quad (18)$$

with shape and scale parameters α_0 and β_0 , respectively. Non-informative priors are attractive in the sense that they allow the data to “speak for themselves” without any subjective input. However, since non-informative priors are typically improper (that is, they integrate to infinity), it is necessary to ensure that the resulting posterior distribution obtained is proper (that is, a proper probability distribution so that quantities such as the mean and variance makes sense and is well defined). Due to page-limit restrictions, the interested reader is referred to the technical report Dass (2009) [13] where propriety of the posterior is established for both models considered here. Further, from the proof, it follows that the posterior for the extended model cannot be proper unless the prior on ν is proper, and that is the reason for choosing the Gamma distribution on ν . Although this limitation is unavoidable, the hyper-parameters α_0 and β_0 are estimated from the database (and not subjectively determined) to allow maximum support from data in this case.

Inference for the posterior can be carried out using Markov Chain Monte Carlo (MCMC) techniques which are popular as Bayesian computational tools. The MCMC technique constructs a Markov Chain of Θ realizations, say, $\Theta_1, \Theta_2, \dots, \Theta_T, \dots$, with stationary distribution given by the posterior distribution (15). As a result, the distribution of Θ_T is approximately the posterior distribution (15) as $T \rightarrow \infty$. The reader is referred to [14]

which gives an excellent introduction to MCMC methods in Bayesian statistics. In the present case, we use a special case of the MCMC procedure called the Gibbs sampler which samples from suitable conditional distributions derived from the full posterior. The reader is referred to Dass (2009) [13] for the details.

A. Convergence Diagnostics

Before samples from the Gibbs sampler can be used to calculate the posterior mean, variance and HPD sets, convergence diagnostics tests has to be performed. A popular method for assessing convergence is with the Gelman-Rubin R statistic (see [14]). A total of $I \geq 2$ initial chains are run starting from different initial estimates of Θ , and the ratio of between chain variance to within chain variance is calculated for each component of Θ based on the samples of the Gibbs chain. Convergence is established if R -statistic corresponding to each Θ component is close to 1.

B. Goodness of Fit

Goodness of fit tests are conducted to determine whether the models proposed earlier are an adequate representation of variability in the fingerprint databases. The subsequent discussion is based on Model II in (13) with necessary adjustments made in the presentation for Model I in (11). Goodness of fit tests are conducted in the following way: Define the residual

$$r(Q, T) = \log p(Q, T) - \mu - \gamma(q(Q), q(T)) - \beta_{ch(Q)} - \beta_{ch(T)} - \alpha_{Finger(Q)} - \alpha_{Finger(T)} \quad (19)$$

for each (Q, T) pair in the database, and the mean residual to be

$$\bar{r}_0 = \frac{\sum_{\text{impostor pairs } (Q, T)} r(Q, T)}{N_0} \quad (20)$$

where N_0 is the total number of impostor pairs in the database. It can be shown that the posterior mean of \bar{r}_0 is 0 under Model II, and therefore, \bar{r}_0 values calculated from Θ samples from the posterior distribution should be close to 0. This is determined by constructing the 95% HPD set for \bar{r}_0 , \mathcal{R}_0 , based on the Θ samples from

Database	No. of fingerprints	No. of different fingers	n_i
IBM	900	269	2 – 8
NIST	4,000	2,000	2

TABLE II. Summary information of the two databases used in the experiments.

the Gibbs chain and checking if 0 belongs to this set. We define

$$L_0 = \begin{cases} 1 & \text{if } 0 \in \mathcal{R}_0, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

Two other goodness of fit measures are derived similarly. Defining

$$r_1(Q, T) = \left(\mu + \gamma(q(Q), q(T)) + \beta_{ch(Q)} + \beta_{ch(T)} \right) \cdot r(Q, T) \quad (22)$$

and

$$r_2(Q, T) = \left(\mu + \gamma(q(Q), q(T)) + \beta_{ch(Q)} + \beta_{ch(T)} + \alpha_{Finger(Q)} + \alpha_{Finger(T)} \right) \cdot r(Q, T), \quad (23)$$

it can be similarly shown that the means of \bar{r}_1 and \bar{r}_2 is zero under Model II. In these cases, the goodness of fit measure translates to

$$L_j = \begin{cases} 1 & \text{if } 0 \in \mathcal{R}_j, \text{ and} \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

for $j = 1, 2$, where R_j is the 95% posterior HPD set based on \bar{r}_j .

V. Experimental Results

Two fingerprint databases, the in-house IBM [2] and the publicly available NIST [15] databases, are considered in our experiments (see Table II). The IBM database consists of fingerprint impressions of 450 pairs of fingers with manually labeled minutiae (ground truth) and minutiae correspondences between each pair. There are $F = 269$ distinct fingers in the IBM database with the number of multiple impressions per finger (n_i) varying from 2 to 8. The ground truth minutiae was denoted by the notation ‘GT’ to represent one mode of minutiae extraction. We also considered the feature extraction and matching algorithm of [6] to obtain

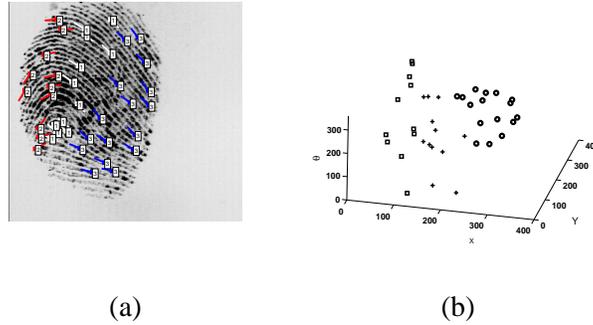


Fig. 4. Mixture model fitting to ground truth minutiae distribution of a fingerprint. Panel (a) and (b) show the clusters in 2-D and 3-D spaces, respectively.



(a) 'Best', Label=1 (b) 'Moderate', Label=2 (c) 'Poor', Label=3

Fig. 5. Examples of fingerprint images in the three categories of quality for 'QUAL'.

the list of extracted (EX) minutiae as our second mode of minutiae extraction. A third state-of-the-art and publicly available minutiae extractor is also considered, namely the 'MINDTCT' minutiae extractor developed by the NIST Image Group. The 'MINDTCT' algorithm is a fingerprint feature extractor algorithm inspired by the Home Office's Automatic Fingerprint Recognition System (see [16]). Executing the algorithm on an input fingerprint gives information on the location and direction of minutiae as well as a measure of reliability ranging from 0 to 100, with higher values indicating better reliability.

The mixture model in (1) was fit to the three sets of minutiae (labeled as 'GT', 'EX' and 'MINDTCT') and was found to be a good fit to all the 450 image pairs (see Figure 4 for an example of fitting the mixture distribution) for all the three minutiae extractors. Based on the mixture distributions, the value of $\log p(Q, T)$ was obtained using equation (8), and then taking logarithms.

Both Models I and II were fit to the observed data consisting of the values of $\log p(Q, T)$. Since the number of

Quality	GT			EX			MINDTCT		
measure	Stat.	Model I	Model II	Stat.	Model I	Model II	Stat.	Model I	Model II
QUAL	R	1.0033 (0.0042)	1.0044 (0.0038)	R	1.0025 (0.0036)	1.0085 (0.0119)	R	1.0288 (0.0162)	1.0028 (0.0057)
	\bar{L}	0.5000 (0.000)	0.9667 (0.1054)	\bar{L}	0.5000 (0.0000)	1.0000 (0.0000)	\bar{L}	0.9000 (0.2108)	1.0000 (0.0000)
NFIQ	R	1.0001 (0.0001)	1.0169 (0.0200)	R	1.0001 (0.0000)	1.0253 (0.0354)	R	1.0331 (0.0094)	1.0088 (0.0138)
	\bar{L}	0.9000 (0.3100)	1.0000 (0.0000)	\bar{L}	0.8500 (0.3375)	1.0000 (0.0000)	\bar{L}	0.9000 (0.3162)	1.0000 (0.0000)

TABLE III. Convergence diagnostics and goodness of fit assessments for the IBM database. Entries for each statistic correspond to the means and standard deviations (in parenthesis) of 10 repeated experiments.

impostor pairs in the IBM database was very large (a total of $900 \times 899 = 809,100$ pairs), the implementation of the Gibbs sampler based on all impostor pairs is computationally time consuming and prohibitive. Alternatively, a total of $N_0 = 10,000$ impostor pairs were randomly selected and Models I and II were fit to this subset of impostor pairs. The sampling procedure was repeated 10 times, and the average of the results of the 10 trials are reported for the two databases.

Estimates of the parameters were computed using the Bayesian framework described in Section IV with pre-specified values of α_0 and β_0 estimated from the database. The Gibbs chain was run up to $T = 1,000$ iterations. Convergence diagnostics (see Section IV-A) performed indicates that the chains (we took $I = 3$) converged from $T = 900$ onwards. See Table III for values of the Gelman-Rubin R -statistic for Models I and II based on the three feature extractors ‘GT’, ‘EX’ and ‘MINDTCT’ and the two quality measures ‘QUAL’ and ‘NFIQ’. The goodness of fit tests to judge the adequacy of the Models I and II uses the average of L_j s for $j = 0, 1, 2$, $\bar{L} = (L_0 + L_1 + L_2)/3$. Note that $0 \leq \bar{L} \leq 1$ with $\bar{L} = 1$ and $\bar{L} = 0$ indicating, respectively, the

Quality	GT			EX			MINDTCT		
measure	Parameter	Model I	Model II	Parameter	Model I	Model II	Parameter	Model I	Model II
QUAL	μ	-9.3018 (0.5670)	-6.9532 (0.1742)	μ	-9.6962 (0.5253)	-7.1833 (0.2872)	μ	-8.2468 (0.1737)	-7.4278 (0.0559)
	σ^2	1.5862 (0.1302)	0.9323 (0.0430)	σ^2	1.8077 (0.2320)	1.0429 (0.0885)	σ^2	0.3377 (0.0226)	0.2602 (0.0067)
	ν	- -	12.756 (1.489)	ν	- -	12.3618 (2.2377)	ν	- -	41.2058 7.4468
NFIQ	μ	-6.7686 (0.0416)	-6.9662 (0.0820)	μ	-6.7347 (0.0461)	-7.0359 (0.0859)	μ	-7.4511 (0.0625)	-7.4095 (0.0276)
	σ^2	1.2941 (0.0540)	0.8969 (0.0267)	σ^2	1.5423 (0.2220)	1.0218 (0.0952)	σ^2	0.2998 (0.0085)	0.2604 (0.0094)
	ν	- -	13.291 (1.095)	ν	- -	13.404 (3.073)	ν	- -	38.97 (5.9070)

TABLE IV. Posterior means for the IBM database corresponding to Models I and II based on 10 repeated trial. Entries are the averages and standard deviations (in parenthesis) over the 10 trials.

NOTE: Only entries corresponding to Model II should be used for further inference since Model I does not fit the database well.

GT				EX				MINDTCT			
r/s	1	2	3	r/s	1	2	3	r/s	1	2	3
1	0.0000	0.0565	0.0600	1	0.0000	0.2110	0.2137	1	0.0000	0.0562	0.0627
2	0.0455	0.0835	0.0882	2	0.2297	0.3066	0.3104	2	0.0651	0.0694	0.0754
3	0.0521	0.0890	0.2978	3	0.2354	0.3105	0.3892	3	0.0701	0.0771	0.1384

TABLE V. The $\gamma_{(r,s)}$ entries for the IBM database for Model II and the quality measure 'QUAL'. Entries are averages of posterior means of the 10 repeated experiments.

GT					EX					MINDTCT							
r/s	1	2	3	4	5	r/s	1	2	3	4	5	r/s	1	2	3	4	5
1	0.000	0.045	0.050	0.059	0.129	1	0.000	0.085	0.089	0.098	0.131	1	0.000	0.028	0.031	0.038	0.053
2	0.043	0.060	0.067	0.081	0.183	2	0.073	0.098	0.104	0.117	0.167	2	0.019	0.034	0.040	0.052	0.085
3	0.056	0.072	0.081	0.105	0.280	3	0.088	0.109	0.117	0.145	0.225	3	0.029	0.044	0.058	0.082	0.145
4	0.089	0.126	0.169	0.241	0.453	4	0.112	0.144	0.172	0.227	0.362	4	0.039	0.059	0.085	0.129	0.235
5	0.168	0.225	0.282	0.396	0.795	5	0.163	0.206	0.247	0.340	0.780	5	0.057	0.081	0.119	0.182	0.386

TABLE VI. The $\gamma_{(r,s)}$ entries for the IBM database for Model II and the quality measure ‘NFIQ’. Entries are averages of posterior means of the 10 repeated experiments.

Extractor	r/s	1	2	3
GT	1	0.0052 (100)	0.0100 (190)	0.0104 (198)
	2	0.0088 (168)	0.0136 (258)	0.0143 (273)
	3	0.0095 (181)	0.0144 (275)	0.1524 (290)
EX	1	0.0024 (100)	0.0261 (1106)	0.0269 (1140)
	2	0.0323 (1367)	0.0769 (3255)	0.0802 (3398)
	3	0.0344 (1458)	0.0803 (3402)	0.1943 (8229)

TABLE VII. PRCs (all entries in the order of 10^{-9}) and their increases as image quality degrades based on the IBM database for feature extractors ‘GT’ and ‘EX’, and the quality extractor ‘QUAL’. The numbers in the parenthesis give the percent increase in PRC from the baseline (1, 1) for Model II.

Extractor	r/s	1	2	3	4	5
MINDTCT	1	0.0143 (100)	0.0348 (243)	0.0369 (258)	0.0389 (272)	0.0418 (292)
	2	0.0297 (207)	0.0456 (319)	0.0498 (348)	0.0527 (368)	0.0582 (407)
	3	0.0311 (217)	0.0491 (343)	0.0537 (375)	0.0582 (407)	0.0660 (462)
	4	0.0336 (235)	0.0520 (364)	0.0577 (403)	0.0694 (485)	0.1075 (752)
	5	0.0373 (260)	0.0569 (398)	0.0628 (439)	0.1043 (730)	0.6105 (4,272)

TABLE VIII. PRCs (all in the order of 10^{-11}) and their increases as image quality degrades based on the NIST database for ‘MINDTCT’ and the quality extractor ‘NFIQ’. The numbers in the parenthesis give the percent increase in PRC from the baseline (1, 1) for Model II. Entries are averaged over all 16 pairs of combinations of fingerprint class for (Q, T) from the class label set $\{W, R, L, A\}$.

best and worst goodness of fit assessment. The values of the R -statistic (with means close to 1 and standard deviations close to 0) indicate that every chain has converged after $T = 900$, and that samples may be collected from the posterior to make inference on the model parameters. The mean and standard deviation values of \bar{L} indicate that Model II is a better fit to the observed data compared to Model I. This can be seen from the fact that the means and standard deviations of \bar{L} are, respectively, closer to 1 and 0 in Model II compared to Model I.

The last 100 samples from the three chains were used to estimate Θ corresponding to Models I and II. The posterior means estimates of μ , σ^2 (and ν) are given in Table IV for the three feature extractors and two quality measures for Model II whereas the estimates of $\gamma_{(r,s)}$ are given in Tables V and VI for Model II (the better fitting model).

Several conclusions can be drawn from the entries of Table IV: First, for the feature extractor ‘MINDTCT’, Model I is also a good fit to the observed data. This can be seen from the high \bar{L} entries in Table III and the large ν values in Table IV; note that large ν values indicate that the correlation in (14) is close to zero, and therefore, Model II is almost equivalent to Model I. However, for the other feature extractors, Model I does not fare as well, as evidenced from the lower \bar{L} and ν values (indicating a higher level of correlation).

Second, ‘MINDTCT’ is the more effective feature extractor compared to ‘EX’ as seen from the μ values in Table IV. The three feature extractors output a different number of minutiae when applied to the same input image. Generally, the ‘GT’ yielded the highest number, followed by ‘MINDTCT’ and finally, ‘EX’. Due to the differing number of minutiae extracted, it turns out that the three feature extractors cannot be compared with each other; the mixture models fitted will be different depending on the number of these detected minutiae. One comparison can be carried out, however. Since ‘MINDTCT’ also outputs a measure of reliability corresponding to each minutiae detected, we sorted the extracted minutiae in decreasing order of reliability and took the top M many, where M corresponds to the number of minutiae extracted by ‘EX’. The lower values of μ corresponding to ‘MINDTCT’ in Table IV in the column for Model II when compared to ‘EX’ indicates that

it is a slightly better minutiae extractor for authentication purposes.

Third, the fact that ‘MINDTCT’ is a better feature extraction algorithm is again demonstrated in Tables V and VI. The increase in the $\gamma_{(r,s)}$ parameters as image quality degrades (corresponding to higher bin labels) is smaller for ‘MINDTCT’ compared to ‘EX’. This indicates more reliability of fingerprint authentication based on ‘MINDTCT’ for the IBM database.

The above methodology was also applied to the NIST Special Database 4 [15]. The NIST fingerprint database is publicly available and consists of 2,000 8-bit gray scale fingerprint image pairs of size 512-by-512 pixels. The first 2,000 images of each finger are denoted by ‘f’ while the second 2,000 images of the same fingers are denoted by ‘s’. Minutiae could not be automatically extracted from 14 images of the NIST database due to poor quality by at least one of the three feature extractors. Thus, the total number of NIST fingerprints used in our experiments is $F = 1,986$. The feature extraction algorithms giving rise to the ‘EX’ and ‘MINDTCT’ minutiae types were used on the fingerprint images. The mixture model in (1) was fit to the two sets of extracted minutiae (from the ‘f’ and ‘s’ images). Goodness of fit tests in [4] demonstrated that the mixture model was a good fit to all fingerprints in the NIST database. Based on the mixture distributions, the value of $\log p(Q, T)$ and the quality label based on ‘QUAL’ and ‘NFIQ’ were obtained.

A similar analysis is carried out for the NIST database and detailed results are reported in Dass (2009) [13]. The conclusions from the analysis are summarized here. Both Models I and II are very good fits to the database except in the case of (‘MINDTCT’, ‘QUAL’) where Model II is clearly a superior fit. High values of ν for both models indicate that the better fitting model is not the one that accounts for correlation between fingers but rather the one that incorporates the additional fixed effects $\beta_{ch(S)}$ terms. Indeed, when Model II was also fitted to the NIST database without the $\beta_{ch(S)}$ fixed effects, it turned out to be worse than Model II with these effects incorporated again for the combination (‘MINDTCT’, ‘QUAL’).

Based on the estimates derived from the IBM and NIST databases, we can compute the PRC corresponding to w matches as a function of the fingerprint image quality. For illustrative purposes, we use $w = 12$ (for

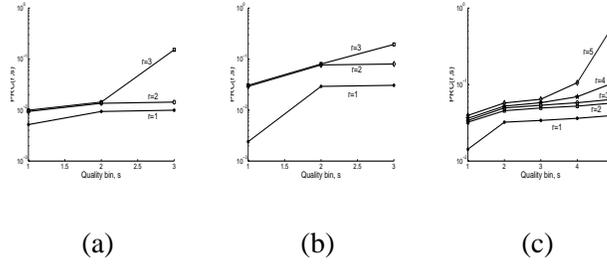


Fig. 6. Figure showing the increasing trend of the PRCs as quality degrades. Panels (a), (b) and (c) are for the combinations ‘GT’ and ‘QUAL’, ‘EX’ and ‘QUAL’, and ‘MINDTCT’ and ‘NFIQ’, respectively.

the 12-point matching criteria) with $m = n = 26$ (this is one of the combinations used by Zhu et al. [4]). The PRCs in Table VII are given for the IBM database for the feature extractors ‘GT’ and ‘EX’. For the NIST database, the PRCs are calculated, as an illustration, for the feature and quality extractor combination MINDTCT and NFIQ (see Table VIII). PRCs for other combinations can be derived similarly. Since additional fixed effects are involved in this case, each combination of fixed effects $(\beta_{ch(Q)}, \beta_{ch(T)})$ is first taken and an estimate of the PRC is obtained. Subsequently, the PRCs are averaged over 16 such combinations of fixed effect pairs resulting from the class labels W, R, L and A . We note from Tables VII and VIII that PRCs deteriorate (become larger) significantly when either one of the fingerprints is of poor quality. In the case when both query and template images are of poor quality, the increases in PRCs are orders of magnitude larger compared to the baseline. Figure 6 plots the PRCs obtained in Tables VII and VIII to show the increasing trend. Panels (a) and (b) correspond to ‘QUAL’ with ‘GT’ and ‘EX’, respectively, as in Table VII while panel (c) is based on ‘MINDTCT’ and ‘NFIQ’ in Table VIII. The B lines in each panel is the plot of $PRC(r, s)$ versus $s = 1, 2, \dots, B$ for every fixed $r = 1, 2, \dots, B$, where $B = 3$ or 5 . $PRC(r, s)$ is determined to be the average of two PRCs corresponding to quality bin (r, s) and (s, r) in the two tables.

VI. Summary and Conclusion

We have outlined a methodology to investigate how PRCs change as a function of the underlying image quality of the pair of fingerprints in question. Based on our analysis, ‘MINDTCT’ appears to be a generally more robust as a minutiae extractor with respect to quality degradation. Intra-finger correlation dictates the selection of model for the IBM database whereas the additional fixed effects (and not the correlation) determines the best fitting model for the NIST database. Therefore, it is recommended to fit the extended form of Model II (with intra-finger correlation and additional fixed effects) for fingerprint databases in general for achieving the best fit.

The increase in PRCs is most significant when both query and template fingerprints are of poor quality. This study cautions the use of poor quality fingerprint images for the purpose of identification. Future work will include modeling of the probability of a genuine match and combining the methodology outlined here to investigate how the genuine probabilities change as a function of the underlying image quality.

VII. Acknowledgments

The author would like to thank Dirk Colbry for help with the computing facilities at HPC. The author would also like to thank Prof. Anil K. Jain for discussions. This research is partially supported by the NSF DMS grant 0706385.

References

- [1] *Daubert v. Merrel Dow Pharmaceuticals Inc*, 509 U.S. 579, 113 S. Ct. 2786, 125 L.Ed.2d 469 (1993).
- [2] Sharath Pankanti, Salil Prabhakar, and Anil K. Jain, “On the individuality of fingerprints,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1010–1025, 2002.
- [3] *United States v. Byron Mitchell*, Criminal Action No. 96-407, U. S. District Court for the Eastern District of Pennsylvania, 1999.
- [4] Y. Zhu, S. C. Dass, and A. K. Jain, “Statistical models for assessing the individuality of fingerprints,” *IEEE Transactions on Information Forensics and Security*, , no. 3, pp. 391–401, 2007.

- [5] J. McMenamin, "U.S. eyeing County case," *The Baltimore Sun*, February 21, 2008.
- [6] Anil K. Jain, Lin Hong, Sharat Pankanti, and Ruud Bolle, "An identity authentication system using fingerprints," *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1365–1388, 1997.
- [7] S. C. Scolve, "The occurrence of fingerprint characteristics as a two dimensional process," *Journal of the American Statistical Association*, vol. 74, no. 367, pp. 588–595, 1979.
- [8] D. A. Stoney and J. I. Thornton, "A critical analysis of quantitative fingerprint individuality models," *Journal of Forensic Sciences*, vol. 31, no. 4, pp. 1187–1216, 1986.
- [9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood for incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B*, vol. 39, no. 1, pp. 1–38, 1977.
- [10] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypothesis*, Springer-Verlag, New York, 2005.
- [11] Y. Chen, S. Dass, and A. K. Jain, "Fingerprint quality indices for predicting authentication performance," *Proc. of Audio- and Video-based Biometric Person Authentication (AVBPA)*, pp. 160–170, 2005.
- [12] E. Tabassi, C. Wilson, and C. Watson, "Fingerprint image quality," *Technical Report 7151*. Online: <http://fingerprint.nist.gov/NBIS>, 2004.
- [13] S. C Dass, "Assessing fingerprint individuality in presence of noisy minutiae," *Department of Statistics & Probability Technical Report, Michigan State University*, 2009.
- [14] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman & Hall/CRC, 1996.
- [15] "NIST: 8-bit gray scale images of fingerprint image groups (FIGS)," Online: <http://www.nist.gov/srd/nistsd4.htm>.
- [16] License 16-93-0026 Home Office Automatic Fingerprint Recognition System (HOAFRS), "Science and technology group," *Home Office, London*, 1993.