Evidential Value of Automated Latent Fingerprint Comparison: An Empirical Approach

Abhishek Nagar, *Student Member, IEEE,* Heeseung Choi, *Member, IEEE,* and Anil K. Jain, *Fellow, IEEE*

***Abstract*—Latent fingerprints are routinely recovered from crime scenes and are compared with available databases of known fingerprints for identifying criminals. However, current procedures to compare latent fingerprints to large databases of full (rolled or plain) fingerprints are prone to errors. This suggests caution in making conclusions about a suspect's identity based on a latent fingerprint comparison. A number of attempts have thus been made to measure the utility of a fingerprint comparison in making a correct accept/reject decision or its evidential value. These approaches, however, either do not represent the state-of-the-art in fingerprint matching due to unrealistic modeling assumptions or they lack simple interpretation. We argue that the posterior probability of two fingerprints belonging to different fingers given their match score, referred to as the Non-match probability (NMP), effectively captures any implicating evidence of the comparison. NMP is computed using state-of-the-art matchers and is easy to interpret. To incorporate the effect of image quality, number of minutiae, and size of the latent on NMP value, we compute the NMP vs. match score plots separately for image pairs (latent and full fingerprints) with different characteristics. Given the paucity of latent fingerprint databases in public domain, we simulate latent fingerprints using two full fingerprint databases (NIST SD-14 and Michigan State Police) by cropping regions of three different sizes. We appropriately validate this simulation using four latent databases (NIST SD-27 and three proprietary latent databases) and two state-of-the-art fingerprint matchers to compute their respective match scores. We also describe the way a latent fingerprint examiner would use the proposed framework to compute the evidential value of a latent-full print pair comparison in practice.**

***Index Terms*—Fingerprint matching, latent fingerprint comparison, individuality, evidential value, Non-match probability, genuine match distribution, impostor match distribution**

I. Introduction

Latent fingerprints are extensively used as forensic evidence in criminal prosecution. This is mainly because i) fingerprint patterns are highly discriminative, and ii) they are routinely found at most crime scenes due to inadvertent contact of the perpetrator's finger tips with various objects in the crime scene. In order to use them as evidence in a court of law, the latent fingerprints are "lifted" from the crime scene and matched either to full (rolled or plain) fingerprints that are captured from the suspect or to reference prints in law enforcement databases. See Figure 1 for a sample latent fingerprint image and its corresponding (mated) full fingerprint.

Typically, latent fingerprint images have significantly poor quality compared to full fingerprints. While full print to full print matching can be done effectively in a lights out mode (fully automatic) by Automatic Fingerprint Identification System (AFIS) (unless the image quality is very poor) [39], latent to full print matching still requires extensive image preprocessing and, in many instances, a manual matching by a latent examiner following a procedure referred to as the ACE-V protocol [15]. Usually, an AFIS is used to filter a large database of reference full prints to a small number of potential mates (typically 50) for further manual examination by latent experts. Despite the ACE-V protocol available for latent matching, there have been a number of cases where an incorrectly identified latent fingerprint resulted in a wrongful conviction [20]. A prominent case in this regard is that of Brandon Mayfield, who was incarcerated for the 2004 Madrid train bombing based on an erroneous latent fingerprint match. The Federal Bureau of Investigation (FBI) later reviewed this case [13] and noted that the reasons for the misidentification included examiner bias due to influence of the knowledge of Mayfield's fingerprint while marking features on the latent fingerprint, and inadequate consideration of fingerprint image quality. Some other cases have also been brought to light by the Innocence Project [40] where the erroneous convictions made based on latent fingerprint matches were later overturned as a result of DNA evidence. The acquitted individuals, however, had already spent many years in the prison.

In light of such misidentifications, it is crucial to measure the accuracy or error rate of a fingerprint comparison and thus the confidence with which the outcome of a fingerprint comparison can be accepted. The latter is usually quantified as the evidential value of a fingerprint comparison. The importance of this evidential value was also established by the Daubert standard set by the United States Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals, 1993* [1]. The Daubert

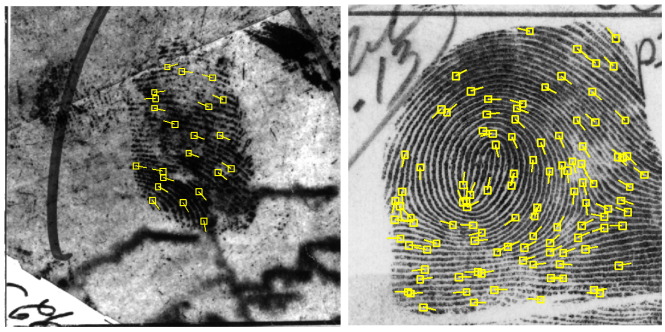


Fig. 1. Example of a latent and the corresponding rolled fingerprint. Based on NIST SD27, latent fingerprints, on average, have around 20 minutiae whereas a rolled fingerprint, on average, has around 150 minutiae.

A. Nagar, H. Choi and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824. A. K. Jain is also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Republic of Korea.

standard requires that the error rate of the forensic analysis be available before the related evidence can be admitted in a court proceeding. An urgent need to properly evaluate the error rates of latent matching was also expressed in an extensive study of forensic techniques prevalent in the United States conducted by the National Research Council (NRC) [32]. The report highlighted that

"In most forensic science disciplines, no studies have been conducted on large populations to determine the uniqueness of marks or features. Yet, despite the lack of a statistical foundation, examiners make probabilistic claims based on their experience. A statistical framework that allows quantification of these claims is greatly needed. These disciplines also critically need to standardize and clarify the terminology used in reporting and testifying about the results and in providing more information."

The NRC study essentially recommends that every forensic science method should undergo substantial research to validate basic premises and techniques, assess limitations, and discern the sources and magnitude of error.

The need to estimate the evidential value of latent fingerprint comparison based on its error rate is urgent, lest undue challenges to fingerprint evidence in court cases affect timely deliverance of justice. The use of fingerprints as evidence was first challenged in 1999 in the case of *U.S. v. Byron C. Mitchell* [2] and since then numerous other court cases have seen calls for motion to exclude fingerprints as evidence. See for example *U.S. v. Llera Plaza* [3], [4], in 2002, *U.S. v. Crisp* [5] in 2003, *State of Maryland v. Bryan Rose* [6] in 2007, and *U.S. v. Hamza Keita* [7] in 2009.

Numerous studies have been undertaken to date to evaluate the evidential value of a fingerprint comparison. See Section II for a discussion on the past studies. These studies can be broadly classified as feature modeling and empirical approaches.[1] The feature modeling approaches aim to statistically model the correspondence between certain features of the two fingerprint images being compared. This statistical model can thus be used to estimate the probability that the correspondences are just due to random chance. The empirical approaches, on the other hand, aim at conducting large-scale experiments for estimating the error rates under different circumstances. In this paper, we mainly explore the empirical approach. Note that it is difficult to accurately model various fingerprint features, as attempted in a number of published studies, that are typically used by state-of-the-art fingerprint matchers as well as identified by latent examiners during fingerprint comparison.

A fingerprint comparison, as required in an empirical study assessing the evidential value of a fingerprint comparison, can be performed either using an automatic matcher or manually by a latent expert. Both these methods of comparing fingerprints have their own implications on the accuracy of the comparison as well as on the evaluation of its evidential value:

1) **AFIS-based fingerprint comparison**: With the advancements in fingerprint matching technology, it is now possible to compare latents with full fingerprints effectively and automatically. The National Institute of Standards and Technology (NIST) conducted an evaluation of available automatic latent fingerprint techniques and reported a rank-1 accuracy of 97.2% on good quality latent fingerprints when matched with a background database of $100,000$ full fingerprints (see [41]) and an accuracy of 62% while matching poor quality latents with a background of 1 million full fingerprints (see [28]). In [41], the manually marked features on latent images were used to search the database and the images with successful retrievals were then used in the fully automatic testing. Further, the throughput requirement in [28] was more restricted compared to [41]. This explains the high matching accuracy reported in [41]. Nevertheless, these promising results from automatic latent matching experiments support the use of AFIS in forensic matching of latent fingerprints. The issue of limited availability of the database for training purposes is also being addressed. See e.g. [21].
2) **Manual fingerprint comparison**: Latent fingerprint examiners typically follow the ACE-V protocol for matching latent fingerprints. While this protocol is considered to be reasonable, it is difficult to conduct large-scale experiments with latent examiners in order to evaluate the error rates associated with the protocol and thus the evidential value of a manual latent fingerprint comparison. In a recent study involving manual fingerprint matching conducted by Ulery et al. [50], 169 latent examiners were asked to match latent-full print pairs from a pool of 744 (520 mated and 224 non-mate) pairs. Still, Ulery et al. concede that, despite the extensive nature, their experiment is not representative of all the latent print examiners and actual casework of latent matching. Further, a proper design of experiments would require that fingerprint images are partitioned based on the image characteristics, but this would further reduce the already small number of image pairs used in their experiment. Evaluation of error rates of the ACE-V protocol is also difficult due to inconsistencies among its different implementations that arise mainly due to its imprecise specification [26]. In a study performed by Dror et. al. [22], the accuracy of latent examiners has also been shown to be affected by extensive use of AFIS to select a list of candidate matches before a manual matching is performed.

Our study mainly focuses on the use of automatic fingerprint matchers because of the difficulty in conducting large-scale experiments with latent fingerprint examiners. We use the Non-Match Probability (NMP) [19] as the quantity that captures

[1] A categorization of approaches that quantify the evidential value of a forensic evidence has also been proposed in [48]. In [48] the existing approaches are categorized into generative and discriminative classes, where a generative approach involves a statistical model of the biometric features while a discriminative approach does not. However, the term generative is typically used to describe a classification approach where the distributions of two classes being distinguished is modeled, whereas a discriminative approach usually only models the decision boundary [37]. In our context, the two classes being distinguished are the genuine and impostor match score distributions, whereas in [48] the generative class encompasses approaches irrespective of whether the corresponding match scores are statistically modeled or not. For this reason, we have proposed a new categorization.

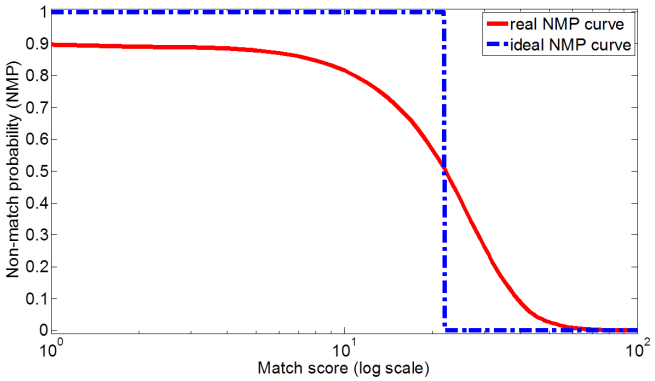


Fig. 2. A typical NMP-curve. An ideal NMP-curve is shown as the dotted blue line. Note that the ideal curve can completely separate mated (having an NMP value 0) and non-mated (having an NMP value 1) fingerprint pairs.

any implicating evidence of a fingerprint comparison. Since the accuracy of AFIS and thus the relationship between the NMP and the match score is dependent upon the characteristics of the fingerprint images, we divide the available fingerprint database into various partitions, each associated with a different covariate, say, a specific range of number of minutiae, image quality and the size (area) of the associated fingerprints. Different NMP vs. match score plots, called the NMP-curves (see Figure 2), are computed based on these database partitions and their relationship with the characteristics of the database is studied. We also compute the significance of evidence associated with an NMP value, which is referred to as the *conclusiveness* of an NMP value. Finally, due to paucity of public domain latent databases, we utilize and validate the use of partial fingerprints as a substitute for latent fingerprints in estimating the NMP-curves.

The main contributions of this paper are as follows:

1) Quantification of the evidential value of a latent fingerprint comparison in terms of *conclusiveness* of an NMP value.
2) An analysis of the evidential value of latent fingerprint comparison as a function of number of minutiae, quality, and latent area or size.
3) Validation of simulation of latent fingerprints using partial prints by comparing the associated NMP-curves.
4) A step-by-step procedure that can be followed by a latent examiner in order to estimate the evidential value of a latent fingerprint comparison.
5) Extensive experiments using two different Commercial Off the Shelf (COTS) fingerprint matchers and four different latent databases.

The rest of the paper is organized as follows. Section II presents a summary of previous approaches to estimate fingerprint evidential value. Section III presents the proposed framework for analyzing the evidential value based on NMP. Section IV presents experimental results. A summary of our work is presented in Section V. A preliminary version of this paper appeared in [19].

II. Background

The various approaches available in literature for formally assessing the identifying information in fingerprints can be broadly divided into two main categories: feature-modeling-based approaches and empirical approaches.

A. Feature-modeling-based Approaches

The first attempt to estimate the evidential value of fingerprints by statistically modeling the fingerprint features was made by Galton [25]. His model required partitioning a fingerprint into 24 non-overlapping square regions whose width was equal to six times the inter ridge distance. He argued that each of these square regions can be correctly reconstructed with a probability of $\frac{1}{2}$ if the information regarding the surrounding ridges is known. This leads to a probability of $\left(\frac{1}{2}\right)^{24}$ that the complete fingerprint can be reconstructed, given the ridge structure in the region surrounding the square regions. Galton further noted that the probability that the correct number of ridges enter and exit the 24 squares is $\frac{1}{256}$ and that the probability of occurrence of a specific type of finger (e.g. whorl, loop, arch, etc.) is $\frac{1}{16}$. This assumption leads to a probability of $\frac{1}{16} \times \frac{1}{256} \times \left(\frac{1}{2}\right)^{24} = 1.45 \times 10^{-11}$ for correctly reconstructing a full fingerprint. This measure of fingerprint individuality has been referred to as the *Probability of Fingerprint Configuration (PFC)* [25]. In later studies, e.g. in [42], the PFC values were characterized by the amount of discriminating information in a fingerprint such as the number of minutiae, fingerprint quality, etc. The PFC value was also viewed as the Probability of False Association (PFA) (See e.g. [14], [30]) which essentially measures the probability that a given fingerprint configuration will perfectly match one of the k available fingerprints. Mathematically,

$$PFA = 1 - [1 - PFC]^k. \tag{1}$$

Note that most of the above approaches involved manual analysis of fingerprints. Champod and Margot [16] were, however, the first to use automatically extracted minutiae for computing PFA. A thorough discussion of other similar approaches is provided in [44].

One of the limitations of the PFC (and PFA) is that it does not take into account the characteristics of a fingerprint comparison e.g. the number of matching minutiae, size of overlap between the two fingerprints being matched, number of non-matching minutiae in the overlapping region, etc. Pankanti et al. [38] first incorporated these match characteristics in quantifying the individuality of fingerprints using the so called *Probability of Random Correspondence (PRC)*. Given a query fingerprint containing n minutiae, they computed the PRC that an arbitrary template fingerprint containing m minutiae will have exactly s mated minutiae with the query. Thus the PRC value is computed as

$$PRC(s) = P(s|I, m, n) \tag{2}$$

where I refers to the impostor pair of fingerprints, i.e. the two fingerprints being compared belong to different fingers.

Pankanti et al. assumed a uniform distribution to model the location and direction of each minutia in a fingerprint

Study	Model description	Database
Pankanti et al. (2002)	Modeled minutiae location and orientation using uniform distribution	668 fingers, 4 impressions per finger
Chen and Moon (2007)	Extended Pankanti et al.'s model by using von-Mises distribution to model the minutiae direction	383 fingers, 1149 fingerprints
Zhu et al. (2007)	Extended Pankanti et al.'s model by using finite mixture models for minutiae location and direction	NIST SD4 (2,000 fingerprints), FVC2002 DB1 (800 fingerprints) and FVC2002 DB2 (800 fingerprints)
Fang et al. (2007)	Extended Zhu et al.'s model by including fingerprint ridges	FVC2002 DB1 (800 fingerprints)
Su et al. (2009)	Extended Zhu et al.'s model by including ridge flow and ridge points	NIST SD4 (2,000 fingers, 2 impressions per finger)
Chen and Jain (2009)	Extended Zhu et al.'s model by including ridges and pores	NIST SD4 (2,000 fingers, 2 impressions per finger)
Su et al. (2010)	Used Bayesian networks to obtain minutiae correspondence	NIST SD4 (2,000 fingers, 2 impressions per finger), NIST SD27 (258 latents and mated full fingerprints)

TABLE I
TECHNIQUES FOR ASSESSING THE EVIDENTIAL VALUE OF FINGERPRINTS BASED ON PRC VALUES.

independently to calculate the PRC value. One limitation of the uniform distribution model is its relatively poor fit to the true minutiae distribution in fingerprints. To address this issue, a number of subsequent studies attempted to improve the model of Pankanti et al. Chen and Moon [17] extended Pankanti et al.'s model by using von-Mises distribution for minutiae direction. Based on the observation that the minutiae tend to form clusters [45], Zhu et al. [51] used finite mixture models for modeling the minutiae distribution. For each fingerprint, a Gaussian distribution was fit to the minutiae locations and a Von-Mises distribution was used for the minutia directions in each component of the mixture. Fang et al. [24] and Su et al. [46] extended this framework by incorporating information regarding the fingerprint ridges. Chen and Jain [18] modeled three different levels of fingerprint features: level 1 (pattern type), level 2 (ridges and minutiae) and level 3 (pores). Su et al. [47] incorporated dependence between neighboring minutiae using Bayesian networks. Despite these developments, there are two main limitations of these model-based studies: i) the matching criteria used, e.g. the number of matching minutiae, is very rudimentary and does not represent the matching criteria of state-of-the-art algorithms, and ii) intra-class variations in fingerprints (see Figure 3), a major source of matching errors, is not explicitly considered in computation of the PRC value.

B. Empirical Approaches

A number of empirical approaches for computing the evidential value have been reported in the literature. Meagher et al. [31] were the first to utilize the FBI's Automated Fingerprint Identification System (AFIS) to compute the evidential value of fingerprints in response to the first legal challenge against fingerprint evidence in the courts, namely *U.S. v. Byron C. Mitchell*. They simulated latent images by partial prints obtained by cropping 50,000 different rolled fingerprints. These partial fingerprint images were compared with the original rolled images to obtain the genuine and impostor match scores. Assuming that the genuine and impostor scores follow a Gaussian distribution, Meagher et al. computed the probability of finding a pair of two unrelated fingerprints, whose match score is greater than the smallest of the genuine match scores observed, as 10^{-97}. One major shortcoming of this study is that the intra-class variation (see Figure 3) was not accounted for since only one image per finger was utilized.

Fig. 3. Two fingerprints belonging to the same finger that appear to have different characteristics due to skin distortion during image acquisition. This illustrates the intra-class variability in fingerprints, a major obstacle in defining quantitative measures of evidential value. Note that placing a simple bounding box around each minutia during comparison is not sufficient to account for the intra-class variation.

Neumann et al. [33], [34], [35] developed a fingerprint matching procedure based on matching a local neighborhood of a small number of minutiae and then converting the resulting similarity value s into a *likelihood ratio (LR)*,

$$LR(s) = \frac{P(s|G)}{P(s|I)}, \tag{3}$$

where I refers to impostor fingerprint pairs (non-mated pairs) and G refers to genuine fingerprint pairs (mated pairs). This likelihood value was proposed as a measure of the evidence captured by a fingerprint comparison. In order to obtain multiple samples of the same minutiae configuration from a fingerprint, which is required in estimating the within class distribution, Neumann et al. artificially applied random non-linear distortion to the minutiae configuration.

In Egli et al.'s approach [23], the corresponding minutiae are manually identified between the given latent and full fingerprint. These corresponding minutiae are then matched using an automatic matcher and the resulting match score is used to compute the likelihood ratio. The genuine score distribution, $P(s|G)$, is obtained by matching the corresponding minutiae from multiple impressions of the fingerprint of interest, whereas $P(s|I)$ is obtained by matching the minutiae selected from the latent with those obtained from non-mate fingerprints. Similar to the techniques developed by Neumann

Study	Features	Database	Matcher	Distribution model	Measure of evidence	Limitations
Meagher et al. (1999)	Cropped fingerprint images	50K distinct fingerprints	AFIS	Gaussian	PRC	Genuine match computed by matching the same fingerprints
Neumann et al. (2006)	Manually matched minutiae triplets	4 fingers, 54 impressions each for intra-class and 818 subjects for inter-class variability	In house matcher	Kernel	LR	Requires large # of impressions per finger
Neumann et al. (2007)	Manually matched configurations of 3-12 minutiae	4 fingers, 54 impressions each for intra-class and 890 fingerprints for inter-class variability	In house matcher	Mixture of Gaussians	LR	Requires large # of impressions per finger
Egli et al. (2007)	Manually matched configurations using AFIS	66 livescans, 88 pseudo-marks of a thumb and 15 tenprint cards of the same donor for intra-class, and 100k background fingerprints for inter-class variability	AFIS	Weibull (genuine), log-normal (impostor)	LR	Experiments conducted on a single finger
Proposed approach	Fingerprint images	NIST SD14 (27,000 fingers, 2 impressions each), MSP Database (144,186 fingers, 2 impressions each), Latent databases (1041 latent images with mates)	Two different COTS matchers	Kernel	NMP	Requires large sample size for database categorization w.r.t. quality, area and no. of minutiae

TABLE II
EMPIRICAL TECHNIQUES FOR ASSESSING THE EVIDENTIAL VALUE OF FINGERPRINTS.

et al., this approach also assumes availability of a multiple samples of the fingerprint of interest in order to compute the genuine match score distribution. Note that very few latent fingerprints from the same finger are available at the crime scenes and legacy fingerprint databases also have only few sample images per finger. In a recent publication, Neumann et al. [36] further developed a framework for incorporating the finger digit information (index, middle, ring, little, or thumb) in estimating the evidential value based on the likelihood ratio.

A notable difference among the various LR approaches published is the method used for estimating the genuine and impostor densities. In [34], a kernel approach was used to estimate the genuine and impostor score densities whereas, in [33], a mixture of Gaussians was used. In [35], a combination of exponential and beta functions of the score values was used to compute the likelihood ratio. Egli et al. [23] used Weibull model for genuine score distribution to accommodate the limited availability of multiple samples from the same finger. For the impostor score distribution, they used the log-normal model.

Tables I and II summarize the different techniques proposed to evaluate the evidential value of fingerprints based on feature modeling and empirical approaches, respectively.

III. NMP: THE EVIDENCE OF A FINGERPRINT COMPARISON

Non-match probability (NMP) is an intuitive quantity that captures the evidence associated with a fingerprint comparison. Given a pair of fingerprints, or their match score, NMP measures the probability that a non-mate decision made for the pair is correct. Mathematically, NMP for a match score s between two fingerprints is given by

$$NMP(s) = P(I|s) = 1 - P(G|s) \tag{4}$$

where I and G correspond to impostor and genuine pair class, respectively. Note that NMP follows a colloquial use of probability. An NMP value of 10^{-6} implies a chance of one in a million that the given pair of prints does not belong to the same finger. Further, the range of valid values for NMP is bounded by the unit interval $[0, 1]$ with a probabilistic interpretation. An NMP value of 0 indicates that the fingerprint pair, i.e., a latent and rolled pair under consideration is definitely a mate whereas a value of 1 indicates that the pair is definitely a non-mate. The relationship between the match score s and its associated NMP value can be graphically represented by a plot, called the NMP-curve (see Figure 2).

Equation (4) provides a compact mathematical representation of the NMP, but a direct computation of NMP using this equation is not feasible due to the large number of distinct possible match score values. However, applying the Bayes theorem makes the computation of NMP tractable:

$$NMP(s) = P(I|s) = \frac{P(s|I)P(I)}{P(s|I)P(I) + P(s|G)P(G)}. \tag{5}$$

The procedure for computing $P(s|G)$ and $P(s|I)$ is detailed in Section IV-A. Note that in eq. (4) the prior probabilities for genuine and impostor classes were implicit in the definition of $P(I|s)$ and $P(G|s)$, but in eq. (5), the two prior probabilities, namely $P(G)$ and $P(I)$, are explicit and reflect any additional evidence that may be available for specific matching scenarios [49]. This is one of the strengths of the NMP measure compared to LR and PRC measures. Further, it is easy to incorporate prior information into an NMP value that has been computed with equal priors using the following equation:

$$NMP_p = \frac{NMP \times P(I)}{NMP \times P(I) + (1 - NMP) \times P(G)} \tag{6}$$

where NMP_p is the NMP value with priors incorporated.

The importance of incorporating prior information can be illustrated by a situation where two latent fingerprints are captured from a crime scene with one fingerprint on the weapon used in the crime and the other on a stray object. If the fingerprint on a stray object matches with the suspect,

it is more likely that the fingerprint on the weapon will match the suspect. This information can be incorporated as prior probability while computing the NMP value associated with this comparison. Note that one of the factors determining this prior probability would be the probability that the multiple latent fingerprints belong to the same person. See [43] for an analysis of latent to latent capabilities of available fingerprint matchers.

It is important to note that NMP can be computed from the match score output by any fingerprint matcher. Furthermore, the following equations can be used to compute the NMP values given LR and PRC values:

$$NMP(s) = \frac{PRC(s) \times P(I)}{PRC(s) \times P(I) + P(s|G)P(G)} \quad (7)$$

$$NMP(s) = \frac{P(I)}{P(I) + LR(s)P(G)}. \quad (8)$$

The above expressions require estimates of densities $P(G)$, $P(I)$, and $P(s|G)$ i.e. the prior values and the genuine score distribution. Note that the computation of NMP from LR (and vice-versa) does not require any density estimation. In fact, computation of LR can be considered as an intermediate step in computation of the more intuitive NMP value.

A. The Extended NMP

The image characteristics play a major role in determining the evidential value of a fingerprint comparison. Consider two different pairs of non-mate fingerprints both with the same match score, but with fingerprints in one pair having significantly poor image quality compared to fingerprints in the other pair. While one may be able to decide with certainty that the good quality pair belongs to the impostor class, similar confidence may not be present in making a decision for the poor quality pair. It is thus natural to assign different NMP values to these two fingerprint pairs. A similar effect may also be observed when considering other fingerprint image characteristics such as the number of minutiae and size of the fingerprint. We thus extend the definition of NMP to take these characteristics into account. The mathematical expression of the extended NMP is given by:

$$NMP(\Phi_{f_1,f_2}) = P(I|\Phi_{f_1,f_2}) \quad (9)$$

where f_1 and f_2 represent the two fingerprints being compared and Φ_{f_1,f_2} is the set of covariates that can be computed from f_1 and f_2 such as image quality, number of minutiae, their match score, etc. Based on eq. (9), an NMP-curve can be obtained by computing the values of NMP(Φ_{f_1,f_2}) for different values of match score s while keeping the remaining elements of Φ_{f_1,f_2} fixed based on a specified criteria. Note that a very restricted criteria is equivalent to estimation of the multivariate NMP function in eq. (9) at a finer resolution. However, it is important to maintain a sufficient number of training fingerprint pairs satisfying the given criteria for a reliable estimation of the NMP values. Section IV-B details the various experiments illustrating the effect of different fingerprint characteristics on the resulting NMP values.

B. Conclusiveness: The Significance of NMP-based Evidence

It is important to quantify the significance of a forensic evidence for its proper acceptance in a court of law. We measure the significance of an NMP value using a quantitative measure, called the *conclusiveness*. The *conclusiveness* of a given NMP value measures the extent to which it deviates from a completely equivocal value of 0.5. The *conclusiveness* (γ) for a given NMP value is computed as:

$$\gamma = |NMP - 0.5| \quad (10)$$

The concept of *conclusiveness* can also be extended to a training database of match scores used to generate an NMP-curve. *Conclusiveness* of a database measures the aggregate discriminative information provided by the associated match scores. It also indicates the closeness of the shape of NMP-curve to a step function. The *conclusiveness* of an NMP-curve is defined as:

$$C = \frac{2}{|S|} \sum_{s \in S} |0.5 - NMP(s)| \quad (11)$$

where S is the set of match scores used to construct the NMP-curve. Note that the value of *conclusiveness* is in the range $[0, 1]$ and is invariant to any translation/scaling of the match scores.

C. Simulation of Latent Fingerprints

Databases of latent prints are scarce and seldom large compared to the available full fingerprint databases. The only latent-full fingerprint database available for public use, that we are aware of, is the NIST Special Database-27 which contains only 258 latent-full print pairs. On the other hand, databases containing thousands of full-full print pairs are available; e.g. the NIST Special Database-14 contains $27,000$ full fingerprint pairs. Due to this paucity of latent-full print databases, we simulate latent fingerprints by cropping small regions from the full fingerprints. In order to validate the use of cropped fingerprints as a substitute for latent fingerprints, we compare the associated NMP-curves generated using a coarsely quantized set of match score values. The resulting NMP-curve is expected to be more reliable due to relatively larger number of samples used to compute each point on the curve. See Section IV-D for more details.

D. The Complete Framework

We present a step-by-step procedure that can be followed by a latent examiner to estimate the NMP value of a given latent-full print pair as follows.

- Simulate a large number of latent fingerprints using any available full fingerprint database by cropping regions of different sizes, quality, and number of minutiae. If sufficiently large number of latent fingerprints and their known mates are available such that the required NMP values can be reliably estimated, this step can be avoided.
- Partition the available set of (simulated) latent-full print pairs based on various fingerprint image characteristics (e.g., image quality and size); compute NMP-curves for

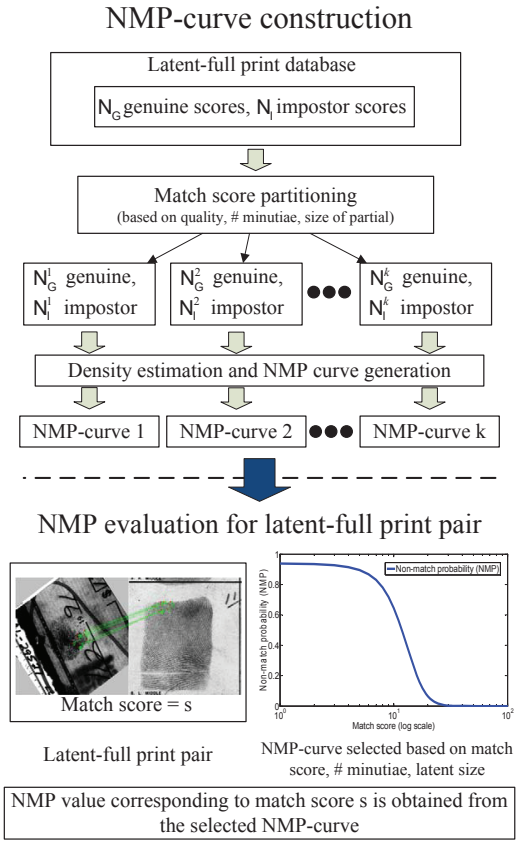


Fig. 4. A schematic diagram for computing the NMP value for a latent-full print pair of interest based on the NMP-curve obtained from a training database of (simulated) latent-full print pairs.

each partition. An effective partitioning of the database ensures that a number of critical covariates are considered and each partition has a sufficient number of latent-full print pairs. Note that the partitions may not be exclusive. The goodness of a partition can be computed based on *conclusiveness*.

- Given a latent-full print pair of interest, select an appropriate NMP-curve based on the covariates used to partition the database.
- Use the match score for the given latent-full print pair to select the NMP value from the associated NMP-curve. The variance of the NMP value is obtained using the procedure detailed in Section IV-A.

Figure 4 shows a schematic diagram for computing the NMP value for a latent-full print pair of interest based on the NMP-curve obtained from a training database of simulated latents. Note that the size of training database and the thus the amount of computation required depends on the precision of the NMP value required in a trial. Further, no manual intervention is required in the above procedure.

IV. Experiments

We use two full (rolled or plain) fingerprint databases (NIST Special Database-14 [10] and the Michigan State Police database [8]) and four latent fingerprint databases (NIST Special Database-27 [10], the West Virginia University (WVU) latent database [9], the CMC latent database [11], and the RS&A latent database [12]) in our experiments. The NIST Special Database-14 contains two rolled impressions for each of the $27,000$ different fingers whereas the Michigan State Police database contains one plain and the corresponding rolled impression for each of the $144,186$ different fingers. The NIST Special Database-27, the WVU latent database, the CMC latent database, and the RS&A latent database contain 258, 449, 134, and 200 latent fingerprints, respectively, along with their mated full prints. Table III lists the characteristics of the different databases we have used and Figure 5 shows sample fingerprint images from these five different databases.

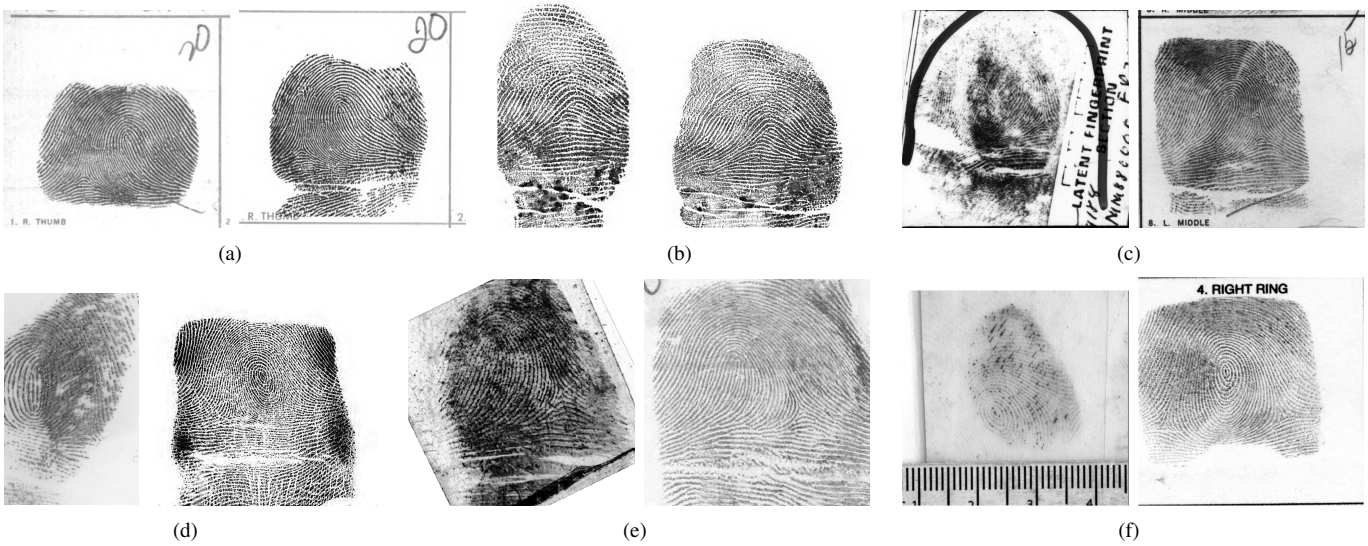
Database	Type	Size
NIST SD 14	full-full pairs	Two rolled impressions for $27,000$ fingers
MSP DB	full-full pairs	Plain and rolled impressions for $144,186$ fingers
NIST SD 27	latent-full pairs	Latent and rolled impressions for 258 fingers
WVU DB	latent-full pairs	Latent and rolled impressions for 449 fingers
CMC DB	latent-full pairs	Latent and rolled impressions for 134 fingers
RS&A DB	latent-full pairs	Latent and rolled impressions for 200 fingers

TABLE III
Details of the two full fingerprint and four latent fingerprint databases used in our experiments.

Fig. 6. Cropped regions of size (b) 400×400, (c) 300×300, and (d) 200×200 from a full fingerprint (a).

In order to simulate partial fingerprints from full fingerprint images, we cropped subimages of different sizes (400×400, 300×300, and 200×200) from the full fingerprint images (see Figure 6). For cropping purposes, we first calculate the region of interest (ROI) of the full fingerprint [27] and then randomly select four points inside the ROI to be used as the centers of the cropping window. We crop four partial images of three different sizes from each full fingerprint. Figure 6 shows a full fingerprint and its cropped images of three different sizes.

These partial fingerprints were matched to the full fingerprints (not used in cropping) to obtain a set of $684,744$ genuine scores for each of the three different sized cropped prints. To limit the number of impostor scores in our analysis, we randomly selected $2,000$ cropped images and compared them with randomly selected non-mate full fingerprints to obtain



(a) (b) (c)

(d) (e) (f)

Fig. 5. Sample latent-full print pairs from the two full fingerprint and four latent databases: (a) NIST SD-14 (two rolled impressions from the same finger), (b) Michigan State Police database (a rolled and a plain impression of the same finger), (c) NIST SD-27 (latent and its rolled mate), (d) WVU latent database (latent and its rolled mate), (e) CMC latent database (latent and its rolled mate), and (f) RS&A latent database (latent and its rolled mate).

2 million impostor scores. Fingerprint feature extraction and matching were performed using two different matchers (COTS 1 and COTS 2). COTS 1 outputs match scores in the range $[0, 1765]$ whereas COTS 2 outputs match score in the range $[0, 21083]$.

A. Density Estimation

We follow three main approaches to estimate the likelihoods $P(s|G)$ and $P(s|I)$: Histogram, kernel density estimation, and parametric density estimation. In the case of the histogram-based technique, the genuine and impostor score histograms are separately normalized prior to their use as probability densities. In the case of kernel density estimation, a Gaussian kernel with bandwidths of 2.0 and 1.5, respectively, was used to estimate the genuine and impostor distributions for the COTS 1 matcher (using the *ksdensity* function in MATLAB). For the COTS 2 matcher, a bandwidth of 2.0 was used to estimate both the genuine and impostor densities. These parameters used in density estimation were empirically selected so that the density estimates are as close to the corresponding match score histograms as possible. In the case of parametric density estimation, Weibull distribution was used to model the genuine match scores and log-normal distribution was used to model the impostor match scores. The choice of these parametric distributions follows [23].

We compute the bias and variance of the NMP-curves as a means of comparison between various density estimation methods. The variance of an NMP-curve can be computed as

$$\Sigma = \frac{1}{|S|}\sum_{s\in S}\left(\frac{1}{B}\sum_{b=1}^{B}(NMP_b(s) - \overline{NMP}(s))^2\right) \qquad (12)$$

where NMP_b corresponds to the NMP-curve associated with the bth bootstrap sample, $\overline{NMP}$ corresponds to the mean NMP-curve obtained by averaging the NMP-curves corresponding to B bootstrap samples and S is the set of match scores used to construct the NMP-curve. In this experiment, 100 bootstrap samples ($B = 100$) and partial prints of size 400×400 were used. The variances of the NMP-curves associated with the three density estimation methods are essentially the same with values of 0.0016, 0.0011, and 0.0008, for histogram, kernel density, and parametric estimates, respectively.

The bias of the kth method of density estimation is computed as

$$\mu_k = \frac{1}{|S|}\sum_{s\in S}|NMP_k(s) - NMP_h(s)| \qquad (13)$$

where NMP_k corresponds to the NMP-curve obtained using the kth method for density estimation and NMP_h is the NMP-curve obtained using the histogram-based density estimate. S is, again, the set of scores used to construct the NMP-curve. The bias for parametric-density-based NMP estimates and kernel-density-based estimates are 0.0161 and 0.0003, respectively (see Figure 7). Due to this large difference in the bias, we use the kernel-density-method-based NMP estimates in further experiments.

B. Analysis of Extended NMP

Here, we study the effect of different fingerprint covariates on the associated NMP-curves.

1) Effect of image size: It is expected that larger partial fingerprints are more discriminative than smaller partial fingerprints. We verified this by obtaining three different NMP-curves, one for each partial print size. Figure 8 shows the NMP-cuves for partial prints of size 200×200, 300×300, and 400×400. There is a direct correlation between the size of the partial print and the *conclusiveness* of the associated NMP-curve.

2) Effect of number of minutiae: It is expected that if there are fewer minutiae in the partial fingerprint, the match score will be less conclusive. To verify this hypothesis, we

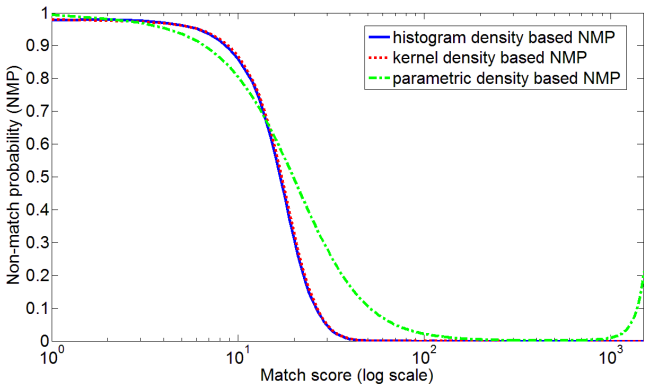


Fig. 7. A comparison of NMP-curves obtained using three different density estimation methods. The NMP-curve obtained using the parametric density estimates deviates largely from that obtained using histogram-based density estimates. These results are based on 400×400 partial prints using COTS 1 matcher. Unless specified otherwise, the results in this paper are obtained using COTS 1 matcher.

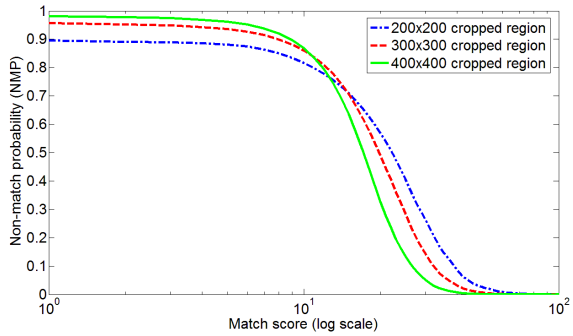


Fig. 8. NMP-curves for various sizes of partial prints. The *conclusiveness* values for 200×200, 300×300, and 400×400 partial prints are 0.988, 0.991, and 0.994, respectively. As expected, the *conclusiveness* values are directly proportional to the size of the partial prints.

divided the partial prints obtained from the two databases (NIST SD 14 and MSP DB) into three groups, each having small, medium, and large number of minutiae. The range of minutiae used for partitioning the database into small, medium, or large number of minutiae category are database specific. This ensures that there are essentially the same number of genuine pairs among different partitions. The *conclusiveness* associated with different partitions based on the number of minutiae are shown in Table IV. As expected, for a given partial print size, the *conclusiveness* of the evidential value is higher for larger number of minutiae. Figure 9 shows the corresponding NMP-curves for partial prints of size 200×200.

3) Effect of image quality: Image quality is a well known covariate of fingerprint matching accuracy [39]. To investigate the effect of image quality on NMP, we partitioned the genuine and impostor match scores into two categories: Good and bad. The good category corresponds to those fingerprint pairs where both the constituent fingerprints are at least of good quality according to the NIST Fingerprint Image Quality (NFIQ) (NFIQ ≤ 3). The remaining fingerprint pairs are assigned to the bad category. Table V lists the number of genuine and impostor pairs used in the analysis. Note that NFIQ assigns one of five quality levels (excellent (1), very good (2), good (3), fair (4), and bad (5)) to a partial or full fingerprint.

Partial print size	No. of minutiae	# Genuine scores	# Impostor scores	Conclu-siveness
200×200	small: [0, 11]	215,113	656,000	0.978
200×200	medium: [12, 16]	209,288	613,000	0.977
200×200	large: [17, 58]	260,343	731,000	0.990
200×200	all: [0, 58]	684,744	2 million	0.988
300×300	small: [0, 24]	213,758	615,000	0.987
300×300	medium: [25, 35]	240,696	746,000	0.992
300×300	large: [36, 115]	230,290	639,000	0.994
300×300	all: [0, 115]	684,744	2 million	0.991
400×400	small: [0, 40]	223,500	561,000	0.992
400×400	medium: [41, 54]	223,745	600,000	0.995
400×400	large: [55, 162]	237,499	839,000	0.996
400×400	all: [0, 162]	684,744	2 million	0.994

TABLE IV
CONCLUSIVENESS OF THE NMP-CURVES BASED ON IMAGE SIZE AND NUMBER OF MINUTIAE.

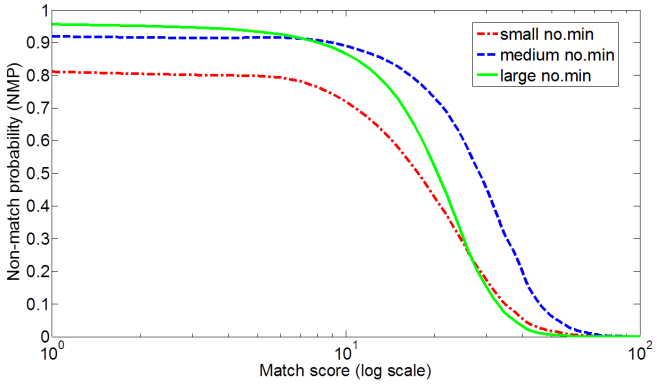


Fig. 9. NMP-curves for 200×200 partial prints with small (0-11), medium (12-16), and large (17-58) number of minutiae. The corresponding *conclusiveness* values are shown in Table IV.

The quality-based NMP-curves are shown in Figure 10. As expected, the good quality fingerprint pairs provide more conclusive NMP values than bad quality fingerprint pairs. For the low match score values, the NMP value for the good quality fingerprint pairs is higher than those corresponding to poor quality fingerprints. This is because if the quality is poor, it is more likely that genuine pairs could lead to low match scores thereby reducing the NMP values corresponding to low match scores.

C. Effect of Prior Information

The prior values, $P(G)$ and $P(I)$, of the genuine and impostor classes also significantly affect the NMP estimates as can be inferred from eq. (6). Figure 11 depicts the relationship between an NMP-curve and the associated values of $P(I)$. Prior values favoring impostor distribution (higher $P(I)$ values) increase the NMP values whereas those favoring genuine distribution (lower $P(I)$ values) decrease the NMP values. Note that prior values allow us to take into account additional information not available in the scores in making a match/non-match decision.

Partial print size	# Genuine scores		# Impostor scores	
	Good	Bad	Good	Bad
200×200	$594,793$	$89,951$	$1,591,577$	$408,423$
300×300	$592,468$	$92,276$	$1,585,496$	$414,504$
400×400	$580,214$	$104,530$	$1,563,917$	$436,083$

TABLE V
NUMBER OF GENUINE AND IMPOSTOR PAIRS USED TO CONSTRUCT THE NMP-CURVES ACCORDING TO DIFFERENT IMAGE QUALITY AND IMAGE SIZES.

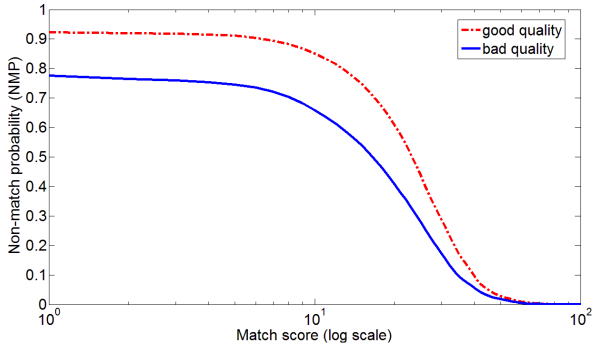


Fig. 10. NMP-curves for the good and bad quality fingerprint pairs for 200×200 partial prints. The *conclusiveness* of NMP curves corresponding to good and bad quality pairs is 0.9894 and 0.9854, respectively.

D. NMP Values for Latent Fingerprints

In practice, the need for estimating the evidential value of a fingerprint comparison usually arises during a forensic examination of a latent fingerprint. It is thus important to validate the proposed technique for real latent fingerprints. Given a latent-full print pair, a database of partial-full print pairs is assigned to it from which the associated NMP value can be reliably obtained. The NMP-curve obtained from this partial-full print database is required to be very similar to the NMP-curve possibly obtained from a database of latent-full pairs that have similar characteristics as the latent-full fingerprint of interest. In this experiment we used the four latent databases: NIST SD-27, WVU latent database, CMC latent database, and RS&A latent database. In addition to the mated full fingerprints available in the four latent databases, the NIST SD-14 was also used as a background database. In our experiments, all the latent fingerprints were manually marked for minutiae whereas minutiae were automatically extracted from the full fingerprint images. Given this data we obtained a set of 1041 genuine and 2 million impostor scores using the two COTS matchers: COTS 1 and COTS 2. A fusion of the match scores obtained by these two matchers using a sum rule based on min-max normalization [29] was also used. See Figure 13 for a comparison of the NMP-curves obtained using the two individual matchers and their score fusion. As expected, the conclusiveness of the NMP-curve using fusion of the two COTS is higher than the individual COTS, demonstrating the benefit of fusion.

Note that the number of genuine scores (1041) in this experiment is much smaller compared to the number of genuine scores $(684,744)$ obtained using the partial-full print matching experiment. Thus, to obtain a reliable estimate of the

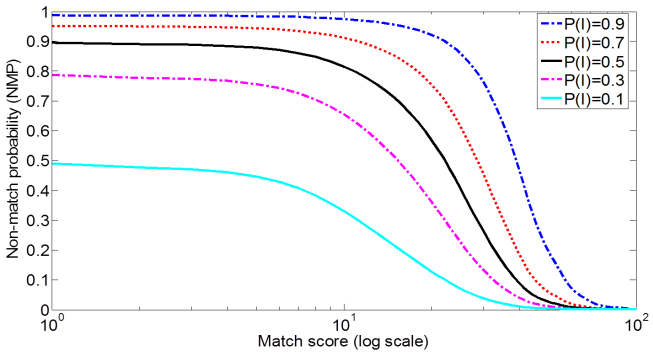


Fig. 11. Effect of impostor prior values, $P(I)$, on NMP-curves for 200×200 partial prints. As $P(I)$ increases, the NMP value also increases monotonically.

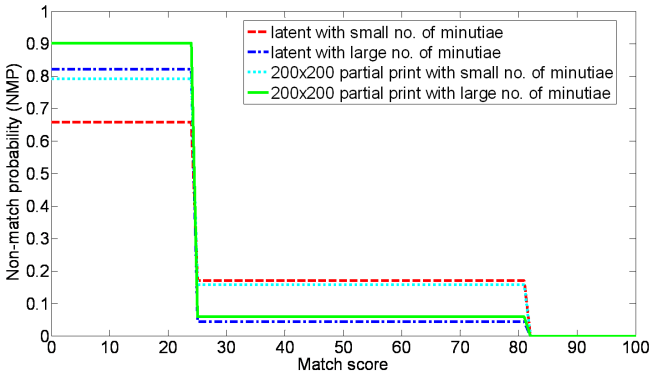


Fig. 12. Comparison between the NMP-curves obtained based on real latents and simulated latents with different number of minutiae.

NMP, we divided the match score values into only three bins such that each bin has an equal number of genuine scores. The genuine and impostor probabilities were then computed for these bins in order to obtain an NMP-curve. The NMP-curve for the partial-full print pairs was also obtained using three bins for a fair comparison. Figure 12 shows a comparison of the NMP-curves associated with real latents to the NMP-curves associated with partial fingerprints with different ranges of the number of minutiae. It can be observed that the NMP-curves for the real latents that have large number of minutiae closely follow the NMP-curve for the partial prints of size 200×200 having large number of minutiae for the middle range of match scores i.e. $[25, 82]$. The difference between the corresponding NMP values is only 0.016. For the match scores in the range $[0, 24]$, the NMP values for the partial-full print pairs are, however, higher than the NMP values for the latent-full print pairs. This is probably due to the generally lower match scores for the latent-full print pairs. In light of these promising results, it is expected that with a more careful partitioning of the partial-full and latent-full print databases, it may be possible to obtain even more similar NMP-curves for these two scenarios.

Figure 14 shows mated latent-full print pairs from the four latent databases used in our experiments. The corresponding values for sum-fusion-rule-based match scores, the NMP, PRC, and LR values are provided in Table VI. Note that the LR values that range in $[0.029, 9.9 \times 10^6]$ are difficult to interpret

Image pair	Match score	# matched minutiae	NMP value	LR value	PRC value
(a)	0.023	5	$0.945 \pm 1.02 \times 10^{-3}$	0.029 ± 0.002	0.187
(b)	0.162	16	$2.5 \times 10^{-5} \pm 2.8 \times 10^{-5}$	$2 \times 10^{4} \pm 2.4 \times 10^{3}$	1.82×10^{-15}
(c)	0.029	2	$0.921 \pm 1.2 \times 10^{-3}$	0.043 ± 0.004	0.231
(d)	0.169	11	$1.1 \times 10^{-5} \pm 2.2 \times 10^{-5}$	$4.5 \times 10^{4} \pm 6.8 \times 10^{3}$	2.92×10^{-9}
(e)	0.046	3	$0.850 \pm 2.4 \times 10^{-3}$	0.088 ± 0.006	0.106
(f)	0.188	12	$7.2 \times 10^{-7} \pm 3.6 \times 10^{-8}$	$6.9 \times 10^{5} \pm 2.2 \times 10^{5}$	4.74×10^{-10}
(g)	0.051	6	$0.832 \pm 2.7 \times 10^{-3}$	0.101 ± 0.006	0.092
(h)	0.235	13	$5.0 \times 10^{-8} \pm 2.8 \times 10^{-9}$	$9.9 \times 10^{6} \pm 6.3 \times 10^{5}$	7.94×10^{-10}

TABLE VI

MATCH SCORES (SUM FUSION OF COTS 1 AND COTS 2 MATCH SCORES), NUMBER OF MATCHED MINUTIAE, NMP AND LR VALUES ALONG WITH 95 PERCENTILE CONFIDENCE INTERVAL, AND THE PRC VALUES COMPUTED FOLLOWING THE MODEL PROPOSED IN [38] CORRESPONDING TO THE SIX LATENT-FULL PRINT PAIRS SHOWN IN FIGURE 14. THESE VALUES ARE BASED ON THE PARTIAL-FULL PRINT DATABASE WITH PARTIAL PRINTS OF SIZE 200×200. NOTE THAT SINCE THE NMP AND LR VALUES NOTED IN THIS TABLE ARE MEAN VALUES BASED ON 100 BOOTSTRAP SAMPLES, THEY DO NOT EXACTLY FOLLOW THE TRANSFORMATION NOTED IN EQ. (8).

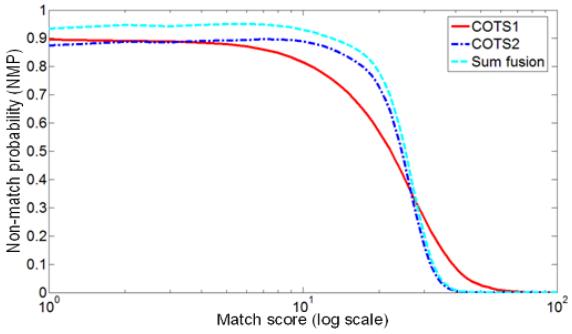


Fig. 13. NMP-curves for two different matchers (COTS 1 and COTS 2) and their fused score for 200×200 partial prints. The *conclusiveness* values for COTS 1, COTS 2 and their min-max-normalization-based sum score-fusion are 0.988, 0.989, an 0.991, respectively. In order to plot the three NMP curves on the same plot, the corresponding impostor match scores are normalized to the same minimum and maximum values.

here. Further, a PRC value of 1.8×10^{-15}, as noted in Table VI, can not be validated since it would require at least 10^{15} impostor matches. NMP values, on the other hand, are empirically obtained and confidence in these values can be easily computed.

We also computed the histogram of NMP values corresponding to the genuine matches from the four latent databases using the two COTS matchers and their min-max-normalization-based sum score fusion. The NMP values are measured using the partial fingerprints of size 200×200. See Figure 15. Note that most of the NMP values are close to 0, indicating that these values correspond to genuine pairs. Further, note that the sum fusion rule in general leads to greater concentration of NMP values towards 0 and the concentration of NMP values at the completely equivocal value of 0.5 is reduced especially compared to the COTS 1 matcher. The noticable concentration around 0.8 corresponds to the poor quality latent that generated very low matching score.

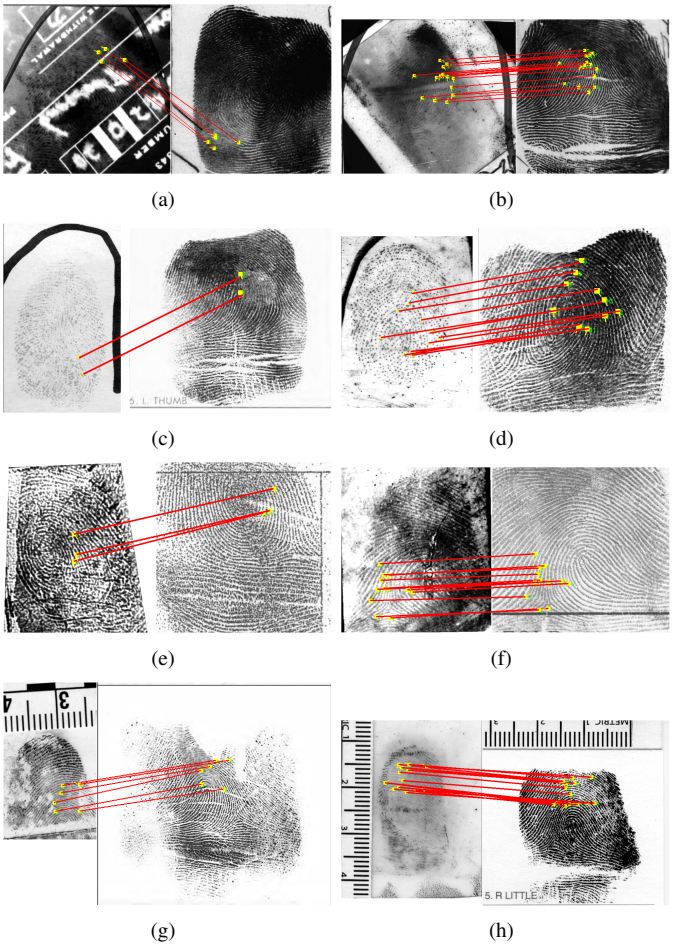


Fig. 14. Mated latent-full print pairs from the four latent databases. (a) & (b) latent-full pairs from NIST SD27, (c) & (d) latent-full pairs from the WVU latent database, (e) & (f) latent-full pairs from the CMC latent database, and (g) & (h) latent-full pairs from RS&A latent database. The corresponding match scores, LR, PRC and NMP values for these fingerprint pairs are shown in Table VI.

E. NMP Values for Top-k Retrievals

Note that the analysis of evidential value performed till this point corresponds to the scenario when a latent fingerprint is directly compared with a full fingerprint. However, in practice, a latent fingerprint obtained from a crime scene is first automatically matched with a large database of full fingerprints and only the top-k retrieved full fingerprints are considered for further matching. We simulated this scenario while estimating the evidential value of latent-full fingerprint comparison. For this experiment, we matched randomly selected $10,000$ query partial fingerprints of size 200×200 with the same number of full fingerprint templates. For each query, only the top 100

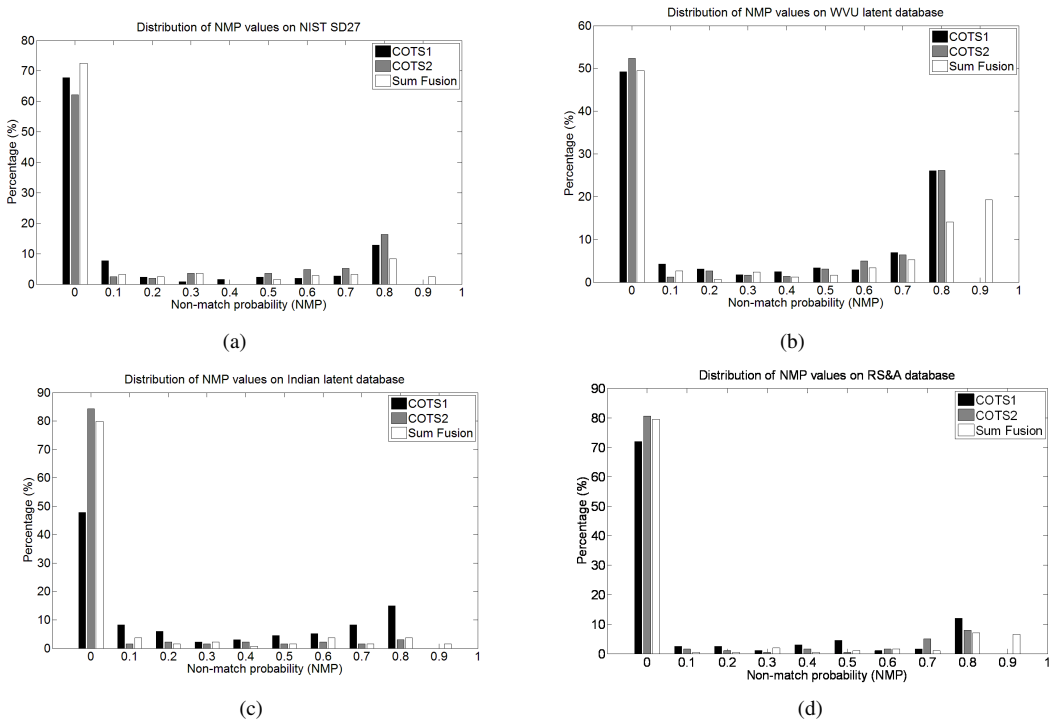


Fig. 15. Comparison of histograms of NMP values corresponding to match scores of true mates obtained using the sum score fusion of two COTS matchers for (a) NIST SD-27, (b) WVU latent database, (c) CMC latent database, and (d) RS&A latent database. COTS 2 failed to generate templates for 7 latent images from NIST SD-27 and 29 images from the WVU latent database. The NMP values are based on partial-full print training database with partial prints of size 200×200.

best matches were used in estimating the NMP-curve. The corresponding NMP-curve along with the NMP-curve obtained under direct comparison with all the full prints is shown in Figure 16. Note that the top-k NMP-curve is shifted to the right compared to the NMP curve with direct comparison score. This is because the match score corresponding to the top-100 retrievals are, in general, expected to be higher than the original distribution of match scores. Further, the top-100 retrieved fingerprints are in general of good quality leading to improved conclusiveness. We also estimated the NMP-cuve in a similar manner for the real latent fingerprints as well. The corresponding curve is shown in Figure 17.

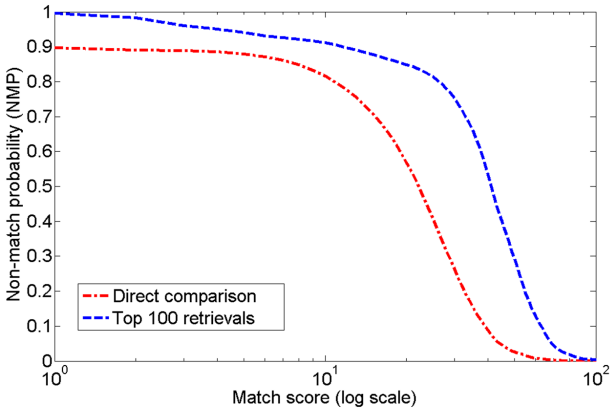


Fig. 16. The NMP-curves for the top-100 retrieved templates. The blue NMP-curve is constructed based on the top 100 retrieved templates when 200×200 partial fingerprints are matched against a database of $10,000$ full fingerprints. The red curve corresponds to the direct-comparison-based NMP-curve obtained using 200×200 partial fingerprints.

V. Summary

In light of the empirically demonstrated non-zero error rates of latent fingerprint matching, and instances of critical errors leading to undue incarceration of innocent individuals, it is crucial to establish the evidential value of a latent fingerprint comparison. We present a framework to capture the evidence of a given fingerprint match score in terms of non-match probability (NMP), namely, the posterior probability that the pair of fingerprints being compared are non-mates. We also studied the variation of NMP values associated with fingerprint databases having specific fingerprint characteristics (image quality, size, and number of minutiae). The NMP values obtained from different partitions of a fingerprint database were compared using a measure, called the *conclusiveness* that estimates the significance of evidence associated with an NMP value. Due to paucity of a large training set of latents, we resort to partial prints obtained by cropping full fingerprints to simulate latents and demonstrate the effectiveness of this simulation. Two full fingerprint databases, four latent databases and two state-of-the-art fingerprint matchers were used in the experiments. We believe the proposed measures of evidential value will allow forensic examiners to present evidence based on latent comparisons in courts of law on a firm footing.

Due to the generic nature of the proposed framework, a number of future studies can be conducted. Below we list some of these directions.

1) We plan to automatically determine the latent image characteristics so that database can be appropriately

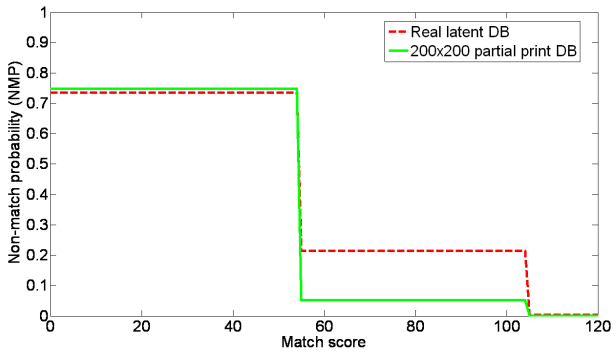


Fig. 17. The NMP-curves for top-100 retrieved templates using latent fingerprints. The green curve is the NMP-curve obtained based on the top-100 retrieved templates when 200×200 partial fingerprints are matched against a database of $10,000$ full fingerprints. The red curve corresponds to the NMP-curve obtained based on the top-100 retrieved templates when the real latent fingerprints were matched against a database of $10,000$ full fingerprints.

partitioned to arrive at the most conclusive estimates of NMP values. Note that these characteristics also severely affect the discriminative capacity of the matching score. Such information can potentially allow design of better ways to combine the outputs of multiple fingerprint matchers leading to higher matching performance. It may also be verified whether by conditioning on these characteristics, the genuine and impostor match score distributions will follow a simple probabilistic model.

2) There is a need to develop an NMP based evaluation technique for matchers and compare it with the traditional Receiver Operating Characteristic (ROC).
3) As one of the major underpinnings of the proposed framework is the empirical validation of an analytical model e.g. the kernel density estimation technique for estimating genuine and impostor match score distributions. We plan to conduct a thorough review and proper selection among available empirical validation procedures.
4) In eq. 7 we show that NMP values can be computed based on the PRC values obtained using a feature modeling approach. We plan to conduct thorough evaluation of the existing feature modeling based approaches under the proposed NMP framework.
5) We would also like to explore the avenue of reliably simulating latent fingerprints for better estimates of NMP.
6) We also plan to consider additional matching scenarios similar to the one described in [21], where the difference between the match scores corresponding to the best retrieval and the average match scores of the top 2 to 10 retrievals is used as the new match score.
7) We would also like to explore the possibility of combining evidence from multiple latent fingerprints captured from a crime scene under the proposed framework.

We are also making attempts to obtain a large database of latents and mated rolled prints to form a larger and diverse training set. Availability of a larger database will significantly improve our ability to perform a more thorough empirical validation.

ACKNOWLEDGEMENT

This work is partially supported by the FBI Biometric Center of Excellence. Anil Jain's research was also partially supported by the WCU (World Class University) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0). Authors would like to thank Karthik Nandakumar and Jianjiang Feng who provided valuable suggestions for improving this manuscript. All correspondence should be directed to Anil K. Jain.

REFERENCES

[1] Daubert v. Merrell Dow Pharmaceuticals. 113 S. Ct. 2786, 1993.
[2] U. S. v. M. Byron, Criminal Action-407 (ED Pa 1999).
[3] U. S. v. Llera Plaza, 179 F Supp 2d 492 (ED Pa 2002).
[4] U. S. v. Llera Plaza, 188 F Supp 2d 549 (ED Pa 2002).
[5] U. S. v. Crisp, 324 F 3d 261 (4th Cir 2003).
[6] State of Maryland v. Bryan Rose, No. K06-0545 (Md., Baltimore Co. Cir. Oct. 19, 2007), http://www.mdd.uscourts.gov/Opinions/Opinions/BrianRoseMem-FINAL.pdf.
[7] U. S. v. Hamza Keita, Criminal Case No. 2008 CF 2 26777.
[8] Forensic Science Division, Michigan State Police. http://www.michigan.gov/msp/0,4643,7-123-1593_3800-15901--,00.html.
[9] Integrated Pattern Recognition and Biometrics Lab, West Virginia University. http://www.csee.wvu.edu/~ross/i-probe/.
[10] NIST Fingerprint Databases. http://www.nist.gov/srd/biomet.cfm.
[11] R&D Centre, CMC Limited. http://www.cmcltd.com/.
[12] Ron Smith & Associates. http://www.ronsmithandassociates.com/.
[13] A Review of the FBI's handling of the Brandon Mayfield Case, 2006. Office of the Inspector General, Oversight and Review Division, U.S. Department of Justice.
[14] L. Amy. Valeur de la preuve en dactyloscopie i. *Journal de la Societe de Statistique de Paris 88*, pages 189–195, 1947.
[15] D. R. Ashbaugh. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC Press, 1999.
[16] C. Champod and P. A. Margot. Computer assisted analysis of minutiae occurrences on fingerprints. In *Proc. International Symposium on Fingerprint Detection and Identification*, page 305, 1996.
[17] J. Chen and Y. Moon. A minutiae-based fingerprint individuality model. In *Proc. IEEE Computer Vision and Pattern Recognition*, June 2007.
[18] Y. Chen and A. K. Jain. Beyond minutiae: A fingerprint individuality model with pattern, ridge and pore features. In *Proc. 2nd International Conference on Biometrics*, pages 523–533, 2009.
[19] H. Choi, A. Nagar, and A. K. Jain. On the evidential value of fingerprints. In *Proc. International Joint Conference on Biometrics*, October 2011.
[20] S. A. Cole. More than zero: Accounting for error in latent fingerprint identification. *Journal of Criminal Law & Criminology*, 95(3):985–1078, 2005.
[21] S. A. Cole, M. Welling, R. Diosi-Villa, and R. Carpenter. Beyond the individuality of fingerprints: a measure of simulated computer latent print source attribution accuracy. *Law, Probability and Risk*, 7(3):165–189, 2008.
[22] I. Dror, K. Wertheim, P. Fraser-Mackenzie, and J. Walajtys. The impact of human-technology cooperation and distributed cognition in forensic science: Biasing effects of afis contextual information on human experts. *Journal of Forensic Sciences*, 2012. To Appear.
[23] N. Egli, C. Champod, and P. Margot. Evidence evaluation in fingerprint comparison and automated fingerprint identification systems - modelling within finger variability. Technical Report 167, Forensic Science International, 2007.
[24] G. Fang, S. N. Srihari, and H. Srinivasan. Generative models for fingerprint individuality using ridge types. In *Proc. 3rd International Symposium on Information Assurance and Security*, pages 423–428, August 2007.
[25] F. Galton. *Finger Prints*. Macmillan, London, 1892.
[26] L. Haber and R. N. Haber. Scientific validation of fingerprint evidence under Daubert. *Law, Probability and Risk*, 7(1):87–109, 2008.
[27] L. Hong, Y. Wan, and A. K. Jain. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(8):777–789, 1998.

[28] M. Indovina, R. A. Hicklin, and G. I. Kiebuzinski. Evaluation of latent fingerprint technologies: Extended feature sets. Technical Report NISTIR 7775, NIST, March 2011.

[29] A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005.

[30] C. Kingston. Probabilistic analysis of partial fingerprint patterns. Ph.D. Thesis, University of California, Berkeley, 1964.

[31] S. B. Meagher, B. Buldowle, and D. Ziesig. 50k fingerprint comparison test. USA v. Byron Mitchell, US District Court Eastern District of Philadelphia. Government Exhibits 6-8 and 6-9 in Daubert Hearing before Judge J. Curtis Joyner, July 1999.

[32] National Research Council. *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press, 2009.

[33] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, and A. Bromage-Griffiths. Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *Journal of Forensic Sciences*, 52(1):54–63, 2007.

[34] C. Neumann, C. Champod, R. Puch-Solis, N. Egli, A. Anthonioz, D. Meuwly, and A. Bromage-Griffiths. Computation of likelihood ratios in fingerprint identification for configurations of three minutiae. *Journal of Forensic Sciences*, 51(6):1255–1266, 2006.

[35] C. Neumann, I. W. Evett, and J. E. Skerrett. Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J. R. Statist. Soc. A*, 175, Part 2:1–26, 2012.

[36] C. Neumann, I. W. Evett, J. E. Skerrett, and I. Mateos-Garcia. Quantitative assessment of evidential weight for a fingerprint comparison i. generalisation to the comparison of a mark with set of ten prints from a suspect. *Forensic Science International*, 207:101–105, 2011.

[37] A. Y. Ng and M. I. Jordan. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. In *Advances in Neural Information Processing Systems (NIPS)*, volume 14, pages 841–848, 2002.

[38] S. Pankanti, S. Prabhakar, and A. K. Jain. On the individuality of fingerprints. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025, August 2002.

[39] C. Wilson et al. Fingerprint vendor technology evaluation 2003: Summary of results and analysis report. Technical Report NISTIR 7123, NIST, June 2004.

[40] Innocence Project. Wrongful convictions involving unvalidated or improper forensic science that were later overturned through DNA testing. http://www.innocenceproject.org/docs/DNA_Exonerations_Forensic_Science.pdf.

[41] M. Indovina et al. ELFT Phase II - an evaluation of automated latent fingerprint identification technologies. Technical Report NISTIR 7577, NIST, 2009.

[42] T. Roxburgh. On evidential value of fingerprints. *Sankhya: Indian J. Statistics*, 1:189–214, 1933.

[43] A. Sankaran, T. Dhamecha, M. Vatsa, and R. Singh. On matching latent to latent fingerprints. In *International Joint Conference on Biometrics*, Washington, DC, October 2011.

[44] D. A. Stoney. *Measurement of Fingerprint Individuality, in Advances in Fingerprint Technology, H. C. Lee and R. E. Gaensslen (eds.)*. CRC Press, Boca Raton, 2001.

[45] D. A. Stoney and J. I. Thornton. A critical analysis of quantitative fingerprint individuality models. *Journal of Forensic Sciences*, 31(4):1187–1216, October 1986.

[46] C. Su and S. N. Srihari. Probability of random correspondence for fingerprints. In *Proc. 3rd International Workshop on Computational Forensics*, pages 55–66, August 2009.

[47] C. Su and S. N. Srihari. Evaluation of rarity of fingerprints in forensics. In *Proc. Neural Information Processing Systems*, December 2010.

[48] C. Su and S. N. Srihari. Generative models and probability evaluation for forensic evidence. In P. Wang, editor, *Pattern Recognition, Machine Intelligence and Biometrics*. Springer, 2011.

[49] F. Taroni, S. Bozza, A. Biedermann, P. Garbolino, and C. Aitken. *Data Analysis in Forensic Science: A Bayesian Decision Perspective*. Wiley, 2010.

[50] B. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts. Accuracy and reliability of forensic latent fingerprint decisions. *Proc. National Academy of Sciences of the United States of America*, 108(19):7733–7738, May 2011.

[51] Y. Zhu, S. Dass, and A. K. Jain. Statistical models for assessing the individuality of fingerprints. *IEEE Trans. on Information Forensics and Security*, 2(3):391–401, 2007.