Latent Fingerprint Value Prediction: Crowd-based Learning

Tarang Chugh, *Student Member, IEEE,* Kai Cao*, Jiayu Zhou,
Elham Tabassi, *Member, IEEE,* and Anil K. Jain, *Life Fellow, IEEE*

***Abstract*—Latent fingerprints are one of the most crucial sources of evidence in forensic investigations. As such, development of automatic latent fingerprint recognition systems to quickly and accurately identify the suspects is one of the most pressing problems facing fingerprint researchers. One of the first steps in manual latent processing is for a fingerprint examiner to perform a triage by assigning one of the following three values to a query latent: Value for Individualization (VID), Value for Exclusion Only (VEO) or No Value (NV). However, latent value determination by examiners is known to be subjective, resulting in large intra-examiner and inter-examiner variations. Furthermore, in spite of the guidelines available, the underlying bases that examiners implicitly use for value determination are unknown. In this paper, we propose a crowdsourcing based framework for understanding the underlying bases of value assignment by fingerprint examiners, and use it to learn a predictor for quantitative latent value assignment. Experimental results are reported using four latent fingerprint databases, two from forensic casework (NIST SD27 and MSP) and two collected in laboratory settings (WVU and IIITD), and a state-of-the-art latent AFIS. The main conclusions of our study are as follows: (i) crowdsourced latent value is more robust than prevailing value determination (VID, VEO and NV) and LFIQ for predicting AFIS performance, (ii) two bases can explain expert value assignments which can be interpreted in terms of latent features, and (iii) our value predictor can rank a collection of latents from most informative to least informative.**

***Index Terms*—Latent value determination, latent matching, latent examiners, crowdsourcing, matrix completion, multidimensional scaling.**

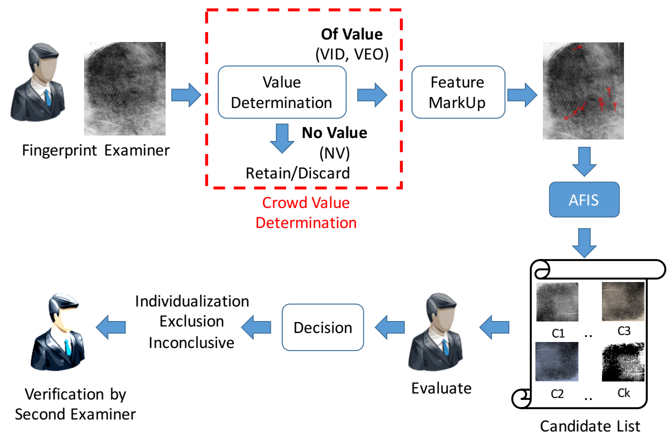

Fig. 1: ACE-V methodology. The proposed latent fingerprint value prediction based on expert crowdsourcing is shown in red block.

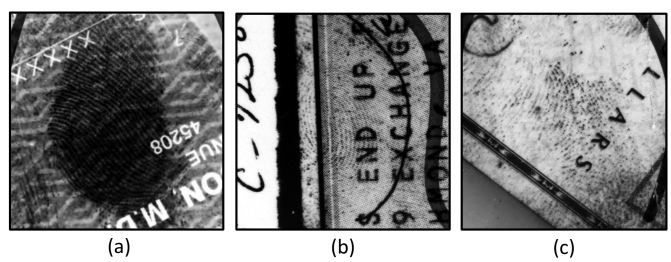

Fig. 2: Latents from NIST SD27 with different values. (a), (b), and (c) were, respectively, determined to be VID, VEO, and NV by examiners in [8]. The corresponding predicted values assigned by the proposed approach are 4.35, 2.34, and 1.26 respectively, where 1 indicates the lowest value and 5 indicates the highest value.

I. Introduction

Latent fingerprints are one of the most crucial sources of evidence in forensic investigations. As such, development of automatic latent fingerprint recognition systems to quickly and accurately identify the suspects is one of the most pressing problems facing fingerprint researchers [2], [3]. Previous publications [4], [5] and NIST evaluations [6], [7] have shown that it is a challenging problem requiring new approaches from machine learning and computer vision.

One of the first steps in the widely practiced Analysis, Comparison, Evaluation and Verification (ACE-V) methodology [9] for latent fingerprint recognition is latent value determination (Fig. 1). During the Analysis stage, a fingerprint examiner performs a triage by assigning one of the following three values to a query latent: Value for Individualization (VID), Value for Exclusion Only (VEO) or No Value (NV) (Fig. 2). Latents deemed to be "of value", which include both VID and VEO types, are then used for comparison to a reference fingerprint database to identify the source of the latent; "no value" or NV latents are simply documented in case files to save examiner's effort in feature markup and comparison [10]. An incorrect value assignment may result in a missed opportunity to find the source of the latent (e.g. when "of value" latents are determined as "no value") or in unproductive use of examiner's effort in feature markup and verification (e.g. when "no value" latents are determined as "of value"). Since state-of-the-art Automated Fingerprint

T. Chugh, K. Cao, J. Zhou and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, 48824. E-mail: {chughtar, kaicao, zhou, jain}@cse.msu.edu

E. Tabassi is with the National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899. E-mail: elham.tabassi@nist.gov

*Corresponding author

A preliminary version of this paper appeared in the proceedings of the 9^{th} IAPR International Conference on Biometrics (ICB), 2016 [1]

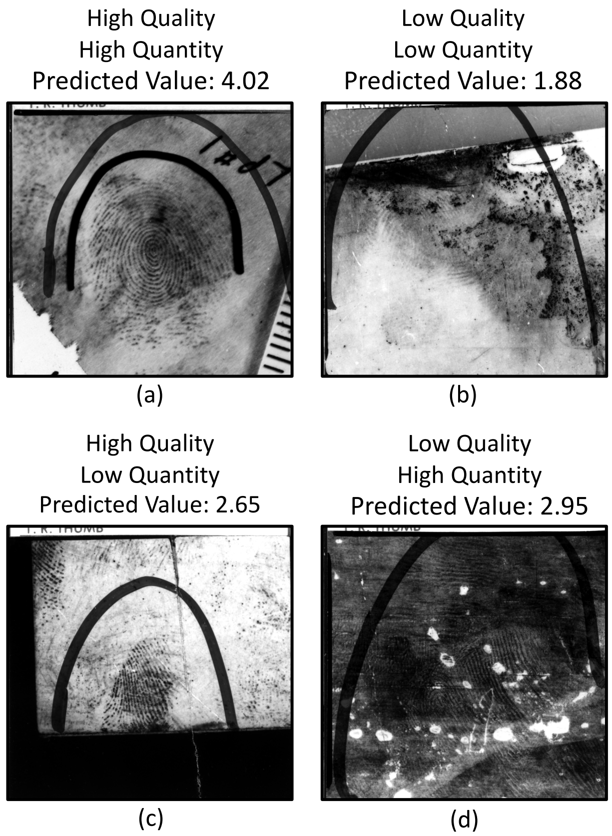

Fig. 3: Difference between latent quality and latent value for four latents from NIST SD27 database. (a) High quality, high quantity latent, (b) low quality, low quantity latent, (c) high quality, low quantity latent, and (d) low quality, high quantity latent as determined by the expert crowd. The predicted value using the proposed approach is also shown for each latent. The value range is [1 - 5] with 1 indicating the lowest value and 5 indicating the highest value.

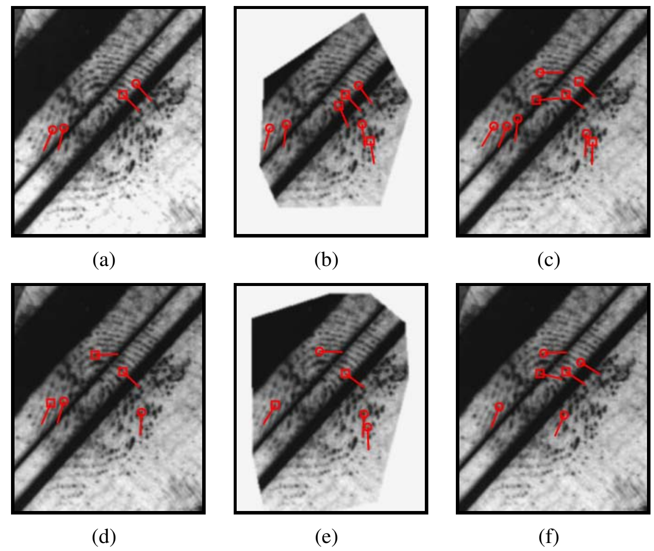

Fig. 4: Variability in minutiae markup, (a) - (f), of the same latent from NIST SD27 by six different examiners. Some examiners also marked Region Of Interest (ROI) in addition to the minutiae. Image from [15].

Identification Systems (AFIS) can process most VID latents[1] in a fully automatic or "lights out" mode with high accuracy, reliable latent value determination is a crucial step.

Latent value determination by examiners is known to be subjective, and large intra-examiner and inter-examiner variations have been reported in [12], [13]. Furthermore, with the growing caseload faced by forensic agencies, there is a need to develop methods for automatic and objective value assignment for latents. To get a sense of the caseload, during November 2016, FBI's IAFIS conducted $21,893$ latent searches with manual markups, and $3,482$ latent searches without manual markups [14]. Automatic latent value assignment rank latents, e.g. found at a crime scene, in terms of their value, for an efficient use of examiners' time and effort.

According to SWGFAST [16], latent value assessments by examiners are based on the quality of features (clarity of the observed features), the quantity of features (e.g., no. of minutiae and friction ridge area), the specificity of features, as well as the relationship between features in a latent. These guidelines are summarized as *quality* and *quantity* of the information present in latent prints in [10], [17], [18]. Note that there is a succinct difference between the quality of a latent and its *value*. Image quality in general [19] is defined as the perceived image degradation based on qualitative (e.g., good, bad, or ugly) or quantitative (e.g., Signal to Noise Ratio or SNR) assessments. In the context of fingerprint images, quality is assessed in terms of the discernibility and reliability of the ridge structure. For example, NIST Fingerprint Image Quality (NFIQ) [20] defines fingerprint image quality as a predictor of AFIS performance. Value of a fingerprint image, on the other hand, incorporates both the quality and the quantity of features. In the case of tenprints (both rolled and slap), the terms quality and value are used interchangeably since they usually have sufficiently large and clear friction ridge areas and consequently, high quality implies high value. However, this is not always true for latent prints. Fig. 3 illustrates this difference between latent quality and latent value based on the inputs from expert crowd.

In spite of the guidelines available for latent value determination [10], the underlying bases[2] that examiners implicitly use for value determination are unknown. In a study on the consensus of latent examiners' value determination [12], [13], Ulery et al. reported that the repeatability (intra-examiner variability) of value determination was 84.6% after a gap of approximately 7 months, while reproducibility (inter-examiner similarity) was only 75.2%. In another study, Ulery et al. [18] modeled the relationship between value determination and feature annotation by examiners and identified low reproducibility of feature markup. Ulery et. al's models did not identify the cause of variations or the actual bases that the examiners adopt for value determination. These bases are needed to learn a

[1] Rank-1 identification rate of 85% is obtained for latents marked as VID in a dataset comprising of NIST SD27 and WVU latents by a state-of-the-art AFIS [11].

[2] In Mathematics, bases refer to a set of linearly independent spanning vectors, *i.e.*, every vector in the vector space is a linear combination of the bases set [21]. Similarly, in this research, bases refer to the underlying dimensions which fingerprint experts implicitly use for latent value determination.

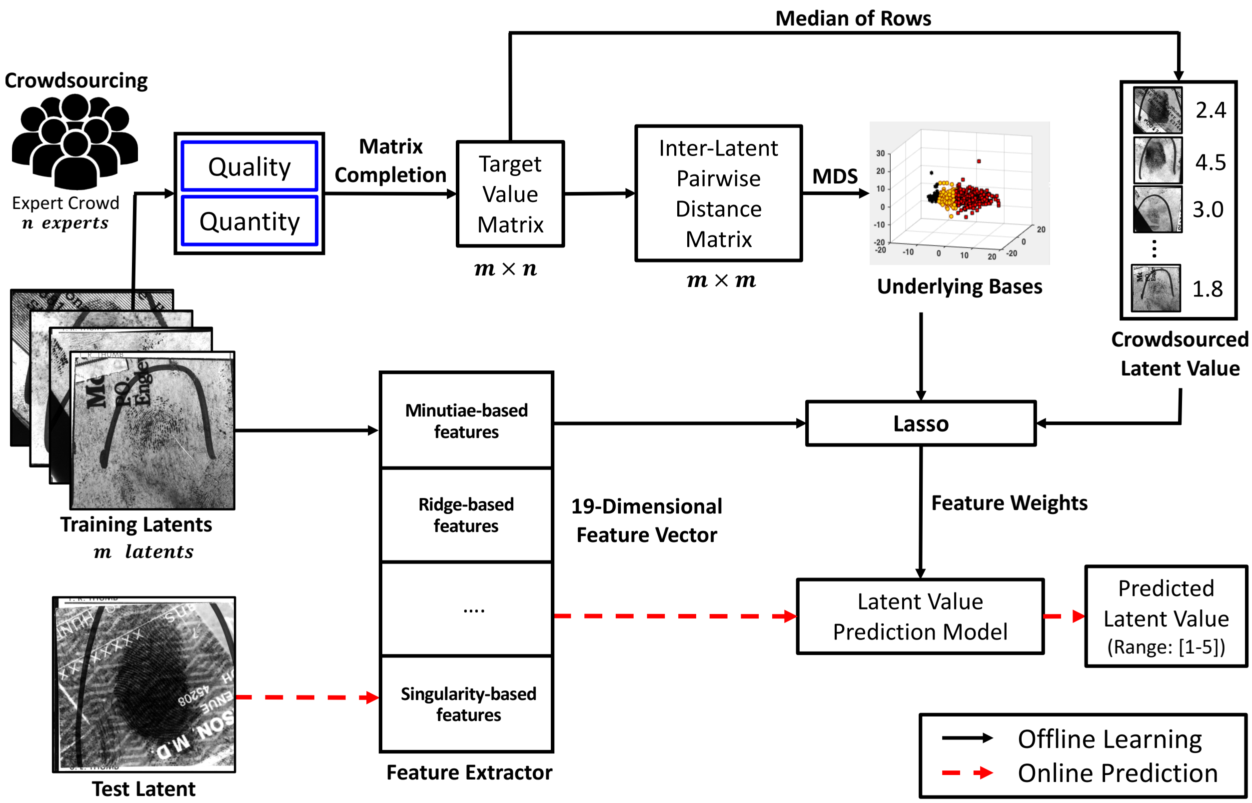

Fig. 5: Overview of the proposed crowdsourcing-based learning approach for latent value assignment.

predictor for quantitative value determination. The examiner subjectivity is also manifested in minutiae markup [15], [22]. Fig. 4 from [15] shows the variability in minutiae markup by six different examiners of the same latent from NIST SD27.

Yoon et al. [23] defined Latent Fingerprint Image Quality (LFIQ) in terms of the average ridge clarity and the total number of minutiae, thereby incorporating both quality and quantity of features. This measure was further extended in [11] by incorporating additional features. A major limitation of [23] and [11] is that they require manually annotated minutiae. Studies on latent value in [24]–[26] also have limitations because they either did not use features explicitly extracted for latents or did not use a latent AFIS for evaluation. We believe that examining the relationship between examiners' value determination (target value) with latent features is necessary to develop a predictor for quantitative value assignment. Another limitation shared by earlier studies is that the target latent value used during training itself may not be reliable because it was established either by a single examiner or by only a few examiners.

We propose a crowdsourcing[3] based framework (Fig. 5) to understand the underlying bases of value assignment by fingerprint experts, and use it to learn a predictor for value assignment. It is well known that the "wisdom of crowd" leads to better decision making [27], [28] as shown in a number of applications [15], [29], [30]. We developed a crowdsourcing tool, called Fingerprint*Mash*, to collect inputs from an expert crowd[4] (Fig. 6). The input consists of quality labels for individual latents and quantity labels for pairwise comparison of latents. We employed Multidimensional Scaling (MDS) [31], a well known ordination and visualization tool, to identify the bases that explain the inter-examiner variations. The relationship between automatically extracted latent features [1] and underlying bases identified by MDS is established using Lasso [32], resulting in a predictor for quantitative latent value assignment. The key contributions of this study are as follows:

1) Design a crowdsourcing tool, Fingerprint*Mash*, to collect quality labels from fingerprint experts for individual latents and quantity labels for pairwise comparisons of latents.
2) Identify the underlying bases that fingerprint experts use for value assignment via MDS and relate them to automatically extracted latent features using Lasso.
3) Learn a prediction model to automatically assign quantitative values, which can be used to rank a collection of query latents. Furthermore, the predicted latent value is shown to have high correlation with the performance of a state-of-the-art latent AFIS.

The major differences between this paper and our preliminary study [1] are as follows: (i) Learning latent value predictor based on target values from a crowd of fingerprint experts as opposed to just one or few experts, (ii) predicted values now are quantitative and multi-valued compared to

[3]https://en.wikipedia.org/wiki/Crowdsourcing

[4]The expert crowd comprises of 13 latent fingerprint examiners and 18 researchers working in the field of latent fingerprints.

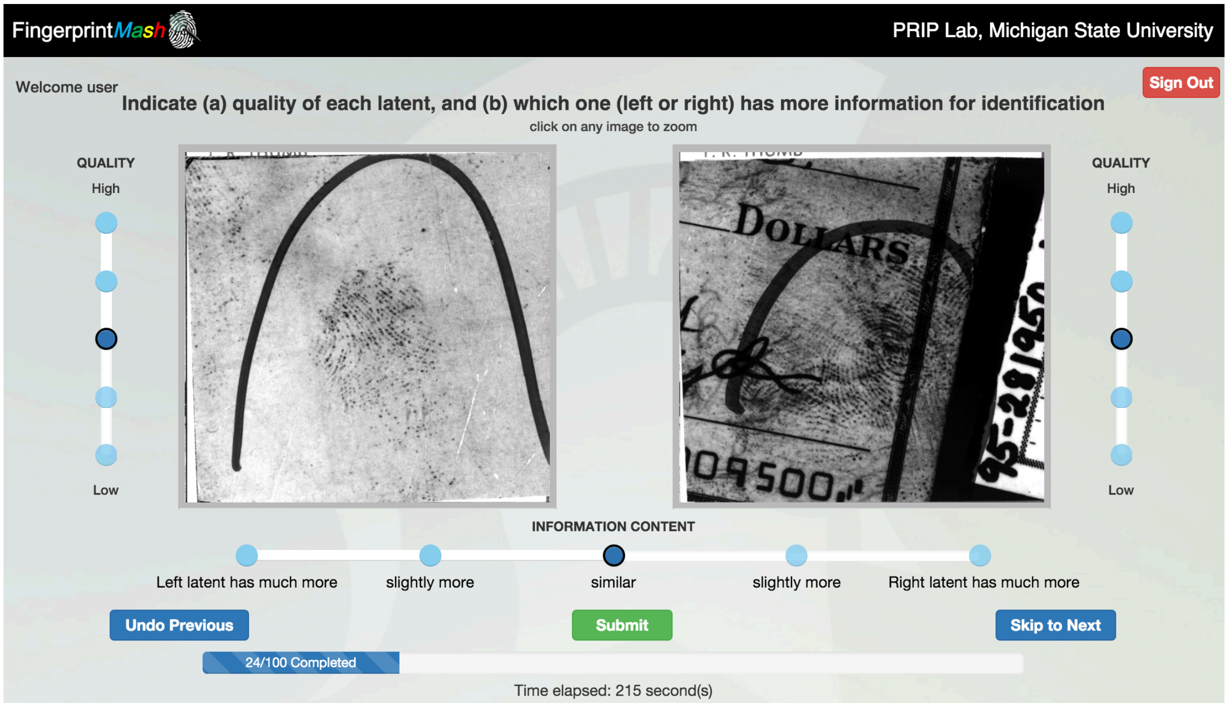

Fig. 6: Interface of the expert crowdsourcing tool, called Fingerprint*Mash*. Each expert assigns a quality label [Low - High] to each latent and provides a pairwise comparison label in terms of quantity of features or information content for the latent pair displayed on the screen. A total of 100 latent pairs, chosen randomly from our database of 516 latents, are presented to each expert.

binary values (VID vs. Not-VID), (iii) identify the underlying bases to explain value assignment in terms of automatically extracted latent features, and (iv) experiments are reported on 4 different latent databases, two from forensic casework (NIST SD27 and MSP) and two collected in laboratory settings (WVU and IIITD MOLF) compared to 2 latent databases in [1].

II. Proposed Framework

The proposed framework involves expert crowdsourcing of latent value assignments via Fingerprint*Mash*[5], inferring missing values of latents not presented to the crowd using matrix completion, identifying underlying bases for expert value assignments using MDS and Lasso, and learning a prediction function to automatically assign values to query latents (Fig. 5). For clarity, the notations used in the paper are summarized in Table I.

A. Crowdsourcing

We requested a pool of fingerprint experts[6] to assign labels to a set of randomly selected 100 latent pairs from a database of 516 latents. These 516 latents (500 ppi) are obtained from two operational latent databases: 258 latents each from NIST SD27 [33] and Michigan State Police (MSP) database[7], which come with their mated rolled prints (Fig. 7). Using the FingerprintMash interface, experts provided the following for each pair of latent images:

- a quality label for each latent on a 5-level scale of low to high, where level-1 represents the worst quality and level-5 represents the best quality, and
- the relative quantity of information, *i.e.* whether the left or the right latent fingerprint contains much more, slightly more or similar quantity of information.

For quality labels, we use a 5-level labelling for each latent in order to be consistent with NFIQ [20]. But, for quantity of information, the pairwise comparison approach is preferred because two experts who agree on the relative preferences of two latents, may assign them different labels, when labelled independently [34]–[36]. To validate the reliability of expert responses, every fifth comparison that is presented contains a randomly selected pair from the set of already compared pairs, providing us a validation set of 20 latent pairs per expert. The positions (left or right) of the latents are also swapped to avoid any bias in expert response. The pairs of latent images presented to the experts neither contained the quality labels nor their source.

With 31 experts in our study, we obtained a total of $6,200$ independent quality labels and $3,100$ quantity labels for pairwise comparisons of latents. Since it is not feasible to expect each expert to label all possible pairs of 516 latent images, two different matrix completion techniques, one for individual latent labels (Section II-D) and the other for pairwise comparisons (Section II-E), were used to infer the missing labels. It has been shown that only $O(r \log m)$ labels per expert are required to accurately recover the ranking list of m items (latents), where $r \ll m$ is the rank of the unknown rating matrix [36] and is determined by the number of bases based on which fingerprint experts assign latent value. In our

[5] Available at http://www.fingerprintmash.org
[6] We sent e-mails to a large pool of fingerprint experts; 31 experts responded.
[7] MSP latent database is not available in the public domain.

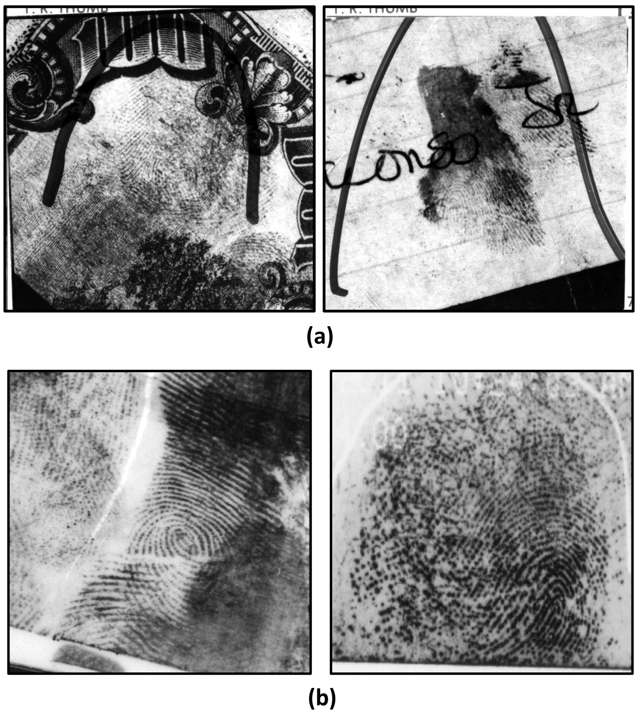

Fig. 7: Two example latents from (a) NIST SD27 and (b) MSP latent databases, each consisting a total of 258 latents.

study involving $m = 516$ latents and maximum possible rank of $r \leq 31$, we require an average of 85 comparisons per expert to recover the ranking list. Given that the number of comparisons provided by each expert in our study is 100, we have sufficient crowd data for matrix completion.

B. Crowd Reliability

One of the important issues in crowdsourcing is to determine the reliability of experts in their annotations [28], [37], [38]. We define the reliability of an expert by consistency between their responses on the 20 repeated latent pairs (40 individual latents) in the validation set. An expert with significantly larger variance in his responses compared to others is deemed unreliable. For analysis, the five level quality labeling [Low - High] is mapped to a numerical scale [1 - 5]. Let $\{NR_{ij}^1\}_{j=1}^{40}$ and $\{NR_{ij}^2\}_{j=1}^{40}$ denote the latent quality labels for the 40 repeated latents labelled by the i^{th} expert at the first and second occurrence, respectively, where $NR_{ij}^1, NR_{ij}^2 \in \{1, 2, 3, 4, 5\}$ denote one of the five latent quality labels. The intra-expert variation (NV_i) of the i^{th} expert is then defined as

$$NV_i = \frac{1}{40} \sum_{j=1}^{40} (NR_{ij}^1 - NR_{ij}^2)^2 \tag{1}$$

Fig. 8 presents the intra-expert variations of quality labels for all the 31 experts in our crowdsourcing task. Three experts (8, 13 and 28) show significantly larger variations compared to others (the average variance is 0.36).

Similarly, let $\{CR_{ij}^1\}_{j=1}^{20}$ and $\{CR_{ij}^2\}_{j=1}^{20}$ denote the pairwise comparisons for the 20 repeated latents pairs labelled by the i^{th} expert at the first and second occurrence, respectively, where $CR_{ij}^1, CR_{ij}^2 \in \{-2, -1, 0, 1, 2\}$ denote one of the five comparison options, where -2 (2) represents the left (right) latent has much more quantity of information present than the right (left) latent, respectively. The intra-expert comparison value variation (CV_i) of the i^{th} expert is then defined as

$$CV_i = \frac{1}{20} \sum_{j=1}^{20} (CR_{ij}^1 - CR_{ij}^2)^2 \tag{2}$$

Fig. 8 shows two experts (13 and 16) exhibit larger intra-expert variation in the quantity labels compared to others (the average variance is 0.24). Overall, the four experts (8,13,16 and 28) show higher variation in their quality and quantity labels

TABLE I: Notations used in this paper.

Notation	Definition
m	Number of latent images
n	Number of fingerprint experts
L_i	i^{th} latent, $i = 1, 2, ..., m$
E_k	k^{th} expert, $i = 1, 2, ..., n$
$Q \in \mathbb{R}^{m \times n}$	Sparse quality matrix collected from expert crowd
$\widehat{Q} \in \mathbb{R}^{m \times n}$	Complete quality matrix after matrix completion
C	Set of triplets that encode all the pairwise comparisons
$\widehat{C} \in \mathbb{R}^{m \times n}$	Complete quantity matrix after matrix completion
$L_j \succ_{E_k} L_i$	Latent L_j is deemed to contain more quantity over latent L_i by expert E_k
$L_j \simeq_{E_k} L_i$	Latent L_j is deemed to have similar quantity as latent L_i by expert E_k
D_v	Dissimilarity matrix between 516 latents based on inferred latent value matrix
$X \in \mathbb{R}^{m \times 20}$	Normalized latent feature matrix consisting of 19 features and 1 constant term for m latents
$Y_i \in \mathbb{R}^m$	i^{th} dimension obtained via MDS
$\widehat{V} \in \mathbb{R}^{m \times n}$	Complete latent value matrix after fusion
$\widehat{v} \in \mathbb{R}^m$	Target latent value vector (median of rows of $\widehat{V}$)

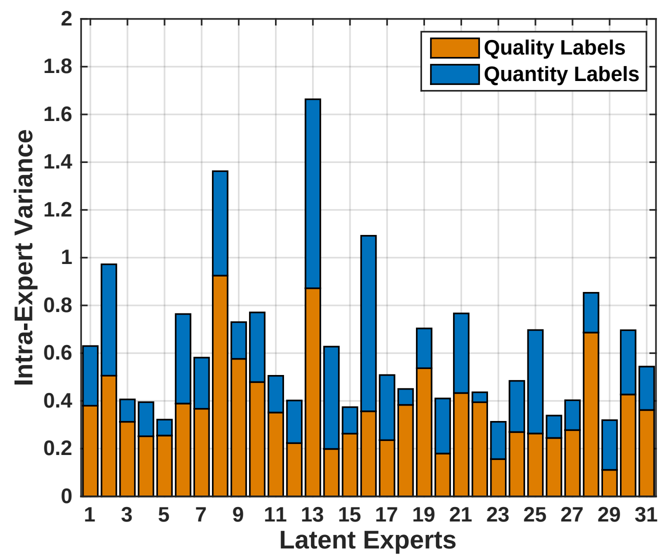

Fig. 8: Intra-expert variations in quality ratings (Eq. (1) and pairwise comparisons (Eq. (2)).

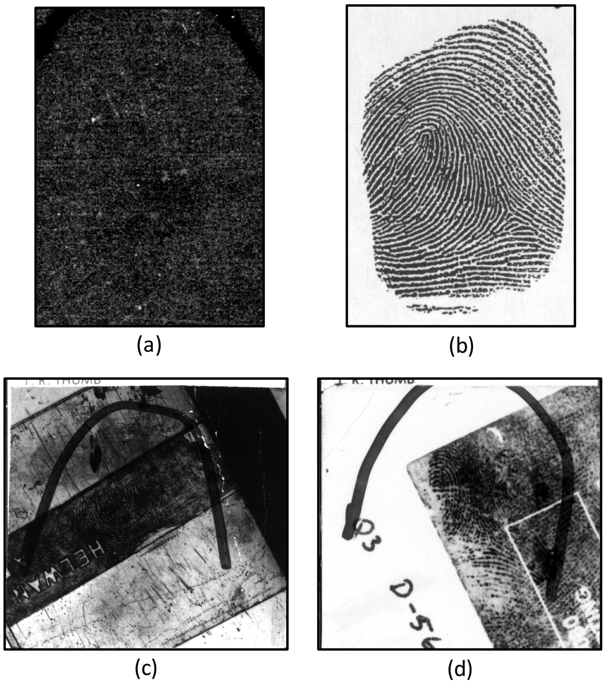

Fig. 9: Inter-expert variations in quality labels. All the experts assigned (a) the lowest quality (*i.e.* 1), and (b) the highest quality (*i.e.* 5). On the other hand, latents in (c) and (d) received both low and high quality labels from the experts.

compared to others. A hypothesis test (p-value approach) on combined intra-expert variations indicated that only expert 13 exhibited a significantly larger variance (p-value < 0.01), and therefore deemed unreliable. After discarding the responses from expert 13, the expert labels [14 - 31] are mapped to [13 - 30] for further analysis for a total of 30 experts.

C. Inter-Expert Variations

On average, 12 experts assigned independent quality labels and 6 experts assigned pairwise comparisons of quantity, to each latent and each pair of latents, respectively. Fig. 9 (a) and (b) show two example latents with zero variation in quality labels as all experts to whom this latent was presented, assigned (a) the lowest quality *i.e.* 1, and (b) the highest quality *i.e.* 5. On the other hand, we observe large variations in quality labels for latents in Fig. 9 (c) and (d) with inter-expert variations of 1.61 and 1.78, respectively. Fig. 10 shows two examples of latent pairs, where latent pair in Fig. 10 (a) has zero variation in pairwise comparisons as all experts to whom this pair was presented, assigned "left latent has much more quantity". On the other hand, the latent pair in (b) was labeled by 4 experts as [-1, 0, 1, 1], where [-1] indicates "left latent has slightly more quantity", [0] indicates "both latents have similar quantity", and [1] indicates "right latent has slightly more quantity", with an inter-expert variation of 0.92.

D. Matrix Completion for Quality Labels

The crowdsourced data for latent quality can be represented by a sparse matrix $Q \in \mathbb{R}^{m \times n}$, where $m = 516$ is the total number of latents in our database and $n = 30$ is the number

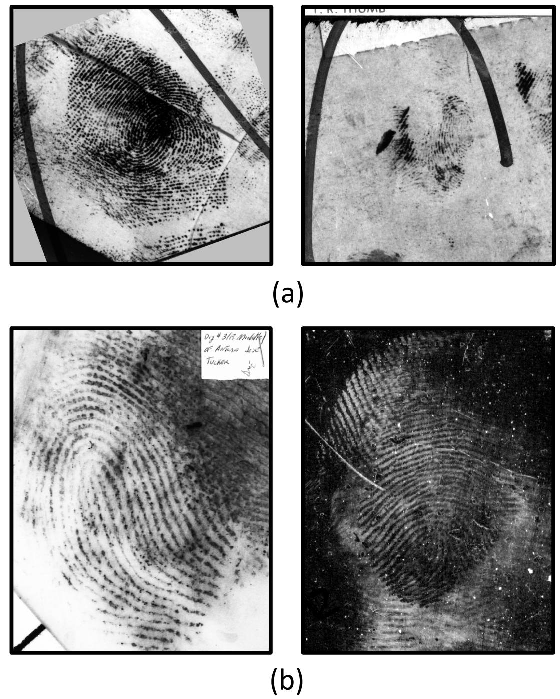

Fig. 10: Inter-expert variations in pairwise comparison labels. Latent pair in (a) was presented to 3 experts and all of them assigned the label "left latent has much more quantity". Latent pair in (b) was labeled by 4 experts as [-1, 0, 1, 1], where [-1] indicates "left latent has slightly more quantity", [0] indicates "both latents have similar quantity", and [1] indicates "right latent has slightly more quantity".

of fingerprint experts who provided reliable quality ratings. The observed rating $Q_{i,j} > 0$ $(i = 1, 2, ..., m; j = 1, 2, ..., n)$ represents the quality for the i^{th} latent assigned by the j^{th} expert. In cases where the i^{th} latent appears more than once for the j^{th} expert, $Q_{i,j}$ is computed as the average of the ratings for that latent. Note that only up to $4,800$ $(30 \times 2 \times 80)^*$ out of a possible total of $15,480$ (30×516) elements in Q are non-zero. Given the sparse quality matrix Q, the task at hand is to infer the complete quality matrix $\widehat{Q}$ using matrix completion [39]. Let Ω be the set of available quality ratings q_{ij} from the sparse matrix Q and $P_\Omega(Q)$ be the orthogonal projector onto the subspace of Ω, which is equal to q_{ij} if $(ij)^{th}$ quality rating $\in \Omega$, zero otherwise. Mathematically, our goal is to find a low-rank matrix $\widehat{Q}$ such that the observed ratings (*i.e.*, indices in Ω) are as close to the inferred ratings as possible. The problem is formulated as minimization of the function: $||\mathcal{P}_\Omega(Q - \widehat{Q})||_F^2$. Since any matrix $\widehat{Q} \in \mathbb{R}^{m \times n}$ of a rank up to K can be represented as a product of two matrices with form $\widehat{Q} = UV$, where $U \in \mathbb{R}^{m \times K}$ and $V \in \mathbb{R}^{K \times n}$, the objective function for inferring the latent quality matrix $\widehat{Q}$ can be formulated as a non-convex optimization problem:

$$\min_{U \in \mathbb{R}^{m \times K}, V \in \mathbb{R}^{K \times n}} \frac{1}{2} ||\mathcal{P}_\Omega(Q - UV)||_F^2, \tag{3}$$

where K controls the rank of $\widehat{Q}$. Popular minimization approaches include nuclear norm minimization [40] and non-linear Gauss-Seidel scheme [41]. We employed the latter for

*Up to 80 unique pairs of latents per expert excluding validation set

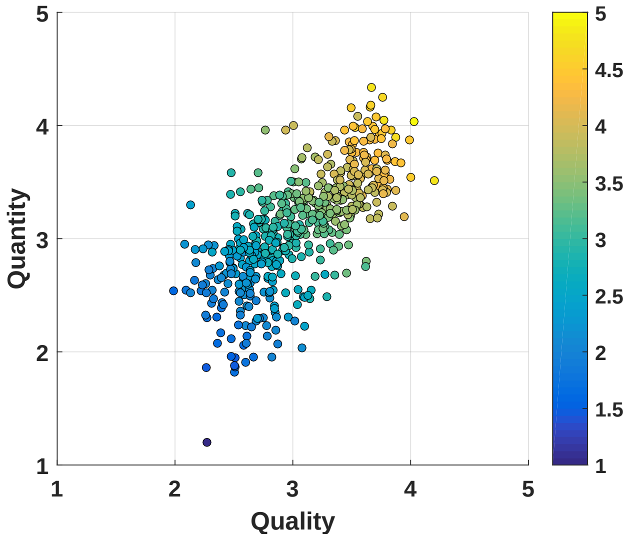

Fig. 11: 2-dimensional scatterplot of quality and quantity (after matrix completion) of the 516 latents. The marker color encodes the latent value in the range $[1-5]$, derived using Eq. 5.

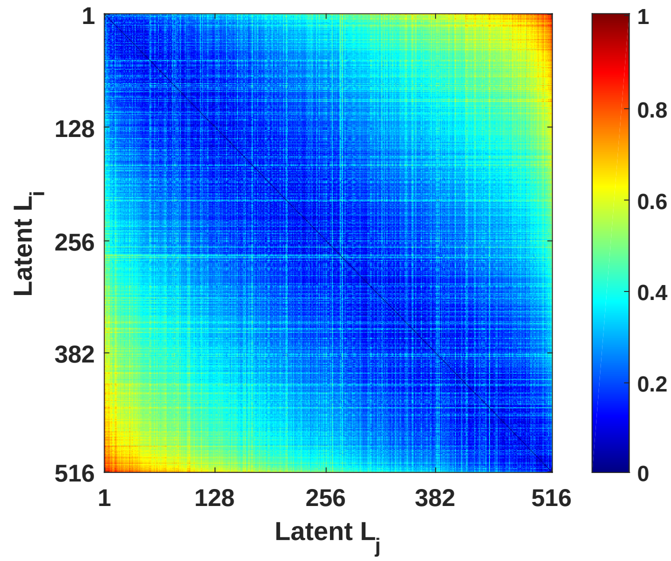

Fig. 12: Heatmap of 516×516 matrix showing dissimilarity between latents, sorted in descending order by the median latent value from expert crowd. Best viewed in color.

its efficiency to solve non-convex models (Eq. 3), even for a large matrix such as Q.

E. Matrix Completion for Quantity Labels

Since the matrix completion algorithms that operate on pairwise comparisons are designed to handle only three ordinal values ($<, =, >$), we merge the "much more" and "slightly more" labels to have a total of three labels instead of five.

Let $(L_i \succ_{E_k} L_j)$ denote that latent L_i contains more quantity over latent L_j assigned by expert E_k and let $(L_i \simeq_{E_k} L_j)$ denote L_i has similar quantity as L_j. We encode each pairwise comparison $(L_i \succ_{E_k} L_j)$ by a triplet (E_k, L_i, L_j), $(L_i \simeq_{E_k} L_j)$ by two triplets (E_k, L_i, L_j) and (E_k, L_j, L_i), and denote the set of triplets that encode all the pairwise comparisons provided by n experts as $C = \{(E_k, L_i, L_j)\}$. Our goal is to infer the quantity labels for each individual latent rated by each expert, from the set of pairwise comparisons C by finding a low-rank matrix $\widehat{C} \in \mathbb{R}^{m \times n}$. This is achieved by minimizing the following objective function:

$$\min_{\widehat{C} \in \mathbb{R}^{m \times n}} \mathcal{L}(\widehat{C}) = \Lambda ||\widehat{C}||_{tr} + \sum_{(i,j,k) \in C} l(\widehat{C}_{i,k} - \widehat{C}_{j,k}), \quad (4)$$

where $\Lambda > 0$ controls the tradeoff between minimizing the rank of $\widehat{C}$ and reducing the inconsistency with respect to the observed pairwise comparisons in C. Note that $||\widehat{C}||_{tr}$ is the trace norm of $\widehat{C}$ which approximates the rank of $\widehat{C}$, and $l(\widehat{C}_{i,k} - \widehat{C}_{j,k})$ is a loss function which measures the inconsistency in inferred comparison between $\widehat{C}_{i,k}$ and $\widehat{C}_{j,k}$, and the rated comparison $(L_i \succ_{E_k} L_j)$. We adopt the approach proposed in [36] to solve Eq. (4) using the hinge loss function.

F. Defining Latent Value

After matrix completion, each expert's latent quality and quantity assignment is represented by the corresponding columns of completed latent quality ($\widehat{Q}$) and quantity ($\widehat{C}$) matrices, respectively. Fig. 11 shows a 2-dimensional scatterplot of quality and quantity (after matrix completion) of the 516 latents in our database. Given that the latent quality and quantity are correlated (Pearson correlation of 0.71 in our experiments), we conducted a grid search to find optimal weights for a linear combination of quality and quantity to define latent value ($\widehat{V}$). The simple average of quality and quantity (with equal weights), defined in Eq. (5), for latent value performed the best in terms of predicting the AFIS performance (See Section III-A for details).

$$\widehat{V} = \frac{\widehat{Q} + \widehat{C}}{2} \quad (5)$$

To understand the underlying bases of latent value assignment, it is crucial to understand the similarity or dissimilarity among latents with respect to value assignments. Let $\widehat{\nu}_i$ ($i = 1, 2, \ldots, 516$), the i^{th} row of $\widehat{V}$, denote the 30-dimensional value vector for i^{th} latent (L_i). The difference between two latents (*e.g.*, L_i and L_j) can be measured by the Euclidean distance between their corresponding value vectors $\widehat{\nu}_i$ and $\widehat{\nu}_j$. Fig. 12 represents the 516×516 latent value dissimilarity matrix (D_v) based on expert ratings, with latents sorted in descending order by the median of rows of $\widehat{V}$. We would like to explain this inter-latent variation in terms of a small number of bases, preferably two or three bases for ease in interpretation and visualization.

G. Bases for Explaining Expert Ratings

We capture the bases for the inter-latent variations provided by the 30 experts using Multidimensional Scaling[8] (MDS) [31]. MDS outputs a d-dimensional embedding of the 516 latents such that the difference between the latent value dissimilarity matrix D_v and the inter-point distance matrix based on the embedding, known as the stress value, is minimized. The dimensionality, d, of the embedding is

[8]We utilize the metric scaling implementation of MDS http://www.mathworks.com/help/stats/mdscale.html.

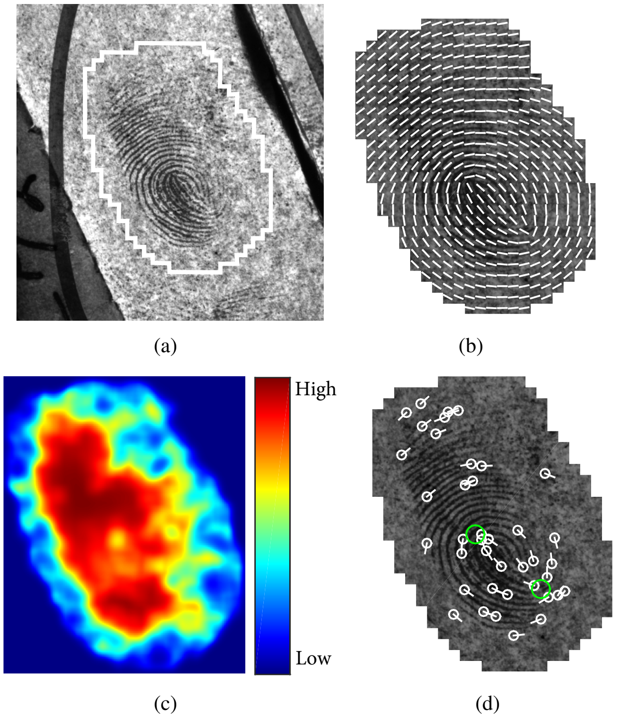

Fig. 13: Automatically extracted latent features. (a) Input latent with manually marked region of interest (ROI), (b) ridge flow overlaid on the cropped latent, and automatically extracted (c) ridge quality map and (d) minutiae (white circles) and core points (green circles).

typically determined based on the relationship between the stress value and d. It is natural that as d increases, the stress value decreases. See Table II. Given that the stress value is close to zero beyond $d = 8$, it confirms the expected low-rank characteristic of the completed value matrix ($\hat{V}$). We consider the top-8 MDS dimensions to learn the underlying bases.

TABLE II: Multidimensional Scaling (MDS) stress value vs. the number of dimensions. As the number of dimensions increase, the stress value decreases [31].

	Dimension (d)									
Stress	1	2	3	4	5	6	7	8	9	10
Value	0.26	0.13	0.08	0.05	0.03	0.02	0.01	0.002	0	0

H. Interpreting Bases in terms of Latent Features

In order to interpret MDS bases, we need to explain the bases in terms of latent features. We automatically extracted a set of $n_f = 19$ features including number of minutiae, ridge clarity, ridge flow, number of core and delta, and minutia reliability from a latent [1]. These latent features are typically used by forensic examiners as well as in designing latent AFIS [4], [5], [9]. See Table III for this list of n_f latent features and Fig. 13 for an illustration of some of these features. Let $X_i \in \mathbb{R}^{(n_f+1)}$ be the min-max normalized [42] feature vector extracted from latent L_i, including a dummy variable 1 used to absorb the bias (intercept) of the regression model [43], m be the number of latents ($m = 516$ here) and $y_k \in \mathbb{R}^m$ denote the k^{th} dimension of the MDS embedding. We explore associations between the normalized feature matrix X and individual MDS dimensions y_k using Lasso [32], which

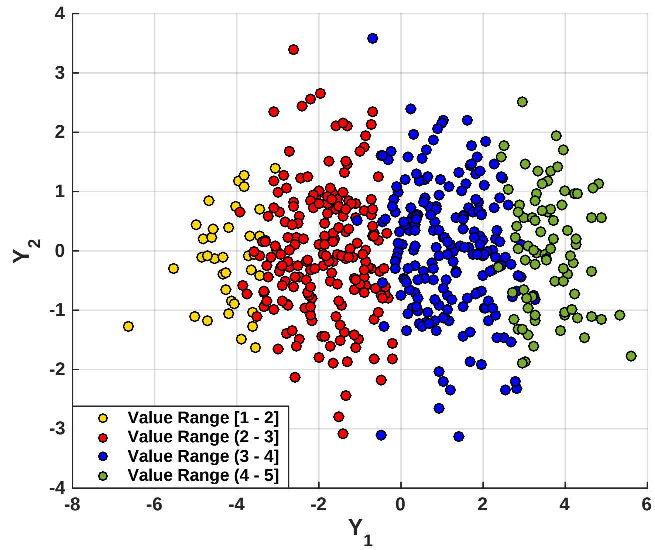

Fig. 14: 2-Dimensional embedding of 516 latents using MDS. For visualization, the latent value is thresholded into four value bins: [1 - 2], (2 - 3], (3 - 4], and (4 - 5].

TABLE III: List of 19 features automatically extracted from a latent.

Feature no.	Description
x_1	Number of minutiae in the latent
$x_2 - x_8$	Sum of reliability of minutiae that have reliability value $\geq t$, $t{=}0, 0.1, ..., 0.6$
x_9	Average area of the triangles in Delaunay triangulation of minutiae
x_{10}	Area of the convex hull of minutiae set
$x_{11} - x_{17}$	Sum of ridge quality of blocks that have quality value $\geq t$, $t{=}0, 0.1, ..., 0.6$
x_{18}	Number of core and delta
x_{19}	Standard deviation of the ridge flow (orientation map) in the latent foreground

selects a subset of features from X that are most relevant to interpret individual MDS dimensions [44]. Formally, the weight vector for the $n_f + 1$ features, β_k, for y_k is obtained by minimizing the following objective function,

$$\min_{\beta_k \in \mathbb{R}^{n_f+1}} \frac{1}{m} ||y_k - X^T \beta_k||_2^2 + \lambda ||\beta_k||_1, \quad (6)$$

where λ controls the sparseness of vector β_k. The optimal value of $\lambda = 0.001$ is learned using five-fold cross validation for parameter tuning [45]. The MDS dimension y_k can be expressed as a linear combination of $n_f + 1$ latent features x_i as follows:

$$y_k = \beta_{k1} \cdot x_1 + \beta_{k2} \cdot x_2 + ... + \beta_{k19} \cdot x_{19} + \beta_{k20} \quad (7)$$

where β_{ki} is the feature weight for x_i and $x_{20} = 1$.

We perform a five-fold cross validation to learn β_k, and observe that only two MDS dimensions in the 8-dimensional embedding receive non-zero weights for the latent features (Table IV). Rest of the MDS dimensions receive non-zero weights only for the constant term. This indicates that a two-dimensional embedding of the inter-latent similarity can adequately explain the difference among expert ratings. Fig. 14

TABLE IV: Average values of feature weights (β_k) learnt using five fold cross-validation to interpret the MDS dimensions. Only two dimensions, $y_k, (k = 1, 2)$, received non-zero weights for the 19 latent features. The corresponding standard deviations based on five fold cross validation are also reported. The 19 features $\{x_1, x_2, ..., x_{19}\}$ are defined in Table III. Non-zero weights are shown in bold; Weights lower than 0.05 are considered zero because they are relatively insignificant in interpreting the MDS dimensions.

Weight Vector	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	1
Avg. (β_1)	**0.68**	**0.28**	**0.31**	**0.22**	**0.16**	0	0	0	0	0	**0.16**	**0.19**	0.02	0	0	0	0	**0.40**	**0.19**	**0.17**
s.d. (β_1)	0.08	0.04	0.04	0.03	0.01	0	0	0	0	0	0.02	0.02	0.01	0	0	0	0	0.03	0.02	0.03
Avg. (β_2)	0	0	0	0	0	0	0	0	**0.69**	**0.26**	0	0	0	0	0	0	0	**0.13**	**0.52**	**0.43**
s.d. (β_2)	0	0	0	0	0	0	0	0	0.07	0.04	0	0	0	0	0	0	0	0.04	0.06	0.04

illustrates the two-dimensional embedding of 516 latents using MDS. The major constituents of the first MDS dimension in terms of latent features are: number of minutiae (x_1), minutia reliability (x_2, x_3, x_4, x_5), and number of core and delta (x_{18}). For the second MDS dimension, friction ridge area (x_9, x_{10}) and std. dev. of the ridge flow (x_{19}) are assigned high weights. As these two MDS dimensions are uncorrelated with Pearson correlation $\rho = 0.08$, they can be interpreted as two independent bases used by fingerprint experts in latent value assignment.

I. Learning Latent Value Predictor

Recall that the main goal of understanding the bases for expert value assignments is to utilize them to learn a predictor for query latent value. Suppose that $y_k \in \mathbb{R}^m$ $(k = 1, ..., d)$ is the k^{th} basis of MDS space and $\widehat{v} \in \mathbb{R}^m$ is the target value vector whose i^{th} element ($\widehat{v}_i$) is the median of the i^{th} row ($i = 1, ..., 516$) of completed value matrix $\widehat{V}$. The median target value is preferred over the mean value, as it is more robust against the outliers. Lasso is used to model the relationship between the MDS bases and the expert assigned latent value ($\widehat{v}$). Formally, this learning problem can be stated as:

$$\min_{\gamma \in \mathbb{R}^d} \frac{1}{m} ||\widehat{v} - Y\gamma||_2^2 + \rho||\gamma||_1, \tag{8}$$

where $Y \in \mathbb{R}^{m \times d}$ is the matrix representation of y_k $(k = 1, ..., d)$, $\gamma \in \mathbb{R}^d$ is the weight vector, $d = 2$, and ρ is the regularization parameter for the regression. Table V presents the average weights and their standard deviations, learnt by minimizing the objective function in Eq. (8) using five-fold cross validation.

Given a query latent, its normalized feature vector X_q (Table III) is extracted. The projection of X_q on the k^{th} basis is computed by $\widehat{y}_k = X_q^T \beta_k$ $(k = 1, 2)$. The value (V_{pred}) assigned to the query latent is then given by

$$V_{pred} = \widehat{Y}^T \gamma \tag{9}$$

where $\widehat{Y} = [\widehat{y}_1, \widehat{y}_2]^T$ and the value of weight vectors β and γ are obtained from Tables IV and V.

In summary, the proposed latent value prediction function can be rewritten as a weighted linear combination of latent features as follows:

$$\begin{aligned} V_{pred} = 1.22\ x_1 + 0.50\ x_2 + \ 0.55\ x_3 + 0.40\ x_4 + 0.29\ x_5 \\ + \ 1.14\ x_9 + 0.43\ x_{10} + \ 0.29\ x_{11} + 0.34\ x_{12} \\ + \ 0.92\ x_{18} + \ 1.19\ x_{19} + 1 \end{aligned} \tag{10}$$

TABLE V: Average and s.d. of the weight vector values for the two-dimensional MDS embedding shown in Fig. 14 using five fold cross-validation. These weights are used for predicting the value of latents in the test set.

	γ_1	γ_2
Avg. value of weight vector γ	1.79	1.64
Std. Dev.	0.10	0.04

Note that the coefficient for x_i is computed as $\sum_j \gamma_j \cdot w_{ij}$, if the i^{th} feature is selected for j^{th} dimension. For example, the coefficient for x_1 is computed as $(1.22 = 1.79 \times 0.68)$ and similarly for other coefficients using weights from Tables IV and V. The offline latent value predictor learning and online latent value prediction for a given query latent are summarized in Algorithms 1 and 2, respectively.

Algorithm 1 Offline Learning of Value Predictor

```
procedure FIVE-FOLD CROSS VALIDATION
input
    Q: Individual latent quality labels
    C: Pairwise comparisons of latent quantity
    X: Normalized latent features
output
    β: Weight vector for latent features
    γ: Weight vector for derived MDS bases
begin:
    Q̂ = MatrixCompletion(Q) Eq. (3)
    Ĉ = MatrixCompletion(C) Eq. (4)
    V̂ = (Q̂ + Ĉ)/2
    [y_1, y_2, ..., y_d] = MDS(V̂)
loop: (k = 1, 2, ..., d)
    β_k = Lasso(y_k, X)
end loop
    β = [β_1, β_2, ..., β_d]
    γ = Lasso(V̂, β)
end
```

The average mean square error between the predicted latent value and the crowdsourced latent value of the five folds is 0.19 with a standard deviation of 0.04. We also evaluated the role of MDS for determining the bases and their interpretation. A direct linear regression (without using MDS) between the 19 latent features and the target latent value achieves a slightly lower average MSE of 0.17 with a standard deviation of 0.04. This indicates that the interpretation of bases in terms of latent

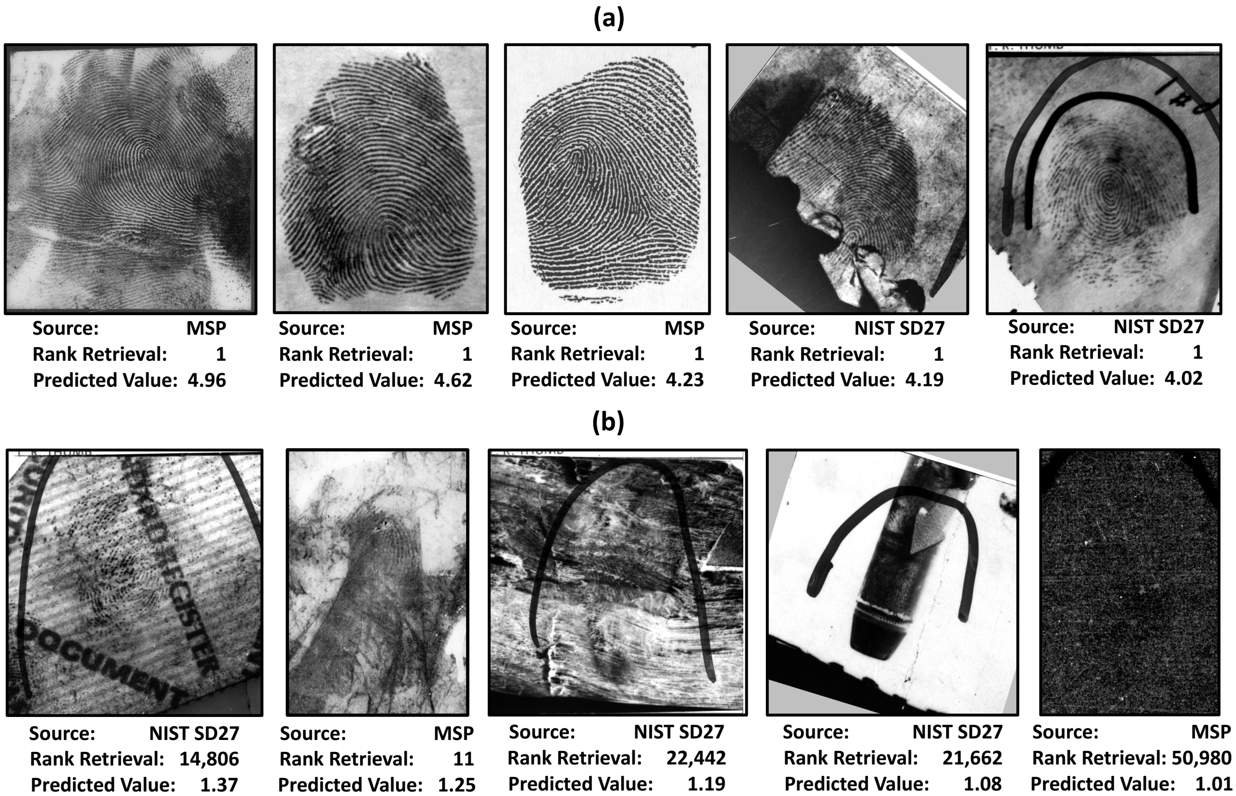

Fig. 15: The (a) top-5 and (b) bottom-5 latents from MSP and NIST SD27 databases based on their predicted value. For each latent, we also report the rank at which it is retrieved using a state-of-the-art latent AFIS (size of reference gallery = 250K rolled prints).

Algorithm 2 Online Query Latent Value Predicton

procedure
input
  β: Weight vector for latent features
  γ: Weight vector for MDS bases
  X_q: Normalized features extracted from query latent
output
  V_{pred}: Predicted latent value
begin:
loop: $(k = 1, 2, ..., d)$
  $\widehat{y}_k = X_q^T \cdot \beta_k$
end loop
  $\widehat{Y} = [\widehat{y}_1, \widehat{y}_2, ..., \widehat{y}_d]^T$
  $V_{pred} = \widehat{Y}^T \cdot \gamma$
end

features does not result in any significant loss of information with the benefit of explaining latent value assignments in terms of two bases. Fig. 15 presents the top-5 and bottom-5 ranked latents based on the predicted latent values for one of the five folds. Note that the AFIS performance depends not only on the value of the query latent, but also on the quality and quantity of information present in the mated rolled print. Fig. 16 presents examples of latents with high predicted value but poor AFIS performance, which can be attributed to poor quality and low quantity of information in their true mates.

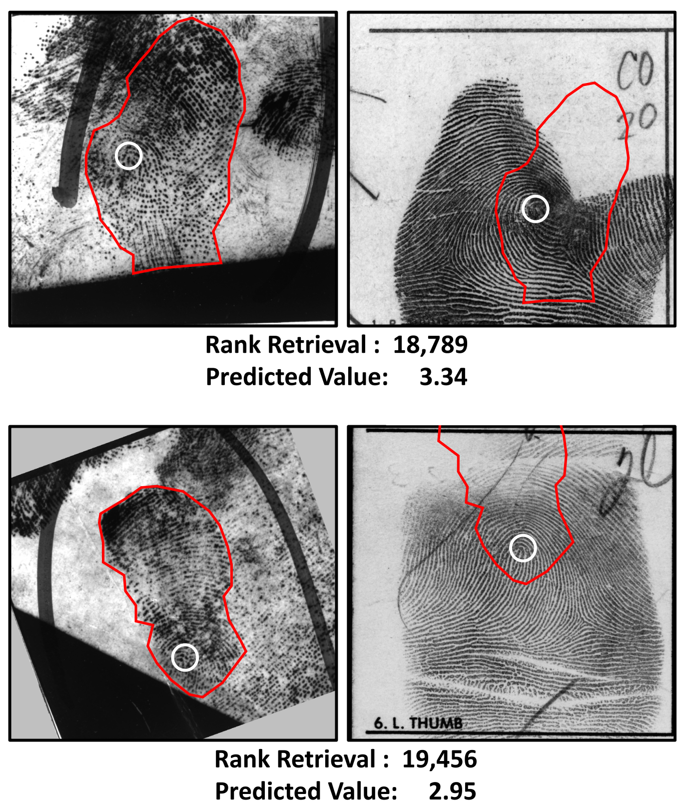

Fig. 16: Example latents from NIST SD27 with high predicted value but poor AFIS performance. The corresponding regions in the latent and its true mate are marked in red; core points are marked in white.

III. Experiments on Latent Value Prediction

We perform two experiments, in order to evaluate (i) the efficacy of latent value from expert crowd as the target value, and (ii) the performance of the proposed latent value predictor on three independent latent databases (MSP400, WVU and IIITD MOLF).

A. Evaluating Target Latent Value from Expert Crowd

We first compare the latent value from expert crowd ($\widehat{V}$) with Latent Fingerprint Image Quality (LFIQ) [11] on NIST SD27 and MSP258 latent databases. Both the latent value from expert crowd and LFIQ scores are continuous values, where the former is bounded in the range of [1 - 5] and the latter are unbounded non-negative values. Following the comparison protocol from [11], the 516 latents are first sorted based on both the median latent value from expert crowd, as well as the LFIQ scores. The sorted latents are then binned into 100 overlapping bins (value index) such that each bin contains 100 latents, using a sliding window for binning with a fixed step size of 4.16. For each of the bins, we compare the number of latents identified at Rank-1 against a reference database of 250,000 rolled prints, including the true mates of 516 latents, by a state-of-the-art latent AFIS[9]. Fig. 17 presents the Rank-1 identification rate of latents with respect to each of the 100 value indices. For each of the value index bins, the percentage of latents whose mates are retrieved at rank-1 is used as an evaluation criterion. A high identification rate is expected for latents with high value index, and low identification rate for latents with low value index, as illustrated by the "ideal" curve (Fig. 17). The ideal curve represents the Rank-1 identification rates across each of the value indices based on the rank of the true mates on the candidate list, *i.e.*, if the latents are sorted in the descending order by the rank retrieval of their true mates. The latent value by expert crowd is performing better than LFIQ in terms of predicting the AFIS performance, as the fused value has a higher correlation (0.98) with ideal curve compared to LFIQ (0.92).

Next we compare the latent value from expert crowd ($\widehat{V}$) with value determination by examiners (VID, VEO, and NV) [8]. As value determinations by examiners are available only for NIST SD27 database, we do not use MSP258 database for this comparison. The numbers of VID, VEO, and NV latents in NIST SD27, reported in [8], are 210, 41, and 7, respectively. Out of 258 latents from NIST SD27, 166 latents are retrieved at Rank-1 using the state-of-the-art latent AFIS. Since the latent value from expert crowd is a continuous value in the range $[1-5]$ and the value determination by examiners is categorical (VID, VEO and NV), we first sort the 258 latents in descending order of the median inferred latent value ($\widehat{v}$). The first 210 latents are considered as VID, the next 41 as VEO and the last 7 as NV latents. This protocol is adopted for a fair comparison between the latent value from expert crowd and the value determination by examiners. Table VI compares the number of latents retrieved at rank-1 using value determination by examiners [8] and the inferred latent value from expert crowd ($\widehat{V}$), with the same exemplar gallery of 250,000 rolled prints. Value determination by expert crowd performs better than value determination by examiners in terms of predicting the AFIS performance. A larger number of VID latents (161) are identified at Rank-1 based on $\widehat{V}$ as compared to the number of VID latents (155) identified at Rank-1 based on value determination by examiners.

[9] One of the top three performing latent AFIS in NIST ELFT evaluation [7]

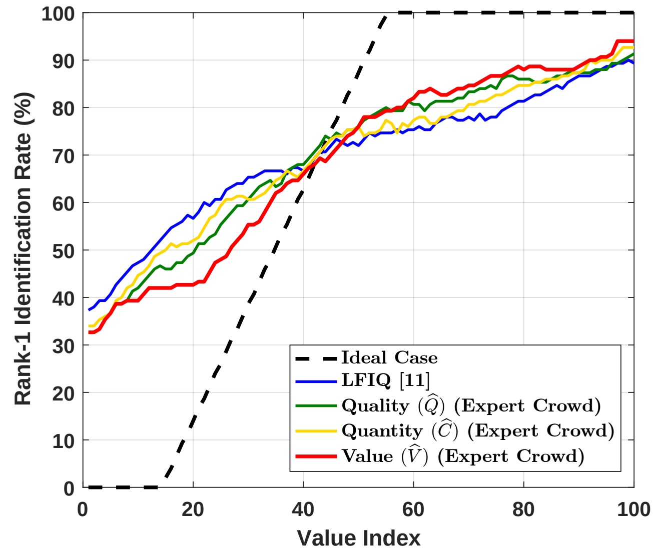

Fig. 17: Rank-1 Identification Rate of a state-of-the-art latent AFIS with respect to different value indices: (i) ideal case where the latents are sorted in the descending order by the rank retrieval of their true mates, (ii) Latent Fingerprint Image Quality (LFIQ) [11], (iii) Quality labels ($\widehat{Q}$) from the expert crowd, (iv) Quantity labels ($\widehat{C}$) from the expert crowd, and (v) Latent Value ($\widehat{V}$) from Expert Crowd (Eq. 5).

Another possible source of target value can be value determination by AFIS [1], but we do not utilize this approach for the following two reasons: (i) the latent value predictor should be independent of AFIS, and (ii) the latent value determination should only be based on query latent and not its mate in the reference print database.

TABLE VI: Number of latents retrieved at Rank-1, using a state-of-the-art latent AFIS, for NIST SD27 latents that were determined as VID (210), VEO (41) and NV (7) by examiners [8]. For value determination by the expert crowd, the median inferred latent value ($\widehat{v}$) is thresholded as VID > 2.16 and NV < 1.66. The range for the median latent value is $[1-5]$.

	VID	VEO	NV
Value Determination by Examiners [8]	155/210	11/41	0/7
Value Determination by Expert Crowd ($\widehat{V}$)	161/210	5/41	0/7

B. Correlation between Predicted Latent Value and AFIS Performance

The success in retrieving the true mate for a query latent not only depends on the latent value but also on the quality of the mate (rolled or slap print). To conduct a controlled experiment, we selected an independent set of 400 latents (not used in crowdsourcing) from MSP latent database whose corresponding rolled mates have the highest NFIQ quality value (i.e. NFIQ = 1). This ensures that the AFIS search performance is primarily affected by the latent value. In a

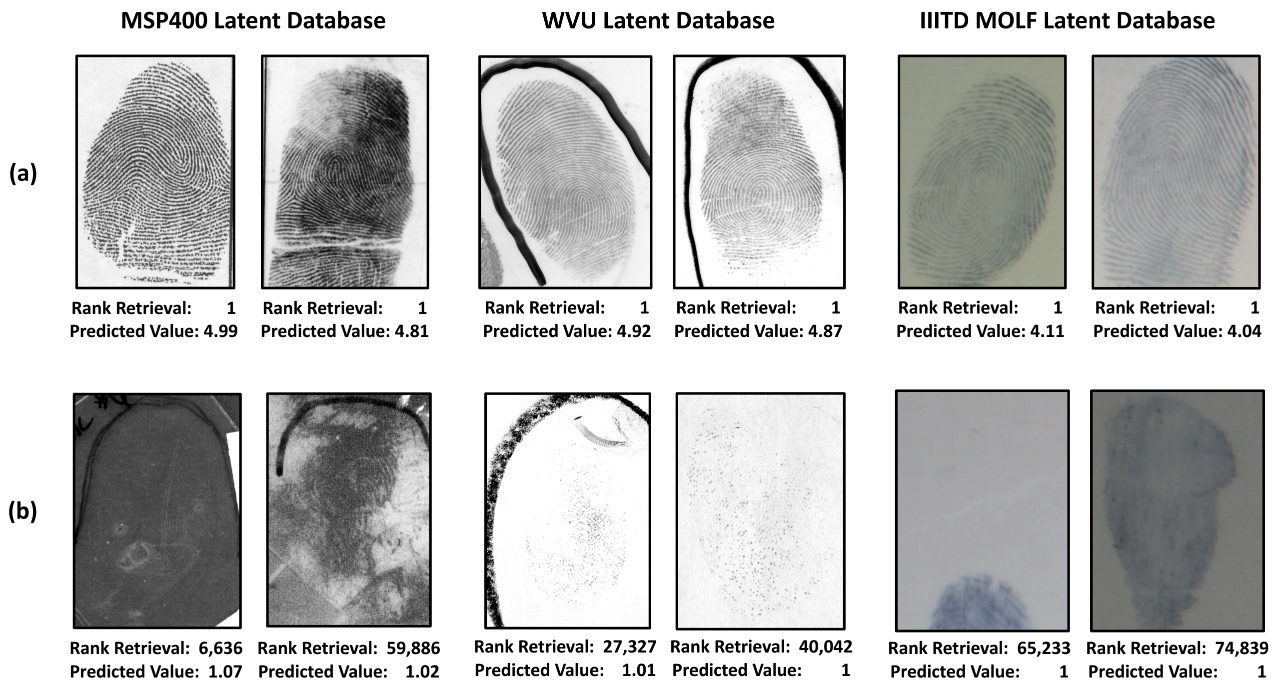

Fig. 18: The (a) top-2 and (b) bottom-2 latents based on their predicted values from (i) an independent set of 400 latents from MSP, (ii) WVU Latent Database, and (iii) IIITD MOLF Latent Database. For each latent, we also report the rank at which it is retrieved using a state-of-the-art latent AFIS (size of reference gallery = 250K rolled prints).

similar fashion, latents with high quality mates (NFIQ = 1) are identified from the WVU latent database [46] containing 449 latents and IIITD MOLF latent database [47] containing $4,400$ latents. The resultant subsets have 139 and $2,926$ latents, respectively, from WVU and IIIT MOLF latent databases. Unlike MSP latent database which was collected during crime scene investigations, WVU and IIITD MOLF latent databases were collected in a laboratory setting. With a background gallery of $250,000$ rolled prints, AFIS comparison scores (with the true mate), LFIQ value [11] and value from the proposed predictor are computed for these three latent databases. Table VII presents the correlation between the predicted latent value by the proposed model and the state-of-the-art AFIS comparison scores. A comparison with the correlation value between LFIQ and AFIS comparison scores suggests that the proposed latent value predictor is better than LFIQ in predicting the AFIS performance. Fig. 18 presents the top-2 and bottom-2 ranked latents based on the latent value from the proposed predictor for each of the three databases.

TABLE VII: Correlation of predicted latent value by LFIQ, and by the proposed method with the performance of a state-of-the-art AFIS for three independent latent databases.

	LFIQ [11]	Proposed Method
MSP Latent Database (400 latents)	0.49	0.70
WVU Latent Database Subset [46] (139 latents)	0.47	0.73
IIITD MOLF Latent Database Subset [47] (2926 latents)	0.42	0.67

IV. Conclusions

Latent value assignment is a crucial step in the widely used ACE-V methodology practiced by examiners in latent fingerprint processing. However, the prevailing approaches for latent value assignment have the following two main limitations: (i) latent values (VID, VEO and NV) assigned by examiners are subjective with low reproducibility, and (ii) directly modeling the relationship between latent features and value determination does not explain inter-examiner variations.

We have proposed a fully automated method to assign quantitative values to latent fingerprints. The main contributions of our approach are: (i) designed and implemented a crowdsourcing-based framework to collect and understand expert latent value assignment from the perspectives of latent quality and quantity (information content), (ii) utilized Multidimensional Scaling (MDS) to identify the underlying bases for expert latent value assignment, (iii) established the relationship between automatically extracted latent features and MDS bases using Lasso, and (iv) learned a predictor based on the underlying bases to assign value to a query latent.

Our experiments involved latent fingerprints from two forensic databases (NIST SD27 and MSP) and two laboratory collected databases (WVU and IIITD MOLF), and 31 fingerprint experts for crowdsourcing. The main conclusions of our study are: (i) the crowdsourced latent value is more robust than prevailing value determination (VID, VEO and NV) in terms of predicting AFIS performance, and (ii) two MDS bases are adequate to explain expert value assignments and can be interpreted in terms of our automatically extracted latent features, and (iii) a collection of latents can be ranked in terms of the predicted value from most informative to least informative.

Our suggestions for future work on latent value prediction include (i) extracting more robust latent features, especially minutiae points, for latent value prediction, and (ii) improving the current prediction model by incorporating feature rarity.

ACKNOWLEDGEMENT

We would like to thank the Michigan State Police (MSP) for sharing the latent fingerprint database and all the fingerprint experts who participated in data collection. This research is supported by the NIST Forensic Research Program.

REFERENCES

[1] K. Cao, T. Chugh, J. Zhou, E. Tabassi, and A. K. Jain, "Automatic Latent Value Determination," in *9th IAPR International Conference on Biometrics (ICB)*, 2016, pp. 1–6.
[2] President's Council of Advisors on Science and Technology, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods*. Executive Office of The President's Council of Advisors on Science and Technology, Washington DC, 2016. [Online]. Available: https://www.whitehouse.gov/administration/eop/ostp/pcast/docsreports
[3] National Research Council, *Strengthening Forensic Science in the United States: A Path Forward*. Statement before the United State Senate Committee on the Judiciary, 2009.
[4] A. K. Jain and J. Feng, "Latent Fingerprint Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, pp. 88–100, 2011.
[5] A. Sankaran, M. Vatsa, and R. Singh, "Latent Fingerprint Matching: A Survey," *IEEE Access*, vol. 2, pp. 982–1004, 2014.
[6] M. Indovina, A. Hicklin, and G. Kiebuzinski, "NIST Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #1]," 2010.
[7] M. Indovina, V. Dvornychenko, R. Hicklin, and G. Kiebuzinski, "ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #2]," *National Institute of Standards and Technology, NISTIR*, vol. 7859, 2012.
[8] R. A. Hicklin, J. Buscaglia, M. A. Roberts, S. B. Meagher, W. Fellner, M. J. Burge, M. Monaco, D. Vera, L. R. Pantzer, C. C. Yeung, and T. N. Unnikumaran, "Latent Fingerprint Quality: A Survey of Examiners," *Journal of Forensic Identification*, vol. 61, no. 4, pp. 385–419, 2011.
[9] D. R. Ashbaugh, *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC press, 1999.
[10] "Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST): Standard for the Documentation of Analysis, Comparison, Evaluation and Verification (ACE-V) (Latent) version 2.0 (2012)."
[11] S. Yoon, K. Cao, E. Liu, and A. K. Jain, "LFIQ: Latent Fingerprint Image Quality," in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.
[12] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts, "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners," *PloS One*, vol. 7, no. 3, p. e32800, 2012.
[13] ——, "Accuracy and reliability of forensic latent fingerprint decisions," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 108, no. 19, pp. 7733–7738, 2011.
[14] "Fact sheet - Integrated Automated Fingerprint Identification System (IAFIS)," https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics/ngi/.
[15] S. S. Arora, K. Cao, A. K. Jain, and G. Michaud, "Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups," in *8th IAPR International Conference on Biometrics (ICB)*, 2015, pp. 363–370.
[16] "Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST): Standards for Examining Friction Ridge Impressions and Resulting Conclusions version 1.0 (2011)."
[17] C. Champod, C. J. Lennard, P. Margot, and M. Stoilovic, *Fingerprints and Other Ridge Skin Impressions*. CRC press, 2004.
[18] B. T. Ulery, R. A. Hicklin, G. I. Kiebuzinski, M. A. Roberts, and J. Buscaglia, "Understanding the Sufficiency of Information for Latent Fingerprint Value Determinations," *Forensic Science International*, no. 1, pp. 99–106, 2013.
[19] B. Keelan, *Handbook of Image Quality: Characterization and Prediction*. CRC Press, 2002.
[20] E. Tabassi, C. Wilson, and C. Watson, "NIST Fingerprint Image Quality," *NIST Res. Rep. NISTIR7151*, pp. 34–36, 2004.
[21] P. R. Halmos, *Finite-Dimensional Vector Spaces*. Springer, 1974.
[22] B. T. Ulery, R. A. Hicklin, M. A. Roberts, and J. Buscaglia, "Interexaminer Variation of Minutia Markup on Latent Fingerprints," *Forensic Science International*, vol. 264, pp. 89–99, 2016.
[23] S. Yoon, E. Liu, and A. K. Jain, "On Latent Fingerprint Image Quality," in *Computational Forensics*. Springer, 2015, pp. 67–82.
[24] A. Sankaran, M. Vatsa, and R. Singh, "Automated Clarity and Quality Assessment for Latent Fingerprints," in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–6.
[25] J. Kotzerke, S. Davis, R. Hayes, L. Spreeuwers, R. Veldhuis, and K. Horadam, "Identification Performance of Evidential Value Estimation for Fingermarks," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2015, pp. 1–6.
[26] M. Olsen, M. Bockeler, and C. Busch, "Predicting Dactyloscopic Examiner Fingerprint Image Quality Assessments," in *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2015, pp. 1–12.
[27] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2005.
[28] O. Dekel and O. Shamir, "Vox Populi: Collecting High-Quality Labels from a Crowd." in *Conference On Learning Theory (COLT)*, 2009.
[29] L. V. Ahn and L. Dabbish, "Labeling Images with a Computer Game," in *Proceedings of the ACM SIGCHI conference on Human factors in computing systems*, 2004, pp. 319–326.
[30] D. Martinho-Corbishley, M. Nixon, and J. Carter, "Analysing Comparative Soft Biometrics from Crowdsourced Annotations," *IET Biometrics*, pp. 1–16, 2016.
[31] J. B. Kruskal and M. Wish, *Multidimensional Scaling*. Sage, 1978, vol. 11.
[32] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, pp. 267–288, 1996.
[33] "NIST SD27 Latent Fingerprint Database," http://www.nist.gov/srd/nistsd27.cfm.
[34] A. Kittur, E. H. Chi, and B. Suh, "Crowdsourcing User Studies with Mechanical Turk," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2008, pp. 453–456.
[35] N. N. Liu, M. Zhao, and Q. Yang, "Probabilistic Latent Preference Analysis for Collaborative Filtering," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM)*, 2009, pp. 759–766.
[36] J. Yi, R. Jin, S. Jain, and A. Jain, "Inferring Users' Preferences from Crowdsourced Pairwise Comparisons: A Matrix Completion Approach," in *First AAAI Conference on Human Computation and Crowdsourcing (HCOMP)*, 2013.
[37] P. Smyth, U. Fayyad, M. Burl, P. Perona, and P. Baldi, "Inferring Ground Truth from Subjective Labelling of Venus Images," *Advances in Neural Information Processing Systems 7 (NIPS)*, pp. 1085–1092, 1995.
[38] F. K. Khattak and A. Salleb-Aouissi, "Quality Control of Crowd Labeling through Expert Evaluation," in *Proceedings of the NIPS 2nd Workshop on Computational Social Science and the Wisdom of Crowds*, 2011.
[39] E. J. Candès and B. Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
[40] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization," *Society for Industrial and Applied Mathematics (SIAM) review*, vol. 52, no. 3, pp. 471–501, 2010.
[41] Z. Wen, W. Yin, and Y. Zhang, "Solving a Low-Rank Factorization Model for Matrix Completion by a Nonlinear Successive Over-Relaxation Algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.
[42] A. Jain, K. Nandakumar, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.
[43] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, 2006.
[44] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, pp. 1705–1732, 2009.
[45] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.
[46] "West Virginia University (WVU) Latent Fingerprint Database," http://www.cse.msu.edu/~rossarun/.
[47] A. Sankaran, M. Vatsa, and R. Singh, "Multisensor Optical and Latent Fingerprint Database," *IEEE Access*, vol. 3, pp. 653–665, 2015.