

# 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**—Value of a latent fingerprint, in forensics community, is determined based on its quality (e.g. ridge clarity) as well as its information content (e.g. no. of minutiae). The prevailing practice is to have a fingerprint examiner assign one of the following three labels to the latent: VID (Value for Individualization), VEO (Value for Exclusion Only) or NV (No Value) prior to comparison. However, examiner value assignment is subjective, with reported issues of inconsistency and irreproducibility in their decisions. We present a predictor for objective and quantitative value assignment to latent fingerprints which is learned based on value assignments by a crowd of fingerprint experts. The contributions of this paper are as follows: (i) development of a crowdsourcing tool, *FingerprintMash*, to collect quality ratings for individual latents, and pairwise comparisons of information content, (ii) use of multidimensional scaling (MDS) to determine the underlying bases (dimensions) of examiner value assignment, (iii) interpretation of the bases using Lasso regression analysis in terms of automatically extracted latent features, and (iv) learning a predictor to automatically assign quantitative value to a query latent. Experiments on two operational law enforcement latent databases (NIST SD27 and MSP) show that only two bases, obtained via MDS, are adequate to explain examiner value assignments. Further, these two bases are found to be correlated with the number of minutiae, the number of singular points (core and delta), ridge clarity, and the friction ridge area extracted from latents. The proposed latent value predictor can be used to triage a collection of latents at a crime scene and is shown to be highly correlated to AFIS performance.

**Index Terms**—Crowdsourcing, latent value determination, latent examiners, matrix completion, multidimensional scaling.

## I. INTRODUCTION

LATENT fingerprint value determination is one of the most important steps in the widely practiced Analysis, Comparison, Evaluation and Verification (ACE-V) methodology [2] for comparison and identification of latent fingerprints (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 database to identify the source of the latent; NV latents are simply documented in case files to save examiner’s effort

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<sup>th</sup> IAPR International Conference on Biometrics (ICB), 2016 [1]

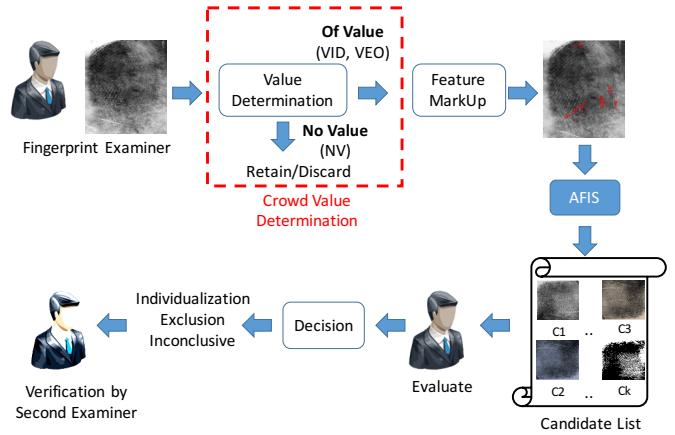


Fig. 1: ACE-V methodology. The proposed latent fingerprint value prediction based on expert crowdsourcing is shown in red block.

in feature markup and comparison [4]. Since state-of-the-art Automated Fingerprint Identification System (AFIS) can process most VID latents<sup>1</sup> in a fully automatic or “lights out” mode with no or little examiner effort, latent value determination is a crucial step. 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”).

Value determination by examiners is presumably based on factors which include both image quality (e.g. friction ridge clarity) and information content or quantity (e.g. number of minutiae) in a latent image [4]. Because latent value determination involves subjective qualitative and quantitative assessments, large intra-examiner and inter-examiner variations in value determination are often observed [7], [8]. Furthermore, with the growing caseload faced by forensic agencies, there is an urgent need to develop a method for automatic, accurate, and objective value assignment for latents. As an example of the latent caseload, it has been reported that during the month of December 2015 alone, FBI’s Integrated Automated Fingerprint Identification Systems (IAFIS) conducted 14,311 latent feature searches, where markups were done by latent examiners, and 2,370 latent image searches (without any manual markups) [9]. Automatic latent value assignment can

<sup>1</sup>Rank-1 identification rate of 85% is obtained for latents marked as VID in a dataset comprising of NIST SD27 and WVU [5].

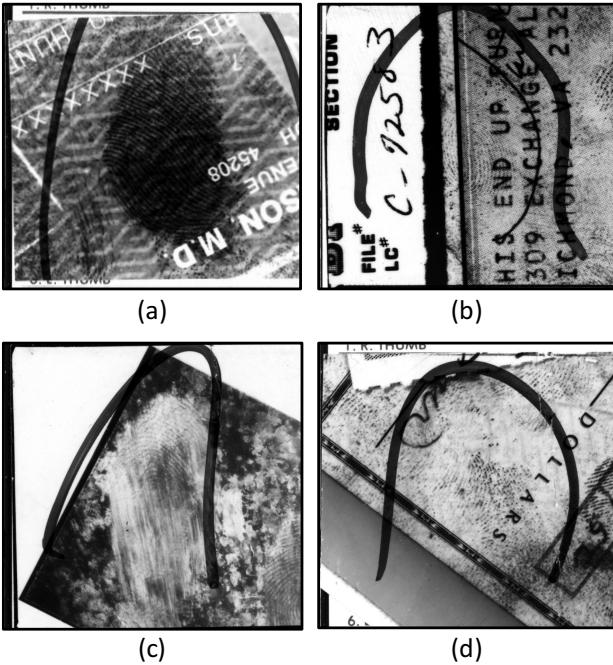


Fig. 2: Latents from NIST SD27 with different values. (a), (b), (c) and (d) were, respectively, determined to be VID, VEO, NV and NV by examiners in [3]. The corresponding predicted values assigned by the proposed approach are 4.33, 2.26, 1.90, and 1.21 respectively, where 1 indicates the lowest value and 5 indicates the highest value.

also be used to rank latents in terms of their quantitative value, thereby allowing for an efficient use of examiners' time and effort.

There is a succinct distinction between the quality of a latent and its value. Image quality in general [10] is defined as the perceived image degradation (typically, compared to an ideal or perfect image) 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) [11] defines fingerprint image quality as a predictor of AFIS performance. Value of a fingerprint image, on the other hand, includes more than just the quality attribute; it also incorporates the quantity, or information content, such as the number of minutiae, the number of singular points, the friction ridge area, etc. 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, a large number of minutiae. However, this is seldom true for latent prints. For this reason, quality and value are defined differently for latent prints with value encompassing both quality of friction ridge features and their information content [4]. Fig. 3 illustrates this difference between quality and information content for latents based on authors' observations.

In spite of the guidelines available for latent value determination [4], the underlying bases or dimensions that latent examiners use for the value determination are not known. Consequently, the prevailing practice of manual value de-

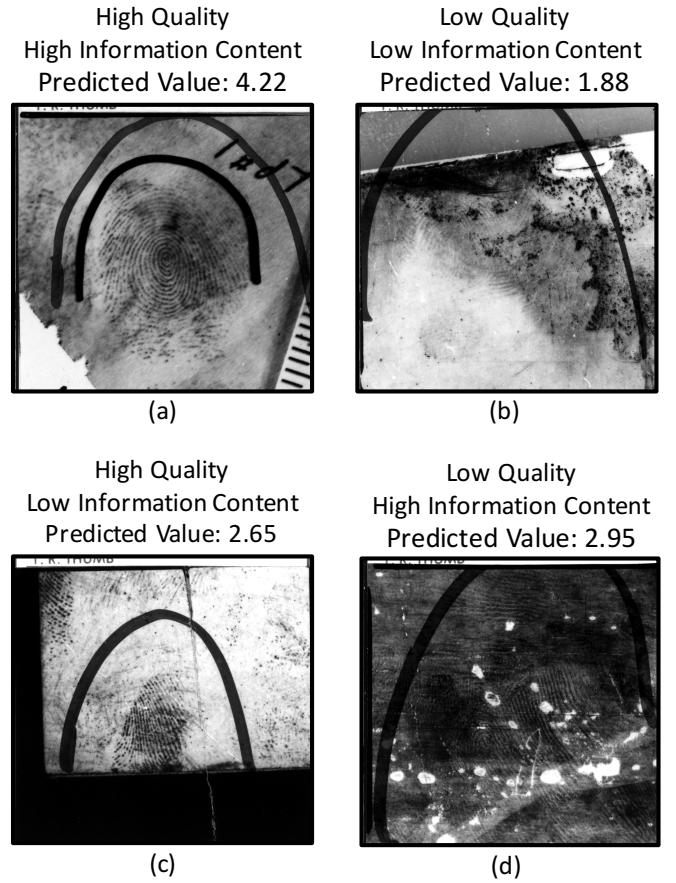


Fig. 3: Difference between latent quality and latent value. (a) High quality, high information content latent, (b) low quality, low information content latent, (c) high quality, low information content latent, and (d) low quality, high information content latent. The predicted value using the proposed 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.

termination is subjective and depends upon examiner's skill and experience [7]. Ulery et al. [12] modeled the relationship between value determination and feature annotation by certified latent examiners and identified two major limitations: (i) low reproducibility of value determination, and (ii) low reproducibility of feature markup. In another study on the consensus of latent examiners' value determination [7], [8], 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%. One limitation of Ulery et al.'s models is that it is not able to identify the actual bases that the examiners adopt for value determination. These bases are needed to learn a predictor for quantitative value determination. The examiner subjectivity is also manifested in minutiae markup [6], [13]. Fig. 4 from [6] shows the variability in minutiae markup by six different examiners of the same latent from NIST SD27.

Instead of modeling the relationship between value determination and feature annotation, Yoon et al. [14] defined a Latent Fingerprint Image Quality (LFIQ) measure in terms of the average ridge clarity, and the total number of minutiae. This measure was further extended in [5] by incorporating

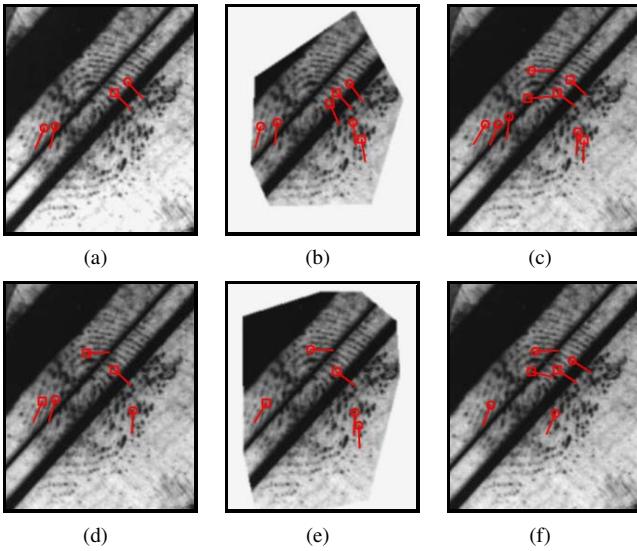


Fig. 4: Variability in minutiae markup of the same latent from NIST SD27 by six different examiners. Some examiners also marked Region Of Interest (ROI) in addition to the minutiae. Image from [6].

additional features, namely connectivity of ridge structures, minutiae reliability, and finger position. A major limitation of [14] and [5] is that they require manually annotated minutiae. In order to alleviate the dependence on manual feature markup, Cao et al. [1] developed methods for automatic latent feature extraction, including minutiae, ridge clarity, core and delta, and ridge flow. However, this approach did not identify the underlying bases that examiners use for assigning latent value. Studies in [15], [16], [17] are of limited value because they either used tenprints instead of latents or did not use a state-of-the-art latent AFIS in their experiments. We believe that linking the examiner value assignments and features extracted from the latent print, is necessary to develop an automatic predictor for quantitative value assignment.

Another limitation shared by all previous studies is that the target latent value used during training itself may not be reliable because it is established by either a single examiner or only a few examiners. Our use of crowdsourcing overcomes this limitation. Crowdsourcing<sup>2</sup> is the process of soliciting inputs from a group of workers, usually online; the collection of all the inputs leads to better decision making [18]. Advantages of crowdsourcing have been shown in a number of application domains, including image tagging [19], audio [20] and 3D video [21] quality annotations. Martinho et al. in [22] utilized crowdsourced soft biometric annotations to identify pedestrians from video surveillance footage. In latent fingerprint domain, Arora et al. [6] showed that utilizing the crowd expertise of the examiners for feature markup significantly boosts the performance of a state-of-the-art latent AFIS.

In this paper, we propose a crowdsourcing based framework for understanding the underlying bases of value assignment by latent examiners. We illustrate this framework in Fig. 5. The proposed framework is used to learn a predictor for value assignment. A crowdsourcing tool, called *FingerprintMash*,

TABLE I: Notations

Notation	Definition
$m$	Number of latent images
$n$	Number of fingerprint experts
$L_i$	$i^{th}$ latent, $i = 1, 2, \dots, m$
$E_k$	$k^{th}$ expert, $i = 1, 2, \dots, n$
$Q \in \mathbb{R}^{m \times n}$	Sparse quality matrix collected from expert crowd
$\hat{Q} \in \mathbb{R}^{m \times n}$	Complete quality matrix after matrix completion
$C$	Set of triplets that encode all the pairwise comparisons
$\hat{C} \in \mathbb{R}^{m \times n}$	Complete information content matrix after matrix completion
$L_j \succ_{E_k} L_i$	Latent $L_j$ is deemed to contain more information over latent $L_i$ by expert $E_k$
$L_j \simeq_{E_k} L_i$	Latent $L_j$ is deemed to have similar information content as $L_i$ by expert $E_k$
$D_q^e$	Dissimilarity matrix between 31 experts based on inferred quality matrix
$D_c^e$	Dissimilarity matrix between 31 experts based on inferred information content matrix
$D_v$	Dissimilarity matrix between 516 latents based on inferred latent value matrix
$X \in \mathbb{R}^{20}$	Normalized latent feature vector consisting of 19 features and 1 constant term
$Y_i \in \mathbb{R}^m$	$i^{th}$ dimension obtained via MDS
$V \in \mathbb{R}^m$	Crowdsourced latent value

is designed to collect inputs from an expert crowd<sup>3</sup>. The input consists of numerical quality ratings for individual latents and pairwise comparison of information content for latent pairs. In order to understand the underlying bases used by latent examiners for latent value assignment, Multidimensional Scaling (MDS) [23], a well known ordination and visualization tool, is used 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 [24], resulting in a predictor for quantitative latent value assignment.

The contributions of this paper are as follows:

- 1) Introduced a crowdsourcing based framework for analyzing the bases of latent value assignment by experts using matrix completion, MDS and Lasso tools.
- 2) Designed a crowdsourcing tool, *FingerprintMash*, to collect numerical quality ratings from fingerprint experts for individual latents and pairwise comparison of information content for a pair of latents.
- 3) Identified the underlying bases that latent experts use for qualitative value assignment via MDS.
- 4) Learned a prediction function to automatically assign quantitative values to query latents.

<sup>3</sup>The expert crowd comprises of 13 latent fingerprint examiners and 18 researchers working in the field of latent fingerprints

<sup>2</sup><https://en.wikipedia.org/wiki/Crowdsourcing>

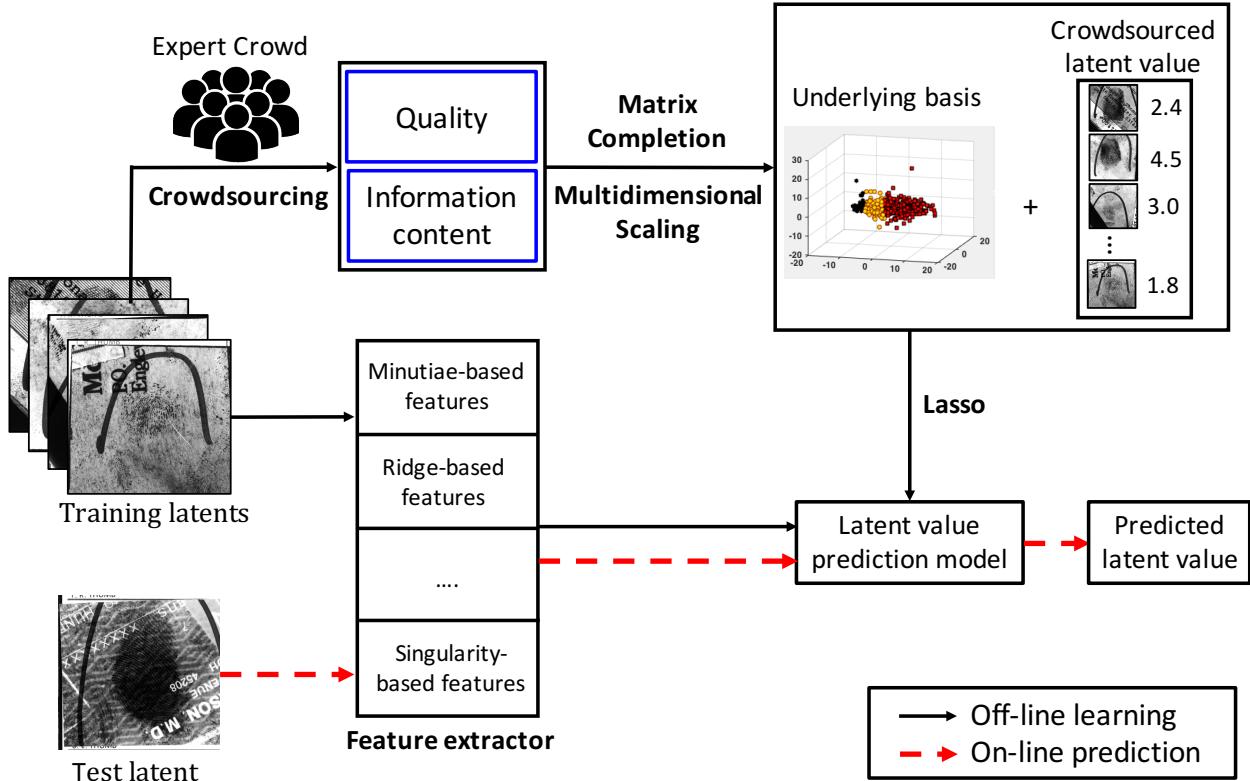


Fig. 5: Overview of the proposed crowdsourcing-based learning approach for latent value assignment.

## II. PROPOSED FRAMEWORK

The proposed framework involves crowdsourcing of latent value assignments via *FingerprintMash*<sup>4</sup>, inferring missing latent values of unseen latents 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, as illustrated in Fig. 5. For clarity, the notations used in the paper are summarized in Table I.

### A. Data Collection

As stated earlier, the value of a latent is presumably based on its image quality and its information content. We adopt a crowdsourcing based framework in which we requested a pool of 31 fingerprint experts<sup>5</sup> to assign values to a set of randomly selected 100 latent pairs from our database of 516 latents. Crowdsourcing has been commonly used to collect numerical ratings (absolute values) or pairwise comparisons (relative preferences) [25]. In our study, we requested the expert crowd to provide both, numerical ratings for quality and pairwise comparisons for information content. To be consistent with NFIQ [11], a five-level quality rating is used for each latent. For information content, the pairwise comparison approach is generally more effective than absolute rating [26] because two workers who agree on the relative preferences, may assign them different numerical ratings. Further, it has been shown that reproducibility of pairwise comparisons is higher [25],

[27]. Fig. 6 presents the interface of *FingerprintMash*, designed for the purpose of collecting both numerical quality ratings and pairwise comparisons for information content.

The 516 latent print images used for data collection are obtained from two operational latent databases: 258 latents each from NIST SD27 [28] and Michigan State Police (MSP) latent database<sup>6</sup> (Fig. 7). All 516 latents are scanned at 500 ppi and have similar image characteristics. We also have their corresponding rolled mates.

For data collection, each participant is presented a total of 100 pairs of randomly selected latent fingerprints for value assignment, one pair at a time. For each pair, the participant indicates:

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

The interface in Fig. 6 allows the workers to skip to the next pair by clicking on the button “Skip to Next” or undo the previous comparison by clicking on the button “Undo Previous”. To validate the reliability of the responses, every 5<sup>th</sup> 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 switched to avoid any bias in worker response.

<sup>4</sup><http://www.fingerprintmash.org>

<sup>5</sup>We sent e-mails to a large collection of experts; 31 experts responded.

<sup>6</sup>MSP latent database is not a publicly available database.

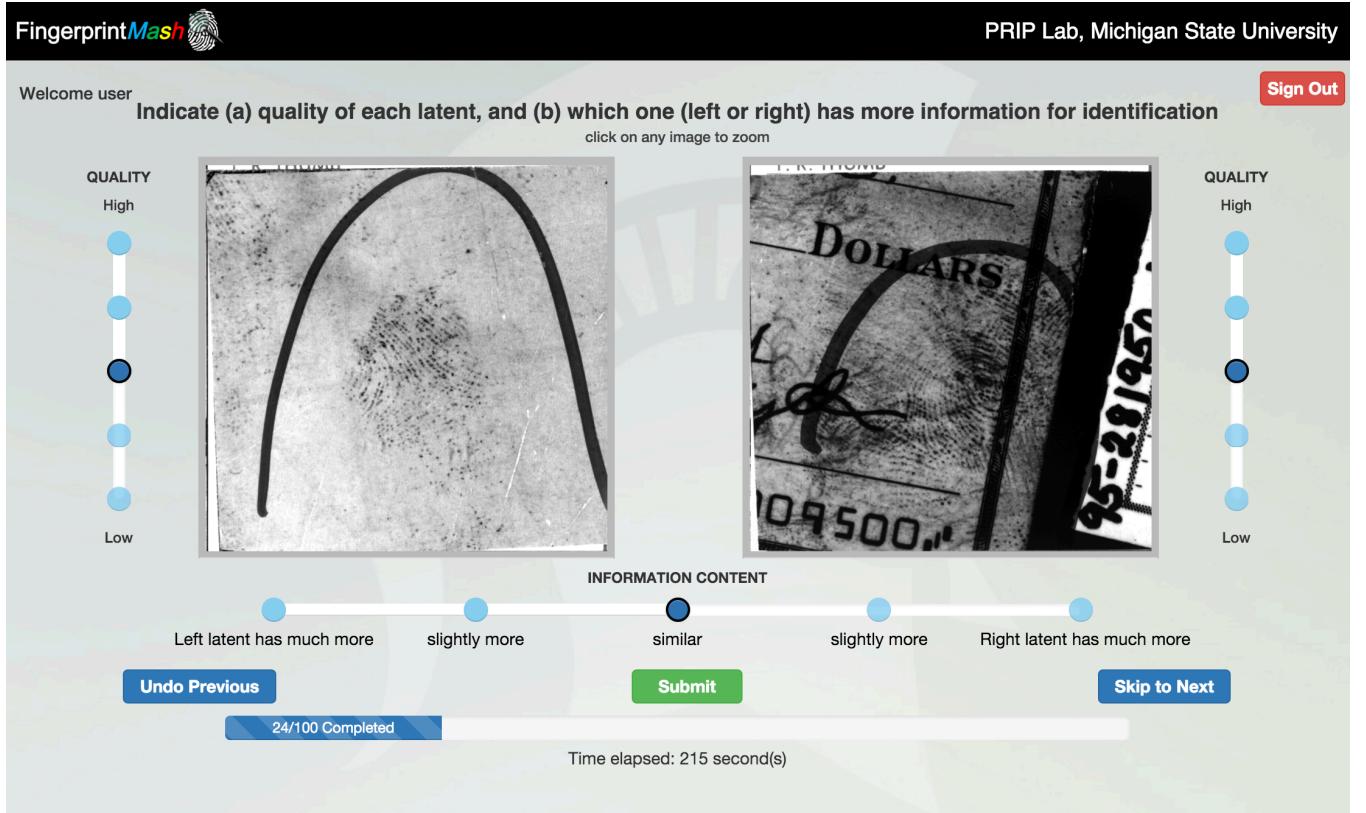


Fig. 6: Interface of the in-house developed crowdsourcing tool, called *FingerprintMash*. Each crowd worker assigns a numerical quality rating (low to high) to each latent and provides a pairwise comparison value in terms of information content for the latent pair displayed on the screen. A total of 100 latent pairs, chosen randomly from our database of 516 latents, are presented to each worker.

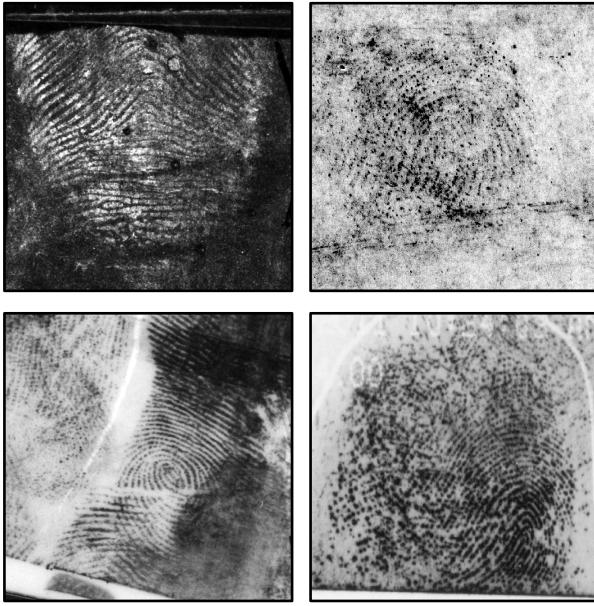


Fig. 7: Examples from MSP database consisting of 258 latents.

It is not feasible to present every pair of latents to all the experts, so the "missing" data needs to be inferred using matrix completion methods. It has been shown that only  $O(r \log m)$  pairwise comparisons per expert are required to accurately

recover the ranking list of  $m$  items (latents), where  $r \ll m$  is the rank of the rating matrix [26]. For  $m = 516$  latents and an estimated rank of  $r < 10$ , 100 comparisons are sufficient to generate the ranking list. With 31 experts in our study, we obtained 6,200 numerical quality ratings and 3,100 pairwise comparisons of information content.

#### B. Expert Reliability

Identifying and removing crowd workers that are not diligent in the assigned task is important [29] [30]. As mentioned in section II-A, a validation set of 20 repeated latent pairs is used to measure each expert's reliability. In case of latent image quality, repetition of 20 latent pairs imply we have 40 repeated latent images to compute each expert's reliability for quality ratings. Let  $\{NR_{ij}^1\}_{j=1}^{40}$  and  $\{NR_{ij}^2\}_{j=1}^{40}$  denote the latent values for the 40 repeated latents rated 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 values. The intra-expert variation ( $NV_i$ ) of the  $i^{th}$  expert is then defined as

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

Fig. 8 presents the intra-expert variations of numerical quality ratings for the 31 experts in our crowdsourcing task. Three experts (8, 13 and 28) show significantly higher variations compared to other experts.

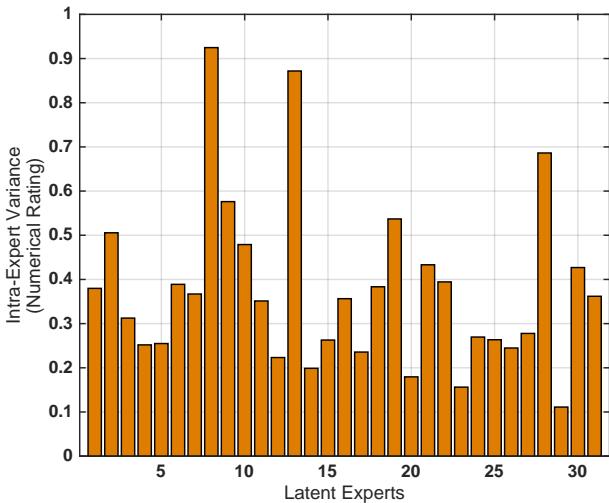


Fig. 8: Intra-expert variations in numerical quality ratings (Eq. (1)).

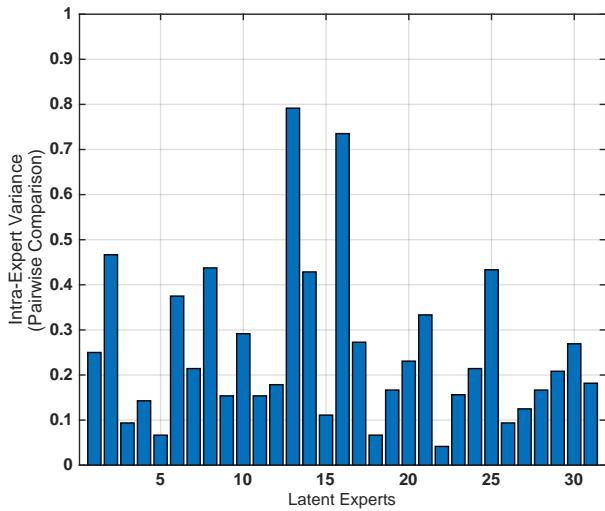


Fig. 9: Intra-expert variations in pairwise comparisons (Eq. (2)).

Let  $\{CR_{ij}^1\}_{j=1}^{20}$  and  $\{CR_{ij}^2\}_{j=1}^{20}$  denote the pairwise comparisons for the 20 repeated latents pairs rated 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 information than the right (left) latent, respectively. The intra-expert comparison value variation ( $CV_i$ ) of the  $i^{th}$  expert is then defined as

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

As shown in Fig. 9, two experts (13 and 16) have significantly higher intra-expert variation in their comparison ratings. Considering the variations of both numerical and comparison ratings, we discarded expert 13 and use crowdsourced data from the remaining 30 experts for learning the value predictor. An empirical analysis showed that any further removal of experts increased the mean square error (MSE) in latent value prediction.

### C. Matrix Completion for Latent Quality

The crowdsourced data for latent image quality can be represented by a two-dimensional 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 of fingerprint experts who provided quality ratings. Every observed value  $Q_{i,j} > 0$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ) represents the latent quality rating for the  $i^{th}$  latent 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 numerical ratings for that latent. Note that only up to 4,800 ( $30 \times 160$ ) out of a possible total of 15,480 ( $30 \times 516$ ) elements in  $Q$  have non-zero elements. The task at hand is to infer the complete quality matrix  $\hat{Q}$  given the sparse quality matrix  $Q$ . In machine learning, this problem is referred to as the matrix completion problem [31]. Since all the crowd workers in this study are fingerprint experts, it is reasonable to assume that there exists a small set of underlying quality functions such that quality matrix can be completed from their combination [32]. Further, for the ratings to be repeatable and reproducible, a small set of common underlying functions has to exist. This implies that the inferred quality matrix  $\hat{Q}$  should be of low rank. Let  $\Omega$  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  $\Omega$ , 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  $\hat{Q}$  such that the observed ratings (i.e., indices in  $\Omega$ ) are as close to the inferred ratings as possible. The problem is formulated as minimization of the optimization function:  $\|\mathcal{P}_\Omega(Q - \hat{Q})\|_F^2$ . Since any matrix  $\hat{Q} \in \mathbb{R}^{m \times n}$  of a rank up to  $K$  can be represented as a product of two matrices with form  $\hat{Q} = UV$ , where  $U \in \mathbb{R}^{m \times K}$  and  $V \in \mathbb{R}^{K \times n}$ , the objective function for inferring the latent value matrix  $\hat{Q}$  can then be formulated as the following non-convex optimization problem:

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

where  $K$  controls the rank of  $\hat{Q}$ . An alternating minimization approach, also known as non-linear Gauss-Seidel scheme, proposed in [33] is adopted to solve this. The models obtained by this approach are empirically shown to be very effective and are frequently used as an alternative to the nuclear norm minimization. As compared to the nuclear norm minimization model [34], it is much faster to solve for the non-convex model (Eq. 3) using this approach, even for a large matrix such as  $Q$ .

### D. Matrix Completion for Information Content

For each latent pair, experts are also requested to provide a pairwise comparison in terms of the information content, as shown in Fig. 6. Again, the task is to use matrix completion to find a complete information content matrix  $\hat{C}$ , given the sparse matrix  $C$ . Since the matrix completion algorithms are primarily designed to handle only three ordinal values ( $<$ ,  $=$ ,  $>$ ), we merge the “much more” and “slightly more” ratings.

Let  $(L_i \succ_{E_k} L_j)$  denote that latent  $L_i$  is regarded to contain more information over latent  $L_j$  rated by expert  $E_k$

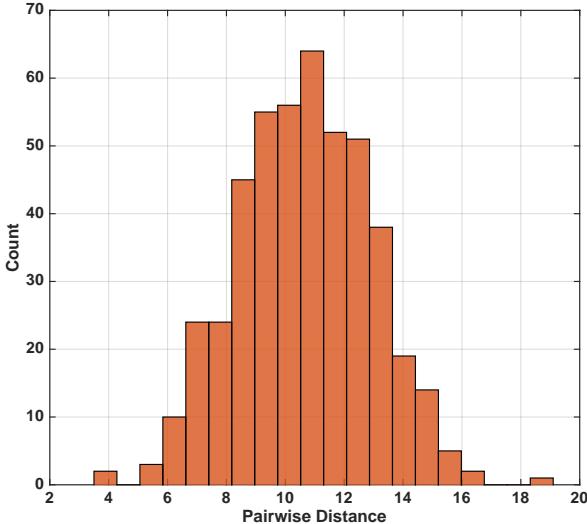


Fig. 10: Histogram of the pairwise distances between experts' value assignments  $\hat{V}$  after matrix completion.

and let  $(L_i \simeq_{E_k} L_j)$  denote  $L_i$  has similar information content 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 information content for each individual latent rated by each expert from the comparative information content  $C$  by finding a low matrix  $\hat{C} \in \mathbb{R}^{m \times n}$ . This is achieved by minimizing the following objective function:

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

where  $\Lambda > 0$  controls the tradeoff between minimizing the rank of  $\hat{C}$  and reducing the inconsistency with respect to the observed pairwise comparisons in  $C$ . Note that  $\|\hat{C}\|_{tr}$  is the trace norm of  $\hat{C}$  which approximates the rank of  $\hat{C}$ , and  $l(\hat{C}_{i,k} - \hat{C}_{j,k})$  is a loss function which measures the inconsistency in inferred comparison between  $\hat{C}_{i,k}$  and  $\hat{C}_{j,k}$ , and the rated comparison  $(L_i \succ_{E_k} L_j)$ . We adopt the approach proposed in [26] to solve Eq. (4) using the hinge loss function.

#### E. Determining the Basis for Expert Assigned Values

After matrix completion, each expert's latent quality and information content assignment is represented by the corresponding columns of completed matrices  $\hat{Q}$  and  $\hat{C}$ , respectively. We define the latent value  $\hat{V}$  as the average of quality  $\hat{Q}$  and information content  $\hat{C}$ :

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

To understand the underlying bases for expert crowd value assignment, it is crucial to understand the relationship among latents with different value assignments. Let  $\hat{v}_i$  ( $i = 1, 2, \dots, 30$ ), the  $i^{\text{th}}$  column of  $\hat{V}$ , denote the value assignment by the  $i^{\text{th}}$  expert  $E_i$ . The difference between two experts (e.g.,  $E_i$  and  $E_j$ ) can be measured by the Euclidean

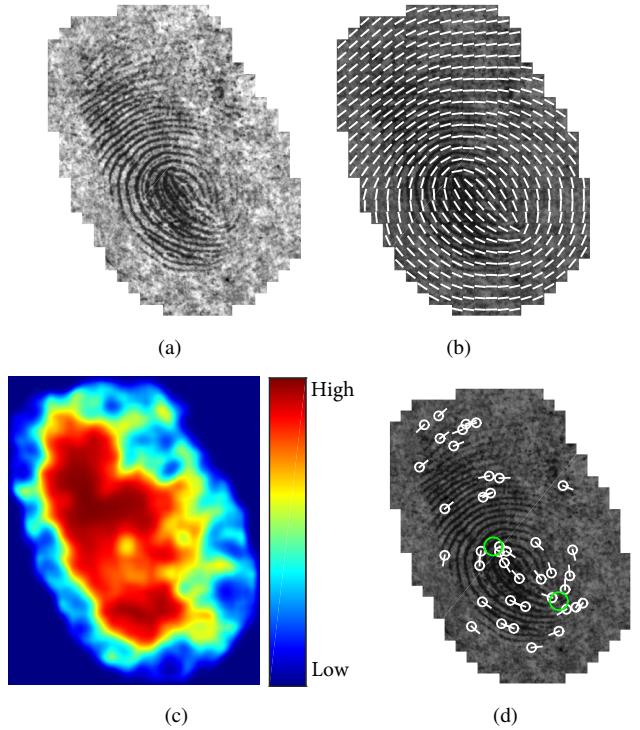


Fig. 11: Automatically extracted latent features. (a) Input latent with manually marked region of interest (ROI) or cropped latent, (b) ridge flow overlaid on the latent image, (c) ridge quality map, and (d) extracted minutiae (white circles) and core points (green circles).

distance between their corresponding value assignments  $\hat{v}_i$  and  $\hat{v}_j$ . While Fig. 10 shows variation in the pairwise distances between the expert value assignments, we would like to explain this variation in terms of a small number of bases or dimensions.

We capture the bases for this variation using Multidimensional Scaling (MDS) [23] whose aim is to place each latent in a  $d$ -dimensional Euclidean space such that the pairwise distances in this new space are as close to the corresponding distances based on expert value assignments. For this reason, MDS can also be viewed as a means of visualizing the expert defined similarity of individual latents embedded in a low dimensional space, typically 2D or 3D. Each MDS dimension is referred to as the basis for understanding the expert value assignments. In order to interpret these bases a set of  $n_f$  features, including minutiae feature, ridge clarity, ridge flow, number of core and delta, and minutia reliability are automatically extracted from a latent [1]. See Table II for this list of  $n_f = 19$  features and Fig. 11 for an illustration of some of the features. Let  $X \in \mathbb{R}^{(n_f+1)}$  be the normalized feature vector, including a dummy variable 1 used to absorb the bias (intercept) of the regression model [35],  $m$  be the number of latents ( $m = 516$  here) and  $y_k \in \mathbb{R}^m$  denote the  $k^{\text{th}}$  dimension of the MDS space. Fig. 12 illustrates the embedding of 516 latents in the three-dimensional MDS space. This embedding allows us to explore associations between the normalized feature vector  $X$  and individual MDS dimensions  $y_k$  using Lasso regression analysis tool [24]. Lasso selects a subset of salient features from  $X$  that are most relevant to

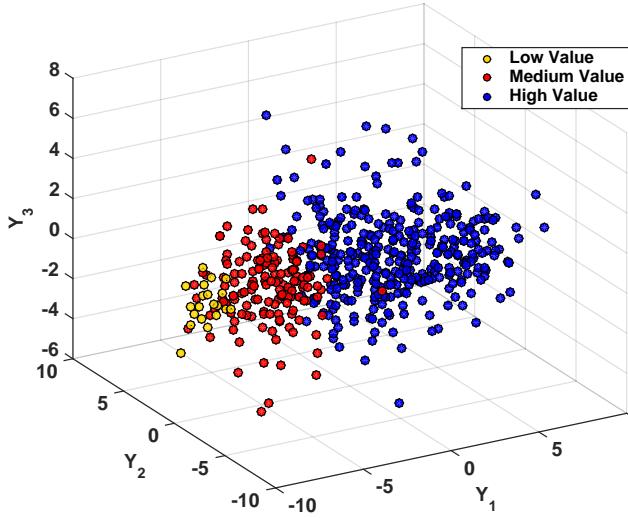


Fig. 12: Representation of 516 latents in the three-dimensional Multidimensional Scaling (MDS) space. The three latent value groups shown in different colors correspond to value ranges: [1, 1.5] for low value, (1.5, 2.5] for medium value, and (2.5, 5] for high value.

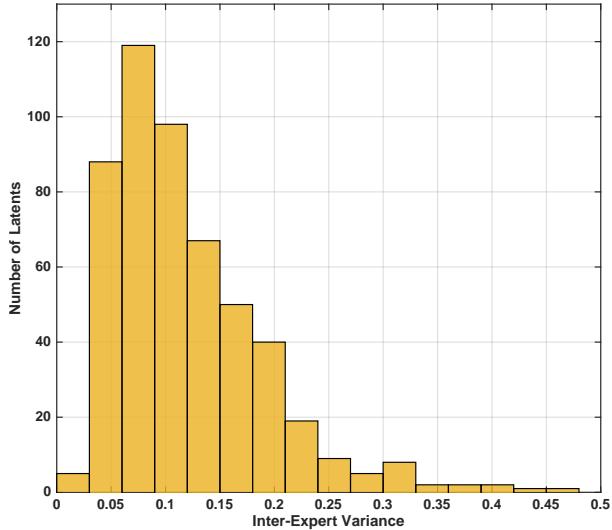


Fig. 13: Histogram of the inter-expert variance of latent values for 516 latents obtained after matrix completion.

interpret individual MDS dimensions and are robust against noisy data [36]. Formally, the weight vector for the  $n_f + 1$  features,  $\beta_k$ , for  $y_k$  is obtained by minimizing the following objective function,

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

where  $\lambda$  controls the sparseness of  $\beta_k$ . The MDS dimension  $y_k$  can be expressed as a linear combination of features  $x_i$ , extracted from a latent, as follows:

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

where  $\beta_{ki}$  is the feature weight for  $x_i$ .

TABLE II: 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 $\geq t$ , $t = 0, 0.1, \dots, 0.6$
$x_9$	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 $\geq t$ , $t = 0, 0.1, \dots, 0.6$
$x_{18}$	Number of core and delta
$x_{19}$	Standard deviation of the ridge flow (orientation map) in the latent foreground

#### F. Predicting Latent Value

Recall that the main goal of understanding the bases for expert assignments is to utilize them to learn a value predictor for query latents. Suppose that  $y_k \in \mathbb{R}^m$  ( $k = 1, \dots, d$ ) is the  $k^{th}$  basis of MDS space and  $\hat{v} \in \mathbb{R}^m$  is the value vector whose  $i^{th}$  element ( $\hat{v}_i$ ) is the median of the  $i^{th}$  row of completed value matrix  $\hat{V}$ . Lasso is again used to model the relationship between the MDS bases and the expert assigned latent value. Formally, this learning problem is given by:

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

where  $Y \in \mathbb{R}^{m \times d}$  is the matrix representation of  $y_k$  ( $k = 1, \dots, d$ ),  $\gamma \in \mathbb{R}^d$  is the weight vector, and  $\rho$  is a parameter to control the sparseness.

Given a query latent, its feature vector  $X_q$  (Table II) is extracted [1] and normalized using min-max normalization [37]. The projection of  $X_q$  on the  $k^{th}$  basis is computed by  $\hat{y}_k = X_q^T \beta_k$  ( $k = 1, \dots, d$ ). The value ( $V_{pred}$ ) assigned to the query latent is then given by

$$V_{pred} = \hat{Y}^T \gamma \quad (9)$$

where  $\hat{Y} = [\hat{y}_1, \dots, \hat{y}_d]^T \in \mathbb{R}^d$  and  $\gamma$  is obtained from solving Eq. (8). The process of offline learning of the value predictor and online latent value prediction for a given query latent is summarized in Algorithms 1 and 2, respectively.

### III. EXPERIMENTS AND DISCUSSIONS

We divide our study into a set of four experiments, in order to (i) understand the variability in expert value assignments, (ii) evaluate the performance of expert value assignment in terms of predicting state-of-the-art AFIS performance, (iii) understand the bases of expert value assignment in terms of automatically extracted features from a latent, and (iv) evaluate the proposed latent value predictor.

#### A. Variations in Expert Value Assignments

As shown in Figs. 8 and 9, respectively, there are significant intra-expert variations in both numerical ratings and pairwise comparisons. More specifically, the average intra-expert variations for numerical ratings and pairwise comparisons are 0.36 and 0.24, respectively, which confirms that pairwise

**Algorithm 1** Offline Learning of Value Predictor

---

```

1: procedure
2: input
3:    $Q$ : Numerical quality rating matrix
4:    $C$ : Pairwise comparisons of information content
5:    $X$ : Latent features
6: output
7:    $\beta$ : Weight vector for latent features
8:    $\gamma$ : Weight vector for derived MDS bases
9: begin:
10:   $\hat{Q} = \text{MatrixCompletion}(Q)$  Eq. (3)
11:   $\hat{C} = \text{MatrixCompletion}(C)$  Eq. (4)
12:   $\hat{V} = (\hat{Q} + \hat{C})/2$ 
13:   $[y_1, y_2, y_3] = \text{MDS}(\hat{V})$ 
14: loop: ( $k = 1, 2, 3$ )
15:    $\beta_k = \text{Lasso}(y_k, X)$ 
16: end loop
17:    $\beta = [\beta_1, \beta_2, \beta_3]$ 
18:    $\gamma = \text{Lasso}(\hat{V}, \beta)$ 
19: end
```

---

**Algorithm 2** Online Latent Value Predicton

---

```

1: procedure
2: input
3:    $\beta$ : Weight vector for latent features
4:    $\gamma$ : Weight vector for MDS bases
5:    $X_q$ : Latent features extracted from query latent
6: output
7:    $V_{pred}$ : Predicted latent value
8: begin:
9: loop: ( $k = 1, 2, 3$ )
10:   $\hat{y}_k = X_q^T \cdot \beta_k$ 
11: end loop
12:   $\hat{Y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3]$ 
13:   $V_{pred} = \hat{Y}^T \cdot \gamma$ 
14: end
```

---

comparisons have better reproducibility and higher reliability than numerical ratings.

On average, 12 experts assigned numerical ratings and 6 experts assigned pairwise comparisons, to each latent and to a pair of latents, respectively. The disparity in the inferred latent values, after matrix completion, obtained from the 31 experts measures the inter-expert variance. Fig. 13 presents a histogram of the inter-expert variance of inferred latent values obtained from an expert crowd of 31 experts. Fig. 14 shows four latents, where latents in Fig. 14 (a) and (b) have zero variation as all experts to whom this latent was presented, assigned (a) the lowest value *i.e.* 1, and (b) the highest value *i.e.* 5. On the other hand, we observe large variations in value assignments for latents in (c) and (d) with inter-expert variations of 1.61 and 1.78, respectively. In order to visualize the inter-expert variations, we apply MDS to the distance matrix of inferred expert ratings. Figures 15 (a) and (b) show a visualization of the 30 experts in a two-dimensional MDS space obtained from completed numerical rating ( $\hat{Q}$ ) and comparison rating ( $\hat{C}$ )

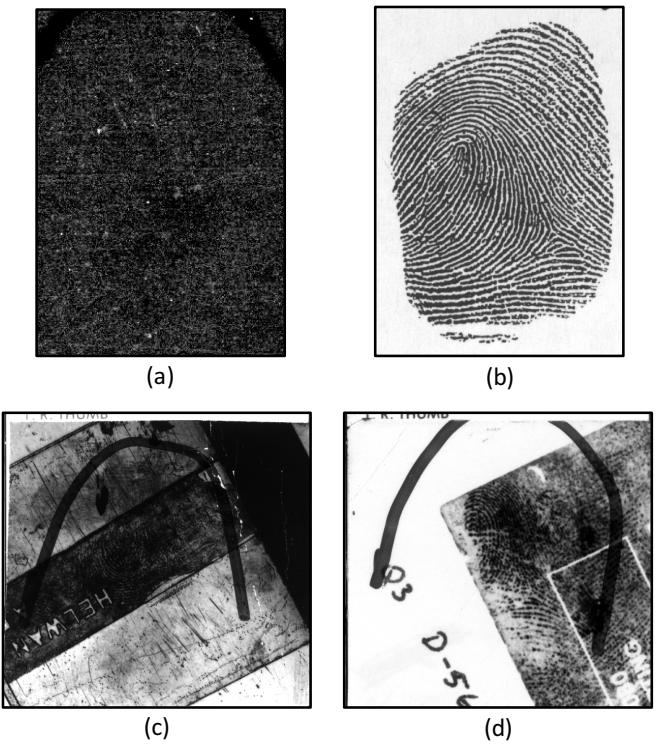


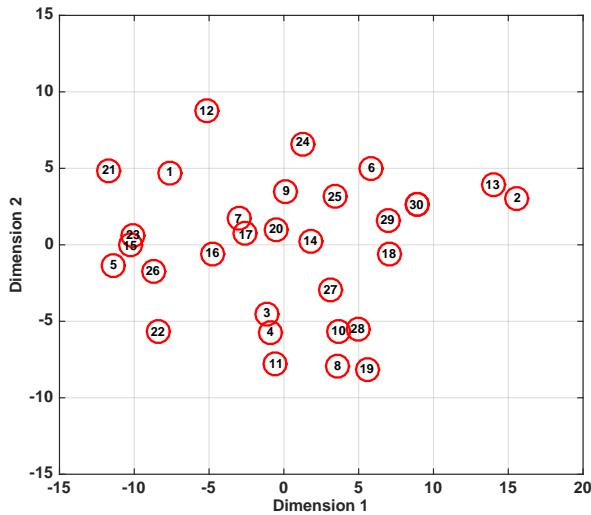
Fig. 14: Inter-expert variations in value assignments. All the experts assigned (a) the lowest value (*i.e.* 1), and (b) the highest value (*i.e.* 5). Latents in (c) and (d) received both low and high value assignments from the experts.

matrices, respectively. The configurations of these 30 experts in Figures 15 (a) and (b) are quite different. For example, expert nos. 15 and 23 are spatially close to each other in Fig. 15 (a) but are far away from each other in Fig. 15 (b). These two plots make it clear that the two types of information that we sought from fingerprint experts, namely individual quality ratings and pairwise comparisons are complementary. This also supports our value definition in Eq. (5) based on an average of quality and information content.

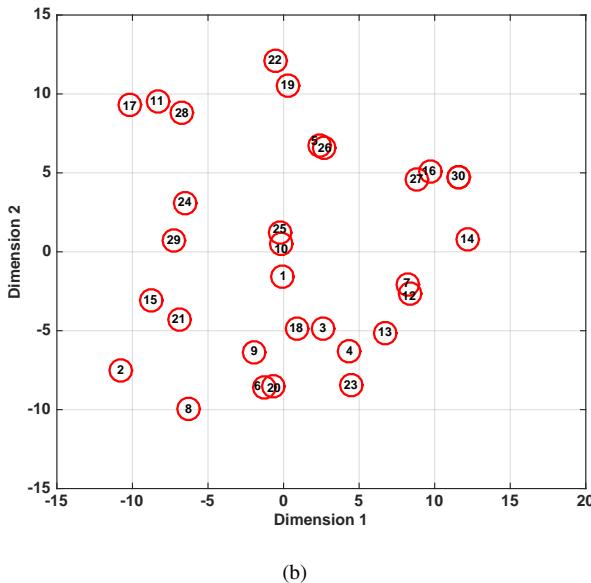
### B. Comparison with Value Determination

There are two possible sources of ground truth for latent value determination, namely, value determination by latent examiners [3] and value determination by AFIS [5], [1]. We do not utilize value determination by AFIS as the ground truth for the following three reasons: (i) the primary focus of this study is to understand the underlying bases used by fingerprint examiners for latent value assignment, (ii) the latent value prediction model learnt would be AFIS dependent, and (iii) the value determination of latents by AFIS also depends on the quality of their true mates. Therefore, we compare the inferred latent value from expert crowd with value determination (VID, VEO and NV) by examiners [3]. As value determinations are available only for NIST SD27 database, we do not use MSP database for this analysis. The numbers of VID, VEO, and NV latents in NIST SD27, reported in [3], are 210, 41, and 7, respectively.

We compare the number of latents retrieved at rank-1 against a reference database of 250,000 rolled prints, including



(a)



(b)

Fig. 15: Visualization of all 30 experts in a two-dimensional MDS space based on their (a) numerical ratings and (b) pairwise comparisons. Different configurations of experts in (a) and (b) indicate that the two types of information sought from them (quality and information content) are complementary. Each expert is represented as a red circle.

the true mates of 258 latents, by a state-of-the-art latent matcher<sup>7</sup>. Out of 258 latents from NIST SD27, 176 latents are retrieved at Rank-1 using this latent matcher. The latents are then sorted in descending order of the median of their inferred values ( $\hat{V}$ ) obtained by solving Eq. (5). The first 210 latents are considered as High Value, the next 41 as Medium Value and the last 7 as Low Value latents. This protocol is adopted for a fair comparison with the value determination from our algorithm. Table III compares the number of latents retrieved at rank-1 using value determination in [3] and inferred value from expert crowd. Expert crowd

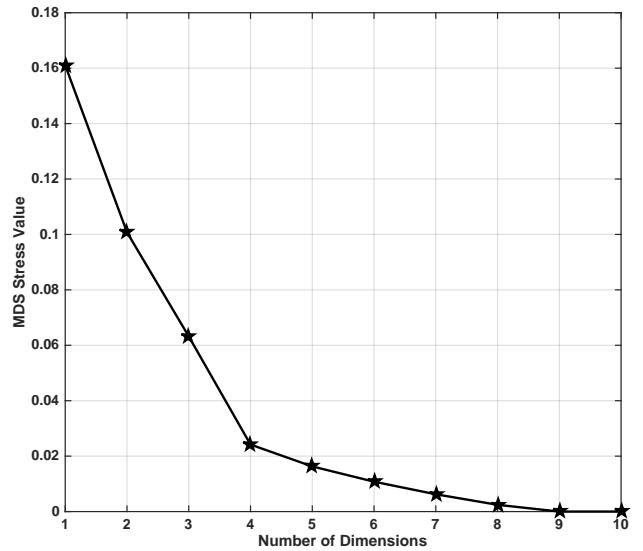


Fig. 16: Multidimensional Scaling (MDS) stress value vs. the number of dimensions. As the number of dimensions increase, the stress value decreases. A “knee” in this plot at  $d = 4$  indicates the low-rank characteristic of the completed latent value matrix ( $\hat{V}$ ).

performs better than value determination as there are more number of High Value latents identified at Rank-1 as compared to VID latents. This implies inferred value from expert crowd is better than value determination in [3] in predicting the AFIS performance.

TABLE III: Number of latents retrieved at Rank-1, using a state-of-the-art latent matcher, for latents determined as VID, VEO and NV by [3], compared to the median inferred latent value assigned by the expert crowd. A total of 176 latents out of 258 latents from NIST SD27 are retrieved at Rank-1. To have the same number of latents in three corresponding value categories for a fair comparison, the quantitative latent value is thresholded as High Value  $> 2.51$  and Low Value  $< 2.13$ . The range for the median inferred latent value is [1.97, 4.04]

Value Determination	VID	VEO	NV
	163/210	13/41	0/7
Expert Crowd (Median Value)	High Value	Medium Value	Low Value
	170/210	6/41	0/7

### C. Underlying bases for Value Assignment

As discussed in section II-E, we learn the feature weights  $\beta_k$  corresponding to each MDS dimension  $y_k$ , ( $k = 1, 2, \dots, d$ ). For a given dimensionality  $d$ , the MDS algorithm outputs  $d$ -dimensional representations of the 516 latents such that the difference between the latent value dissimilarity matrix  $D_v$  and the inter-point distance matrix of the output representation, known as the stress value, is minimized. We utilize the metric scaling implementation of MDS<sup>8</sup> which minimizes the stress value given by:

$$\text{Stress}(S) = \left( \frac{\sum \sum (f(x_{ij}) - d_{ij})^2}{\sum \sum d_{ij}^2} \right)^{1/2} \quad (10)$$

<sup>8</sup>Utilized MATLAB implementation (mdscale) <http://www.mathworks.com/help/stats/mdscale.html>

<sup>7</sup>One of the top performing latent matchers in NIST ELFT evaluation [38]

TABLE IV: Average feature weight vectors ( $\beta_k$ ) learnt using five fold cross-validation to predict the three MDS dimensions (Fig. 12),  $k = 1, 2, 3$ . The corresponding standard deviations are also reported. The 19 features  $\{x_1, x_2, \dots, x_{19}\}$  are defined in Table II. Non-zero weights are in bold; Weights lower than 0.05 are considered zero because they are relatively insignificant.

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. ( $\beta_1$ )	<b>0.68</b>	<b>0.28</b>	<b>0.31</b>	<b>0.22</b>	<b>0.16</b>	0	0	0	0	0	<b>0.16</b>	<b>0.19</b>	0.02	0	0	0	0	<b>0.40</b>	<b>0.19</b>	<b>0.17</b>
s.d. ( $\beta_1$ )	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. ( $\beta_2$ )	0	0	0	0	0	0	0	0	<b>0.69</b>	<b>0.26</b>	0	0	0	0	0	0	0	<b>0.13</b>	<b>0.52</b>	<b>0.44</b>
s.d. ( $\beta_2$ )	0	0	0	0	0	0	0	0.09	0.04	0	0	0	0	0	0	0	0.04	0.06	0.05	
Avg. ( $\beta_3$ )	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	<b>0.99</b>	
s.d. ( $\beta_3$ )	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0.02	

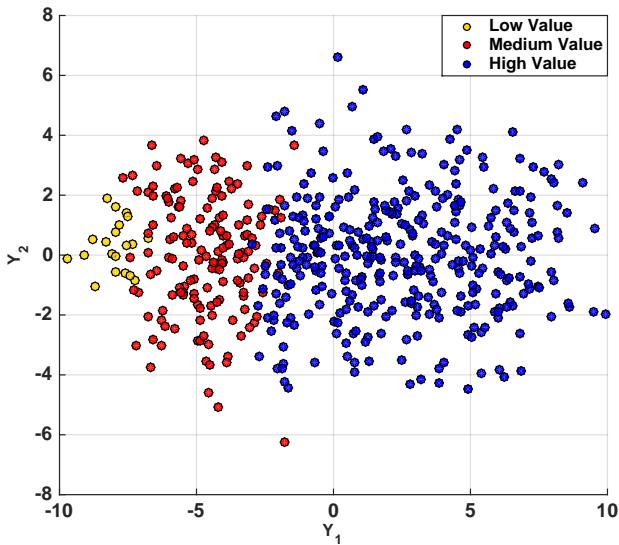


Fig. 17: Representation of 516 latents in two-dimensional MDS space.

TABLE V: Average and s.d. of the weight vector values for a three-dimensional MDS space (Fig. 12). These weights are used for predicting the value of latents in the test set.

Avg. Weight Vector $\gamma$	$\gamma_1$	$\gamma_2$	$\gamma_3$
	1.79	1.66	1.28
Std. Dev. ( $\gamma$ )	0.24	0.17	0.11

where  $f(x_{ij})$  is a monotonic function of input dissimilarities ( $x_{ij}$ ) and  $d_{ij}$  refers to the inter-latent euclidean distance. Fig. 16 presents the relationship between the MDS stress values against the number of dimensions. Fig. 12 illustrates the representation of 516 latents in three-dimensional MDS space using the dissimilarity matrix ( $D_v$ ) obtained from inferred latent values  $\hat{V}$ .

We perform a five-fold cross validation to evaluate the performance of our value predictor. Each fold contains approx. 413 latents for training and approx. 103 latents for testing. The value vector  $\hat{V}$  is considered as the ground truth or target value. We aim to learn a function which can predict the quantitative latent values for test latents with lowest mean square error. To learn  $\beta_k$ , for each dimension  $y_k$ , ( $k = 1, 2, 3$ ), we minimize the objective function presented in Eq. (6). The

optimal value of  $\lambda$  in Eq. (6) is learned using five fold cross validation. Table IV presents the feature weights for the three MDS dimensions obtained by performing Lasso on the inferred latent value assignment  $\hat{V}$  for the training set. As observed in Table IV, the third MDS dimension contains non-zero weight<sup>9</sup> only for the constant term, which indicates that only two bases are adequate to explain the inter-latent similarity. Fig. 17 illustrates the MDS representation of all the 516 latents in the two-dimensional space. The major constituents of the first dimension in terms of latent features that we extracted are: number of minutiae ( $x_1$ ), number of core and delta ( $x_{18}$ ), and minutia reliability ( $x_2, x_3, x_4, x_5$ ). For the second MDS dimension, friction ridge area ( $x_9, x_{10}$ ) and std. dev. of the ridge flow ( $x_{19}$ ) obtain high weights. As these two MDS dimensions are highly uncorrelated with  $\rho = 0.08$ , they can be interpreted as two independent bases used by fingerprint experts in latent value assignment.

#### D. Predicting Latent Value Assignment

As explained in section II-F, we perform Lasso on the MDS dimensions  $Y$  to learn the weight vector  $\gamma$  using five-fold cross-validation. Table V presents the average weights and their standard deviations, learnt by minimizing the objective function presented in Eq. (6). For each of the latents in the test set of each fold, we extract the latent image feature vector  $X_q$  in Table II. Using the weight vectors  $\beta_k$  ( $k = 1, 2, 3$ ), we compute each of the underlying bases by  $\hat{y}_k = X_q^T \beta_k$ . The predicted latent value is then computed by  $V_{pred} = \hat{Y}^T \gamma$ , where  $\hat{Y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3]^T$ . Therefore, the latent value prediction function can be rewritten as a linear combination of feature weights as follows:

$$\begin{aligned} V_{pred} = & 1.22 x_1 + 0.50 x_2 + 0.55 x_3 + 0.39 x_4 + 0.29 x_5 \\ & + 1.15 x_9 + 0.43 x_{10} + 0.29 x_{11} + 0.34 x_{12} + 0.20 x_{13} \\ & + 0.16 x_{14} + 0.94 x_{18} + 1.20 x_{19} + 2.31 \end{aligned} \quad (11)$$

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 average mean square error<sup>10</sup> between the predicted latent value and the crowdsourced latent value of the five folds

<sup>9</sup>Non-zero weights are in bold; Weights lower than 0.05 are considered zero because they are relatively insignificant.

<sup>10</sup><http://www.mathworks.com/help/images/ref/immse.html>

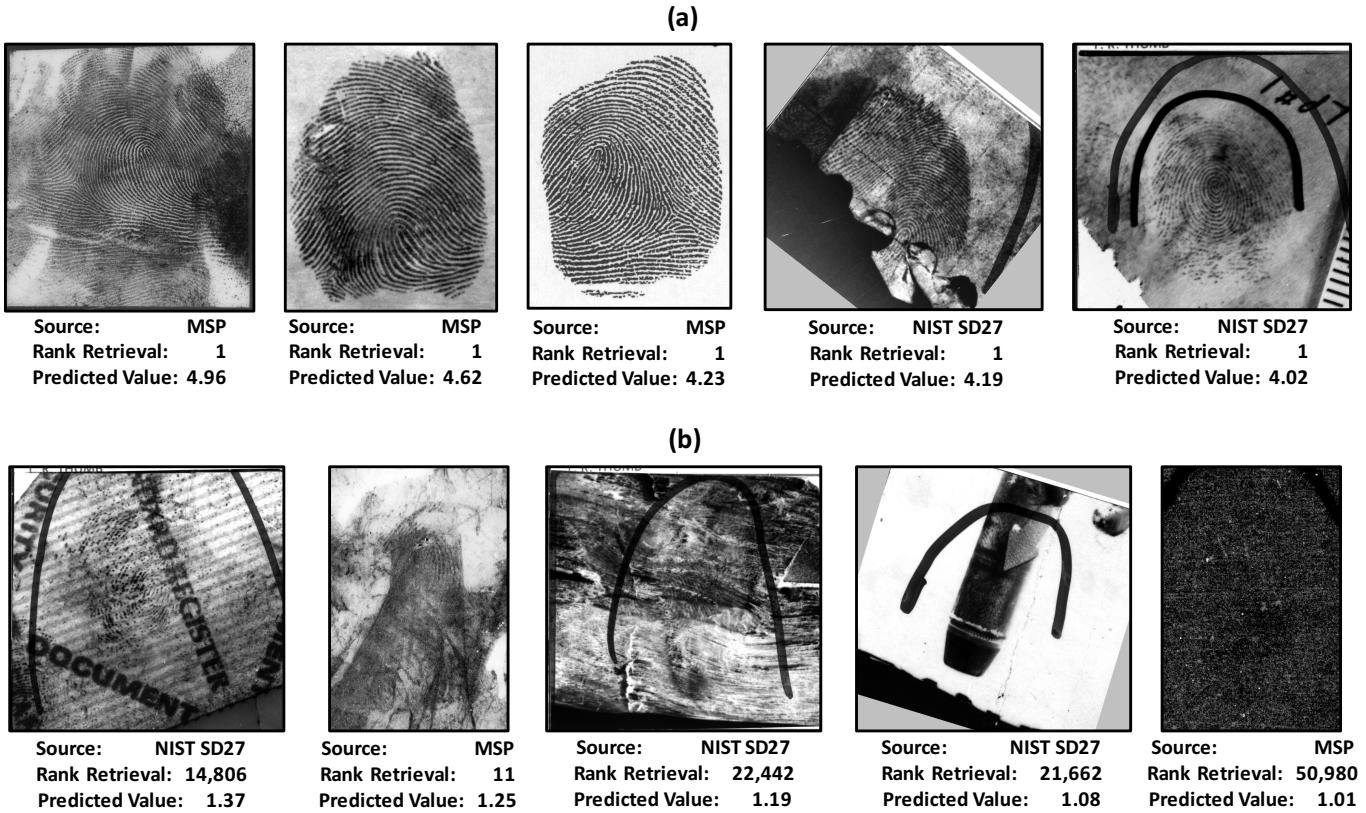


Fig. 18: The (a) top-5 and (b) bottom-5 latents based on their predicted value. For each latent, we also report the rank at which it is retrieved using a state-of-the-art COTS latent matcher (size of reference gallery = 250K rolled prints).

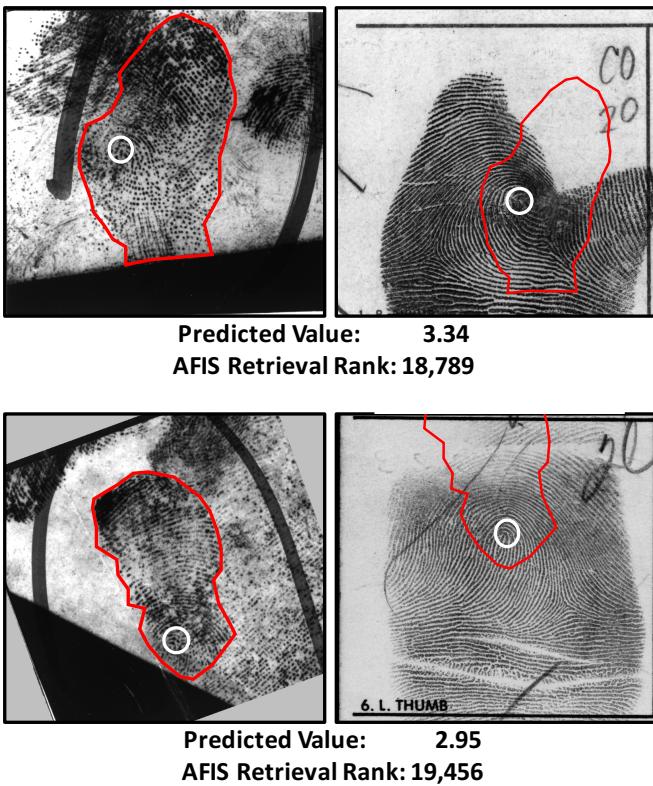


Fig. 19: Example latents with fairly high predicted value but poor AFIS performance. Latent ridge area and its correspondence on their true mate are presented in red; core points are marked in white.

is 0.19 with a standard deviation of 0.04. Fig. 18 presents the top-5 and bottom-5 ranked latents based on the predicted latent value for one of the five folds. The AFIS performance not only depends on the value of the query latent, but is also affected by the quality and information content of rolled prints. Fig. 19 presents examples of latents with reasonably high predicted value but poor AFIS performance, which can be attributed to poor quality and missing information in their true mates. The proposed algorithm is capable of generating a ranked list of unseen latents based on their value in a fully automatic or “lights out” mode. Such a ranking will enable efficient use of examiners’ efforts and resources.

#### IV. CONCLUSIONS

We have proposed a fully automatic method to assign quantitative values to latent fingerprints. Latent value assignment is a crucial step in the widely used ACE-V methodology. However, the prevailing approaches for latent value assignment or modeling have the following two main limitations: (i) latent values (VID, VEO and NV) assigned by examiners used in ACE-V methodology are subjective with low reproducibility, and (ii) directly modeling the relationship between latent feature set and value determination, can not explain inter-examiner variations and, more importantly, the underlying bases that examiners use for value determination. In order to address these two limitations, we have (i) developed a crowdsourcing-based framework for understanding expert latent value assignment from the perspectives of latent quality

and information content, (ii) used MDS to identify the underlying bases for expert latent value assignment, (iii) modeled the relationship between latent feature set and MDS bases using Lasso, and (iv) predicted quantitative latent value based on the underlying bases.

Our experiments involved 516 latent images, from two different sources, and 31 fingerprint experts who participated in our study. Based on our results, we conclude that (i) the crowdsourced latent value is more robust than prevailing value determination (VID, VEO and NV) in terms of predicting AFIS identification performance, and (ii) two MDS bases are adequate to explain human value assignments. These two bases themselves are uncorrelated and can be interpreted in terms of latent features that we can automatically extract.

Our suggestions for future work on latent value prediction include (i) extracting robust latent features, especially minutiae points, for latent value prediction, (ii) improving the current model by incorporating feature rarity, and (iii) evaluating the proposed method on an independent large operational latent database.

#### 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 Forensic Science Program, NIST.

#### 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] D. R. Ashbaugh, *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC press, 1999.
- [3] 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.
- [4] "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)."
- [5] 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.
- [6] 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.
- [7] 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.
- [8] ——, "Accuracy and reliability of forensic latent fingerprint decisions," *Proceedings of the National Academy of Sciences (PNAS)*, vol. 108, no. 19, pp. 7733–7738, 2011.
- [9] "Fact sheet - Integrated Automated Fingerprint Identification System (IAFIS)," <https://www.fbi.gov/about-us/cjis/fingerprints-biometrics/ngi/next-generation-identification-monthly-fact-sheet>.
- [10] B. Keelan, *Handbook of Image Quality: Characterization and Prediction*. CRC Press, 2002.
- [11] E. Tabassi, C. Wilson, and C. Watson, "NIST Fingerprint Image Quality," *NIST Res. Rep. NISTIR7151*, pp. 34–36, 2004.
- [12] B. T. Ulery, R. A. Hicklin, G. I. Kiebzinski, 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.
- [13] 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.
- [14] S. Yoon, E. Liu, and A. K. Jain, "On Latent Fingerprint Image Quality," in *Computational Forensics*. Springer, 2015, pp. 67–82.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2005.
- [19] 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.
- [20] F. Ribeiro, D. Florencio, C. Zhang, and M. Seltzer, "CrowdMOS: An Approach for Crowdsourcing Mean Opinion Score Studies," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 2416–2419.
- [21] P. Hanhart, P. Korshunov, and T. Ebrahimi, "Crowd-based Quality Assessment of Multiview Video Plus Depth Coding," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 743–747.
- [22] D. Martinho-Corbisley, M. Nixon, and J. Carter, "Analysing Comparative Soft Biometrics from Crowdsourced Annotations," *IET Biometrics*, pp. 1–16, 2016.
- [23] J. B. Kruskal and M. Wish, *Multidimensional Scaling*. Sage, 1978, vol. 11.
- [24] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, pp. 267–288, 1996.
- [25] 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.
- [26] 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.
- [27] S. Negahban, S. Oh, and D. Shah, "Iterative Ranking from Pair-wise Comparisons," in *Advances in Neural Information Processing Systems 25 (NIPS)*. MIT Press, 2012, pp. 2483–2491.
- [28] "NIST SD27 Latent Fingerprint Database," <http://www.nist.gov/srd/nistsd27.cfm>.
- [29] 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.
- [30] O. Dekel and O. Shamir, "Vox Populi: Collecting High-Quality Labels from a Crowd," in *Conference On Learning Theory (COLT)*, 2009.
- [31] E. J. Candès and B. Recht, "Exact Matrix Completion via Convex Optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.
- [32] E. Schenck and C. Guittard, "Towards a Characterization of Crowdsourcing Practices," *Journal of Innovation Economics & Management*, no. 1, pp. 93–107, 2011.
- [33] 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.
- [34] 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.
- [35] C. M. Bishop, "Pattern recognition," *Machine Learning*, vol. 128, 2006.
- [36] P. J. Bickel, Y. Ritov, and A. B. Tsybakov, "Simultaneous Analysis of Lasso and Dantzig Selector," *The Annals of Statistics*, pp. 1705–1732, 2009.
- [37] A. Jain, K. Nandakumar, and A. Ross, "Score Normalization in Multimodal Biometric Systems," *Pattern Recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.
- [38] M. Indovina, V. Dvornychenko, R. Hicklin, and G. Kiebzinski, "ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation #2]," *National Institute of Standards and Technology, NISTIR*, vol. 7859, 2012.