To appear in the Proceedings of the 8th International Conference on Biometrics (ICB),
Phuket, Thailand, May 19-22, 2015

Crowd Powered Latent Fingerprint Identification: Fusing AFIS with Examiner Markups

Sunpreet S. Arora, Kai Cao and Anil K. Jain
Department of Computer Science and Engineering
Michigan State University
East Lansing, Michigan 48824
Email: {arorasun, kaicao, jain@cse.msu.edu}

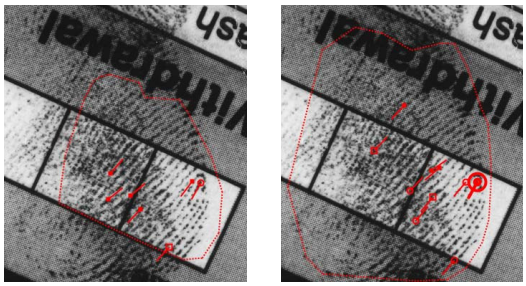
Gregoire Michaud
Forensic Science Division
Michigan State Police
Lansing, Michigan 48913
Email: michaudg@michigan.gov

Abstract

Automatic matching of poor quality latent fingerprints to rolled/slap fingerprints using an Automated Fingerprint Identification System (AFIS) is still far from satisfactory. Therefore, it is a common practice to have a latent examiner mark features on a latent for improving the hit rate of the AFIS. We propose a synergistic crowd powered latent identification framework where multiple latent examiners and the AFIS work in conjunction with each other to boost the identification accuracy of the AFIS. Given a latent, the candidate list output by the AFIS is used to determine the likelihood that a hit at rank-1 was found. A latent for which this likelihood is low is crowdsourced to a pool of latent examiners for feature markup. The manual markups are then input to the AFIS to increase the likelihood of making a hit in the reference database. Experimental results show that the fusion of an AFIS with examiner markups improves the rank-1 identification accuracy of the AFIS by 7.75% (using six markups) on the 500 ppi NIST SD27, 11.37% (using two markups) on the 1000 ppi ELFT-EFS public challenge database, and by 2.5% (using a single markup) on the 1000 ppi RS&A database against 250,000 rolled prints in the reference database.

1. Introduction

One of the most challenging problems in fingerprint recognition is comparing latent prints to rolled/slap (reference) fingerprints. Comparison of latents to reference prints by state-of-the-art AFIS does not typically yield satisfactory results. This is because many operational latents (i) are partial prints with relatively small friction ridge area, (ii) have poor contrast and clarity with significant distortion, and (iii) have significant background noise [11]. Therefore, a latent examiner is, in general, needed to mark features on a latent before submitting a query to an AFIS, and for subsequently reviewing the top-K (usually K = 20-50) retrievals to determine if the latent hit against a reference print. This manual intervention procedure is summarized in the ACE-V protocol [5].



(a) (b)

Figure 1: Two markups (by two different examiners) for a latent image from 1000 ppi ELFT-EFS database. A state-of-the-art AFIS was unable to make a hit for the latent image in lights-out mode (score of 0 with the true mate in the reference database). However, feeding the AFIS with the markups shown in (a) and (b) resulted in the mated print being retrieved at rank-1 and rank-129, respectively.

The NIST Evaluation of Latent Fingerprint Technologies, Extended Feature Sets (ELFT-EFS) 2 [9] reported that the likelihood of finding a hit in the reference database improves when an AFIS is provided with a markup[1] (see Figure 1). The identification accuracy of the best AFIS operating in the lights-out mode[2] is reported to be 67.2% in identifying 1,066 latent prints against reference prints from 100,000 subjects. However, the above accuracy improves to 70.2% when the AFIS is fed with both the latent image and the extended feature set (EFS) markup provided by NIST.

The above performance gain, however, depends on the precision of the markup being fed to the AFIS [10]. Imprecise markups can result in the mated reference print being returned at a lower rank amongst the retrieved candidates [7] [14] compared to the image alone being fed to the AFIS. Furthermore, markups for the same latent by different examiners can differ significantly and, con-

[1] Markup, in this paper, refers to the latent image with features marked by a human examiner.

[2] In the lights-out mode, AFIS automatically extracts features and compares the latent to the reference prints, without any human intervention.

sequently, different markups may lead to difference in identification performance of the AFIS (see Figure 1).

To overcome the aforementioned limitations, we propose a latent identification framework where the AFIS and latent examiners operate in synergy to improve the latent identification accuracy. In this framework, a latent is first submitted to the AFIS to be matched in the lights-out mode. Based on the output of the AFIS, the likelihood that the AFIS hit against a reference print at rank-1 is determined using a variant of the criterion described in [4]. If the likelihood of the AFIS making a hit at rank-1 is low, the latent is crowdsourced to a pool of latent examiners for marking features. In this manner, the collective "wisdom" of several latent examiners is utilized to obtain multiple markups for a latent only when required. The manual markups are then used in conjunction with the AFIS for improving the latent identification accuracy.

The proposed framework is based on the conjecture that combining markups obtained from different examiners with the automated encoding of the AFIS can benefit the identification performance of the AFIS. The conjecture stems from the classic pattern recognition theory that a group of experts with diverse and complementary skills can collectively solve a difficult problem, on average, better than each individual expert [6] [8]. Each latent examiner, as well as the AFIS, can be viewed as an expert for latent markup. Because manual markups obtained from different latent examiners lead to different candidate lists being retrieved by the AFIS, their expertise is rather diverse. Thus, a combination of AFIS with examiner markups should boost the identification performance of the AFIS.

State-of-the-art AFIS typically generate multiple templates (encodings) from an input latent. These templates are then individually compared with the reference prints and the resulting comparison scores are combined to generate a single candidate list. Our method can be viewed analogous to generating multiple templates, albeit based on feature markup by multiple latent examiners, and fusing them with the multiple templates internally generated by an AFIS.

To evaluate the proposed framework, we crowdsourced markups for the NIST Special Database 27 (NIST SD27) latents [2] to six certified latent print examiners affiliated to Michigan State Police. We also conduct experiments using two individual markups provided in the ELFT-EFS public challenge database [1] and one individual markup provided in the RS&A database [3]. We compute the efficacy of the proposed criterion to compute the likelihood that the AFIS makes a hit for the latent at rank-1. The proposed criterion is able to reduce the number of latents that need to be crowdsourced for manual markup from 258 to 151 for NIST SD27, from 255 to 151 for ELFT-EFS database, and from 200 to 35 for RS&A database (at a significance level of 0.05) without impacting the overall hit rate. Our experimental results on a reference database of 250,000 rolled prints show that by fusing the scores from lights-out comparison with the scores obtained using the examiner markups, the rank-1 identification accuracy of the AFIS improves by 7.75% on 500 ppi NIST SD27 (using six markups), by 11.37% on 1000 ppi ELFT-EFS database (using two markups), and by 2.5% on 1000 ppi RS&A database. Our experimental results indicate that markups obtained from different latent examiners contain complementary information which, in turn, helps to boost the identification performance of an AFIS.

The contributions of this paper are:

- A systemic way to combine the AFIS with examiner markups to boost the latent hit rate,
- A crowd powered latent identification framework where a latent is crowdsourced to a pool of examiners for obtaining multiple markups,
- A criterion to automatically determine when crowdsourcing is required, and
- A method to dynamically determine how many crowd experts are needed.

2. Collective Wisdom of Multiple Examiners

Harnessing the "collective wisdom" of the crowd is a commonly used methodology for performing relatively simple tasks (e.g., image labeling, product recommendations). For instance, recommendation systems in Netflix[3] and Amazon[4] use the collective preferences of a large number of customers when recommending movies or products to a specific customer. Expert crowdsourcing is another concept which has recently gained prominence [13] [16]. This involves dynamically assembling a team of expert crowd workers for accomplishing specialized tasks. We extend these concepts to latent fingerprint identification in the following manner.

Given a latent, an AFIS operating in lights-out mode is first used to compare it to reference prints in the background database. Based on the score distribution of the top-K candidate matches output by the AFIS, we ascertain whether manual markup is needed to boost the identification performance. If it is determined that manual markup is needed, the latent markup is crowdsourced to a pool of latent examiners. The obtained markups are input to AFIS individually to generate multiple scores for each reference print in the database. These individual markup and lights-out scores are then fused to boost the identification accuracy of the AFIS[5] (see Figure 2).

2.1. Expert crowdsourcing framework

Let a query latent image be denoted by I_L, and the set of reference print images in the database be denoted by I_R. Let the total number of reference prints in the

[3] http://www.netflix.com

[4] http://www.amazon.com/

[5] Rank-level fusion was also investigated for fusing the lights-out identification results with the identification results obtained for different markups. However, score-level fusion outperformed rank-level fusion in our experiments.

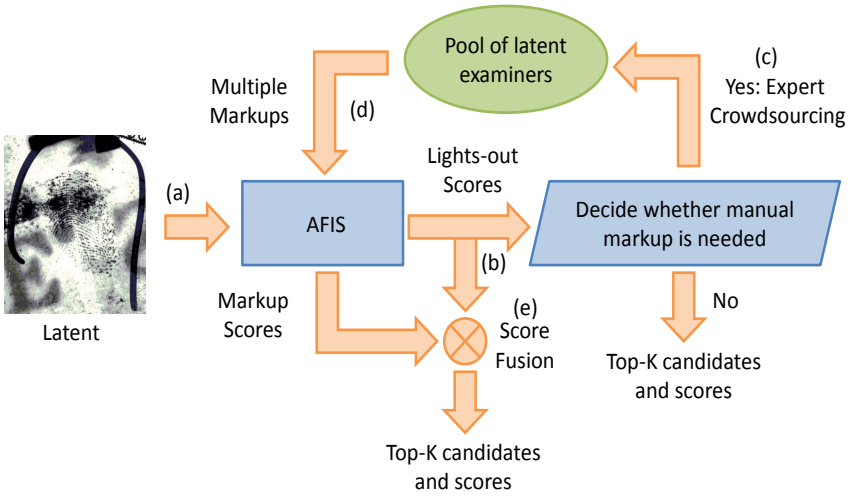


Figure 2: The proposed crowd powered latent fingerprint identification framework. (a) Latent is fed to an AFIS, (b) it is determined whether manual markup is needed, (c) markups are obtained via expert crowdsourcing, (d) multiple markups are fed to the AFIS, and (e) AFIS scores in (b) are fused with the multiple markup scores in (d).

database be N, and the i^{th} reference print be denoted by $I_R(i)$. The latent I_L is compared against the set of reference prints I_R using an AFIS operating in lights-out mode to generate a set of similarity scores S_{LO}:

$$S_{LO}(i) = S(I_L, I_R(i)); \forall i = \{1, 2, \dots N\}. \quad (1)$$

where $S(I_L, I_R(i))$ is the similarity between I_L and $I_R(i)$ output by the AFIS.

A parametric probability distribution model is fit to the distribution of the top-K scores from the set S_{LO}, and a variant of the method described in [4] is used to ascertain whether manual markup is required (see Section 2.2). If it is determined that manual markup is not required, the set of top-K candidates and their corresponding scores are directly output for validation by a latent examiner. Otherwise, the latent image I_L is crowdsourced to a pool of latent examiners for providing manual markups.

Let P be the number of examiners that provide manual markups. Denote these markups as I_M, where the j^{th} markup is $I_M(j)$. Each of the P markups are individually input to the AFIS to obtain similarity scores against the set of reference prints I_R,

$$S_{Mj}(i) = S(I_M(j), I_R(i)); \quad (2)$$
$$\forall j = \{1, 2, \dots, P\}, \forall i = \{1, 2, \dots, N\}.$$

Here $S_{Mj}(i)$ denotes the similarity score by comparing the j^{th} latent markup to the i^{th} reference print.

Finally, we fuse the lights-out scores S_{LO} with the similarity score of each markup S_{Mj} for every reference print to obtain a combined score S_F,

$$S_M(i) = \otimes\{S_{Mj}(i)\}, \quad (3)$$
$$\forall j = \{1, 2, \dots, P\}, \forall i = \{1, 2, \dots, N\};$$
$$S_F(i) = S_{LO}(i) \otimes S_M(i); \forall i = \{1, 2, \dots, N\}. \quad (4)$$

Here, $\otimes$ is the score fusion operator, and S_M denotes the score obtained by fusing the scores of the P different markups. The top-K fused scores from the set S_F, and the corresponding candidates based on the fused scores are then output to the latent examiner for evaluation.

2.2. When to crowdsource?

Although expert crowdsourcing has its advantages, crowdsourcing every latent to a pool of latent examiners is costly in terms of time and effort. If it can be established that the likelihood of AFIS making a hit at rank-1, for a given latent, is fairly high, then expert crowdsourcing is not utilized for that latent. While latent quality can be an indicator of this likelihood, to our knowledge, there is no existing satisfactory indicator of latent quality [15]. Therefore, we base our decision on the order statistic of the top-K candidate scores returned by the AFIS [4].

Let the set of top-K scores returned by the the AFIS be denoted as $X = \{X_{(1)}, X_{(2)}, \dots X_{(K)}\}$ where $X_{(i)}$ denotes the rank-i score. An exponential distribution model is fit to the set X. A hypothesis test is conducted to determine whether there is an upper outlier present in the distribution of the top-K scores. The null hypothesis (H_0) and the alternative hypothesis (H_1) are defined as follows:

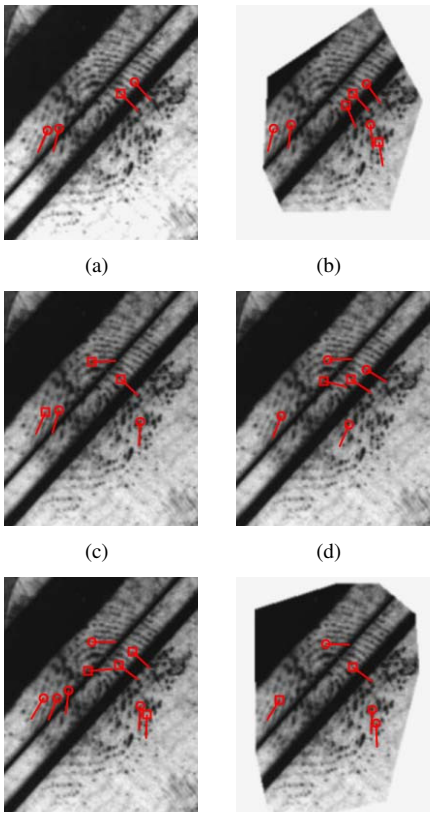
H_0: All scores in the set X are i.i.d from an exponential distribution.

H_1: Rank-1 score $X_{(1)}$ is an upper outlier of the score distribution.

The test statistic Z for testing H_0 against H_1 is defined as:

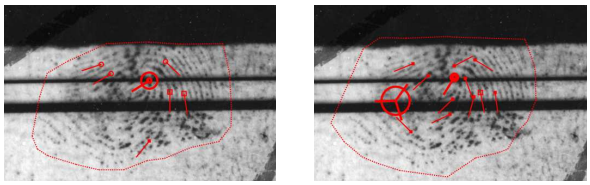
$$Z = \frac{X_{(1)} - X_{(2)}}{S_K}; \; S_K = \sum_{i=1}^{K} X_{(i)}. \quad (5)$$

The critical value of the test $z(\alpha)$ at significance level α is:



(a) (b) (c) (d) (e) (f)

Figure 3: Markups by six different latent examiners for a latent image in the 500 ppi NIST SD27.



(a) (b)

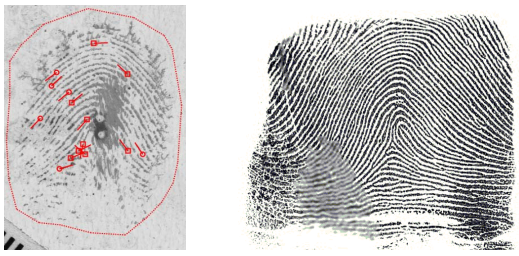
Figure 4: Markups by two examiners for a latent in the 1000 ppi ELFT-EFS public challenge database.

$$z(\alpha) = 1 - \alpha^{\frac{1}{K-1}}. \tag{6}$$

The value of the test statistic $Z = z$ should be greater than the critical value $z(\alpha)$ when rank-1 score $X_{(1)}$ is an outlier. Thus, we define an indicator random variable I_C which takes the value 1 when expert crowdsourcing is needed, and 0 when it is not needed:

$$I_C = \begin{cases} 0, & z > z(\alpha), \\ 1, & \text{otherwise} \end{cases} \tag{7}$$

In other words, if the rank-1 score is indeed an upper outlier, we are sufficiently confident that lights-out identification retrieved the mated reference print at rank-1. Therefore, the query latent does not need markups from latent examiners.



(a) (b)

Figure 5: Markup for a latent image (a) in the 1000 ppi RS&A database. The mated reference print of the latent is shown in (b).

Database	#Latents	Resolution	Latent Type	#Markups
NIST SD27	258	500 ppi	operational	6*
ELFT-EFS**	255	1000 ppi	operational	2*
RS&A	200	1000 ppi	collected in lab	1

*The scope of this research is to investigate how best to combine independent markups. Therefore, juried markups, although available, are not used because they involve the expertise of multiple examiners.
** ELFT-EFS database contains 255 latents from NIST SD27 rescanned at 1000 ppi.

Table 1: Summary of the latent databases used.

Examiner	1	2	3	4	5	6
No. of markups	253	255	255	255	253	257

Table 2: Number of latents markups provided by each of the six examiners (out of 258) for the NIST SD27 latents.

2.3. How many experts are enough?

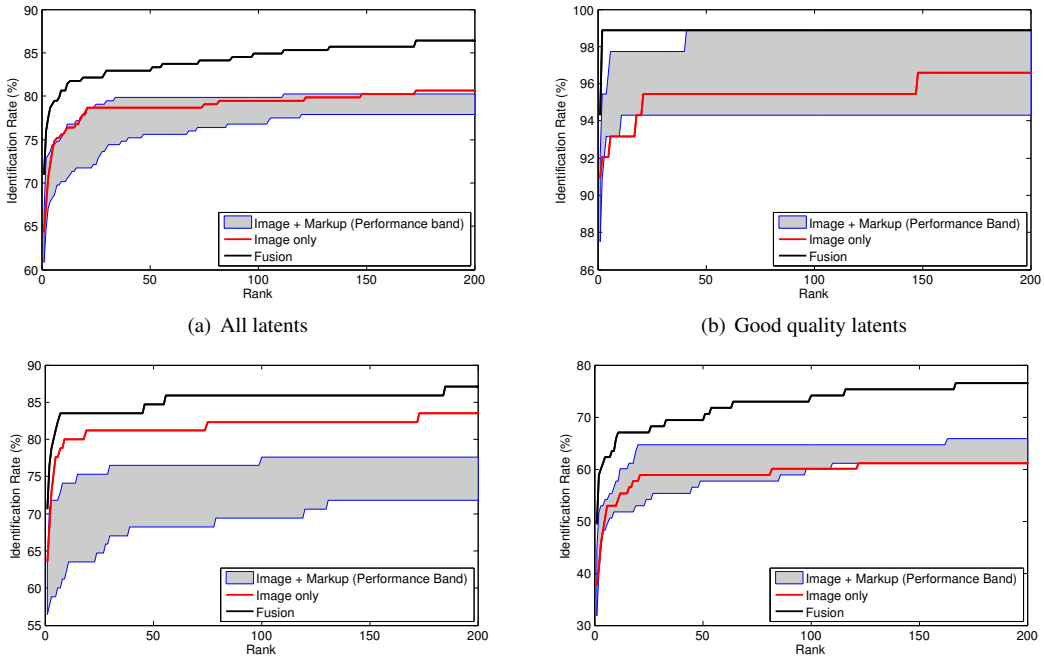
A priori information about latent examiners (e.g., years of experience, the number of cases solved) is often known and can be utilized while crowdsourcing latent markup. Assume that the latent examiners can be rated based on such prior information. When additional markup is required for a latent, instead of crowdsourcing the latent to every examiner, it can be first sent to the best examiner to obtain a markup. The best examiner's markup can then be fused with the lights-out AFIS, and the decision whether additional markup is needed made. Subsequently, the latent can be sent to the next best examiner, if required. Such a greedy (sequential) strategy can dynamically determine the number of examiners needed for providing markups, in turn, reducing the required cost and effort [12].

3. Experimental Details

A state-of-the-art AFIS, which was one of the top performing AFIS in the NIST ELFT-EFS 2 evaluation [9], is used for conducting all identification experiments.

3.1. Databases

The proposed latent markup crowdsourcing framework is evaluated on three different latent databases (summarized in Table 1), the NIST SD27 [2], ELFT-EFS [1] and the RS&A [3]. In addition to the mated reference prints of the latents available from these databases,



(a) All latents

(b) Good quality latents

(c) Bad quality latents

(d) Ugly quality latents

Figure 6: Identification performance (CMC curves) of the AFIS on NIST SD27 when (i) operating in lights-out mode (Image only), (ii) fed with markup from a single examiner (Image + Markup), and (iii) fusion of lights-out and 500 ppi markups from all six examiners (Fusion) for (a) all 258 latents, (b) 88 good quality latents, (c) 85 bad quality latents, and (d) 85 ugly quality latents. The size of the reference database is 250K rolled prints, including the true mates of latents from NIST SD27. The performance band of the latent examiners indicates the maximum and minimum accuracy obtained using an individual examiner markup at different ranks.

we use rolled prints, provided by Michigan State Police, to enlarge our reference database to 250,000 rolled prints for all the experiments reported here.

3.2. Latent Markup

Independent feature markups for NIST SD27 latents were obtained from six certified latent print examiners affiliated to Michigan State Police. The average feature markup time is about 5 min. per latent (around 20 hours for all 258 latents). Examiners were specifically asked to mark minutiae, ridge counts between minutiae and/or region of interest (ROI) on the latents. However, not all examiners marked all 258 latents (see Table 2). Figure 3 shows sample markups obtained from the six examiners for a latent in NIST SD27. Some examiners marked ROI while others did not. For each latent in the ELFT-EFS database, at least two independent feature markups are available with the database. Standard EFTS-LFFS feature markups (minutiae, ridge counts between minutiae, singular points and ROI) are used in our experiments. Note that latent examiners, in general, do not mark extended features on a latent because it is a challenging (ambiguous) and time consuming process. Hence, our experiments are in accordance with the general markup protocol being followed by examiners in law enforcement agencies. Figure 4 shows sample markups for a latent from the ELFT-EFS database. Only a single markup is available in the RS&A database [3] which is utilized in our experiments (Figure 5).

3.3. Experiments

To evaluate the efficacy of the proposed expert crowdsourcing framework, we perform the following set of experiments.

3.3.1 Lights-out Matching

The Cumulative Match Characteristic (CMC) curves of the AFIS in the lights-out mode on NIST SD27 are marked as Image only in Figure 6 (a). The rank-1 identification accuracy is 64.34%. Notice the reduction in identification performance of the AFIS on bad and ugly quality latents as compared to the good quality latents in the NIST SD27 (Figures 6 (b)-(d)).

Figure 7 (Image only) shows the CMC curves for lights-out identification on ELFT-EFS database. The rank-1 identification rate is 65.10%. Figure 8 (Image only) shows the CMC curve for lights-out identification on the RS&A database. The rank-1 identification accuracy obtained on the RS&A database is 87.50%. This is much higher than the accuracy obtained on the NIST SD27 and ELFT-EFS databases because the latents in the RS&A database were collected in a laboratory and are comparatively of better quality.

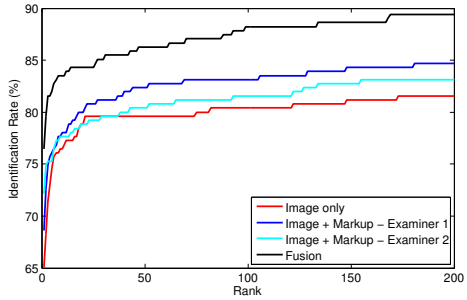


Figure 7: Identification performance (CMC curves) of the AFIS when (i) operating in lights-out mode (Image only), (ii) fed with an individual 1000 ppi markup (Image + Markup), and (iii) fusion of lights-out AFIS scores with the scores obtained using the two 1000 ppi markups (Fusion) for all 255 latents in the ELFT-EFS database against a reference database of 250K rolled prints.

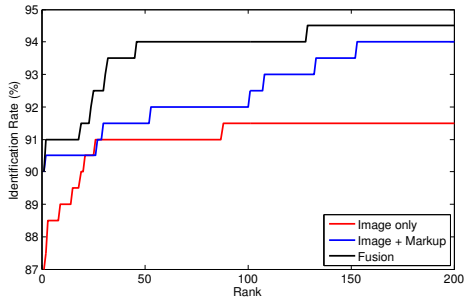


Figure 8: Identification Performance (CMC curves) of the AFIS when (i) operating in lights-out mode (Image only), (ii) fed with the single available markup (Image + Markup), and (iii) fusion of lights-out with examiner markup (Fusion) for the 200 latents in the RS&A database against a reference database of 250K rolled prints.

3.3.2 Matching Individual Examiner Markups

The Image plus Markup performance band in Figure 6 (a) indicates the identification accuracy of the AFIS on the NIST SD27 when fed with individual 500 ppi markups. The best rank-1 identification accuracy obtained using an individual markup is 66.67%. Note that the lights-out performance is within the performance band of the examiners. As expected, the identification accuracy is higher for good quality latents, compared to the bad and ugly quality latents (Figures 6 (b)-(d)).

Figure 7 shows the performance of the AFIS when fed with 1000 ppi markups available for the ELFT-EFS database. The best individual rank-1 identification accuracy obtained is 72.16%. On the RS&A database, on the other hand, the rank-1 identification accuracy obtained using the single available markup is 90% (Figure 8).

3.3.3 Fusing Multiple Examiner Markups

Since we have six different markups available for the NIST SD27 latents, we fuse the scores obtained using different markup combinations, and then compute the average accuracy of the AFIS when fed with different subsets of examiner markups. Several different score level fusion strategies were investigated. Simple sum fusion rule provided the best performance. No score normalization is necessary here since all the scores are being generated by the same AFIS. Table 3 shows that while identification performance of the AFIS improves with additional markups, there is a saturation after 3 or 4 markups per latent. For the NIST SD27 with 258 latents, each 1% improvement in performance, say at rank-1, corresponds to roughly two or three latents being promoted to rank-1.

Combination	Rank-1	Rank-50	Rank-100
One examiner	63.11	77.13	78.23
Two examiners	68.04	80.88	81.96
Three examiners	69.42	82.15	83.29
Four examiners	70.00	82.71	83.98
Five examiners	70.80	83.14	84.56
All six examiners	70.93	82.95	84.88

Table 3: Identification accuracy (%) of the AFIS, on average, on the NIST SD27 against 250K reference prints when fed with markups from different subsets of latent examiners.

3.3.4 Fusing lights-out AFIS with Multiple Markups

The CMC curves plotted in Figure 6 show that the rank-1 identification accuracy of the AFIS increases by 7.75% on the NIST SD27 by fusing the scores obtained using the six markups with the scores obtained from lights-out identification. On the other hand, a performance improvement of 11.37% is observed when fusing the scores obtained from the two individual markups for the ELFT-EFS database with the lights-out scores. Figures 9 and 10, respectively, show an example of a successful and failure case using fusion of the AFIS with the examiner markups.

For the RS&A database, although fusion of lights-out match scores with markup scores does not seem to benefit in terms of the rank-1 identification accuracy in comparison to only using the manual markup, significant performance improvement is observed for higher ranks (see Figure 8).

3.3.5 Determining the need for crowdsourcing

To measure the efficiency of the test based on order statistic for determining the need for crowdsourcing manual markup, we compute the (i) number of latents where markup is not needed and the mated print was not retrieved at rank-1, and (ii) number of latents where markup is ascertained but the mated print was retrieved at rank-1 (Table 4). The value of K used here is 200. For case (i) we found that the rank of the mated print did not decrease after fusion of lights-out with markup scores. This demonstrates the efficacy of the order statistic based test.

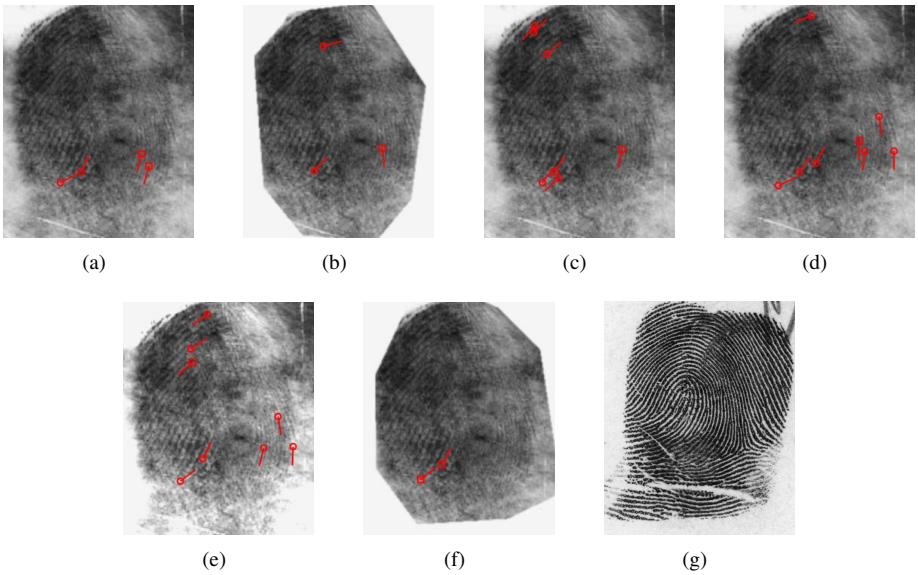


Figure 9: An example latent for which the mated reference print is retrieved at a higher rank after fusing the six crowdsourced markups with the AFIS. In the lights-out mode, the AFIS could not match the latent to the mated print shown in (g) (score=0). The rank of the mated print using the individual markups by the six examiners shown in (a)-(f) is 80, - (score=0), 45, 7, 57 and 12971, respectively. The mated print is retrieved at rank-2 using the combination of the AFIS with the six markups.

Significance level (α)	#Latents requiring markup			# Latents not requiring markup when mated reference print is not at rank-1			# Latents requiring markup when mated reference print is at rank-1		
	NIST SD27	ELFT-EFS	RSA	NIST SD27	ELFT-EFS	RSA	NIST SD27	ELFT-EFS	RSA
0.01	166	166	46	0	0	2*	74	74	22
0.05	151	151	35	0	0	2*	59	59	11
0.1	137	137	33	0	0	2*	45	45	9

*The mated reference prints are incorrectly labelled for these latents; does not impact the accuracy of the AFIS.

Table 4: Number of latents where markup is required, markup is not required when mated reference print is not at rank-1, and markup is required despite the mated reference print being retrieved at rank-1 for NIST SD27, ELFT-EFS, and RS&A databases. The number of latents in these three databases is 258, 255, and 200, respectively.

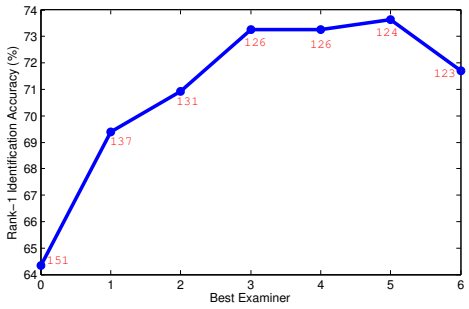


Figure 11: Identification accuracy of the AFIS using greedy crowdsourcing for the 258 NIST SD27 latents. Starting with best examiner, a significance level of 0.05 is used to decide if markup from the next best examiner is needed. Numbers of latents given to the next best examiner are indicated in red. Due to the preponderance of low quality prints in NIST SD27, the rank-1 identification accuracy tapers off after three examiner markups.

3.3.6 Greedy crowdsourcing

To test the benefit of using the greedy sequential strategy to dynamically determine the number of examiners required, we rated the individual examiners based on their skill set. This was estimated based on the AFIS performance obtained on the markups they provided. Figure 11 shows performance improvement when individual examiners are selected in decreasing order of their skill set. After fusing the three markups from the top three examiners, additional markups have negligible impact on the overall identification accuracy. Also, utilizing more number of markups does not necessarily improve the overall accuracy. Overall, 151 latents in the NIST SD27 require examiner markups based on the lights-out AFIS results (at a significance level of 0.05). 137, 131, 126, 126, 124, and 123 latents need markups after fusion of lights-out AFIS with best-1, best-2, best-3, best-4, best-5, and all six examiner markups, respectively.

4. Conclusions and Future Work

Matching poor quality latents to reference prints is one of the most challenging problems in fingerprint recognition. In order to match latents to reference prints with high accuracy, we propose a crowd powered latent identification paradigm which involves a symbiosis of human examiners with AFIS. Given a latent print, it is first compared against reference prints using an AFIS. Based

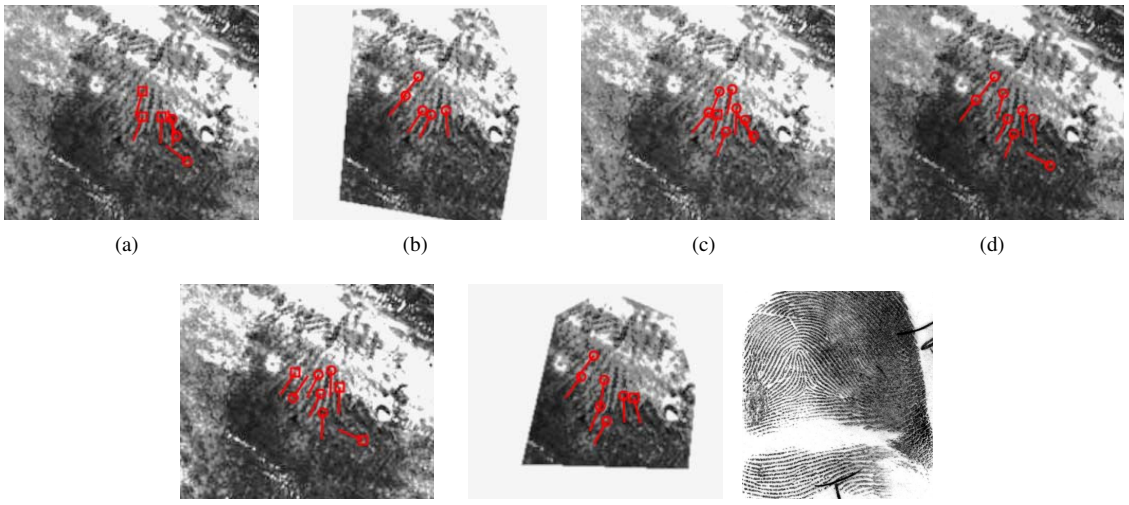


Figure 10: An example latent for which the mated reference print is retrieved at a lower rank after fusing the crowdsourced markups with the AFIS. In the lights-out mode, the AFIS retrieved the mated print shown in (g) at rank-1. The rank of the mated print in (g) using the individual markups by the six examiners shown in (a)-(f) is 54, 1171, 3426, 595, 22 and 8450, respectively. The mated print is retrieved at rank-26 using the combination of the AFIS with the six markups.

on the output of the lights-out match, an automatic decision is made to determine if manual feature markups from latent experts would be beneficial. If it is determined that additional markup would help, the latent print is crowdsourced to a pool of latent examiners. The manual feature markups are fed to the AFIS and the comparison scores from lights-out AFIS and those from manual markups input to AFIS are combined to boost the identification accuracy. Experimental results obtained on three different latent databases (NIST SD27, ELFT-EFS and RS&A), against a reference database of 250,000 rolled prints, demonstrate that a significant performance improvement can be obtained using the proposed crowd powered framework. It would be desirable to validate the proposed framework against a larger reference database of a few million reference prints. We also plan to explore ensemble-based meta algorithms such as bagging and boosting to improve the matching performance of AFIS.

References

[1] ELFT-EFS public challenge dataset. http://biometrics.nist.gov/cs_links/latent/elft-efs/ELFT-EFSPublicChallenge_2009-04-23.pdf.

[2] NIST SD27 latent database. http://www.nist.gov/srd/nistsd27.cfm.

[3] Ron Smith and Associates. http://www.ronsmithandassociates.com/.

[4] S. S. Arora, E. Liu, K. Cao, and A. K. Jain. Latent Fingerprint Matching: Performance Gain via Feedback from Exemplar Prints. *IEEE TPAMI*, 36(12):2452–2465, 2014.

[5] D. R. Ashbaugh and C. Press. *Quantitative-Qualitative Friction Ridge Analysis: An Introduction to Basic and Advanced Ridgeology*. CRC press, 1999.

[6] T. G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15. Springer, 2000.

[7] I. E. Dror, K. Wertheim, P. Fraser-Mackenzie, and J. Walajtys. The Impact of Human–Technology Cooperation and Distributed Cognition in Forensic Science: Biasing Effects of AFIS Contextual Information on Human Experts. *Journal of forensic sciences*, 57(2):343–352, 2012.

[8] L. K. Hansen and P. Salamon. Neural network ensembles. *IEEE TPAMI*, 12(10):993–1001, 1990.

[9] M. Indovina, V. Dvornychenko, R. Hicklin, and G. Kiebuzinski. ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation# 2]. *NISTIR*, 7859, 2012.

[10] M. Indovina, R. Hicklin, and G. Kiebuzinski. ELFT-EFS Evaluation of Latent Fingerprint Technologies: Extended Feature Sets [Evaluation# 1]. *NISTIR*, 7775, 2011.

[11] A. K. Jain and J. Feng. Latent Fingerprint Matching. *IEEE TPAMI*, 33(1):88–100, 2011.

[12] A. K. Jain and D. Zongker. Feature selection: Evaluation, application, and small sample performance. *IEEE TPAMI*, 19(2):153–158, 1997.

[13] D. Retelny, S. Robaszkiewicz, A. To, W. Lasecki, J. Patel, N. Rahmati, T. Doshi, M. Valentine, and M. S. Bernstein. Expert crowdsourcing with flash teams. In *ACM Symposium on UIST*, 2014.

[14] B. T. Ulery, R. A. Hicklin, J. Buscaglia, and M. A. Roberts. Repeatability and reproducibility of decisions by latent fingerprint examiners. *PloS one*, 7(3):e32800, 2012.

[15] S. Yoon, K. Cao, E. Liu, and A. Jain. LFIQ: Latent fingerprint image quality. In *IEEE BTAS*, pages 1–8, 2013.

[16] Z. Zhai, P. Sempolinski, D. Thain, G. Madey, D. Wei, and A. Kareem. Expert-citizen engineering: "crowdsourcing" skilled citizens. In *IEEE DASC*, 2011.