Improving Automated Latent Fingerprint Identification using Extended Minutia Types

Ram P. Krish\textsuperscript{a}, Julian Fierrez\textsuperscript{b,}\textsuperscript{*}, Daniel Ramos\textsuperscript{b}, Fernando Alonso-Fernandez\textsuperscript{c}, Josef Bigun\textsuperscript{c}

\textsuperscript{a}School of Electronic Engineering, Dublin City University, Ireland
\textsuperscript{b}School of Engineering, Universidad Autonoma de Madrid, Spain
\textsuperscript{c}School of Information Technology, Halmstad University, Sweden

Abstract

Latent fingerprints are usually processed with Automated Fingerprint Identification Systems (AFIS) by law enforcement agencies to narrow down possible suspects from a criminal database. AFIS do not commonly use all discriminatory features available in fingerprints but typically use only some types of features automatically extracted by a feature extraction algorithm. In this work, we explore ways to improve rank identification accuracies of AFIS when only a partial latent fingerprint is available. Towards solving this challenge, we propose a method that exploits extended fingerprint features (unusual/rare minutiae) not commonly considered in AFIS. This new method can be combined with any existing minutiae-based matcher. We first compute a similarity score based on least squares between latent and tenprint minutiae points, with rare minutiae features as reference points. Then the similarity score of the reference minutiae-based matcher at hand is modified based on a fitting error from the least square similarity stage. We use a realistic forensic fingerprint casework database in our experiments which contains rare minutiae features obtained from Guardia Civil, the Spanish law enforcement agency. Experiments are conducted using three minutiae-based matchers as a reference, namely: NIST-Bozorth3, VeriFinger-

\textsuperscript{*}Corresponding author

\textit{Email addresses:} ram.krish@dcu.ie (Ram P. Krish), julian.fierrez@uam.es (Julian Fierrez), daniel.ramos@uam.es (Daniel Ramos), feralo@hh.se (Fernando Alonso-Fernandez), josef.bigun@hh.se (Josef Bigun)
We report significant improvements in the rank identification accuracies when these minutiae matchers are augmented with our proposed algorithm based on rare minutiae features.

Keywords:
Latent Fingerprints, Forensics, Extended Feature Sets, Rare minutiae features

1. Introduction

Fingerprints left at a crime scene, referred to as latent prints, are the most common type of forensic science evidence and have been used in criminal investigations for more than 100 years, but comparing latent fingerprints is not an easy task. This is mainly attributed to the poor quality of the latent fingerprints obtained from the crime scenes. When a latent fingerprint is found, the criminal investigators first search for the suspect in their criminal database using an Automated Fingerprint Identification System (AFIS) to narrow down their manual work. If there is a match, then the individual is linked to the crime under investigation. Individualization (identification or match) is the decision yielded by a forensic examiner about the latent fingerprint belonging to a particular individual. This is the outcome of the Analysis, Comparison, Evaluation and Verification (ACE-V) \[1\] methodology currently followed in friction ridge examination.

In order to improve the matching efficiency, the concept of “Lights-Out System” was introduced for latent matching [2]. A Lights-Out System is a fully automatic identification process with no human intervention. Here, the system should automatically extract the features from the latent fingerprint and match it against the tenprints (exemplars) stored in the AFIS database to obtain a set of possible suspects with high degree of confidence. In general, latent fingerprints are partial in nature and are of varying quality (see Figure [1]), mostly distorted, smudgy, blurred, etc. These factors lead to high number of unreliable extracted features in fully automatic mode, and make it difficult for AFIS to
perform well. AFIS do not commonly use all the discriminatory features that could be derived from a fingerprint, mainly due to the limitations of automatic and reliable extraction of all types of discriminatory features. The accurate performance of feature extraction and matching algorithms of AFIS in forensic scenario is of great importance to avoid erroneous individualization.

Current practice in latent AFIS technology involves marking the latent fingerprint features manually by forensic examiners and then using both the latent fingerprint image and the manually marked features to search in the AFIS for a list of suspects. Sankaran et al. [3] reviewed existing automated latent fingerprint matching algorithms and their limitations. To avoid the burden of manual marking and with the hope of fully automating the latent AFIS in Lights-Out mode, the US National Institute of Standards & Technology (NIST) conducted a public evaluation of commercial AFIS performance in Lights-Out mode. This was a multi-phase open project called Evaluation of Latent Fingerprint Technologies (ELFT) [4]. The results of various phases of ELFT are summarized in Table 1. The reported accuracies from Phase-I and Phase-II of ELFT cannot be directly compared as the database and the quality of the latents were different. As one of the results of the ELFT initiative [5], it is shown that in some practical conditions manual intervention may be needed, as fully automated procedures

Figure 1: Subjective quality classification of latent fingerprint images in NIST Special Database 27 (NIST-SD27).
are not yet robust enough, especially when challenging latent fingerprints are considered. The procedures of marking the minutiae, determining the subjective quality of latents, etc. may need to be carried out manually.

<table>
<thead>
<tr>
<th>Phase of ELFT</th>
<th>Database size</th>
<th>Rank-1 accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase-I [6]</td>
<td>100 latents compared against 10,000 rolled prints</td>
<td>80.0%</td>
</tr>
<tr>
<td>Phase-II, Evaluation-1 [7]</td>
<td>835 latents compared against 100,000 rolled prints</td>
<td>97.2%</td>
</tr>
<tr>
<td>Phase-II, Evaluation-2 [8]</td>
<td>1,114 latents compared against 100,000 rolled prints</td>
<td>63.4%</td>
</tr>
</tbody>
</table>

Table 1: Summary of NIST Evaluation of Latent Fingerprint Technologies (ELFT) results.

AFIS commonly use only a limited number of features automatically extracted from the fingerprints using a feature extraction algorithm. On the other hand, forensic examiners use a richer set of features during their manual comparisons. This could be a possible reason why manual comparisons outperform AFIS comparisons [9]. Any features that are not commonly used by commercial AFIS are generally termed as Extended Feature Sets (EFS) [7]. The use of EFS by forensic examiners in manual comparisons is much debated, mainly due to non-repeatability by another examiner to validate the previous decision.

Two major problems in friction ridge analysis that raised the attention of the forensic community about 10 years ago were identified as follows in [10]:

1. Latent AFIS searches are commonly limited by an over simplified feature set.
2. During the latent examiner comparison, there are no standard format to document the features used in comparison decision. This leads to problems with future reference or interchange with other forensic examiners.
The SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study, and Technology) drafted a memo to NIST noting that forensic examiners use features that are not currently addressed in fingerprint standards. The ANSI/NIST Standard Workshop II charted the Committee to Define an Extended Finger-print Feature Set (CDEFFS). The CDEFFS included 45 members from various federal agencies, the forensic community, AFIS vendors, and academia [10]. The purpose of CDEFFS was to define a standard to completely represent the distinctive information in the fingerprint which are quantifiable, repeatable and develop a clear method of characterizing information for: 1) AFIS searches initiated by forensic examiner, and 2) forensic examiner markup and exchange of latent fingerprints.

Fingerprint features are categorized into three levels as well as a feature category called “other” to be used for friction ridge examination. Level-One considers general overall direction of the ridge flow. Level-Two describes the path of specific ridges. Level-Three are the shapes of the ridge structure. “Other” features describe temporary features or imperfections in ridges [11]. Some of the extended fingerprint features defined by CDEFFS under each of the three level categories [10] [12] are summarized in Table 2. Figure 2 shows some extended features and typical minutia features (ridge-endings and bifurcations) in an exemplar fingerprint from NIST Special Database 27 (NIST-SD27).

In this work, we propose a method to improve the identification accuracy of minutiae-based matchers for partial latent fingerprints by incorporating reliably extracted rare minutia features. Most minutiae-based fingerprint matchers use only two prominent ridge characteristics namely ridge-endings and bifurcations. We propose an algorithm that will modify the similarity scores of minutiae-based matchers based on the presence of rare minutia features like fragments, enclosures, dots, interruptions, etc. The decision for a match or non-match is automatically estimated based on least squares fitting of an affine transformation between the latent minutiae set and the tenprint minutiae set.

We show a significant improvement in the overall rank identification accuracies for three minutiae-based matchers (NIST-Bozorth3, VeriFinger-SDK and
Type of category | Extended feature set
---|---
Level-One Details | Ridge flow map, local ridge quality, pattern classification (whorl, arch, tentarch, left/right loop etc), singular points (core, delta), core-delta ridge count.
Level-Two Details | Minutiae-ridge relationship, ridge curvature, feature relationship, unusual/rare minutiae, scars, creases, incipient ridges, dots.
Level-Three Details | Pores, edge shapes, ridge/furrow width.

Table 2: Extended features defined by CDEFFS categorized into respective fingerprint feature details (not a comprehensive list).

MCC-SDK) when their similarity scores are modified using our proposed algorithm which incorporates rare minutia features. A preliminary version of this work [13] modifies the scores based on the probability of occurrence of rare minutiae features. Here, we propose a method which avoids dependencies on such probability of occurrence as they may vary depending on the size of the database.

The main contributions of this work are as follows:

1. A methodology to adapt any minutiae-based matcher by incorporating information from rare features.
2. A specific algorithm to align the latent minutiae pattern and the tenprint minutiae pattern using rare minutiae.
3. Experimental demonstration of the performance improvement of minutiae-based matchers when incorporating information from rare features.
4. We finally present also various population statistics about rare minutia features present in a realistic forensic casework database obtained from Spanish law enforcement agency (Guardia Civil).

In the following sections, we review related works in the individualization of
fingerprints, and describe: the database and statistics of rare minutia features, the proposed algorithm to modify the similarity scores based on rare features, experiments, results and conclusions.

2. Related works

To use EFS in automated systems, reliable feature extraction algorithms are mandatory. Many law enforcement agencies follow a 500 ppi scanning resolution for fingerprint images to be used in AFIS. With such a resolution, it is difficult to reliably extract many of the extended features automatically. Due to advances in fingerprint scanning technologies, SWGFAST during the ANSI/NIST Fingerprint Standard Update Workshop in 2005 proposed 1000 ppi as the minimum scanning resolution for fingerprint images. This proposal was hugely supported by the forensic community. To test the feasibility of including EFS in latent AFIS, NIST conducted another multi-phase commercial algorithm evaluation called Evaluation of Latent Fingerprint Technologies - Extended Feature Sets.
Table 3: Rank-1 identification accuracy of NIST Evaluation of Latent Fingerprint Technologies - Extended Feature Sets (ELFT-EFS).

<table>
<thead>
<tr>
<th>ELFT-EFS</th>
<th>Database size</th>
<th>Rank-1 accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation-1 [14]</td>
<td>1,114 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints</td>
<td>66.7%</td>
</tr>
<tr>
<td>Evaluation-2 [15] [16]</td>
<td>1,066 latents compared against 1,000,000 rolled prints and 1,000,000 plain prints</td>
<td>71.4%</td>
</tr>
</tbody>
</table>

ELFT-EFS was conducted in a “Semi Lights-Out” mode, unlike the “Lights-Out” mode for ELFT. The main purpose of ELFT-EFS was to determine the effectiveness of forensic examiner marked latent fingerprint features on the latent identification accuracy. NIST conducted two evaluations for ELFT-EFS and the best achieved Rank-1 identification accuracy for each of the evaluations is summarized in Table 3.

As in ELFT, the results of different evaluations in ELFT-EFS cannot be directly compared because the database used were not exactly the same [15] [14]. In [15], it is reported that though the highest measured accuracy achieved by an individual matcher at Rank-1 was 71.4%, and approximately 82% of the latents were correctly matched at Rank-1 when multiple matchers were combined. This corroborates the potential for additional accuracy improvement when combining multiple algorithms [17]. Nevertheless, these NIST evaluations show that the performance of state of the art latent fingerprint recognition technologies are not satisfactory.

An extensive study on extended fingerprint feature sets is reported by Jain [9]. This includes several extended features from Level-One, Level-Two and Level-Three. It was concluded in [9] that manual intervention is strongly recommended.
while using EFS, as well as extended features from Level-One and Level-Two are highly recommended to be incorporated in latent AFIS. Extended features such as ridge flow map, ridge wavelength map, ridge quality map, and ridge skeleton have shown significant improvements in latent identification accuracies. Level-One and Level-Two details used in [9] [18] are insensitive to image quality, and do not rely on high resolution images. To incorporate Level-Three EFS such as pores, dots, incipients, etc, it is essential to improve the quality of enrolled fingerprints.

The use of pores as extended features was studied in high resolution 1000 ppi images by Zhao et al. [19] and Jain et al. [20]. Dots and incipients were studied by Chen et al. [21]. Out of pores, dots and incipients, pores resulted in better performance [19]. Even though high resolution 1000 ppi images were used, live scan images resulted in easy detection of pores automatically, which was not the case with inked fingerprint images. Pore extraction based on skeletonized and binary images was studied by Stosz et al. [22, 23] and Kryszczuk et al. [24]. These techniques were demonstrated effective only on very good quality high resolution fingerprint images scanned approximately at 2000 ppi [22]. These methods were more sensitive to noise, and the performance degrades for poor quality of fingerprint images and low resolution images.

A local image quality based method applied on extended fingerprint features for high resolution 1000 ppi fingerprint images was reported by Vatsa et al. [25]. A fast curve evolution algorithm was used to quickly extract extended features such as pores, ridge contours, dots and incipient ridges. Together with other Level-One and Level-Two details as proposed in Jain et al. [18], these extended features were used to generate a quality-based likelihood ratio to improve the identification performance.

Score level fusion of different algorithms using various extended fingerprint features was reported by Fierrez et al. [26]. Features like singular points, ridge skeleton, ridge counts, ridge flow map, ridge wavelength map, texture measures were studied by analyzing the correlation between them using feature subset-selection techniques. Combination of features showed significant improvement
in the performance of the system.

When only partial fingerprints are available, pre-alignment of partial minutiae set and full minutiae set based on orientation fields of respective fingerprints helps in reducing the minutiae search space of full fingerprint relative to the size of partial fingerprint. Such reduction in the size of minutiae search space has been shown to improve the performance by Krish et al. [27], [28], [29]. This approach has shown significant improvement in the system performance especially for poor quality latent fingerprints.

Si et al. [30] combined local and global approaches for minutiae matching. Their proposed method estimates a dense deformation field between two fingerprints to remove the negative impact of distortion. The dense deformation field aligns not only minutiae but also ridges. By fusing minutiae and image correlation, they improved the matching performance significantly.

Cao and Jain [31] proposed the use of minutiae descriptors that are learned via a ConvNet together with minutiae and texture information to improve the latent fingerprint recognition. They performed score level fusion of their proposed algorithm with commercial minutiae-based matchers to improve the rank identification accuracies. We followed a similar methodology to show the importance of rare minutiae features in improving the rank identification accuracies of minutiae-based matchers.

3. Database

Similar to the related works discussed in Sect. 1, we first tried to use SD27 data from NIST in our experiments. Regretfully, the public distribution of SD27 data from NIST is now discontinued. Additionally, copies already distributed of SD27 lack ground truth information of the existing rare features, therefore we decided to generate and make public a new dataset similar to SD27 but in this case including ground truth information generated by forensic experts.

\[1\] The dataset is available here: [https://atvs.ii.uam.es/atvs/gcdb_features.html](https://atvs.ii.uam.es/atvs/gcdb_features.html)
Figure 3: Minutia types used by Guardia Civil. Names corresponding to individual minutia type numbers can be found in Table 4.

The database used in this work was obtained from Guardia Civil, the Spanish law enforcement agency. The Guardia Civil database (GCDB) is a realistic forensic fingerprint casework database. Apart from having typical minutiae feature types (ridge-endings, bifurcations), GCDB also comprises rare minutiae types like fragments, enclosures, dots, interruptions, etc [32]. A comprehensive list of rare minutia features used by Guardia Civil are shown in Figure 3 and the corresponding minutiae type names are listed in Table 4.

GCDB used in this work consists of 268 latent and tenprint (exemplar) pairs of fingerprint images (scanned at 500 ppi) and minutia sets. All the minutiae in the latent fingerprint images were manually extracted by forensic examiners of Guardia Civil. The corresponding mated minutiae in the tenprints were also manually established. This includes the typical (ridge-endings and bifurcations) minutiae and the rare minutiae. These are called matched minutia set, i.e, the minutiae sets for which a one-to-one correspondence is established between the latent and the mated tenprint. Here, the number of minutiae in the latent and the corresponding mated tenprint minutiae set are the same.

To generate an ideal minutiae set (i.e, all possible minutiae) for the tenprint, we used the minutiae extractor module from VeriFinger-SDK [33]. We performed a Gabor filtering based global post-processing to remove any spurious minutiae that are outside the foreground. In particular, Gabor filtering is used to obtain the Region of Interest (RoI) similarly to [34]. The spurious minutiae generated by VeriFinger outside the RoI are then eliminated. Veri-
iFinger extracts only the typical minutiae features from the fingerprint image. We then added the rare minutiae from the GCDB tenprint minutiae set into the post-processed VeriFinger generated minutiae set for the tenprints. In this case, the number of minutiae between the latent and the corresponding mated tenprint minutiae set are not equal, the latent minutiae set is only a subset of the tenprint minutiae set. The average number of minutiae in the latents was 13 and that of tenprints was 125. There exists automated algorithms to perform segmentation for latent fingerprints \[35\], but in this work, we relied on the manually extracted minutiae alone.

The original latent minutia sets provided by Guardia Civil and the post-processed VeriFinger generated minutia sets are used in all our experiments. To represent some rare minutiae, multiple points were needed. For example, to represent a deviation two points are needed (see type 3 in Figure 3), and to represent an assemble three points are needed (see type 13 in Figure 3). Whenever multiple points are needed to represent a rare minutia, we mapped them to a single point representation by taking the average of locations of all points, and minimum orientation among all the orientations.

From the 268 latent fingerprint minutia sets, we estimated the probability of occurrence \((p_i)\) of various minutia types. The probability \((p_i)\) for each minutia type present in GCDB is listed in Table 5. In the 268 latent fingerprints of GCDB, we noticed only seven types of rare minutia features. They are listed in

<table>
<thead>
<tr>
<th>No</th>
<th>Minutiae type</th>
<th>No</th>
<th>Minutiae type</th>
<th>No</th>
<th>Minutiae type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ridge Ending</td>
<td>6</td>
<td>Interruption</td>
<td>11</td>
<td>Circle</td>
</tr>
<tr>
<td>2</td>
<td>Bifurcation</td>
<td>7</td>
<td>Enclosure</td>
<td>12</td>
<td>Delta</td>
</tr>
<tr>
<td>3</td>
<td>Deviation</td>
<td>8</td>
<td>Point</td>
<td>13</td>
<td>Assemble</td>
</tr>
<tr>
<td>4</td>
<td>Bridge</td>
<td>9</td>
<td>Ridge Crossing</td>
<td>14</td>
<td>M-structure</td>
</tr>
<tr>
<td>5</td>
<td>Fragment</td>
<td>10</td>
<td>Transversal</td>
<td>15</td>
<td>Return</td>
</tr>
</tbody>
</table>

Table 4: List of minutia types used by Guardia Civil. Numbering with respect to Figure 3.
<table>
<thead>
<tr>
<th>No</th>
<th>Minutiae Type</th>
<th>Probability ($p_i$)</th>
<th>Number of occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ridge-ending</td>
<td>0.5634</td>
<td>1902</td>
</tr>
<tr>
<td>2</td>
<td>Bifurcation</td>
<td>0.3620</td>
<td>1222</td>
</tr>
<tr>
<td>3</td>
<td>Deviation</td>
<td>0.0015</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>Bridge</td>
<td>0.0024</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>Fragment</td>
<td>0.0444</td>
<td>150</td>
</tr>
<tr>
<td>6</td>
<td>Interruption</td>
<td>0.0021</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>Enclosure</td>
<td>0.0204</td>
<td>69</td>
</tr>
<tr>
<td>8</td>
<td>Point</td>
<td>0.0036</td>
<td>12</td>
</tr>
<tr>
<td>10</td>
<td>Transversal</td>
<td>0.0003</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5: The probability of occurrence of minutia types present in the 268 latent fingerprints of GCDB (total number of minutiae observed = 3376). The numbers correspond to minutia types in Figure 3.

Other rare minutia types are not found in the current database used in this study. This is particularly due to the fact that we are dealing with highly partial latent fingerprints, with an average of 13 minutiae per latent.

In related works, minutiae frequency for 20 different minutiae types were reported for full fingerprints obtained from a population of 200 Spanish individuals \[36\] and 278 Argentinian individuals \[37\]. The statistics of the various minutiae types obtained in those works (for ca. 2,000 different fingerprints, more than 80,000 minutiae) are similar in many cases to the statistics obtained in the partial latent fingerprint database presented in the present work if we compare corresponding types of features, especially for the most frequent minutiae types. For the comparison, please note that the naming conventions differ among works. Here we follow the naming conventions of Guardia Civil, e.g.
Type 6 in Fig. 3 is called “ Interruption”, whereas this kind of feature is called “ Break (BR)” in works [36] and [37]. Comparing for example our statistics from Table 5 to Table 1 of [36], and using approximated values, our Type 1 is 56% vs Type E in [36] is 50%, our Type 2 is 36% vs Type B+C is 40%, our Type 7 is 2% vs Type ENBG+ENSM is 2%, etc. A few other types, specially the rarest, have larger differences in their statistics, most probably because of the sample size, e.g., our Type 3 is ca. 0.1% vs Type O in [36] is 0.7%.

Given our limited sample size, please note that the statistics from Table 5 are not pretending to be representative of large-scale populations, especially for the rarest features. It would be necessary a really large dataset to generate statistically significant frequencies for all the reported types, which is out of the scope of the present paper.

4. Algorithm

We assume in our algorithm that the rare features won’t be always repeatable and won’t be always labelled uniformly over multiple captures (either manually or automatically), e.g., due to acquisition variations or other image quality factors [38, 39]. Similarly to minutiae matchers able to cope with variations between matching fingerprint images, our algorithm is also able to cope with such intra-variability. A comprehensive study of such variability factors is out of the scope of the present work. Here we only focus in analyzing to what extent those rare minutiae can improve existing AFIS, using realistic data in which such intra-variations between multiple captures are naturally present. The latent fingerprints of GCDB are highly partial in nature, with an average of 13 minutiae per latent. To make an appropriate alignment between the latent minutia points and the tenprint minutia points (with an average of 125 minutia points) requires a reliable reference point. We choose the rare minutia features as reference points to perform the alignment.

Let $L$ and $M$ be the representation of latent and tenprint minutia sets respectively. Each minutia is represented as a quadruple $m = \{x, y, \theta, t\}$ that
indicates the \((x, y)\) location as coordinates, the minutia angle \(\theta\) and the minutia type \(t\):

\[
L = [m_1 \ m_2 \ ... \ m_p], \quad m_i = [x_i \ y_i \ \theta_i \ \ t_i]^T, \quad i = 1...p
\]

\[
M = [m'_1 \ m'_2 \ ... \ m'_q], \quad m'_j = [x'_j \ y'_j \ \theta'_j \ \ t'_j]^T, \quad j = 1...q,
\]

where \(p\) and \(q\) are the number of minutiae in \(L\) and \(M\) respectively. If \(t > 2\), then the minutia is of rare type (from Table \([4]\), and \([\cdot ]^T\) denotes transpose.

The algorithm to generate weighted similarity scores from a minutiae matcher is described in two stages. Similarity scores of minutiae matchers are modified only if they contain rare minutia features.

The first stage of the algorithm estimates the least square fitting error for an affine transformation of the latent minutiae set onto a tenprint minutiae set. The second stage of the algorithm modifies the similarity score generated by the minutiae-based matcher based on the fitting error. Other works related with modifying the similarity score based on pre-alignment are reported in \([40, 27, 41]\). The sequence of steps involved in generating the modified score of the minutiae matcher using our proposed algorithm is summarized in Figure \([4]\).

**Stage-1 : Least Square Fitting Error**

**Step 1:** To find the affine transformation between \(L\) and \(M\), it is first needed to establish a one-to-one correspondence between minutiae from \(L\) and minutiae from \(M\). Let the subset of minutiae from \(M\) which establishes correspondence with \(L\) be denoted as \(M_s\).

**Step 2:** Superimpose one rare minutia point of \(L\) onto the corresponding rare minutia point of \(M\), only if they both are of the same type (if there are multiple rare minutia points, take any). This step compensates the translation between latent and tenprint. If the type of the rare minutia between \(L\) and \(M\) differs, or \(M\) does not contain any rare minutiae, then the similarity score generated by our algorithm is fixed to a value 0.25 inside a range \((0,1)\). That similarity score will then be combined with the score provided by the AFIS system (see Figure \([4]\)). Note that we align based on rare minutiae of the same
Figure 4: Sequence of steps in estimating the modified similarity score of a reference minutiae-based matcher. Note that the AFIS matcher that we want to improve runs in parallel and independently of our proposed algorithm based on rare features. Our proposed algorithm generates a new score that is fused to the standard output score provided by the AFIS matcher.

type between latent and tenprint. Provided that rare minutia are more unique than standard minutia and that the baseline AFIS that we are trying to improve most probably won’t be exploiting such specific rare types, we expect that our combination approach shown in Figure 4 will fuse complementary information resulting in improved performance.

Step 3: To establish the correspondence between latent and tenprint minutia points, we choose the minutia points from $M$ that are close to the minutia points of $L$. The Euclidean distance is calculated between the minutia pairs to determine whether the pairs are close or not.

Step 4: To compensate for rotation alignment, we rotate the latent in the range $[-45^\circ, +45^\circ]$ with respect to the superimposed rare minutiae, and estimate the Euclidean distance for each rotation step of size $1^\circ$.

Step 5: The optimal rotation is the one for which the average sum of distances between closest pairs is minimum.
**Step 6:** After the alignment, all those minutia pairs which are within a threshold distance are considered to be mated pairs, and a one-to-one correspondence is established between them. As a result, we obtain a subset $M_s$ of the tenprint minutiae $M$. After establishing the correspondence, the number of minutiae between $L$ and $M_s$ are the same.

**Step 7:** Once the correspondence is established, we find the least square fitting error for the affine transformation between the latent minutia points and the subset of tenprint minutiae set.

For $\hat{L}$ and $\hat{M}_s$, which are the modified version of $L$ and $M_s$ with only the $(x, y)$ locations as minutia representation augmented with a value 1, i.e.,:

$$\hat{L} = [\hat{m}_1 \hat{m}_2 ... \hat{m}_p]; \quad \hat{m}_i = [x_i \ y_i \ 1]^T; \quad i = 1...p$$

$$\hat{M}_s = [\hat{m}'_1 \hat{m}'_2 ... \hat{m}'_p]; \quad \hat{m}'_j = [x'_j \ y'_j \ 1]^T; \quad j = 1...p,$$

we are looking for some affine transformation matrix

$$A = [a_{jk}]_{j,k=1...3} \tag{1}$$

and some translation vector

$$\tau = [\tau_1 \ \tau_2 ... \ \tau_p]; \quad \tau_1 = \tau_2 = ... = \tau_p = [\delta_x \ \delta_y \ 1]^T; \tag{2}$$

such that

$$\hat{M}_s \approx A\hat{L} + \tau \tag{3}$$

where $[\delta_x \ \delta_y]$ is the translation needed to superimpose the rare minutia of latent minutia set $L$ and tenprint minutia set $M$.

**Step 8:** Find the least square fitting error between $\hat{L}$ and $\hat{M}_s$ defined as follows:

$$E^{L,\hat{M}_s} = \frac{1}{p} \sum_{i=1}^{p} ||\hat{m}'_i - A\hat{m}_i - \tau_i||_2^2 \tag{4}$$

where $||\hat{m}'_i - A\hat{m}_i - \tau_i||_2$ is the $L_2$ norm.
For a match comparison, we expect this fitting error to be small, whereas for a non-match comparison, the fitting error is expected to be large.

If there are multiple matching rare minutiae feature between \( L \) and \( M_s \), then \( \hat{E}^{L,M_s} \) is calculated for all such minutiae types. The fitting error for such a comparison is chosen to be the minimum of all the fitting errors calculated.

**Stage-2: Modified scores**

**Step 9:** Using a standard minutia matcher, generate the similarity score \( S_m \) between \( L \) and \( M \). Assuming that the similarity score values generated by the minutiae matcher is normalized in range \([0, 1]\). If by default the minutiae-based matcher do not generate normalized scores, then they are explicitly normalized.

The fitting error \( \hat{E}^{L,M_s} \) is a dissimilarity score, i.e., lower the fitting error, comparison is more similar. The fitting error is normalized in the range \([0, 1]\) and then they are converted into a similarity score \( \hat{E}^{L,M_s} \) as follows:

\[
\hat{E}^{L,M_s} = 1 - \hat{E}^{L,M_s}
\]

(5)

\( S_m \) and \( \hat{E}^{L,M_s} \) are similarity scores in the range \([0, 1]\). First level of score modification is done by score fusion of \( S_m \) and \( \hat{E}^{L,M_s} \) based on mean rule to obtain \( S'_m \) as follows:

\[
S'_m = \frac{S_m + \hat{E}^{L,M_s}}{2}
\]

(6)

**Step 10:** Further to the score modification based on fusion, similarity score \( S'_m \) is again modified based on a fitting error threshold \( E_t \) to obtain \( S''_m \) as follows:

\[
S''_m = \begin{cases} 
S'_m \times \alpha & \text{if } \hat{E}^{L,M_s} > E_t, \\
S'_m \times \beta & \text{otherwise.}
\end{cases}
\]

(7)

where \( \alpha \) and \( \beta \) are constants used to reward and penalize the fused score \( S'_m \) respectively.
If $\hat{E} \hat{L}, \hat{M}_s > E_t$, then the comparison is deemed to be a match, and the fused score is further rewarded by multiplying with a constant $\alpha$ to obtain $S''_m$. If $\hat{E} \hat{L}, \hat{M}_s \leq E_t$, the comparison is deemed to be a non-match, and the fused score $S'_m$ is penalized by multiplying it with constant $\beta$ to obtain $S''_m$.

Thus, we obtain a modified similarity scores $S''_m$ for a particular minutiae matcher by incorporating information from rare minutia features based on the fitting error obtained using our approach.

5. Experiments

GCDB consists of 268 latent and corresponding 268 mated tenprint images and minutiae. Only 151 of them contains rare minutiae in the minutiae set, and the remaining consisted only of bifurcations and ridge-endings. We performed all our experiments on the 151 latent and tenprint minutiae set to establish the importance of rare minutiae using our proposed algorithm in improving the rank identification accuracy of minutiae-based matchers which uses only typical minutiae. To generate similarity scores based on typical minutia features, we used three minutiae matchers namely: NIST-Bozorth3 [42], VeriFinger-SDK [33] and MCC-SDK [43, 44, 45]. When reporting the rank identification accuracies in our experiments, there are 151 match comparisons and $151 \times 150$ non-match comparisons.

NIST-Bozorth3 is a minutiae based fingerprint matcher that is specially developed to deal with latent fingerprints and is publicly available. This matcher is part of the NIST Biometric Image Software (NBIS) [42], developed by NIST. VeriFinger is a commercial SDK that is widely used in academic research. MCC-SDK is a well known minutiae matcher made available for research purposes. Both NIST-Bozorth3 and MCC-SDK are publicly available minutiae matchers, whereas VeriFinger is not. We report the performance accuracy and improvement of all the matchers using Cumulative Match Characteristic (CMC) curves.
5.1. Experiment 1: Fitting Error probability distribution

The least square fitting error probability density estimates for both match and non-match comparisons are shown in Figure 5. We can observe that the fitting error itself is discriminatory enough, having separate peaks for both match and non-match distributions. This supports the methodology followed in our algorithm. The following experiments also support this fact.

5.2. Experiment 2: Score Fusion

The fitting error is a similarity score in the range $[0, 1]$ from Eq. (5). The similarity scores obtained from minutiae-based matchers are normalized in the range $[0, 1]$. These scores are fused based on sum rule (Step 9 of Algorithm).

Figures 6(a), 6(b) and 6(c) show the CMC curve before and after score fusion for minutiae-based matchers NIST-Bozorth3, VeriFinger-SDK and MCC-SDK respectively. Typical minutiae-based similarity scores were very poor for NIST-Bozorth3 and VeriFinger-SDK at Rank-1 identification and beyond. NIST-Bozorth3 achieved only 34.44% Rank-1 identification accuracy when only typical minutiae (ridge-endings and bifurcations) were used. VeriFinger-SDK achieved 31.13% Rank-1 identification accuracy under similar configuration.
The performance of MCC-SDK is far better than other two matchers, and achieved 78.15% Rank-1 identification accuracy when only using typical minutiae. In general, MCC-SDK is one of the best performing local structure based minutiae matcher.

The similarity score $S_m$ of the minutiae-based matcher is modified by fusing
with the fitting error $\tilde{E}_{L,M_s}$ using mean rule to obtain $S'_m$ (Eq. (6)). This score fusion significantly improves the rank identification accuracies for all minutiae-based matchers. The Rank-1 identification improved from 34.44% to 51.00% for NIST-Bozorth3, 31.13% to 62.25% for VeriFinger-SDK, and 78.15% to 92.72% for MCC-SDK. The improvement obtained at this stage is significant and consistent at all the ranks (Figures 6(a), 6(b) and 6(c)).

Table 6 summarizes the Rank-1 accuracy for NIST-Bozorth3, VeriFinger-SDK and MCC-SDK before and after score fusion.

Table 6: Rank-1 identification (in %) for NIST-Bozorth3, VeriFinger-SDK and MCC-SDK before and after score fusion.

<table>
<thead>
<tr>
<th>Matcher</th>
<th>Before score fusion (Rank-1)</th>
<th>After score fusion (Rank-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST-Bozorth3</td>
<td>34.44</td>
<td>51.00</td>
</tr>
<tr>
<td>VeriFinger-SDK</td>
<td>31.13</td>
<td>62.25</td>
</tr>
<tr>
<td>MCC-SDK</td>
<td>78.15</td>
<td>92.72</td>
</tr>
</tbody>
</table>

5.3. Experiment 3: Score modification based on fitting error threshold

The fitting error threshold $E_t$ plays a crucial factor in our proposed algorithm. Based on the fitting error, a given comparison can be concluded whether it is a match comparison or a non-match comparison. If the fitting error $\tilde{E}_{L,M_s}$ is greater than a threshold $E_t$, then the comparison is concluded a match comparison. This is because $\tilde{E}_{L,M_s}$ is also viewed as a similarity measure. When the comparison is deemed to be a match comparison based on threshold $E_t$, then the fused score $S'_m$ is rewarded by multiplying with a constant $\alpha$. Similarly, when the comparison is deemed as non-match, then fused score $S'_m$ is penalized by multiplying with a constant $\beta$. In our experiments, we empirically chose $\alpha = 2$ and $\beta = 1$ (Step 10 of Algorithm, Eq.(7)).

Figures 7(a), 7(c) and 7(e) show Rank-1 identification accuracies (Y-axis) obtained for various fitting error thresholds (X-axis) for NIST-Bozorth3, VeriFinger-
Table 7: Rank-1 identification (in %) for NIST-Bozorth3, VeriFinger-SDK and MCC-SDK when the fused scores are modified based on fitting error threshold. Results are reported on the subset of mated latent-tenprint pairs where rare features are found (151 out of 268 pairs).

<table>
<thead>
<tr>
<th>Matcher</th>
<th>Before score fusion (Rank-1)</th>
<th>After score fusion (Rank-1)</th>
<th>Threshold based modification (Rank-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST-Bozorth3</td>
<td>34.44</td>
<td>51.00</td>
<td>74.83</td>
</tr>
<tr>
<td>VeriFinger-SDK</td>
<td>31.13</td>
<td>62.25</td>
<td>77.48</td>
</tr>
<tr>
<td>MCC-SDK</td>
<td>78.15</td>
<td>92.72</td>
<td>96.03</td>
</tr>
</tbody>
</table>

SDK and MCC-SDK respectively. The threshold is varied in the range 0.8 to 1.0, as most of the fitting error similarity values are concentrated in this range (see Figure 5). The best fitting error threshold for each of the matchers are obtained heuristically. Same database is used for both this heuristic estimation as well reporting the performance improvement of the systems. This is particularly due to the objective of looking for best possible performance the system can achieve as compared to only score fusion. Consequently, this also helps to understand the discriminating capability of the fitting error, supporting Figure 5.

For NIST-Bozorth3, we notice that for a threshold value of 0.92 (see Figure 7(a)), the system achieves 74.83% Rank-1 identification accuracy. Using our proposed algorithm, we were able to significantly improve the Rank-1 identification accuracy of NIST-Bozorth3 from 34.44% to 74.83%. Figure 7(b) show the CMC curve for NIST-Bozorth3 when the fused scores are modified based on the fitting error threshold. We observe a significant and consistent improvement of rank identification accuracies for NIST-Bozorth3.

Figure 7(d) show the CMC curve based on optimal threshold for VeriFinger-SDK. For the optimal threshold value of 0.93 (see Figure 7(c)), VeriFinger-SDK achieved Rank-1 identification accuracy of 77.48% which is a significant improvement from 31.13%. We also observe that, using our proposed algorithm, we achieve significant and consistent improvement in rank identification accuracy for VeriFinger-SDK when the rare minutiae information is incorporated.

Similarly, Figure 7(f) show the best CMC curve for MCC-SDK. For an op-
Figure 7: CMC curve showing consistent improvement in the rank identification accuracies for NIST-Bozorth3, VeriFinger-SDK and MCC-SDK when the fused scores are further modified based on fitting error threshold. Left column show the parameter optimization by looking for optimal threshold. Right column show the improvement in rank identification after applying optimal threshold.
Table 8: Rank-1 identification (in %) for the 3 considered baseline fingerprint matchers before and after applying the proposed improvement based on rare features, on the full set of 268 latent-tenprint pairs included in GCDB database.

<table>
<thead>
<tr>
<th>Matcher</th>
<th>Baseline AFIS (Rank-1)</th>
<th>Improved Matcher (Rank-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIST-Bozorth3</td>
<td>25.37</td>
<td>36.94</td>
</tr>
<tr>
<td>VeriFinger-SDK</td>
<td>31.72</td>
<td>59.33</td>
</tr>
<tr>
<td>MCC-SDK</td>
<td>80.60</td>
<td>89.93</td>
</tr>
</tbody>
</table>

timal threshold value of 0.95 (see Figure 7(e)), MCC-SDK achieved the best Rank-1 identification of 96.03%, as well as consistent improvement in rank identification using our proposed algorithm. Table 7 summarizes the Rank-1 identification for all the minutiae-based matchers used in our experiments on the subset of mated latent-tenprint pairs where rare features are found (151 out of 268 pairs). Table 8 finally shows the performance over the full set of 268 latent-tenprint pairs, in which 151 out of them (the set that includes rare features) are benefited from our proposed algorithm as indicated in Table 7, and the remaining 117 are processed only with the baseline AFIS under consideration (i.e., the similarity score $S_m$ in Figure 4 is the only one used).

6. Conclusions

We discussed the challenges faced by latent fingerprint identification. One of the crucial challenges faced by AFIS is on how to improve the rank identification accuracies when only partial fingerprints are available. We proposed an algorithm that incorporates information from reliably extracted rare minutiae features to improve the rank identification accuracies for minutiae matchers.

The usefulness of the proposed algorithm is demonstrated on three widely used minutiae-based matchers, NIST-Bozorth3, VeriFinger-SDK and MCC-SDK. All the three matchers showed significant improvement in the rank identification accuracies when their similarity scores were modified based on the fitting error proposed in our methodology. We conclude that even if we have only few
number of minutiae in a partial latent, presence of reliably extracted rare minutia features makes the comparison more robust. In our experiments, we used the rare minutia features that were manually extracted by forensic examiners. Developing more robust automatic extraction of rare minutiae can significantly improve the current state of the art in AFIS adapted for latent fingerprints. Also, future work may exploit the differences between minutiae types to further improve the performance of standard AFIS, e.g., by incorporating weighting factors in the alignment or matching steps [40] when combining those AFIS with auxiliary approaches like the one presented in this paper.

Acknowledgments

R.K. was supported for the most part of this work by a Marie Curie Fellowship under project BBfor2 from European Commission (FP7-ITN-238803). This work has also been partially supported by Spanish Guardia Civil, and project CogniMetrics (TEC2015-70627-R) from Spanish MINECO/FEDER. The researchers from Halmstad University acknowledge funding from KK-SIDUS-AIR project and the CAISR program in Sweden.

References


URL http://www.neurotechnology.com/verifinger.html


