DOI: http://dx.doi.org/10.1109/CCST.2013.6922077

On the importance of rare features in AFIS-ranked latent fingerprint matched templates

Ram P. Krish, Julian Fierrez, Daniel Ramos and Ruifang Wang
Biometric Recognition Group - ATVS, EPS - Univ. Autonoma de Madrid
C/ Francisco Tomas y Valiente, 11 - Campus de Cantoblanco - 28049 Madrid, Spain
Email: {ram.krish, julian.fierrez, daniel.ramos, ruifang.wang}@uam.es

***Abstract*—In this paper, we introduce an algorithm to generate a score from the matched templates derived by the forensic examiner at the ACE-V stage. Such a score can be viewed quantitatively as a measure of confidence of the forensic examiner for the given latent and impression prints. This quantitative measure can be used in statistics-based evidence evaluation frameworks. Together with the description and evaluation of new realistic forensic casework driven score computation, we also exploit this experimental framework to show the importance of type attributes for minutiae in terms of its discriminating ability in forensic scenarios. We derive the conclusion that together with reliably extracted typical minutiae features, the presence of rare minutiae features helps to improve the measure of confidence of the forensic examiner at the ACE-V stage.**

I. Introduction

The friction ridge examination currently followed in forensic domain is known as ACE-V (*analysis, comparison, evaluation and verification*). But recently this procedure has been criticized for lacking a proper methodology in capturing any uncertainty involved in the decision yielded by a forensic examiner. The perception and the decision making ability among forensic examiners vary, e.g, the decision made by a novice examiner is not always consistent with the decision made by an experienced examiner for the same casework [1]. One of the most popularly cited examples where an erroneous individualization was made is with the Brandon Mayfield case [2]. Other similar cases of erroneous individualization have been reported in [3].

The latent fingerprints which are the unintentionally left impressions of the friction ridge skins obtained from the crime scene followed by chemical processing of the latent print and photographing it are of poor quality in nature [4] [5] [6]. From such poor quality latent fingerprint images, reliable manual feature extraction depends on the perception, decision making ability and experience of the forensic examiner. The ultimate objective of the examiner here is the identification of the perpetrator from a population of suspects based on these poor quality images. The first step usually followed is to manually extract the minutiae features from the latent fingerprint images (Stage 1 in Fig.1) and then search against a large database of known suspects using an Automated Fingerprint Identification System (AFIS). The AFIS generates a set of possible suspects on a rank basis using a similarity score (Stage 2 in Fig.1). Once there is a shortlist of suspects, a forensic examiner will manually compare the latent fingerprint with each of the shortlisted impression fingerprints following the ACE-V methodology (Stage 3 in Fig.1) to yield a decision as to

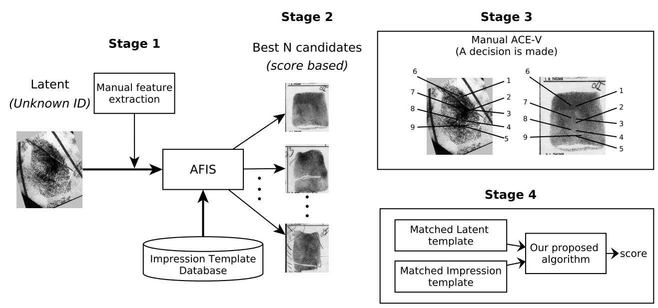

Fig. 1: Stage 1 to Stage 3 captures the Latent fingerprint examination methodology currently practiced. In Stage 4, we propose our framework to generate a score from matched template obtained from Stage 3.

whether the given latent and impression print match, do not match or the comparison is inconclusive.

The Stage 3 in Fig.1 (ACE-V) comprises of the following four phases [7] [2]:

1) *Analysis :* The examiner looks for sufficiency of the details present in the given latent print. This comprises of checking for ridge clarity, quantity of Level 1, Level 2 and Level 3 details.
2) *Comparison :* Once the latent print passes the analysis phase, many useful friction ridge details are extracted manually and are compared against one or more exemplar/reference fingerprints shortlisted by an AFIS to determine whether they are in agreement.
3) *Evaluation :* Based on the conclusions derived from the analysis and comparison phases, the forensic examiner yields a decision as *individualization (identification or match), exclusion (non-match) or inconclusive* for the given latent and impression fingerprint image pair.
4) *Verification :* In this phase, another qualified forensic examiner reexamines the decision made by the previous examiner by following the above three phases once again.

The latent examiner following the ACE-V methodology not only uses minutiae configuration (Level 2 features), but also extended information like general ridge flow (Level 1 features), number of ridges between two minutiae, presence of creases and other Level 3 features [2] [6]. Several studies have been

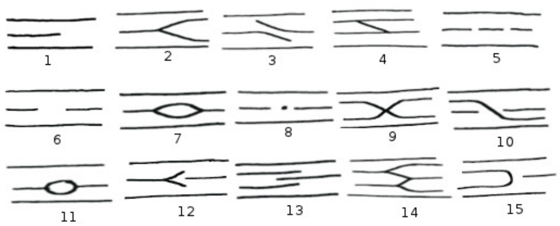

Fig. 2: Minutiae types used in Guardia Civil Database. Names corresponding to individual type numbers can be found in Table I.

TABLE I: List of typical and rare minutiae in Guardia Civil Database. Numbering with respect to Fig.2.

No	TypeName	No	TypeName	No	TypeName
1	Ridge Ending	6	Interruption	11	Circle
2	Bifurcation	7	Enclosure	12	Delta
3	Deviation	8	Point	13	Assemble
4	Bridge	9	Ridge Crossing	14	M-structure
5	Fragment	10	Transversal	15	Return

made towards including extended features like ridge quality map, ridge flow map, skeletanized images, dots and incipient ridges [6] as well as use of minutiae based descriptors [13] to improve the performance of automated latent fingerprint matchers. Importance of rarity in the configuration of fingerprint features were emphasized by latent print examiners while making decisions at the ACE-V stage [2].

There is no scientific framework in use at the criminal justice system to characterize the uncertainty involved in the ACE-V procedure, as well as to express the strength of opinion of the forensic examiner quantitatively [2]. Such a requirement has been articulated in several influential reports [2] like the NRC 2009 report [8] and the NIST Human Factors report [7]. The new paradigm coming forward in this regard [9] avoids hard identification decisions by considering evidence reporting methods that incorporate uncertainty and statistics. Among all the methods of evidence evaluation, the likelihood ratio is receiving greater attention [10] [2].

In this work, we propose an algorithm to generate a score from the matched templates derived by the forensic examiner at the ACE-V level (Stage 4 in Fig.1). Such a score can be viewed as a measure of confidence of the examiner quantitatively in place of a logical decision (*individualization, exclusion or inconclusive*). Such a score can be used in a statistics-based evidence evaluation framework to quantify the fingerprint evidence. We exploit this proposed experimental framework to study the importance of reliably-extracted type information for minutiae by a forensic examiner. Together with typical minutiae features, the importance of rare minutiae features has been established based on its discriminating ability in a forensic scenario.

The remainder of the paper is organized as follows. We first explain in detail about the real forensic casework databases used in this study, then the method developed to generate a score as a measure of confidence for a forensic examiner. We then present the experimental protocol and results, followed by a discussion on the discriminating capability of rare minutiae features.

II. Database

Two different forensic databases were used in this study. One is the publicly available NIST Special Database (SD) 27, and the other one is acquired from Guardia Civil, the law enforcement agency of the Government of Spain. NIST SD27 minutiae template database is broadly classified into two: 1) *ideal*, and 2) *matched* minutiae database. The *ideal* minutiae set for latents was manually extracted by a forensic examiner without any prior knowledge of its corresponding impression image. The *ideal* minutiae for impressions was initially extracted using an AFIS, and then these minutiae were manually validated by at least two forensic examiners. The *matched* minutiae templates contains those minutiae which are in common between the latent and its mated impression image. There is a one-to-one correspondence in the minutiae between the latent and its mate in the matched template. This ground truth was established by a forensic examiner looking at the images and the *ideal* minutiae following an ACE-V procedure. For NIST SD27 database, only the ideal-latent templates had type information for each minutiae in addition to location and orientation attributes. The other three datasets (ideal-impression, matched-latent and matched-impression) do not have type information but only location and orientation attributes.

The Guardia Civil database (GCDB) is similar to the NIST SD27 database except that all the templates in ideal and matched sets in GCDB have type information. Apart from having typical minutiae types (*ridge-endings, bifurcations*), GCDB also comprises rare minutiae types like *fragments*, *enclosures*, *points/dots*, *interruptions, etc* [11]. Please refer to Fig.2 and Table I for a comprehensive list of all minutiae feature types present in GCDB. Table II shows the statistics of various types of minutiae features present in the 258 template pairs available in GCDB. Rest of the minutiae types were not observed so far in this collection of GCDB.

We will follow the notation GCDB-M and NIST-SD27-M to denote the matched template database of GCDB and NIST SD27 respectively. In Fig.3, we show a latent fingerprint and its corresponding impression with typical features and some of the rare features annotated with their correspondences. The latent and impression images used here were taken from NIST SD27 database, and the typical and rare minutiae features were manually annotated on them.

III. Algorithm

We propose an algorithm to generate a score for the templates in GCDB-M. This algorithm can be adapted to templates from NIST-SD27-M by discarding the weights for type information when calculating *typeError* explained in the algorithm. The various stages involved in the computation of the score are as follows:

Fig. 3: Typical and some rare minutiae features on a latent and its mated impression fingerprint. The latent image G004L8U and the impression image G004T8U were taken from NIST SD27 database.

TABLE II: Statistics of typical and rare minutiae present in Guardia Civil Database. Numbering with respect to Fig.2.

No	Contribution	No	Contribution	No	Contribution
1	56%	4	0.265%	7	2.058%
2	36.38%	5	4.515%	8	0.332%
3	0.166%	6	0.232%	10	0.0332%

A. Alignment and correspondence

Since the framework is developed to deal with matched databases, we expect that for genuine matches, superimposing the centroids of both latent and impression minutiae points with appropriate rotation alignment would lead to an approximate fitting of point patterns based on mated pairs with minimum overall fitting error, and for impostors it would lead to a high fitting error.

As typical minutiae features are the majority with 92% (see Table II), we only use typical features to estimate the rotation parameters. By rotating the latent template over the impression template w.r.t centroid in a range of $[-45°, +45°]$, we find the closest matching minutiae pairs, and add their distance. The rotation for which the average sum of closest pairs is the minimum is considered to be the best rotation alignment for their approximate pattern fitting.

After the alignment, all those minutiae pairs which are within a threshold distance are considered to be mated pairs, and their correspondences are established.

B. Fitting and Orientation errors

Once the correspondences are established for all the typical minutiae features, the scores are computed hierarchically looking at each of minutiae attributes, namely *location, orientation* and *type* information. Scores based on *type* information are discussed in the next subsection.

For all the typical minutiae which established correspondences based on optimal rotation, we find a fitting error using an affine transformation for the mated minutiae patterns by least square fitting. This score is denoted as *fittingError*, which is averaged w.r.t total number of mated minutiae pairs.

Again for all the mated minutiae pairs, we sum up all the orientation differences of corresponding minutiae and average this sum of degrees w.r.t total number of mated pairs. When averaging the orientations, the circularity of degrees are taken care of. This score is denoted as *orientationError*.

C. Type errors

If the mated pairs disagree w.r.t *type* information, which otherwise are mated based on only location and orientation attributes, we associate a penalty for such type of mismatches. The penalty for each typical minutiae type is a constant factor estimated from Table II. This score is denoted as *typeError*. This is possible because the *type* information for both latent and impression are estimated manually by a forensic examiner, and we assume type information is available here.

Based on the alignment estimated using *typical* minutiae, we also look for the presence of *rare* minutiae correspondences. If they are within a location and orientation threshold, then they constitute mated pairs, and thus correspondence is established. As the percentage of occurrence of rare minutiae is very small, around 8%, we only estimate *typeError* for rare minutiae. The penalty for each rare minutiae type is a constant factor estimated from Table II.

D. Final Score

Since all the individual scores we have generated are of different nature, namely *fittingError* in distance, *orientationError* in degrees, *typeError* in probability based cost, these scores are combined using logistic regression to generate the final score [12]:

$$\begin{aligned} finalScore = &(\alpha \times fittingError) \\ &+ (\beta \times orientationError) \\ &+ (\gamma \times typeError) \end{aligned} \tag{1}$$

where α, β, γ are the logistic regression coefficients for each classifier respectively.

This final score can be viewed as a measure of confidence of the forensic examiner numerically, otherwise the forensic examiner only have a logical decision at the stage of ACE-V. Note that the $finalScore$ is a dissimilarity score, so the higher the score the higher the distance between a match and non-match.

IV. Experiments

A. Experimental protocol

The total number of latent fingerprint templates in GCDB-M is 258, with their corresponding matched impression fingerprint templates. This size of GCDB is equivalent to the publicly available NIST-SD27-M. This way, we could do some performance comparisons between databases, unbiased in terms of partitioning for train and test dataset sizes. For training the logistic regression coefficients, we used 129 template pairs and 129 for testing.

We performed experiments to study the performance of the developed approach by comparing the degree of overlap between matching and non-matching scores in the matched databases. Various parameters like the distance and orientation thresholds were finetuned to minimize this degree overlap. We also exploit the developed framework to understand the importance of minutiae feature types, as well as the contributions of rare minutiae features in terms of discriminating power. All experiments have 129 matching and 129×128 non-matching test comparisons.

B. Results using location and orientation

In this experiment, we only used the location and orientation attributes, because NIST-SD27-M does not have type attributes for the minutiae. This helps in evaluating the performance of the developed algorithm in a more transparent way, comparing its behavior in both GCDB-M and NIST-SD27-M. We obtained good discrimination for both GCDB-M and NIST-SD27-M, with only 3.8760% and 6.9767% score overlaps respectively.

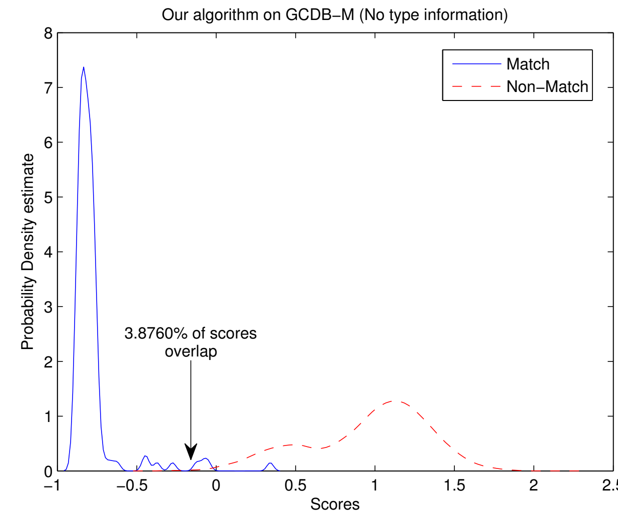

Fig. 4: Performance of our algorithm on GCDB-M. No type information of GCDB-M is used.

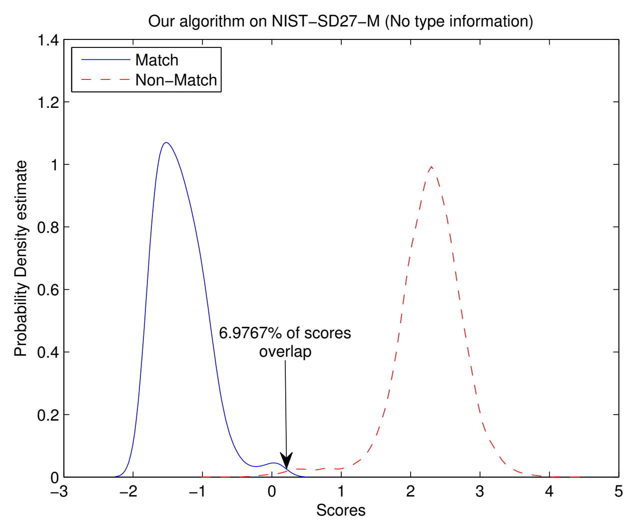

Fig. 5: Performance of our algorithm on NIST-SD27-M. No type information is present in the database.

Fig.4 and Fig.5 shows the degree of overlap of scores on both GCDB-M and NIST-SD27-M using our proposed algorithm respectively. The average number of minutiae per template for GCDB-M is 12 and that of NIST-SD27-M is 21. Referring to Fig.4 and Fig.5, we were able to obtain good discrimination between match and non-match.

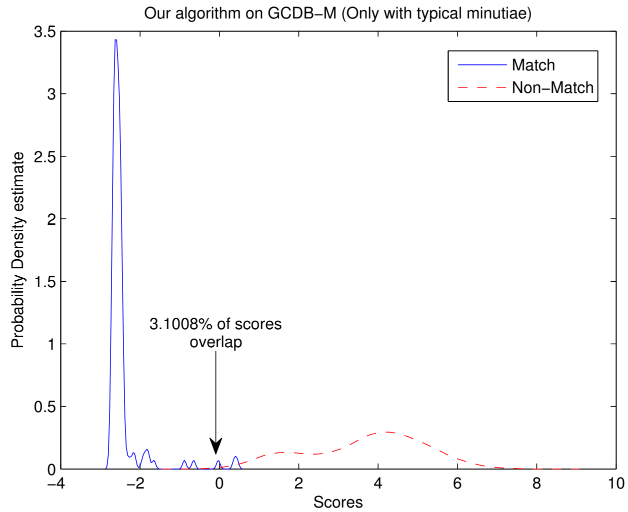

Fig. 6: Performance of our algorithm on GCDB-M when only typical minutiae features are used.

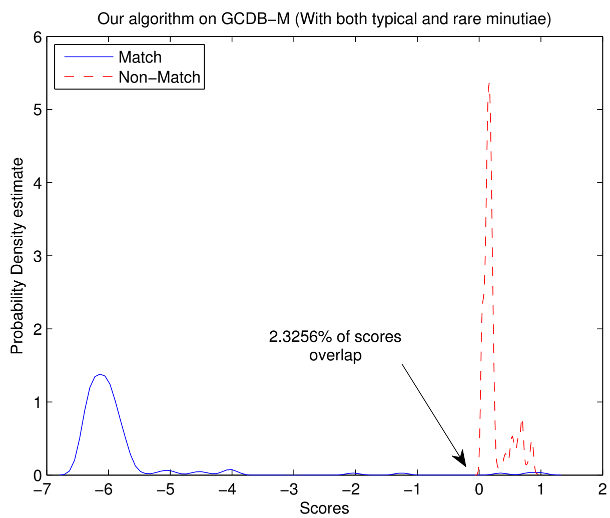

Fig. 7: Performance of our algorithm on GCDB-M when both typical and rare minutiae features are used.

C. Results using location, orientation and type

In this experiment to study the importance of type information, we only used the GCDB-M database. Together with the location and orientation attributes, we tested the performance of the algorithm by first adding only typical minutiae features and then both typical and rare minutiae features. The weight for each minutiae type is a constant factor estimated from the values corresponding to the minutiae as shown in Table II.

Fig.6 and Fig.7 show the degree of overlap of scores on GCDB-M when using only typical minutiae features, and when using both typical and rare minutiae feature respectively. We obtained a score overlap of 3.1008% when only typical features were used as compared against 3.8760% of score overlaps when no type information was used (refer Fig.4). This helps us to understand the importance of reliably extracted typical minutiae features by a forensic examiner.

Typical minutiae features constituted of 92% of minutiae present in GCDB-M. Together with typical minutiae features, when rare minutiae features are combined, we see a further improvement from 3.1008% to 2.3256% of score overlaps. Also, comparing visually Fig.6 and Fig.7 we can see that match and non-match distributions when consisting rare features (Fig.7) are much more separated than when not (Fig.6). From this, we conclude that rare minutiae features if present improve the discrimination ability, and consequently improving the measure of confidence of the forensic examiner.

V. Discussions

We developed a framework to generate a score from matched templates generated by the forensic examiner as a result of ACE-V. This score can be viewed as a measure of confidence of the forensic examiner numerically in place of a logical decision. We tested the algorithm on GCDB-M as well as on the publicly available NIST-SD27-M database. We also exploited this framework to understand the importance of minutiae type information in forensics, as well as the importance of rare minutiae type.

The discriminating ability of type attributes for minutiae is clearly evident from the above experiments. We started by discarding type information in the first experiment and then incrementally added the type information from typical to rare minutiae to see if there is any improvement in the discrimination ability and we found that a reliably extracted type information help significantly in discriminating match and non-match templates at the ACE-V stage.

A deeper analysis about how to exploit the developed score computed as a confidence measure for the forensic examiner, as well as an implementation of the likelihood ratio approach based on these scores to express the strength of opinion of forensic examiners quantitatively is in order.

Acknowledgment

R.K. and R.W. are supported by Marie Curie Fellowships under project BBfor2 (FP7-ITN-238803). This work has also been partially supported by Spanish Guardia Civil, Cátedra UAM-Telefónica, and projects Bio-Shield (TEC2012-34881) and Contexts (S2009/TIC-1485).

References

[1] J.R. Vanderkolk, Chapter 9, Examination Process, *The Fingerprint Sourcebook*, U.S Department of Justice, 2011.

[2] S.N. Srihari, *Quantitative Measures in Support of Latent Print Comparison: Final Technical Report: NIJ Award Number: 2009-DN-BX-K208*, University at Buffalo, SUNY, 2013.

[3] S:A. Cole, More than zero: Accounting for error in latent fingerprint identification, *Journal of Criminal Law and Criminology*, 985-1078, 2005.

[4] F. Alonso-Fernandez, J. Fierrez and J. Ortega-Garcia, Quality Measures in Biometric Systems, *Security Privacy, IEEE*, 2012.

[5] C. Champod, C.J. Lennard, P. Margot and M. Stoilovic, Fingerprints and other ridge skin impressions, *CRC*, 2004.

[6] A.K. Jain and J. Feng, Latent Fingerprint Matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 88-100, 2011.

[7] Expert Working Group on Human Factors in Latent Print Analysis. *Latent Print Examination and Human Factors: Improving the Practice through a Systems Approach*, NIST, 2012.

[8] National Academy of Sciences, *Strengthening the Forensic Sciences in the United States: A Path Forward*, National Academies Press, 2009.

[9] M.J. Saks and J.J. Koehler, The Coming Paradigm Shift in Forensic Identification Science, *Science*, 892-895, 2005.

[10] I. Evett, et al, Expressing evaluative opinions: A position statement, *Science and justice*, 2011.

[11] F.S. Santamaria, A New Method of Evaluating Ridge Characteristics, *Fingerprint and Identification Magazine*, 1955

[12] F. Alonso-Fernandez, J. Fierrez, D. Ramos and J. Gonzalez-Rodriguez, Quality-Based Conditional Processing in Multi-Biometrics: application to Sensor Interoperability, *IEEE Transactions on Systems, Man and Cybernetics Part A*, 2010.

[13] J. Feng, Combining minutiae descriptors for fingerprint matching, *Pattern Recognition*, 342-352, 2008.