Semantic Conformance Testing Methodology for Finger Minutiae Data

Dana Lodrova[1], Christoph Busch[2,3], Elham Tabassi[4],
Wolfgang Krodel[5], Martin Drahansky[1]

1 - Brno University of Technology - Faculty of Information Technology, CZ
2 - Hochschule Darmstadt - CASED, DE
3 - Gjøvik University College - NISlab, NO
4 - National Institute of Standards and Technology, US
5 – Bundeskriminalamt - ZD23/AFIS, DE

email: ilodrova@fit.vutbr.cz

Abstract: This paper proposes a methodology to measure the semantic conformance rate of standardized biometric minutia interchange records. The paper proposes a fingerprint modality specific assertion test. A conformance test based on this methodology can attest for a given algorithm or software under test that the generated minutiae templates are a faithful representation of the input signal (i.e. fingerprint image). The test methodology is based on ground truth data that has been composed by dactyloscopic experts. As individual experts assessment yields slightly diverging coordinates a clustering algorithm is proposed that merges a set of manually placed minutia into one ground truth data set. The methodology is evaluated on ten-print fingerprint images and the NIST baseline minutia extraction algorithm.

1 Introduction

Many large scale biometric systems require compact storage of biometric references. The reference should represent a biometric characteristic and be compliant to an interoperable standardized format. The reference should be a faithful representation of a biometric characteristic (e.g. fingerprint). Also since for enrolment and verification different feature extraction algorithms could be used, it is necessary that a biometric reference is an interoperable representation of the biometric characteristic and therefore compliant to an interoperable standardized format. For fingerprint recognition systems the compact coding of minutia data provides interoperability among systems, where the reference is stored in tokens with limited storage capacity [iso05]. Examples for such systems are the European Citizen Card [ecc07] or the U.S. PIV Card [nist07]. The essential features of a fingerprint minutia template are locations, type (ridge endings and ridge bifurcations) and directions. This data is the relevant information for almost every fingerprint comparison subsystem.

As different vendors apply different concepts and algorithms to identify minutiae locations, directions and types, automatically generated minutiae are scattered around the truth (real) minutiae data. That means, in order to achieve sufficient interoperability and acceptable overall performance among different implementations, conformance testing is an essential process. ISO/IEC FDIS 29109-1 has categorized conformance testing into three levels [iso09a]. Level 1 focuses on basic data field testing. Level 2 is a syntactic test and inspects whether the data fields are filled with meaningful values [iso09b]. Level 3, however, is a semantic test, which inspects whether a generated interchange record is a faithful representation of the initial biometric data (e.g. fingerprint image) [bus09]. Level 3 conformance test is important because without accurate representation of biometric data, desirable interoperability and performance could not be achieved.
In this paper we focus on Level 3 conformance testing for finger minutia data. The basic idea of our method was presented in [bus09]. This paper contains an extension of the proposed method and augments new methodology for clustering of minutiae, which is required for the computation of conformance rates. Furthermore we describe an implementation and present preliminary results.
This paper is organized as follows. The second section describes challenges associated with minutiae detection. In section 3 we propose a methodology for computation of semantic conformance rates. The fourth section describes a clustering algorithm needed to merge ground truth data provided by multiple experts. Conclusion remarks and future work are in section 6.

2 Challenges associated with minutia detection

When minutia extractors are applied to a fingerprint images the following three situations can occur that may cause a challenge for the comparison subsystem:

Imprecisely placed minutiae

Imprecise detection of a minutia may be associated with:

- inaccurate minutia position
  (some distance can be tolerated),
- false minutia type,
- inaccurate minutia direction
  (some delta angle can be tolerated),
- wrong (different) minutia quality [1]

Probably the most frequent defect is the wrong minutia type (see Fig. 1). Ridge ending is detected as ridge bifurcation or vice versa, mostly because of noise around this minutia or due variations of the papillary line grey value. On the other side, some vendors intentionally do not set the type of minutiae properly.

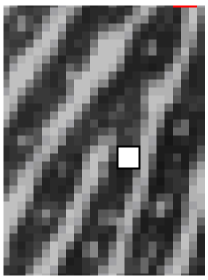


Fig. 1: Wrong minutia type: ridge bifurcation detected as ridge ending (square).

Problematic minutia detection inside the fingerprint area

Automatically detected minutiae can be in a number of problematic locations:

[1] In the absence of a standardized quality algorithm – investigation of minutia quality is not considered in this work.

- scars,
- “papillary dots”,
- dirt or hair glued on finger,
- skin diseases (for example eczema or tubercle),
- bent skin,
- written text or drawings inside the fingerprint area.

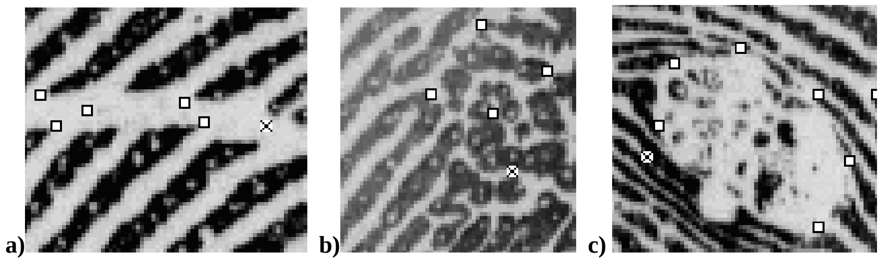


Fig. 2: Minutiae detected in problematic locations in the fingerprint area: a) bent skin, b) papillary dots, c) tubercle, (square: ridge ending, cross: ridge bifurcation, extractor: NIST mindtct).

Problematic minutiae detection outside the fingerprint area or at the borders
Some minutiae extraction algorithms detect minutiae at the border of the fingerprint area or even outside. This is a consequence of improper foreground/background masking and can be caused by dirt and drawings or characters in the background. Fig. 3a shows one false minutia (ridge ending) in the background noise and a further false minutia (ridge bifurcation) in some background drawing (present in the scanned ten-print card)

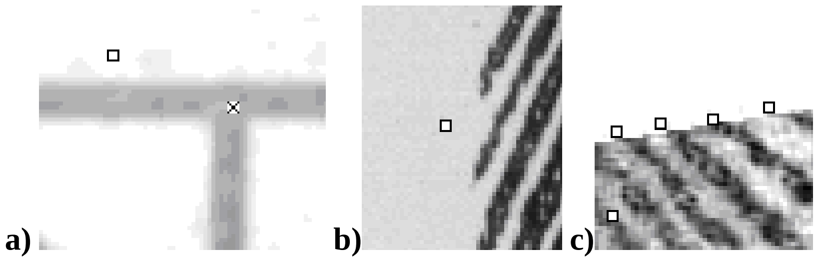


Fig. 3: Minutiae detected: a) outside the fingerprint area or b), c) at the borders.

3 Semantic conformance testing methodology

In order to determine whether or not a minutia extractor is conformant to some ground truth, we propose three conformance rates[2]. The ground-truth minutiae (GTM) placements, as explained in section 4, are the cluster center of various manual expert minutiae placements.

[2] Conformance can be stated, if the conformance test yields a conformance rate above a defined threshold.

The first rate cr_{gtm} indicates to which extent automatically placed minutia are located in the vicinity of the ground truth. If no automatically generated minutia (AGM) is found within the tolerance limits of a ground truth minutia (GTM), the minutia conformance score is valued 0. Otherwise the *i*-th minutia specific score mcs_i yields some value in the range [0, …, 1], where a cost-factor (punishment) p represent other defects. The conformance rate is given by

$$cr_{gtm} = \frac{\sum_{i=1}^{ngtm} mcs_i}{ngtm} \tag{3.1}$$

where *ngtm* is the number of minutiae (GTM) in the ground truth database. The minutia conformance score is given by:

$$mcs = \begin{cases} 0 & if\ d \geq tol_d \\ 1 - p & otherwise \end{cases}, \quad tol_d = \frac{W}{4} \tag{3.2}$$

where d is the Euclidean distance between a GTM and the nearest AGM. W is the space between parallel skeletonized ridges. We intentionally chose tol_d to be W/4, since this is the maximal possible radius around a GTM, such that two neighbored GTM areas will not overlap each other. This situation is illustrated in Fig. 4.

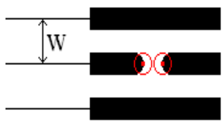


Fig. 4: Ridge space *W* and minutia tolerance

A punishment p reduces the *mcs* due to differences in the orientation or due to a different minutia type.

$$p = p_{\Delta\theta} + p_{\Delta t} \tag{3.3}$$

$$p_{\Delta\theta} = \frac{\left|\theta_{gtm} - \theta_{agm}\right| * 0{,}5}{\pi} \tag{3.4}$$

$$p_{\Delta t} = \begin{cases} 0{,}25 & if\ t_{gtm} \neq t_{agm} \\ 0 & otherwise \end{cases} \tag{3.5}$$

We intentionally chose different punishments for different deficiencies, as the impact on the observed biometric interoperability performance is strongest for the inaccuracies in minutia location, less relevant for the inaccuracies in minutia angle determination and least relevant for a diverting minutia type.

Frequently minutia extractors mislabel the minutia type, i.e. a ridge bifurcation is detected as ridge ending and vice versa. In this case not only the type is different, but also the delta $\Delta\Theta$ between angles might be close to π. We assume that it is not justified to punish one defect twice. Thus if we detect that one minutia is labeled as ridge ending and the other as ridge bifurcation, we automatically increase the angle of *agm* by π.

The second conformance rate is cr_{agm}, which describes the proportion of false minutiae wrongly placed outside or at the borders of the fingerprint area.

$$cr_{agm} = \frac{\sum_{i=1}^{nagm} mps_i}{nagm} \tag{3.6}$$

$$mps = \begin{cases} 0 & if\ agm\ is\ outside\ the\ fingerprint\ area \\ 0{,}5 & if\ agm\ is\ at\ the\ borderline \\ 1 & otherwise \end{cases} \tag{3.7}$$

where *nagm* is the number of AGMs.

The third conformance rate is cr_{amf}, which represents the automated extracted minutiae focus with respect to the fingerprint area. This can be understood as the proportion of minutiae inside the fingerprint area for which no mate was found in the set of GTMs:

$$cr_{amf} = 1 - \frac{niagm}{nagm} \tag{3.8}$$

In Eq. (3.8) *niagm* is the number of focused AGMs inside the fingerprint area, which does not correspond to any GTM.

4 Ground truth minutia data

Conformance testing based on the proposed methodology requires a ground truth database with a large set of minutiae.

4.1 Collecting of ground truth data

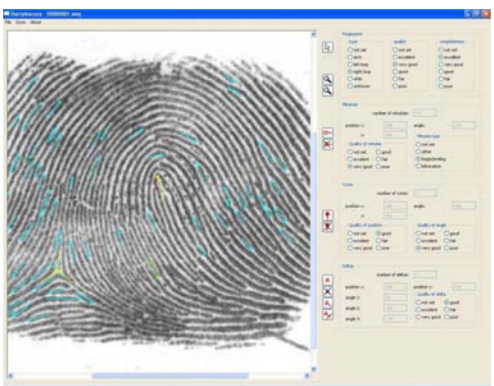


Fig. 5: GUI for dactyloscopic experts.

To collect the GTM database, we provide a graphical user interface for dactyloscopic experts (see screenshot in Fig. 5), which supports measuring of location, type, angle and quality in an image. Further information, e.g. on cores and deltas, pattern type and signal quality, is determined for future use.

Information set by experts is stored in an internal *.gtm file format. Its encoding scheme follows the ISO 19794-2 standard, where possible.

Example of *.gtm file format:

```
Width                   : 832 px
Height                  : 768 px
Fingerprint type        : R
Fingerprint quality     : 2
Fingerprint completeness: 1
```

```
Number of minutiae: 3
-----------------------------------------------------------
 id:  type,    x  ,   y  , angle, quality of minutiae
-----------------------------------------------------------
  0:     2,   527,   234,    81,   90
  1:     1,   452,   358,   104,   70
  2:     0,   360,   170,   187,   10

Number of cores    : 1
-----------------------------------------------------------------
 id:   x  ,   y  , quality of position, angle, quality of angle
-----------------------------------------------------------------
  0:   388,   165,                  90,   213,   70

Number of deltas   : 1
--------------------------------------------------------
 id:   x  ,   y  , angle, angle, angle, quality of delta
--------------------------------------------------------
  0:   342,   341,    66,   231,    66,   70
```

4.2 Clustering scattered data from experts

The minutia measurements by experts can be expected to be similar in many cases but will be scattered. Thus it is required to cluster the scattered data (individual *.gtm files from *n* contributing experts) and to compose the ground truth data as an input to our process, which generates conformance rates (see Fig. 6).

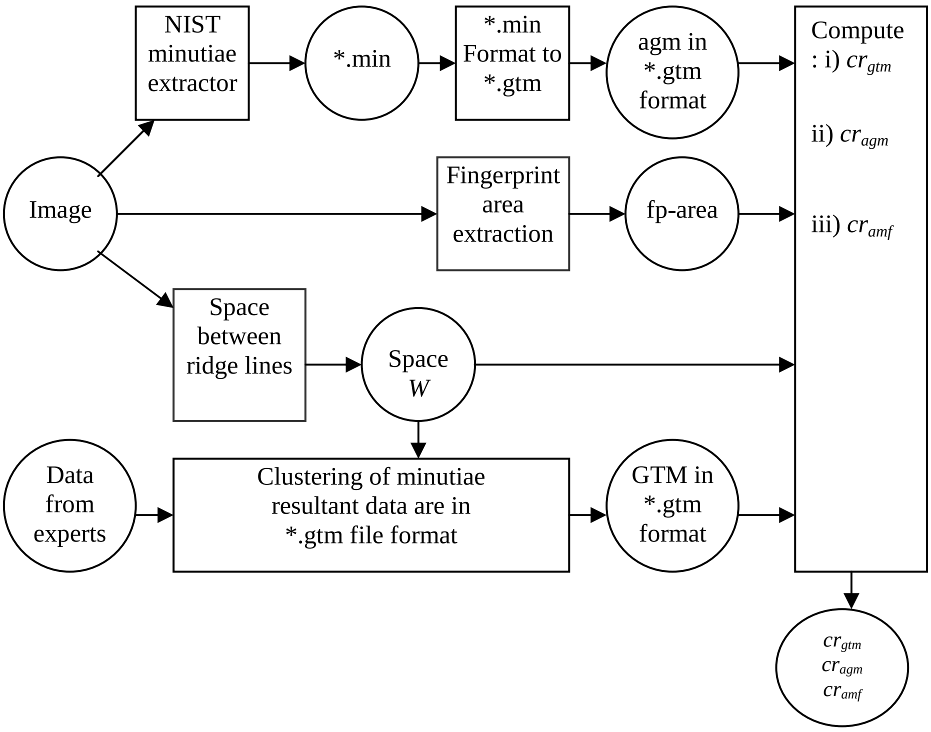


Fig. 6: Process workflow to determine conformance rates. For a sample evaluation the NIST mindtct minutiae extraction algorithm has been submitted to the conformance testing methodology. Circles represent files/values and squares represent software components.

The first processing step is to analyze cluster of minutiae *gtms* in an image where *gtms* are marked by different experts. Then we mark the fingerprint area of the image and compute space between ridges (*W*). The same image is also processed by the minutiae extraction algorithm under the test, in our case, the NIST mindtct algorithm [nbis] was used for illustration purposes. These information sources influence the resulting conformance rates.

The clustering algorithm that analyzes the minutia measurements from various experts and computes a ground truth minutia (GTM) as cluster center is a non-trivial task, as the target number of clusters is not known. To solve this task we propose a new algorithm, which is inspired by the Apriori algorithm [wk09] and by hierarchical clustering generally. At first, the *gtmi* data sets from *n* experts are stored into an array of minutiae (in this case a `struct` with values regarding position, angle, type, quality, expert ID and a Boolean marker "processed"/"not used"). Next we create an array of minutiae pairs. We create a pair from each two minutiae, if the following conditions are satisfied:

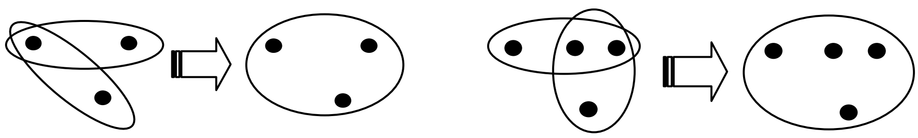
- Each minutia has been placed by a different expert
- The distance between minutiae is less or equal than $W/2$
  (all minutiae will be inside a circle with radius $W/4$)

When we are creating a pair of two minutiae, we mark both minutiae as processed and then insert a newly created pair to the array of pairs only if such pair is not already included in the set.

Then we similarly create an array of triplets. We create a triplet from all pairs of minutiae pairs (created in the previous step), which satisfy the following conditions:

- Minutiae pairs have **one** identical (joint) minutiae
- Each minutia in a new **triplet** candidate has been placed by a different expert
- The distance of all minutiae pairs from new **triplet** candidate is less or equal than $W/2$ (all minutiae will be inside a circle with radius $W/4$)

Thus we have added the first condition and require that the minutia pairs have one identical minutia that will establish the link for the triplet creation (see Fig. 7a).



a) **b)**

Fig. 7: Minutiae clustering: a) creation of triplet from two pairs, b) creation of quadruples.

The process step for creation of quadruples is almost identical:

- Minutiae triplets have **two** identical (joint) minutiae (see Fig. 7b)
- Each minutia in a new **quadruple** candidate has been placed by a different expert.
- The distance of all minutiae triplets from the new **quadruple** candidate is less or equal than $W/2$ (all minutiae will be inside a circle with radius $W/4$).

Then we continue the creation of n-tuples until n is equal to the number of experts ($nexp$).

In order to determine each cluster center it is necessary to compute an average minutiae position in the cluster, as well and an average angle and type. There are two possible methods to derive the average minutia positions, which implement a straightforward sum

$$X_{GTM} = \frac{\sum_{i=1}^{ngtm} x_i}{ngtm}, \quad Y_{GTM} = \frac{\sum_{i=1}^{ngtm} y_i}{ngtm} \tag{4.1}$$

and a minimum / maximum approach, as given in Eq. (4.2).

$$X_{GTM} = \frac{\min(x) + \max(x)}{2}, \quad Y_{GTM} = \frac{\min(y) + \max(y)}{2} \tag{4.2}$$

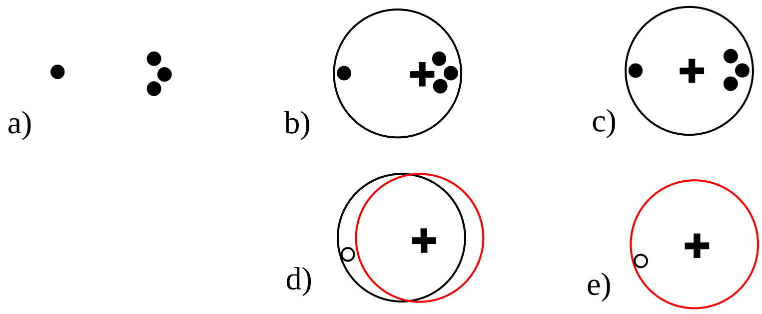


Fig. 8: Comparison of two methods for computation of the cluster center. Eq. 4.1 is in parts b) and d), eq. 4.2 in parts c) and e). Black dots are minutiae from experts; crosses are computed centers of cluster and white dots are tested agm.

The impact of the two methods is illustrated in Fig.8. As one can see, the first method shows stronger robustness w.r.t. outliers. As only one expert measured the minutia to be on the left side and the other three experts opted for the right side, the cluster center will tentatively be located on the right hand side. The advantage of this choice is that the ground truth data will show stronger robustness and reliability, while at the same time the risk that an automated generated minutia will be rejected corresponds to the likelihood that the minority opinion eventually represents the ultimate truth. However we have chosen the first averaging method since experts are only human beings, their hands can shake or they might be distracted while measuring the minutia position.

In the same line it is necessary to compute the average minutia type. We assign a ground truth minutia type if more than 2/3 of the experts vote for one type and we can state consensus[3]. Otherwise the minutia type is set to UNKNOWN and punishment for wrong minutia type can not be used.

---

[3] According to ISO directives a majority of 2/3 in a ballot manifests consensus.

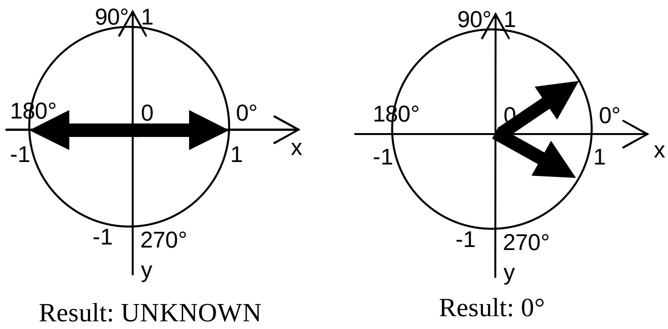


Fig. 9: Computation of average angle.

The computation of the mean direction requires an additional consideration. It might happen that one expert measures a specific minutia direction to be 180° while a second expert measures the same direction with 0°. Furthermore there might be a situation, in which three experts conclude in three completely different opinions (e.g. 0°, 120° and 240°). In such a case it is appropriate to set the ground truth direction to UNKNOWN. We compute an average direction by first converting all angles to directional vectors with length 1. Thus each endpoint (x_m and y_m coordinates) is located on the unit circle. Next we compute the mean $\underline{x}$ and $\underline{y}$ coordinate and take them as endpoint of the resultant direction vector, which might have a length smaller 1. If the resultant vector's length is less than 1/3, then the resultant direction will be UNKNOWN, otherwise we just convert resultant vector into a ground truth direction. We also set the direction to UNKNOWN in such cases, where the minutia type is UNKNOWN, as we consider a consensus regarding the minutia type to be a precondition for a reliable ground truth minutia direction.

4.3 Reliability of clusters

For the computation of the conformance rates of equations (3.1) – (3.8) it is essential to consider the reliability of each GTM. Such GTM reliability in turn depends on the quality of a cluster that created the GTM. The quality of a cluster is impacted by two factors. On one hand the number of experts that detected the minutia. If an image has been processed by 20 experts and only two of them have found this concrete minutia (and maybe those attributed a low minutia quality), then we cannot consider the mean minutia to be reliable. On the other hand if the concrete minutia is detected by 18 experts (and maybe all of them attributed a good minutia quality) then we can consider cluster center to be a reliable GTM. In order to distinguish unreliable minutia from reliable minutia we consider the quality of a cluster as defined in equation (4.3):

$$quality\ of\ cluster = \frac{\sum_{i=1}^{ncl} q_i}{nexp},\ quality\ of\ cluster \in \langle 0-100 \rangle \tag{4.3}$$

where q_i is the minutia quality of the *i*-th minutia in the cluster, *ncl* is the number of minutia in that cluster and *nexp* is number of experts processing this image. For example if all experts detected this minutia with minutia quality 50, then the quality of this cluster is 50. This is the same result as if this concrete minutia would be detected only by half of the experts but with minutia quality 100.

5 Methodology evaluation

For evaluation purposes, we used 17 images from NIST SD14, SD29 database, which were processed by 11 experts from the German Federal Criminal Police Office (BKA). The average space between parallel ridge lines and the fingerprint area were computed manually.

In Fig. 10 you can see the example of measured minutiae from experts mapped into the original image. Squares are ridge endings and triangles are minutiae of type "other". As you can see, the experts are quite consistent in their measurement (minutia placement and types), but there are still some problematic cases (e.g. two minutiae of "other" type in the top/left corner of the image).

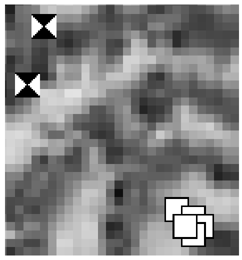


Fig. 10: Minutiae positions and types (8 experts; squares are ridge endings, symbol of two bl;ack triangles indicate minutia of "other" type).

One possible problems is e.g. a very short ridge line (dot). Some experts mark the beginning and end (two ridge endings) of this short ridge line and other experts mark the center of the dot specified the minutia type "other". Other problem can be e.g. minutiae, where experts cannot decide if there are ridge endings or bifurcations.

Finally we can see in Fig. 11 the results of the clustering algorithm – the cluster centers. The shape of minutiae has the same meaning as in the figures. The clustering method is very reliable in cases where experts' opinions are consistent.

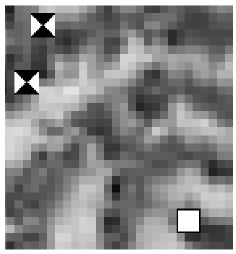


Fig. 11: Location and minutia type of cluster centers (squares are ridge endings, symbol of two bl;ack triangles indicate minutia of "other" type).

If experts are not consistent in their opinions and measured minutia locations are spread more widely, then it happens that instead of one cluster center there are two or even more of them. In order to limit the ground truth database to just the most reliable minutiae it was necessary to decide, which threshold value should be used for the "quality of cluster". On the one hand it is not reasonable to keep a cluster that has been created from only one expert's opinion if we have a large number of experts. On the other hand, the threshold value should not be too high, such that there will be too few clusters and eventually the conformance rate would be computed on very few GTMs.

In order to identify a suitable threshold for the quality of clusters, we compute all conformance rates for all images for threshold values between 0 and 50. Next we compute average values and their standard deviations (see Fig. 12). As a threshold value we choose the value, where both conformance rates (cr_{gtm} and cr_{amf}) have the same value. Thus for this sample data set the threshold value was chosen as 37. All computed conformance rates can be found in Tab. 1.

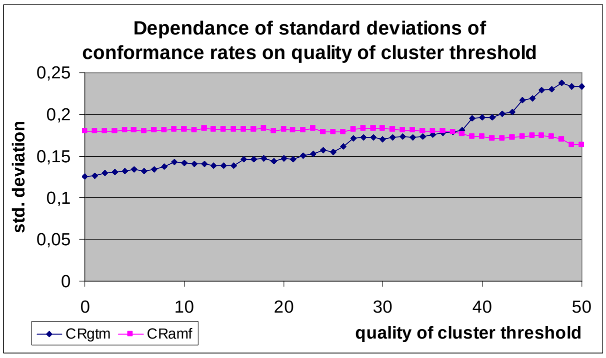


Fig. 12: Standard deviation of conformance rates vs. quality of cluster threshold.

Tab. 1: Results for the chosen threshold of cluster quality (37) .

	cr_{gtm}	cr_{agm}	cr_{amf}	*ngtm*	*nagm*
average	0,353	0,885	0,662	59	100
std. deviation	0,179	0,066	0,178		

Fig. 13 shows cluster centers, i.e. ground truth minutiae (*gtms*) that pass the quality threshold of 37. Previously figured problems have been resolved, because the problematic cluster centers, which caused these problems, are not included because they did not pass the cluster quality of 37.

One possibly problematic situation remains. For some minutiae there is more than one cluster center. In this case the AGM can belong to one of such clusters or all of them and this can have an influence on the cr_{gtm} conformance rate. Theoretically it can happen that also two ridge endings will be vis-à-vis and the minutiae from experts will be set so that the resultant clusters will partly overlap each other. If the AGM will be placed so that it can belong to both of them, this would be a greater problem than the previous situation.

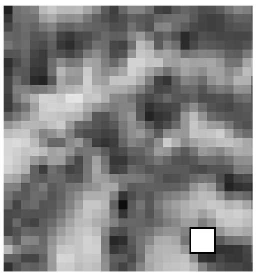


Fig. 13: Positions and types of cluster centers, which pass the quality threshold 37.

As a solution of this problem we propose to try clustering of clusters and then set the rule that one AGM can belong to one cluster only. This will of course be the cluster where AGM has the lower punishment.

6 Conclusion and future work

In this paper we have proposed a methodology for Level 3 Conformance Testing for finger minutiae data. We have also implemented the proposed method and the preliminary evaluation is yielding promising results. For illustrative purposes we have conducted a conformance test for the NIST mindtct algorithm. The preliminary tests show that this methodology works well; nevertheless more extensive tests with several 100 images will be conducted in the near future. However there are still a number of open issues, which need to be addressed in future research: i) inclusion of a conformance rate for cores and deltas in the methodology, ii) quality controlled semi-automated definition of the fingerprint area, iii) quality controlled semi-automated definition of the average space between the ridge lines in an the image, iv) determination and validation of thresholds for every conformance rate such that minutiae extractor will be conformant only if the extractor exceeds all thresholds and v) validation of the clustering of clusters or clustering approach in accordance with the minutiae type.

7 Acknowledgement

We thank all the forensic experts at BKA that devoted their time to generate the ground truth database and thus make this work possible.

References

[bus09] C. Busch, D. Lodrova, E. Tabassi, W. Krodel: Semantic Conformance Testing for Finger Minutiae Data, Proceedings of IEEE IWSCN 2009, Trondheim, pp. 17-23, ISBN 978-82-997105-1-0, May 2009.

[ecc07] CEN TC 224 WG15 Identification card systems: European Citizen Card, 2007.

[iso05] International Standards ISO/IEC IS 19794-2: Information technology – Biometric data interchange formats – Part 2: Finger minutiae data, 2005.

[iso09a] International Standards ISO/IEC FDIS 29109-1 Information Technology - Conformance Testing Methodology for Biometric Interchange Formats defined in ISO/IEC 19794 – Part 1: Generalized Conformance Testing Methodology, Feb. 2009.

[iso09b] International Standards ISO/IEC FCD 29109-1 Information Technology - Conformance Testing Methodology for Biometric Interchange Formats defined in ISO/IEC 19794 – Part 2: Finger minutiae data, Feb. 2009.

[nist07] National Institute of Standards and Technology: Biometric Data Specification for Personal Identity Verification, NIST Special Publication 800-76-1, 2007, http://csrc.nist.gov/publications/nistpubs/800-76-1/SP800-76-1_012407.pdf.

[wk09] http://en.wikipedia.org/wiki/Apriori_algorithm. Last visited July 2009.

[nbis] NIST Biometric Image Software http://fingerprint.nist.gov/NBIS/index.html.