

A quality score honoring approach to semantic conformance assessment of minutiae-based feature extractors

Sebastian Abt, Christoph Busch, and Harald Baier
Hochschule Darmstadt – University of Applied Sciences
Center for Advanced Security Research Darmstadt (CASED)
{sebastian.abt, christoph.busch, harald.baier}@h-da.de

Abstract: The use of biometric systems is steadily increasing, which leads to heterogeneity and, thus, possibly interoperability issues. In order to address such issues, standards are developed by ISO/IEC Joint Technical Committee 1. In the case of minutiae-based fingerprint recognition, for instance, ISO/IEC 19794-2 defines biometric data interchange records that vendors should adhere to. To validate adherence to this standard, ISO/IEC 29109-2 specifies means to perform syntactic conformance testing. Yet, a specification of semantic conformance testing is missing and is currently being discussed as a working draft amendment of ISO/IEC 29109-2. In order to contribute to this development, this paper proposes a general semantic conformance testing framework. Especially, a formal semantic conformance computation model is proposed that can be used to assess a biometric systems vendor's semantic conformance according to a ground-truth data set and to compare different testing approaches. Furthermore, an instance thereof is proposed, that honors minutia quality scores when computing semantic conformance rates. Initial evaluation shows, that semantic conformance rates computed using this quality-honoring approach correlate with inter-vendor performance measures we would expect in a real-world scenario.

Keywords: Biometrics, Fingerprints, Minutiae, Interchange records, Semantic conformance testing.

1 Introduction

According to the International Biometric Group's (IBG) Market and Industry Report 2009–2014 [Int09], the annual industry revenues earned with biometrics in 2009 was at 3.4 billion US dollars and is predicted to increase to 9.3 billion US dollars until 2014, with the most predominant biometric technology currently being fingerprint recognition. According to [Int09], the fingerprint recognition technology market share in 2009 was more than 65%. Naturally, such an emerging market results in lots of different biometric systems vendors researching and developing proprietary solutions and competing for market share. Hence, in order to avoid vendor lock-ins within large installations of biometric systems (e.g. in border control and law enforcement) and to guarantee interoperability between different vendor's biometric systems, standardization of a biometric characteristic's digital representation, i.e. biometric samples or biometric templates, is elementary. Currently such standardization is mainly driven by the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) and especially their Joint

Technical Committee 1 (JTC1), which develops information and communication technology standards for business and consumer applications. In ISO/IEC 19794 family of standards, biometric data interchange records (BDIRs) for different biometric characteristics are standardized. Specifically, ISO/IEC 19794-2 [ISO05] specifies a biometric data interchange format for finger minutiae data, i.e. specific points that can be found within an individual's fingerprint and can be used to distinguish individuals. Derived from the informal definition given in ISO/IEC 19794-2 [ISO05], a minutia $m \in \mathcal{M}$, with \mathcal{M} denoting the set of all minutiae, can formally be defined as 5-tuple $m = \langle x, y, \theta, t, q \rangle$, with (x, y) being coordinates of a cartesian coordinate-system induced by a biometric sample, $0 \leq \theta \leq 255$ being the minutiae direction measured in units of $360/256$ degrees, t being a minutiae type (e.g. ridge line ending, ridge line bifurcation, or unknown), and $0 \leq q \leq 100$ being a quality value expressing the confidence of this minutia's attributes. However, currently no standard way of determining minutia quality is defined.

In order to verify the conformance of BDIRs produced by biometric systems to its base standard, ISO/IEC 29109 family of standards specifies conformance testing methodologies for biometric data interchange records as defined in ISO/IEC 19794 family. Among this ISO/IEC 29109 family of standards, ISO/IEC 29109-1 [ISO09a] defines a general framework for conformance testing and ISO/IEC 29109-2 [ISO09b] specifies a conformance testing methodology for finger minutiae records according to ISO/IEC 19794-2 [ISO05].

ISO/IEC 29109-1 [ISO09a] currently defines three levels of conformance testing. Level 1 (data format conformance) and level 2 (internal consistency checking) types of tests can roughly be categorized as being syntactic conformance tests which ensure that BDIRs exported by one biometric system can be imported and interpreted by another biometric system and vice versa. Level 3 conformance testing on the other hand can be regarded as semantic conformance testing which, after a BDIR has passed syntactic testing, shall ensure that a BDIR created by an implementation under test (IUT) is a faithful representation of the input biometric data record (IBDR). Syntactic, i.e. levels 1 and 2, conformance testing for finger minutiae records is currently covered in ISO/IEC 29109-2:2010. However, this version of the standard yet lacks semantic conformance, i.e. level 3, testing, which we contribute to in this paper.

The remainder of this paper is structured as follows: After this section introduced the topic, section 2 will give an overview on related work in this field. Section 3 will propose a semantic conformance testing framework by introducing a semantic conformance computation model. In section 4, a quality score honoring instance of this computation model will be proposed that can be used to compute an IUT's semantic conformance rate and its evaluation will be discussed in section 5. Finally, section 6 summarizes the results and concludes.

2 Related Work

Semantic conformance testing of minutiae-based feature extractors is a rather new field of research with limited amount of earlier work. In [BLT⁺09], Busch et al. propose a

semantic conformance testing methodology and especially highlight the importance of a reference data set, i.e. ground-truth data, in order to perform semantic conformance testing. In this work, reference data is manually assembled by dactyloscopic experts from the german federal criminal police office (BKA). Further, two measures, cr_{gtm} and cr_{agm} , are proposed which can be used to assess per-sample semantic conformance by measuring the ratio of ground-truth minutiae, i.e. minutiae marked by human experts, for which a mate was found and the ratio of automatically extracted minutiae, i.e. extracted using a specific vendor's feature extractor, placed within the fingerprint area, respectively. Based on this initial work, in [LBT⁺09] Lodrova et al. enhance the measures proposed in [BLT⁺09] by further adding a score cr_{amf} measuring the ratio of mated automatically extracted minutiae within the fingerprint area. Additionally, a hierarchical clustering algorithm to compile ground-truth minutiae from scattered expert markup is proposed in [LBT⁺09]. Using this clustering algorithm a ground-truth data set consisting of pairs (P_i, R_i) of biometric samples P_i , i.e. fingerprint images, and corresponding biometric references R_i is compiled from the scattered expert data. An additional minutiae-clustering algorithm based on the DBSCAN clustering algorithm has been proposed in [ABN10].

3 Semantic Conformance Testing Framework

In ISO/IEC 29109-1 [ISO09a], semantic conformance testing – or content checking – currently is defined *modality-independent* and *informally* as “a conformance testing methodology that tests that a BDIR produced by an IUT is a faithful representation of the IBDR subject to the constraints of the parameters in the metadata records” [ISO09a]. This type of definition may seem appropriate for use in case of ISO/IEC 29109-1 [ISO09a], as it is precise enough at an intuitive level on the one hand and on the other hand leaves enough room for a more precise refinement in subsequent or modality-dependent standards. From a modality-dependent point of view, however, the objectives of semantic conformance testing are more or less defined by the word *faithful* with no further detailed elaboration. Hence, in the following we propose a definition of *faithful representation* and propose a general and formal semantic conformance computation model, which is currently missing in literature.

3.1 A Minutiae-specific Definition of Faithfulness

Intuitively, a faithful representation of a biometric characteristic can be understood as any kind of biometric template or biometric data interchange record that contains exactly the same kind and amount of features that as well can be found on the true biometric characteristic. Unfortunately, however, transforming each and every feature of a specific biometric characteristic into a biometric template usually is not possible in an unbiased way as this at least is influenced by varying, usually non-linear, physical effects during data capture (e.g. moisture of the skin, pressure, temporary scars, noise and dust on the capture plate, etc.). Hence, we define a faithful representation as *a biometric template resulting from a*

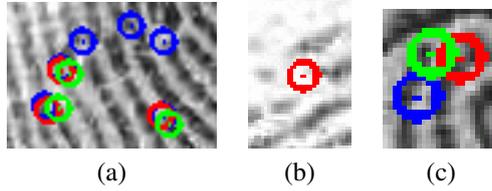


Figure 1: Illustration of minutiae as detected using three different feature extractors (red, green, blue) upon the same input biometric sample: (a) some minutiae not being detected by red and green extractors; (b) addition of spurious minutia; (c) minutiae mis-location around its true position.

noise-free and linear transformation applied to the input biometric sample's traits.

In the above proposed definition, a faithful representation is defined by the transformation of the input biometric sample's traits into a biometric template. We require this transformation to be both, noise-free and linear, which we explain using the concept of minutiae as features. Let \mathcal{M} be the set of all minutiae, $P_i = \{m_1, \dots, m_j\} \subseteq \mathcal{M}$ denote an input biometric data record consisting of j minutiae and $T_{k,i} = A_k(P_i) = \{m'_1, \dots, m'_n\} \subseteq \mathcal{M}$ denote an biometric data interchange record, or template, produced by vendor k 's feature extraction process A_k upon input of P_i . Then, for a feature extraction process to create a faithful representation of an IBDR P_i we require that

1. each specific true minutia $m_j = \langle x_j, y_j, \theta_j, t_j, q_j \rangle \in P_i$ equals its corresponding mated minutia $m'_n = \langle x'_n, y'_n, \theta'_n, t'_n, q'_n \rangle \in T_{k,i}$, i.e. $\forall \psi \in \{x, y, \theta, t\} : \psi = \psi'$, and that
2. $|P_i| = |T_{k,i}|$ holds, i.e. that the feature extraction process neither *misses* any minutiae, nor *adds spurious* minutiae.

3.2 Semantic Conformance Computation Model

However, feature extractors usually do not produce faithful BDIRs as defined above [TGSW09, NPL07]. Instead, often the number of minutiae detected as well as the values of specific minutiae attributes vary (cf. figure 1). The resulting difference between an indeed faithful representation $R_i \subseteq \mathcal{M}$ of P_i , i.e. a biometric reference, and $T_{k,i}$ can be used to formally define a general model for semantic conformance rate computation.

Let \mathcal{M} be the set of all minutiae, \mathcal{A} be the set of some vendor implementations under test and let $A_k \in \mathcal{A}$ denote a specific biometric systems vendor's implementation under test that, given a specific input biometric data record $P_i \subseteq \mathcal{M}$ as input, produces a biometric data interchange record $T_{k,i} = A_k(P_i) \subseteq \mathcal{M}$ as output. Further, let $GTM = \{(P_1, R_1, \omega_1), \dots, (P_{N_{GTM}}, R_{N_{GTM}}, \omega_{N_{GTM}})\}$ denote a ground-truth data set consisting of N_{GTM} triplets of input biometric data records P_l , corresponding known faithful biometric data interchange records R_l and weights ω_l . Additionally, in order to be able to define a computation model, we make use of a helper function $\mathcal{F} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$,

which takes a biometric data interchange record $T_{k,i} = A_k(P_i)$ as generated by an IUT A_k and a corresponding reference record $R_i \in (P_i, R_i, \omega_i) \in GTM$ as input and outputs a real valued number $f_{k,i}$. This helper function \mathcal{F} is used to quantify the *faithfulness* $f_{k,i}$ of BDIR $T_{k,i}$ as created by IUT A_k upon input of P_i according to reference R_i . Given these definitions, we define a semantic conformance computation model as four-tuple $SCM = (\mathcal{A}, GTM, \mathcal{F}, CR_{max})$ which is used to compute semantic conformance rates $CR(A_k)$ as follows:

$$\forall A_k \in \mathcal{A} : CR(A_k) = \frac{1}{N_{GTM}} \sum_{i=1}^{N_{GTM}} \omega_i \cdot \mathcal{F}(R_i, A_k(P_i)), \quad (1)$$

Put informally, the resulting conformance rates $CR(A_k)$ as computed according to $SCM = (\mathcal{A}, GTM, \mathcal{F}, CR_{max})$ are the average of the faithfulness $f_{k,i}$ of the biometric data interchange records $T_{k,i}$ as produced by A_k for the input biometric data records P_i and reference records R_i , weighted according to ω_i .

In order to be able to compare different instances of the above mentioned computation model SCM and in order to obtain reasonable results we require instances of this model to have the following properties:

1. The output values $CR(A_k)$ of the model, i.e. a semantic conformance rate belonging to vendor k 's feature extractor, must fall within the interval $[0, CR_{max}]$, i.e. $0 \leq CR(A_k) \leq CR_{max}$. A higher value $CR(A_k)$ indicates a higher semantic conformance, while a value of 0 indicates no semantic conformance at all.
2. The weighting parameters ω_i specified by the ground-truth data set GTM have to be from the interval $[0, 1]$, i.e. $0 \leq \omega_i \leq 1$.
3. Derived from requirements one and two, the values $f_{k,i}$ computed by the helper function \mathcal{F} must fall within the interval $[0, CR_{max}]$ as well, i.e. $0 \leq f_{k,i} \leq CR_{max}$.
4. The helper function \mathcal{F} used to quantify the faithfulness of a BDIR $T_{k,i}$ according to the given GTM has to be continuous.

The notion of semantic conformance rates according to the above given formal definition of a semantic conformance computation model deviates from the one given in Busch and Lodrova's work. Busch, Lodrova et al. [BLT⁺09, LBT⁺09] define semantic conformance rates for a specific input biometric data record and as a means to calculate semantic conformance rates propose three different measures, $cr_{gtm}, cr_{agm}, cr_{amf}$. In contrast, we define semantic conformance rates for a specific biometric systems vendor's implementation under test and for its computation make use of a function \mathcal{F} measuring the faithfulness of a biometric template $T_{k,i}$ created by vendor k 's algorithms upon input of a biometric sample P_i according to a provided reference R_i . Hence, semantic conformance rates in the sense of Busch, Lodrova et al. relate to what in this work is introduced as measures of faithfulness. The reason for this difference in notion is that we believe that an IUT specific conformance rate seems to be more useful, as having a semantic conformance rate indicating the vendor's implementation under test capabilities to generate faithful representations

of the input biometric data record enables the possibility to certify IUTs according to a conformance rate specific threshold CR_0 . In other words, using the notion of IUT specific conformance rates introduced in this paper we can attest a vendor k 's conformance if $CR(A_k) > CR_0$ holds.

In order to map the notion and formulae proposed in [BLT⁺09, LBT⁺09] to the proposed computation model, we can define a helper function \mathcal{F}_{BL} as follows:

$$\mathcal{F}_{BL}(R_i, T_{k,i}) = \lambda_{gtm} cr_{gtm}(R_i, T_{k,i}) + \lambda_{agm} cr_{agm}(T_{k,i}) + \lambda_{amf} cr_{amf}(R_i, T_{k,i}), \quad (2)$$

with $\lambda_{gtm}, \lambda_{agm}, \lambda_{amf}$ being conformance rate, in the sense of Busch et al., specific weights and $\lambda_{gtm} + \lambda_{agm} + \lambda_{amf} = 1$. Using this function, for a given ground-truth data set GTM and a set of vendor implementations under test \mathcal{A} , we derive an instance $SCM_{BL} = (\mathcal{A}, GTM, \mathcal{F}_{BL}, 1)$ of the semantic conformance computation model proposed above. Unfortunately, score fusion is not discussed in [BLT⁺09, LBT⁺09] and hence reasonable values for $\lambda_{gtm}, \lambda_{agm}, \lambda_{amf}$ yet have to be determined.

4 Quality-honoring Conformance Rate Computation

In the preceding section a general semantic conformance computation model $SCM = (\mathcal{A}, GTM, \mathcal{F}, CR_{max})$ has been introduced, which heavily depends on a ground-truth data set GTM and a function $\mathcal{F} : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ in order to compute semantic conformance rates $CR(A_k)$ for implementations under test $A_k \in \mathcal{A}$. What is still missing to be able to compute conformance rates is the discussion of a function $\mathcal{F}(R_i, T_{k,i})$ that can be used to quantify the faithfulness of a biometric template $T_{k,i} = A_k(P_i)$ generated by A_k upon input of P_i according to reference R_i . Ideally, this function should be able to quantify effects due to the following misbehavior as identified in the MINEX [Nat06] and MITIT [NPL07] projects:

1. *Misplacement of minutiae*, i.e. incorrect minutia attributes (x, y, θ, t) . Assessment of minutia quality q is excluded in MINEX and MITIT as there's currently no standard process defined which can be used for quality determination.
2. *Placement of spurious minutiae*, i.e. false addition of artificial minutiae.

As mentioned earlier, semantic conformance testing fundamentally bases on the notion of faithfulness. In section 3.1, faithfulness has been defined using minutiae attributes. Hence, this section develops a semantic conformance computation model that quantifies *minutiae misplacement* as well as *placement of spurious minutiae* in order to determine faithfulness of a biometric template $T_{k,i}$ with regard to biometric reference R_i . Additionally, this instance of SCM leverages minutiae quality scores, which are currently being neglected by other approaches. As this model is inspired by SCM_{BL} and basically is a quality-honoring version of it, this model is further called $SCM_{QBL} = (\mathcal{A}, GTM, \mathcal{F}_{QBL}, 1)$.

In order to assess faithfulness of a biometric template $T_{k,i}$ with regard to a reference R_i , SCM_{QBL} makes use of the following measures:

Minutiae misplacement Let \mathcal{M} denote the set of all minutiae, $T_{k,i} \subseteq \mathcal{M}, R_i \subseteq \mathcal{M}$, $m_j = (x_j, y_j, \theta_j, t_j, q_j) \in R_i$ denote the j -th minutiae in R_i and $m'_j = (x'_j, y'_j, \theta'_j, t'_j, q'_j) \in T_{k,i}$ denote an automatically extracted minutia spatially closest to m_j . Then minutiae misplacement is quantified using a helper function $\gamma_1(R_i, T_{k,i})$ that is defined as

$$\gamma_1(R_i, T_{k,i}) = \frac{1}{|R_i|} \sum_{j=1}^{|R_i|} \left(1 - (1 - \text{faith}(m_j, m'_j)) * \frac{\exp(q'_j/100)}{\exp(1)}\right)^2, \quad (3)$$

with $0 \leq \text{faith}(m_j, m'_j) \leq 1$ being a helper function used to quantify the faithfulness of minutia m'_j with respect to ground-truth minutia m_j defined as

$$\text{faith}(m_j, m'_j) = \begin{cases} 0, & \text{if } d_2(m_j, m'_j) > \text{tol}_d \\ f_j, & \text{otherwise} \end{cases}, \quad (4)$$

and f_j being the faithfulness score of minutiae m'_j that is closest to m_j , i.e. with euclidean distance $d_2(m_j, m'_j) \leq \text{tol}_d$. We choose $\text{tol}_d = W/2$, with W denoting the average ridge width in analogy to [LBT⁺09]. f_j is composed of three scores $s_j^{\Delta d}, s_j^{\Delta \theta}, s_j^{\Delta t}$ as follows:

$$f_j = \frac{s_j^{\Delta d} + s_j^{\Delta \theta} + s_j^{\Delta t}}{3}. \quad (5)$$

The scores $s_j^{\Delta d}, s_j^{\Delta \theta}, s_j^{\Delta t}$ are computed as follows:

$$s_j^{\Delta d} = \frac{\text{tol}_d - d_2(m_j, m'_j)}{\text{tol}_d}, \quad (6)$$

$$s_j^{\Delta \theta} = \frac{\pi - \min\{2\pi - |\theta_j - \theta'_j|, |\theta_j - \theta'_j|\}}{\pi}, \text{ and} \quad (7)$$

$$s_j^{\Delta t} = \begin{cases} 1, & \text{if } t_j = t'_j \\ 0, 25, & \text{if } t_j \neq t'_j \text{ and } t_j \text{ is unknown} \\ 0, & \text{otherwise} \end{cases}. \quad (8)$$

Spurious minutiae Let \mathcal{M} be the set of all minutiae, let $R_i \subseteq \mathcal{M}$ be a biometric reference belonging to biometric sample P_i , let $T_{k,i} = A_k(P_i)$ denote a biometric template generated by vendor k 's feature extraction algorithm upon input of biometric sample P_i and let $m = (x, y, \theta, t, q) \in R_i$, $m' = (x', y', \theta', t', q') \in T_{k,i}$ be ground-truth and automatically extracted minutiae, respectively. Further, let $S_{k,i} \subseteq T_{k,i}$ denote the subset of automatically extracted minutiae that do not fall within a tolerance-bound defined by tol_d of any ground-truth minutiae, i.e. $S_{k,i} = \{m' \in T_{k,i} | \nexists m \in R_i : d_2(m, m') \leq \text{tol}_d\}$. In essence, $S_{k,i}$ denotes the subset of spurious minutiae. Then, spurious minutiae are quantified as follows:

$$\gamma_2(R_i, T_{k,i}) = 1 - \frac{1}{|T_{k,i}|} \sum_{j=1}^{|S_{k,i}|} \frac{q'_j}{100}, \quad (9)$$

with q'_j specifying the quality score associated with j -th minutia $m'_j \in S_{k,i}$.

Using these rates, \mathcal{F}_{QBL} is computed according to

$$\mathcal{F}_{QBL}(R_i, T_{k,i}) = \lambda_1 \gamma_1(R_i, T_{k,i}) + \lambda_2 \gamma_2(R_i, T_{k,i}). \quad (10)$$

Terms $\gamma_1(R_i, T_{k,i})$ and $\gamma_2(R_i, T_{k,i})$ are quality-honoring modifications of rates $cr_{gtm}(R_i, T_{k,i})$ and $cr_{amf}(R_i, T_{k,i})$, respectively, with some further considerations. Most significantly, both measures utilize minutiae quality scores to weight penalties induced by misplaced or spurious minutiae. In order to achieve this, the automatically extracted minutia's quality value q' is divided by 100, leading to penalty factors in the interval $[0,01; 1]$. Thus, a higher quality score, i.e. higher confidence, leads to a higher-weighted penalty or a lower faithfulness score. If a specific minutia's quality value is set to 0, indicating that no quality information is available, the minutia is regarded to have a quality value of 100. The reason for this is that if a vendor does not provide minutia quality information, all minutiae have to be assumed to be of same high confidence. Hence, in these cases, penalties will be weighted maximally. Furthermore, neither measure distinguishes between minutiae placed within the fingerprint area, at the border or in the background area as this information is not available in minutiae-based templates and hence should neither affect comparison scores, nor has impact on a biometric template's faithfulness according to our definition.

In order to determine $\gamma_2(R_i, T_{k,i})$, i.e. to quantify minutiae misplacement, three scores are computed that assess spatial differences ($s_j^{\Delta d}$), deviations in minutiae angles ($s_j^{\Delta \theta}$) and differences in minutiae types ($s_j^{\Delta t}$). All those scores fall within the interval $[0; 1]$, with the extrema 0 denoting maximum difference and 1 denoting no difference. Besides $s_j^{\Delta t}$, all scores are continuous. $s_j^{\Delta t}$ measures differences in minutiae types. If the type t of a ground-truth minutiae m equals the type t' of an automatically generated minutia m' , i.e. if $t = t'$, then $s_j^{\Delta t}$ reaches maximum. If the type value of the ground-truth minutia m is set to unknown, or other, than $s_j^{\Delta t}$ is set to 0,25. The rationale behind this is that if dactyloscopic experts, or other reference feature extractors, are not able to correctly determine a minutia's type, the implementation under test should not be penalized rigorously as there's no reference. In any other case, $s_j^{\Delta t}$ is set to 0.

The score f_j basically is an unweighted measure of minutiae faithfulness, i.e. similarity of two minutiae m_j, m'_j . It is composed of the above mentioned scores $s_j^{\Delta d}, s_j^{\Delta \theta}, s_j^{\Delta t}$. For determining this measure, the average of those three scores is computed, i.e. all three scores influence f_j in the same manner. This decision was made by purpose as while according to experts from the german federal criminal police office, cartesian coordinates of minutiae are more distinctive than minutiae angles, which are more distinctive than minutiae type, their influence on the faithfulness according to the definition introduced in section 3.1 should be the same. However, if further research may suggest to assign different weights to the scores $s_j^{\Delta d}, s_j^{\Delta \theta}, s_j^{\Delta t}$, equation (4) can easily be adapted.

In order to compute $\gamma_2(R_i, T_{k,i})$, the penalty imposed due to differences in minutiae attributes, i.e. $(1 - faith(m_j, m'_j))$, is weighted according to the mated automatically generated minutiae's quality scores. This is done by using the weight factor $\frac{\exp(q'_j/100)}{\exp(1)}$. Further on, the thus resulting weighted penalty is subtracted from 1, i.e. to derive a minutia's quality-weighted faithfulness, and the the result is raised to the power of 2. The reason for this is that the non-linearity induced by the raise to the power of 2 is used to dilate

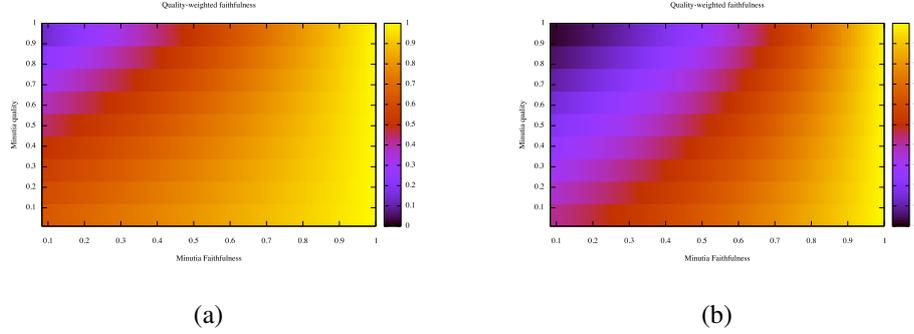


Figure 2: This figure illustrates the impact of quality scores and minugia faithfulness on the resulting quality-weighted faithfulness score. Part (a) illustrates scores without dilation due to the raise by power of two, part (b) depicts results using the formula given in formula (3).

the faithfulness scores achieved. This especially leads to a higher resolution of low-valued scores. The resulting score distribution for varying quality and minugia faithfulness parameters is depicted in figure 2. Especially, figure 2 (a) illustrates minugia faithfulness if the scores would not have been raised to the power of 2, while figure 2 (b) shows the dilated result according to the formula given in equation (3). As can be seen, due to this dilation a higher resolution of low-quality values, i.e. upper left corner of the images, is achieved.

Finally, \mathcal{F}_{QBL} makes use of weight parameters λ_1, λ_2 which specify the influence of $\gamma_1(R_i, T_{k,i})$ and $\gamma_2(R_i, T_{k,i})$, respectively, on the conformance rates $CR(A_k)$. For the time writing, we suggest assigning both a value $\lambda_1 = \lambda_2 = 0.5$ as the different measures' impact on semantic conformance is unclear. However, these parameters can be fine-tuned using an appropriate numeric optimization process if further insight is available.

5 Evaluation

In order to draw conclusions on SCM_{QBL} , first an impression of inter-vendor performance has to be obtained, which will be discussed in the upcoming section. After that, plausibility of SCM_{QBL} will be assessed and its performance will be evaluated with regard to SCM_{BL} .

5.1 Inter-Vendor Performance

Three software development kits, $A_{V_A}, A_{V_B}, A_{V_C}$ from different biometric systems vendors that claim conformance to ISO/IEC 19794-2 were available for testing. Using these SDKs, ISO/IEC 19794-2 conforming biometric templates have been generated and equal error rates have been computed for every comparator, reference extractor, probe extrac-

Avg	A_{V_A}	A_{V_B}	A_{V_C}	IUT	nnEER	$CR_{QBL}(\cdot)$	$CR_{BL}(\cdot)$
A_{V_A}	0.0415	0.0459	0.0493	A_{V_A}	0.0476	0.6214	0.6285
A_{V_B}	0.0455	0.0428	0.0519	A_{V_B}	0.0488	0.5133	0.6295
A_{V_C}	0.0495	0.0516	0.0376	A_{V_C}	0.0506	0.4039	0.6192

(a) (b)

Table 1: (a) lists average EERs of feature extractors A_{V_A} , A_{V_B} , A_{V_C} across those vendor’s comparators; rows denote reference generators, columns denote probe generators. (b) lists nnEERs computed according to equations (11) – (13) and conformance rates of IUTs A_{V_A} , A_{V_C} , A_{V_C} as computed according to SCM_{BL} and SCM_{QBL} .

tor permutation using the FVC2000 DB1¹ data set. In order to approximate performance achieved using a specific feature extractor’s templates in an heterogeneous environment and to rank the feature extractors according to that, we compute average equal error rates over all three comparators. The average EER values are comparator independent and listed in table 1 (a). Based on these values, we calculate feature extractor specific average *non-native equal error rates* (nnEER) as follows:

$$nnEER_{V_A} = \frac{EER_{V_A,V_B} + EER_{V_A,V_C} + EER_{V_B,V_A} + EER_{V_C,V_A}}{4} \quad (11)$$

$$nnEER_{V_B} = \frac{EER_{V_B,V_A} + EER_{V_B,V_C} + EER_{V_A,V_B} + EER_{V_C,V_B}}{4} \quad (12)$$

$$nnEER_{V_C} = \frac{EER_{V_C,V_A} + EER_{V_C,V_C} + EER_{V_A,V_C} + EER_{V_B,V_C}}{4} \quad (13)$$

The rationale of this is that nnEERs should give an impression of equal error rates in heterogeneous environments – those scenarios we actually are interested in. From a higher nnEER, we conclude lower performance and vice versa. As these values are computed on comparator independent EERs, i.e. values listed in table 1 (a), the resulting nnEER values (cf. table 1 (b)) are assumed to give a rough ranking of the feature extractors’ performances, independent of template comparators’ performances. Thus, based on the given data, A_{V_C} seems to be the worst feature extractor, while A_{V_A} seems to be the best and A_{V_B} the second best performing feature extractors.

5.2 SCM_{QBL} Evaluation

In order to empirically assess the plausibility of semantic conformance rates computed according to SCM_{QBL} , the following specific tests were performed. In all cases, parameters λ_1 and λ_2 (cf. equation (10)) were set to 0.5:

1. *Equality test*: Conformance rates were computed using the same set of biometric templates both, as reference and as probe. The result of this test is that for all cases a maximum semantic conformance rate of 1 is computed. This is the expected

¹<http://bias.csr.unibo.it/fvc2000/>

outcome as if biometric templates under test equal the references in ground-truth, all biometric templates are faithful.

2. *Empty template test*: Conformance rates were computed using a modification of the ground-truth data set as probe data set. In this modified probe data set, no biometric template contains a single minutia. This results in a conformance rate of 0, which is the expected outcome.
3. *Single mate test*: Again, a modification of the ground-truth data set is used as probe data set. The probe data set is modified in such that all but one minutiae are synthetically misplaced to not mating with any minutiae in the reference data set. I.e., every biometric probe contains a single faithful minutia and many spurious minutiae. All quality values were additionally set to 100. This resulted in a rather small conformance rate (0.0163), which is expected for this type of test.

Based on these empirical analyses we argue that SCM_{QBL} computes plausible semantic conformance rates, such that higher similarity of biometric references and biometric probes leads to higher semantic conformance rate and vice versa.

Further to this basic plausibility testing, we try to show that semantic conformance rates computed by SCM_{QBL} correlate with expected real-world inter-vendor performance. Especially, this means that we try to show that a vendor ranking induced by conformance rates correlates with the nnEER-based ranking given in table 1 (b). For this, semantic conformance rates are computed for all feature extractors using the data set described in [BLT⁺09], with parameters λ_1, λ_2 being set to 0.5, each. The results of this computation are given in table 1 (b). As can be seen, conformance rates resulting from SCM_{QBL} correlate with nnEER rates, i.e. $CR_{QBL}(A_{V_A}) > CR_{QBL}(A_{V_B}) > CR_{QBL}(A_{V_C})$.

In order to get an impression of the performance of SCM_{QBL} in contrast to SCM_{BL} , the same computations have been performed for SCM_{BL} . The results of this are listed in column $CR_{BL}(\cdot)$ of table 1 (b). Conformance rates $CR_{BL}(\cdot)$ are derived by specifying $\lambda_{gtm} = \lambda_{agm} = \lambda_{amf} = 1/3$. Based on these results SCM_{QBL} seems to outperform SCM_{BL} for the given data set and feature extractors, as the ranking induced by SCM_{BL} does not correlate with nnEER values.

6 Summary and Conclusions

In this paper, a semantic conformance testing framework for minutiae-based feature extractors has been proposed. Especially, a general formal semantic conformance computation model $SCM = (\mathcal{A}, GTM, \mathcal{F}, CR_{max})$ has been derived from informal descriptions found in ISO/IEC 29109 family of standards, which can be used to develop and compare different instances thereof. The basic underlying concept of this general model is the notion of faithfulness, which has been formally defined for minutiae-based biometric interchange records specified in ISO/IEC 19794-2 [ISO05]. This computation model has been used to derive a quality-honoring instance, which has been compared to an existing approach. Initial results show that the quality-honoring approach SCM_{QBL} proposed in this

paper outperforms SCM_{BL} , as the first one correlates with inter-vendor performance that would be expected in real-world scenarios which the latter does not. However, the authors are aware that the number of SDKs available for testing was limited and further analyses should be carried out. For future work, the author's propose to perform in detail analyses of the *minutiae misplacement problem (MMP)*, i.e. the impact of minutiae misplacement on comparison scores, as this not only seems to be one of the most pre-dominant issues with automatic feature extraction, but also can easily be measured and used for semantic conformance testing.

References

- [ABN10] Sebastian Abt, Christoph Busch, and Claudia Nickel. Applikation des DBSCAN Clustering-Verfahrens zur Generierung von Ground-Truth Fingerabdruck-Minutien. In *Proceedings of the Special Interest Group on Biometrics and Electronic Signatures (BIOSIG) 2010*, pages 95–106, 2010.
- [BLT⁺09] Christoph Busch, Dana Lodrova, Elham Tabassi, Wolfgang Krodel, and Martin Drahan-sky. Semantic Conformance Testing for Finger Minutiae Data. In *Proceedings of IEEE International Workshop on Security and Communication Networks (IWSCN) 2009*, pages 17–23, 2009.
- [Int09] International Biometric Group. Biometrics Market and Industry Report 2009–2014, 2009.
- [ISO05] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 19794-2 Information technology – Biometric data interchange formats – Part 2: Finger minutiae data. International Organization for Standardization, 2005.
- [ISO09a] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 29109-1 Information technology – Confor-mance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 – Part 1: Generalized conformance testing methodology. International Organiza-tion for Standardization, 2009.
- [ISO09b] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 29109-2 Information technology – Confor-mance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 – Part 2: Finger minutiae data. International Organization for Standardization, 2009.
- [LBT⁺09] Dana Lodrova, Christoph Busch, Elham Tabassi, Wolfgang Krodel, and Martin Dra-hansky. Semantic Conformance Testing Methodology for Finger Minutiae Data. In *Pro-ceedings of the Special Interest Group on Biometrics and Electronic Signatures (BIOSIG) 2009*, pages 31–42, 2009.
- [Nat06] National Institute of Standards and Technology. MINEX: Performance and Interoperabil-ity of the INCITS 378 Fingerprint Template. NIST Interagency Report 7296, 2006.
- [NPL07] NPL Management Ltd. / MTIT Project. MTIT Deliverable 6.2: Research report on minu-tiae interoperability tests. <http://www.mtitproject.com/DeliverableD62.pdf>, 2007.
- [TGSW09] Elham Tabassi, Patrick Grother, Wayne Salamon, and Craig Watson. Minutiae Inter-operability. In *Proceedings of the Special Interest Group on Biometrics and Electronic Signatures (BIOSIG2009)*, pages 13–30, 2009.