Characterizing Minutia Extractors for Semantic Conformance Testing

Terry P. Riopka and Limin Ma

Abstract—**Recent effort has been expended toward the development of a methodology for semantic conformance testing of standardized biometric interchange records. Specifically, the primary motivation is to evaluate the degree to which a generated compact interchange record such as a minutia template is a faithful representation of the original digital representation of the biometric characteristic (i.e. image of a finger pattern). In and of itself, such an evaluation would seem to have intrinsic value, as it is almost obvious that this would be of paramount importance to fingerprint minutia extractors. However, in this paper, we provide empirical data that suggests some caution in coming to such a conclusion. We also propose several other approaches to evaluating minutia extractors that might augment their characterization for the purpose of comparison and evaluation.**

I. Introduction

FINGERPRINTS are a powerful biometric because of the uniqueness and relative stability of the feature signature of the human fingerprint pattern. The last decade has seen a dramatic improvement in the ability of automated image analysis systems to quickly, accurately and consistently extract features from fingerprint patterns. These factors have contributed to the success of fingerprints as a biometric for person identification systems.

It is generally acknowledged that greatest interoperability between fingerprint recognition systems is achieved by the interchange of fingerprint image data. However, the use of image data can be costly with respect to storage requirements, transmission time, image compression/de-compression time and minutia extraction time. The use of minutia templates offers a more resource-efficient and cost-effective alternative.

It has been shown [2] that minutia templates consisting of only minutia location and angle data can provide sufficient information for effective matching. Furthermore, a significant degree of vendor interoperability can be achieved using such templates, with an acceptable level of error [2][3]. In fact, the National Institute of Science and Technology (NIST) now provides on-going testing of INCITS 378 fingerprint template interoperability via the Minutiae Interoperability Exchange Test (MINEX), which periodically adds minutia extraction and matching algorithms to a list of vendor algorithms that conform to the interoperability criteria.

Manuscript received June 4, 2010. This work was supported by Aware, Inc.

T. P. Riopka is Director of Research at Aware, Inc., Bedford, MA 01730 USA (phone: 781-687-0473; fax: 781-276-4001; e-mail: riopka@kantbelievemyeyes.com).

L. Ma is a senior imaging scientist with Aware, Inc., Bedford, MA 01730 USA. (e-mail: lma@aware.com).

978-1-4244-7580-3/10/$26.00 ©2010 IEEE

As an adjunct to MINEX, recent effort has been expended to develop a methodology for semantic conformance testing of fingerprint minutiae data [1]. The primary motivation is to evaluate the degree to which a generated compact interchange record such as a minutia template is a "faithful representation of the initial digital representation (e.g. the fingerprint image) of the biometric characteristic (e.g. the finger pattern)" as outlined in [1]. Such a goal seems to have intrinsic worth, as the current standard for the fingerprint interchange record involves a very limited set of features deemed to constitute the "faithful representation" of the biometric signature, specifically: the location, type and angle of minutiae. Because of its compactness, the accuracy of the construction of that representation directly impacts its utility with respect to matching. Consequently, the more accurate the representation, the more likely the biometric can be successfully matched.

Despite advances in scanning devices, fingerprint biometrics are still inherently noisy due to a multitude of factors: image deformation due to finger pressure, image superposition from finger platen residue-images, and/or image distortion as a result of trying to image a 3D object onto a 2D flat plane. All of these can contribute to producing a flawed reproduction of the biometric impacting the signal's fidelity. In addition, inherent characteristics of the finger such as scars, dried or cracked fingers can further degrade the fingerprint pattern, creating a significant challenge for automated minutia extraction algorithms.

The goal of creating a perfect representation (or at least one consistent with human experts) with respect to the standard features is clearly desirable. However, if we assume an imperfect result (as we must), there will always be some balance between true and false detections, as there is in any pattern recognition problem. In other words, we know what the optimal balance is (only true minutiae are found), but what is the range for the acceptable sub-optimal solution? Furthermore, how can that range be objectively delineated given the close relationship between fingerprint image quality and the difficulty of minutia extraction? This paper seeks to provide insight into these questions by treating the minutia extraction algorithm as a system requiring characterization, analogous to studying the response of a dynamic system to that of an impulse response.

In this research, we provide empirical data that emphasizes the complexity of the minutia extraction problem and provide some insight into some additional possible methods for the characterization of minutiae extractors.

The paper is organized as follows. Section II gives a brief introduction to the latest proposal for semantic conformance testing of finger minutia data. Section III describes the method used to obtain ground truth minutia data from the synthetic images used for the experiments presented in this paper. Section IV offers an analysis and discusses the results. Section V proposes a new interoperability test based on conformance ratios. Finally, the paper concludes with some speculation and suggestions for future research.

.

II. SEMANTIC CONFORMANCE TESTING

Proposed measures for semantic conformance of finger minutiae data involve several ratios, specifically:

1. a (modulated) ratio of true minutiae found by the extractor to the actual number of true minutiae, referred to as the first conformance rate (cr_{gtm})
2. a (modulated) ratio of the number of false minutia found by the extractor outside of the fingerprint area to the total number of minutiae found, referred to as the second conformance rate
3. the ratio of true minutiae found by the extractor to the total number of minutia found in the fingerprint area, referred to as the third conformance rate (cr_{amf}).

For the experiments in this paper, we use the ratios cr_{gtm} and cr_{amf} as metrics related to the operation of the extractors and as metrics for fingerprint images. For example, a fingerprint image can be considered to have a certain cr_{gtm} or cr_{amf} value as "measured" by a given extractor.

The qualifier "modulated" is used above to refer to the additional penalty terms associated with minutia type and angle error, neither of which are used in this paper.

for future research.

.

III. GROUND TRUTH AND EXPERIMENT DESIGN

Ground truth was obtained as follows. Using the program SFinGe, developed by Cappelli, Maio and Maltoni [4][5], a set of 100000 identities, each with 100 different impressions with no synthetic perturbations (except for rotation) were generated. This yielded a raw black and white fingerprint pattern used as the initial template for synthetic fingerprint generation, one that could be relatively easily analyzed for minutia. The 100 "clean" patterns for each identity were generated so that a statistical analysis of the minutia locations could lead to a more robust estimation of the exact location of the ground truth minutiae. Furthermore, because the more realistic "noisy" versions of each of these identities (generated later) could vary in the size of the impression and rotation, it was important to obtain a "map" of the entire list of possible ground truth minutia. In this way, the correct region in the "global-template" for ground truth minutiae could be identified when registering the synthetically noisy image with the global template.

Once the global-templates were generated, the same random number generator seeds were used to obtain a set of 10 impressions of varying quality for each of the 100000 identities. The distribution, in terms of NFIQ value is shown in table 1.

TABLE I
NFIQ DISTRIBUTION

NFIQ	Count	Percent(%)
1	525698	52.57
2	228923	22.89
3	195973	19.60
4	8983	0.90
5	40043	4.00

Table 1: Quality distribution of the synthetic database generated for the experiments.

For a given extractor, all 10x100000 images could then be analyzed, minutia located, registered with the global-template, sorted and binned according to the measured values of cr_{gtm} and cr_{amf} . Binning was done in increments of 0.1 for each of the ratios, for a total of 10x10 = 100 bins for all combinations of cr_{gtm} and cr_{amf} .

IV. RESULTS AND DISCUSSION

In our experiments, 3 MINEX-compliant extractors (referred to as C1, C2 and C3) and one non-MINEX-compliant minutia extractor (referred to as N1) were tested. Figure 1 shows the distribution of the fingerprint images as binned using each of the four extractors. Darker areas in the images indicate regions of the plot with larger numbers of fingerprint images. Note that the same fingerprint image can be binned differently depending on the cr_{gtm} and cr_{amf} values obtained for a given extractor.

The similarity of the MINEX-compliant extractors is quite apparent. The operating region of the MINEX extractors, not surprisingly, is concentrated in the upper regions of the plot (high cr_{gtm} and high cr_{amf}). It is interesting to note the relatively consistent location of the peak cr_{gtm} hovering around 0.7. No doubt this peak cr_{gtm} location could be increased, but at the cost of increased false minutiae. One would expect minutia extractors to strive for high numbers of true minutia, yet try to minimize the number of false minutiae.

Figure 2 shows corresponding performance plots for the same extractors (each using their own matchers). Here, dark regions indicate poor performance, light regions indicate good performance. C1 and C2 are similar, although C1 is somewhat skewed to the left of the diagonal, while C2 is somewhat skewed to the right. C3 is definitely skewed to the left. These essentially show how the matchers respond to the minutia extractors, as well as give insight into the matching strategy

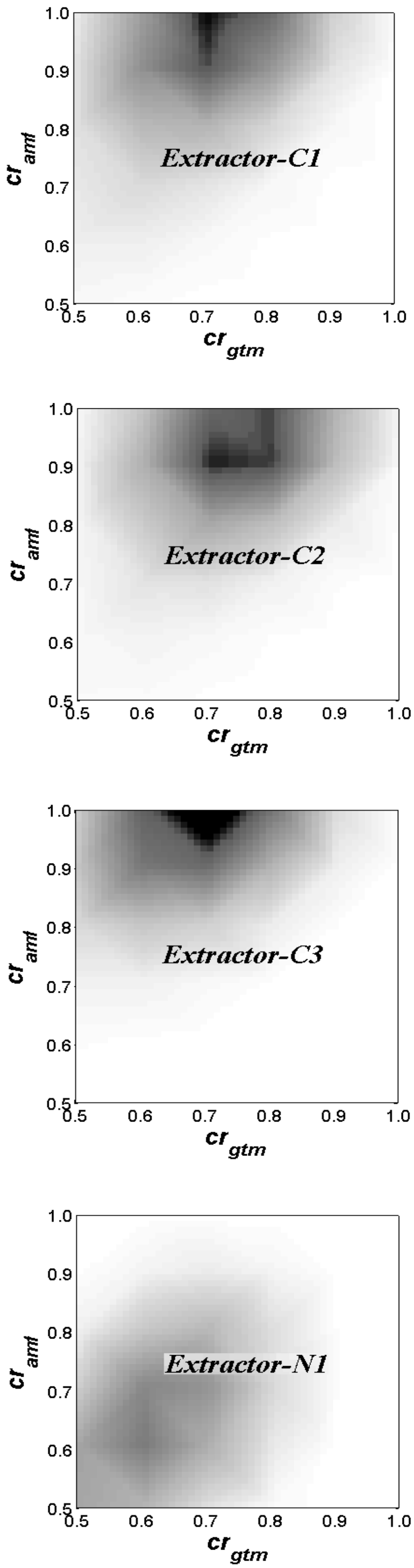

Figure 1: Conformance plots for MINEX-compliant extractors (C1-C3) and a non-compliant extractor (N1). Darker areas indicate higher image counts.

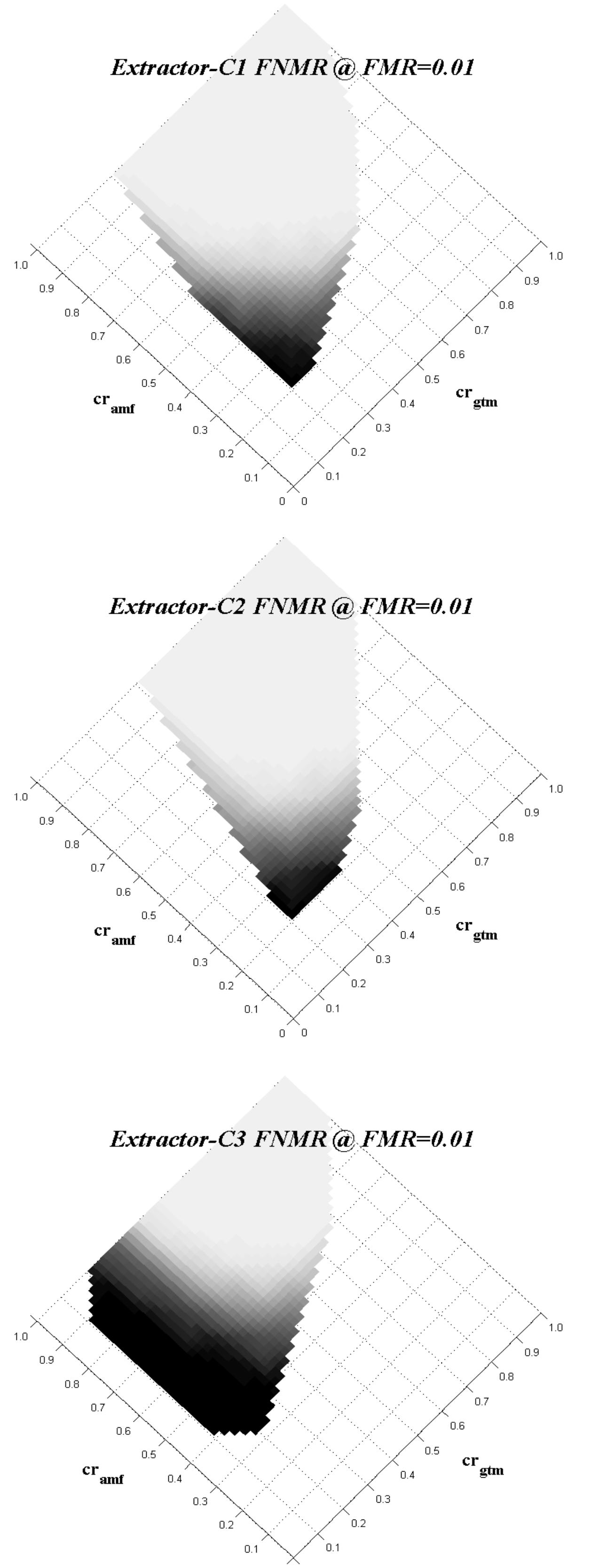

Figure 2: Performance plots for MINEX compliant extractor/matcher combinations for all bins containing more than 100 different identities. Darker areas indicate higher FNMR (poorer performance).

used by each. For example, the skew of C2 to the right shows a slight bias towards the region where cr_{amf} is lower but cr_{gtm} is slightly higher (in comparison to C1), indicating that the extractor seems to produce a greater degree of true minutiae at the cost of additional false minutiae. C3 appears to perform quite well, but the matcher is apparently very sensitive to image quality. This is reflected in the very poor performance of C3 on low quality images (see figure 3 for NFIQ5). The extractor appears to be tuned to the matcher, which seems to require high quality templates to perform well.

A reasonable operating region is where minutia quality filtering is relatively strict, resulting in an extractor that is very discriminating with respect to the quality of minutiae – the likelihood of false minutiae is low, but the caveat is a corresponding decrease in the total number of true minutiae found. Matchers operating in this region would be expected to perform quite well on high quality fingerprints (where the number of true good quality minutia is high), but might falter on low quality fingerprints where a strict filter might severely limit the number of minutiae that would be made available for matching. This seems to be the case with C3.

Less desirable is operating in the region where cr_{gtm} is high and the cr_{amf} ratio is very low. This corresponds to an operating point with a very large number of false minutiae relative to the total minutiae found. Even though the number of true minutiae found is high, the confusion caused by the excessive number of false minutiae degrades matcher performance. None of the extractors tested seemed to operate in this region.

Performance plots of subsets of IDs based on NFIQ are shown in figure 3. Darker areas indicate higher FNMR (poorer performance). The lightest (negative) regions are bins for which not enough data was available for performance measurements. Not surprisingly, performance worsens as NFIQ increases and for cells corresponding to lower conformance rates. What is notable is the sensitivity of performance to levels of the conformance rate. A mere change of 0.1 in either fraction results in a significant change in performance. The graphs clearly show appropriate trends in both directions, *i.e.* performance tends to degrade monotonically with a decrease in either conformance rate.

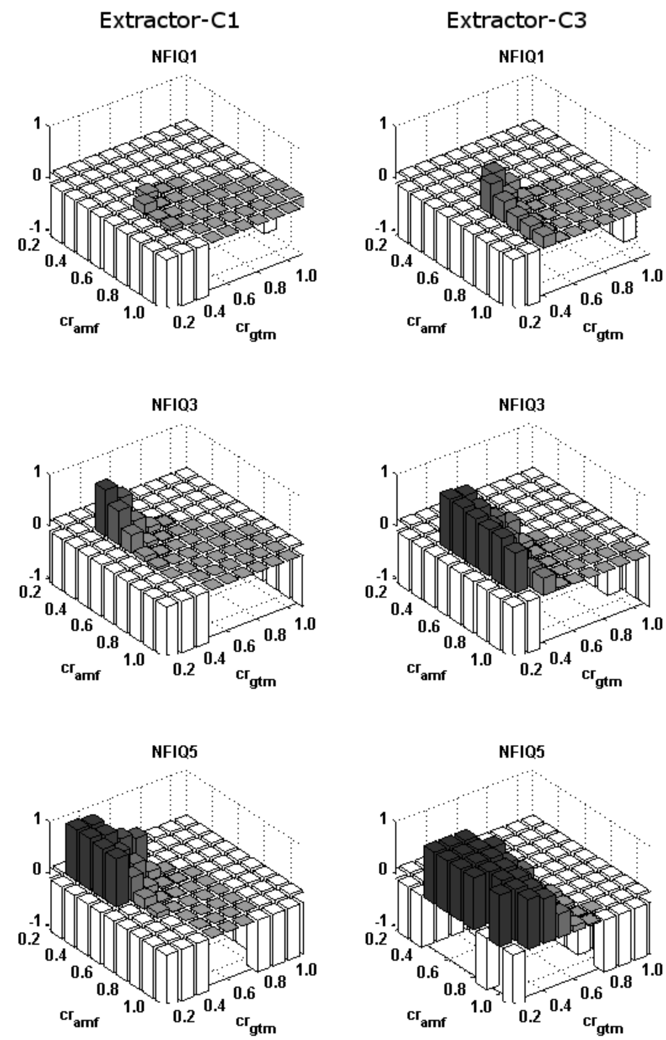

Figure 3: Performance plots for C1 and C3 using their respective matchers for subsets of IDs based on NFIQ at an FMR of 0.01.

V. An Interoperability Test for Extractors

If we consider the conformance rate of an extractor on a single image and take that as a "measurement" of the quality of the image with respect to that extractor, it is clear that this measurement may be different for different extractors. However, the difficulty of extracting minutiae and the likelihood of being wrong are as much a function of the quality of the image as they are a function of the extractor, perhaps more so, given the relative homogeneity of the technology. As far as interoperability is concerned, the degree to which two or more extractors can agree on the "quality" of a set of images, as "measured" by the conformance rate could be a useful metric.

To this end, an analysis was done to determine the fraction of the database (commonality ratio) for which two or more extractors "agree" on the measured ratios of cr_{gtm} and cr_{amf} within some tolerance. In this case, the tolerance was varied from 0 to 4 bin increments, corresponding to increments of 0.1, i.e. from 0 to 0.4.

In each case, a reference extractor is first selected. One by one, the [cr_{gtm} , cr_{amf}]ratio combination of each image in the reference is compared to the [cr_{gtm} , cr_{amf}]ratio of the same image in all other extractors under consideration. If the other ratio combinations are within the specified tolerance, that image is counted and contributes to the computation of the commonality ratio. This was repeated using each of the extractors as a reference. The final graphs in the following figures is the average commonality ratio over all such graphs obtained using each of the extractors as a reference. Standard deviations were negligible and are not shown because they

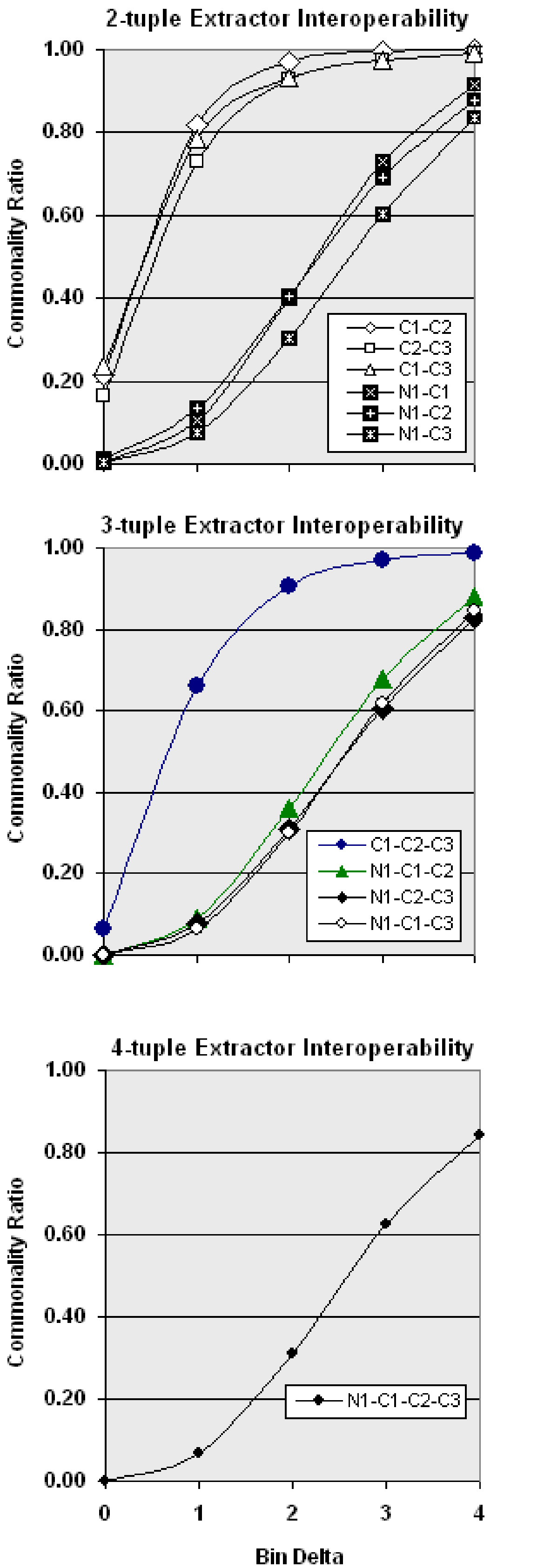

Figure 4: Extractor interoperability graphs showing how various combinations of extractors might be compared in various interoperability scenarios.

were too small to be depicted.

Figure 4 shows three different scenarios. The 2-tuple interoperability graph shows a hypothetical situation where one might be interested in determining which pair of extractors is most interoperable. Not surprisingly, all pairs in which the members are MINEX-compliant seem to perform comparably with respect to the commonality ratio. As an example interpreting the plot, the C1-C2 combination appears to be most interoperable with C1 and C2 agreeing on approximately 80% of the images at a tolerance of 1 bin (+/-0.1) . This goes up to about 95% for a tolerance of 2 bins (+/- 0.2).

The 3-tuple interoperability graph shows a hypothetical situation where one might be interested in determining which triplet of extractors is most interoperable. The clear winner, as expected, is the MINEX-compliant triplet. Note that this result really does not explicitly say anything about performance. For example, a very good extractor could be developed that might operate entirely within the [1.0, 1.0] region. Clearly, its performance would be outstanding, but it would undoubtedly fail the interoperability test.

The final 4-tuple interoperability graph shows a hypothetical situation where one might be testing an extractor for membership in a group *e.g.* testing N1 in the MINEX group. In such a case, some criterion, such as a set of thresholds, could be selected to determine membership.

Note that such an interoperability test might be significantly less sensitive to the precise quality distribution of the database used for testing.

VI. Conclusion

The empirical data presented in this paper show the utility of automated ground truth for minutia detection given synthetic fingerprint generation. Despite its artificiality, it can have value in helping to characterize fingerprint algorithms. Specifically, this investigation has shown that conformance rates can be used to measure the degree of interoperability between minutia extractors and may be useful in studying their behavior independent of matchers. What seems evident is the complexity of that behavior, suggesting the need for more complex characterization. Minutia extractors perform complex tradeoffs between true and false detections that may not be adequately summarized in a single number.

References

[1]. Busch, C., Lodrova, D., Tabassi, E. and Krodel, W., “Semantic Conformance Testing for Finger Minutiae Data”, in *Proceedings of the IEEE International Workshop on Security and Communication Networks (IWSCN)*, Trondheim, ISBN 978-82-997105-1-0, pp. 17-23, 2009.

[2]. Grother, P. J. et al., “MINEX: Performance and Interoperability of the INCITS 378 Fingerprint Template”, National Institute of Standards and Technology. NISTIR 7296, 2006.

[3]. Grother, P. J. et al., “ISO/IEC 19795-4 Biometric Performance Testing and Reporting: Performance and Interoperability Testing of Data Interchange Formats”, JTC1:: SC37:: WG5, 2006.

[4]. R. Cappelli, A. Erol, D. Maio and D. Maltoni, “Synthetic Fingerprint-image Generation”, in *Proceedings 15th International Conference on Pattern Recognition (ICPR2000)*, Barcelona, vol.3, pp.475-478, September, 2000.

[5]. R. Cappelli, D. Maio and D. Maltoni, “Synthetic Fingerprint-Database Generation”, in *Proceedings 16th International Conference on Pattern Recognition (ICPR2002)*, Québec City, vol.3, pp.744-747, August 2002.