

Fingerprint Quality Assessment: Matching Performance and Image Quality

Zhigang Yao, Jean-Marie Le Bars, Christophe Charrier, Christophe Rosenberger
e-mail: {jean-marie.lebars; christophe.charrier}@unicaen.fr
{zhigang.yao; christophe.rosenberger}@ensicaen.fr

Abstract This article chiefly focuses on Fingerprint Quality Assessment (FQA) applied to the Automatic Fingerprint Identification System (AFIS). In our research work, different FQA solutions proposed so far are compared by using several quality metrics selected from the existing studies. The relationship between the quality metric and the matching performance is specifically analyzed via such a comparison and an extra discussion based on the sample utility. This study is achieved via an interoperability analysis by using two different matching tools. Experimental results represented by the global Equal Error Rate (EER) further reveal the constraint of the existing quality assessment solutions in predicting the matching performance. Experiments are performed via several public-known fingerprint databases.

1 Introduction

The disadvantage of biometric recognition systems is chiefly attributed to the imperfect matching in contrast with traditional alphanumeric system. Because of this, sample quality is more important for image-based biometric systems, and so is fingerprint image used for the AFIS. Matching of fingerprint images is generally divided into three classes: correlation-based, image-based and minutiae matching, among which the last one is acknowledged as the primary solution so far [12]. In this case, good quality sample is basically a prerequisite for extracting reliable and sufficient minutia points, and is hence the essential factor for the overall matching performance. The effect of sample quality to the matching performance is defined as the utility of a biometric sample [1]. Therefore, most of the FQA approaches (or fingerprint quality metrics) rely on two aspects: subjective assessment criteria of the pattern [9] and sample utility. In addition, most of the quality metrics are also evaluated in terms of the utility. [3]. However, this property is limited by matching configurations, *i.e.* sample utility varies as the matching algorithm changes because

no matching approach proposed so far is perfect or robust enough in dealing with different image settings though their resolution is similar to each other (normal application requires gray-level images of 500-dpi according to the ISO).

This chapter compares the existing solutions of the FQA in terms of a methodological categorization [30]. Such a comparison analyzes whether those quality metrics based on multi-feature are really able to take the advantages of the employed features. Similarly, quality assessment approaches rely on a prior-knowledge of matching performance still need discussion, especially the prediction to the matching performance. Our work gives a study of these potential problems in an experimental manner. Each of the selected quality metrics in this chapter represents a typical solution in the existing studies.

This chapter is organized as follows: Section 2 presents a brief review of the categorization of the existing FQA solutions. In section 3, the description of trial fingerprint quality metrics is given. Experimental results are given in 4. Section 5 concludes the paper.

2 Background

Yao *et al.* [30] categorize prior work in FQA into several classes in terms how this problem is solved. Typical FQA solutions can be summarized as:

1) Single feature-based approaches: these could be further divided into solutions rely on the feature itself or a regularity [18] observed from the employed feature. For instance, standard deviation [14] at block-wise is a brief factor which somehow measures the clarity and differentiates the foreground block of fingerprint. Some studies also obtain relatively good result by using a single feature, such as the Pet's hat wavelet (CWT) coefficients [17] and the regularity of fingerprint Discrete Fourier Transform (DCT) [7], Gabor feature [19], etc. These features also represent the solution of FQA in different domain. In addition, the 'relatively good result' here means that those solutions perform well in reducing the overall matching performance because we believe that the evaluation of a quality metric is basically a biometric test which involves both genuine matching and impostor matching errors.

2) FQA via segmentation-like operations: these kinds of solutions are divided into two vast classes at first, including global-level and local-level approaches. Mostly, local-level approaches estimate a quality measure to a fingerprint block in terms of one or several features or indexes, such as directional information, clarity, etc [5, 16, 14, 11]. Some other local-level approaches choose to determine whether a block is a foreground at first [23], and then give a global quality measure to the fingerprint image. This type of solutions implemented globally are further divided as non-image quality assessment and image-based approach. Yao *et al.* [30] propose one FQA approach by using only minutiae coordinates, meaning that no real image information is used in assessing fingerprint quality. Image-based solutions are basically achieved by performing a segmentation at first, and then estimate the quality of the foreground area according to one or more measurements [30].

3) FQA approaches by using multi-feature: these could be carried out by using either fusion or classification. For example, some studies combine several quality features or indexes together via a linear (or weighted) fusion [16, 6, 25, 8]. The linear fusion is basically used for a specific scenario because coefficient is a constraint of this kind of solution. Similarly, fusion of multiple features or experts outputs could also be achieved via other more sophisticated approaches such as Bayesian statistics [20], Dezert-Smarandache (DS) theory [2], etc. The effectiveness of the fusion algorithm itself and differences between multiple experts outputs impact on the fused result. For instance, it is quite difficult to look out an appropriate way to fuse results generated by two different metrics, where one gives continuous output and another generalize a few discrete numbers. This chapter considers only FQA problem of the AFIS rather than any multi-modal, score/cluster-level fusion and some fusion related issues.

FQA via multi-feature classification [16, 15] basically employs one (or more) classifier(s) to classify fingerprint image into different quality levels. Obviously, this kind of solution depends on the classifier itself. In addition, the robustness and the reliability of the prior-knowledge used by learning-based classification or fusion also impacts on the effectiveness of the quality metric, particularly when generalizing a common solution such as the State-of-the-Art (SoA) approach [24].

In addition, some studies propose to use knowledge-based feature by training a multi-layer neural network [18]. However, it is essentially an observed regularity of the learnt feature and external factors such as classifier and tremendous training data set are also required.

According to the discussion above, one can note that fingerprint quality is still an open issue. Existing studies are mostly limited in these kinds of solutions, where learning-based approaches are chiefly associated with the prior knowledge of matching performance which is debatable for a cross-use. Grother *et al.* [12] have introduced that quality is not linearly predictive to the matching performance. This chapter gives an experimental study to analyze this problem by comparing FQA approaches selected from each of the categorized solutions.

3 Trial Measures

In order to observe the relationship between the quality and the matching performance, several metrics carried out by using each of the categorized solution are employed in this study, given as follows.

3.1 Metrics with Single Feature

As mentioned in Section 2, we first choose one quality metric generalized by using a single feature. The selected metric is implemented via the Pet's Hat continuous

wavelet, which is denoted as the CWT as mentioned in Section 2. The CWT in a window of W is formulated as

$$C_{wt} = \sqrt{\frac{\sum_W |c_i|}{W}} \quad (1)$$

where c_i is wavelet coefficient and the windows size depends on the image size, for example, 16 pixels for gray scale images of 512dpi. In our study, the CWT is implemented with two default parameters, a scale of 2 and angle of 0. We choose

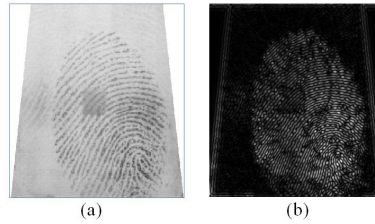


Fig. 1: Example of CWT (b) of a fingerprint image (a).

this quality metric because it outperforms the SoA approach in reducing the overall error rate for some different image settings. **Note** that the resolution of fingerprint image is about 500-dpi, which is the minimum requirement of the AFIS [10].

3.2 Segmentation-based Metrics

Fingerprint segmentation is one way to separate the foreground area (ridge-valley pattern) from the background (vacuum area) formed by input sensor(s). This operation is in some measure equivalent to the quality assessment of a fingerprint image because the matching (or comparison) is mainly dependent on the foreground area. It is reasonable that a fingerprint image with relatively clear and large foreground area can generate a higher genuine matching score than those characterized in an opposite way. In this case, many studies use segmentation-based solutions to perform quality assessment. This section gives two metrics based on segmentation-like operations to show how foreground area is important to quality assessment. The first one is an image-independent quality metric and the second is dependent on the image pixel.

3.2.1 FQA via informative region

The image-independent approach employed in this chapter is known as the MQF [29] which uses only the coordinates information of the minutiae template of the associated fingerprint image. Figure 2 gives a general diagram of this quality metric.

As depicted by the diagram (Figure 2), the convex-hull and Delaunay triangulation are used at first for modeling the informative region of a fingerprint image in terms of the detected minutiae points. Next, some unreasonable-looking triangular areas marked by pink are removed from the informative region. The remaining area of the informative region hence represents the quality of the associated fingerprint or the minutiae template [29].

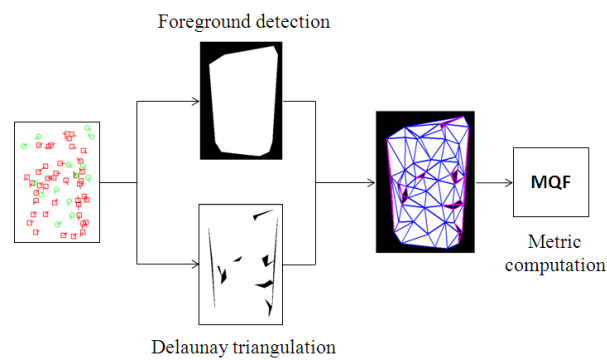


Fig. 2: Diagram of the framework of the MQF.

This quality metric is chosen because it is a new solution of the FQA and it outperforms the SoA approach in some cases though only minutiae coordinates are used. The details of this metric can be found in the reference article and are not given here.

3.2.2 FQA via Pixel-pruning

Another segmentation-based quality metric is denoted as MSEG [30] which performs a two-step operation to a fingerprint image, including a coarse segmentation and a pixel-pruning operation. The pixel-pruning is implemented via categorizing fingerprint quality into two general cases: desired image and non-desired image. Figure 3 illustrates such a categorization.

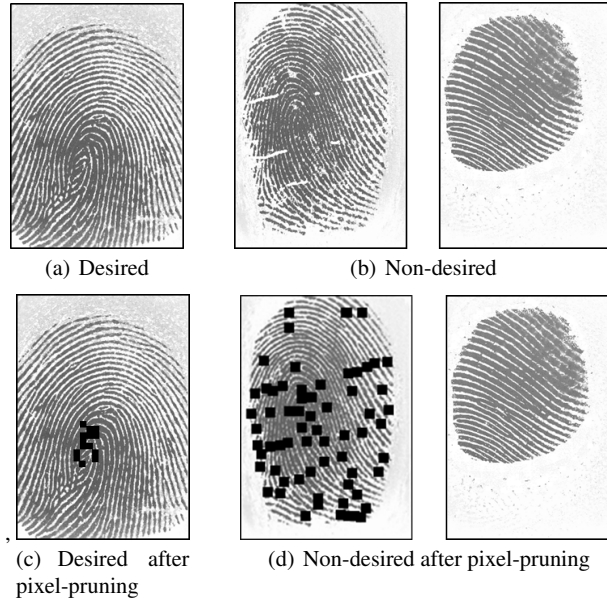


Fig. 3: Demonstration of pixel-pruning approach.

Obviously, an AFIS basically prefers keeping images like Figure 3(a) because it is more probably to give reliable and sufficient feature. Figure 3(b) shows two images that are not desired subjectively because the left one has some tiny quality problems and the right one is relatively small and both may lead to low genuine matching or high impostor matching. In this case, a better quality assessment can be done if one can make a clearer difference between the desired image and the non-desired image. The MSEG employs a gradient measure of image pixel to prune pixels of non-desired image as much as possible. Figure 3(c) and 3(d) illustrate the result of pixel-pruning operation of two kinds of images.

3.3 FQA via multi-feature

Similarly, we also choose two quality metrics that rely on multi-feature and both are implemented via a prior-knowledge of matching performance. By using this kind of solutions, an experimental comparison can be made between different approaches, especially one can find that solutions based on multiple features do not really take the advantages of the employed features because of the effect generated by the variation of image specifications, so is the employed prior-knowledge generated from the big data [21]. The first one is classification-based approach which is the SoA solution known as NFIQ [24]. The NFIQ estimates a normalized matching score of a fingerprint sample by sending a set of quality features (11 features) to a neural

network model. The NFIQ algorithm remapped the estimated matching score into five classes denoted by integers from 1 to five where 1 indicates the best quality level.

On the other hand, we choose quality metric based on multi-feature fusion which is actually a No-reference Image Quality Assessment (NR-IQA) [22] solution used for FQA by integrating multiple features with a set of weighted coefficients. The selected approach is denoted as Qabed [8], which is basically defined as

$$Q = \sum_{i=1}^N \alpha_i F_i, \quad (2)$$

where N is the number of quality features F_i ($i = 1, \dots, N$), α_i are the weighted coefficients obtained by optimizing a fitness function of a genetic algorithm. The fitness function is defined as a correlation between linearly combined quality value and genuine matching score [13]. Maximizing such a linear relation is somehow equivalent to the concept that quality predicts matching performance. The weighted coefficient is dependent on a training set of fingerprint samples. We choose this approach because it performs well in predicting the matching performance in comparison with the SoA quality metric.

4 Experimental Results

Some existing studies propose to calculate correlation between different metrics [4] for comparing the behavior of them. However, this is not completely observable, because there is no explicit linear relation among every group of variant quality metrics. Generally, this kind of measure is to observe the similarity between two different variables such as wavelet coefficients. In this case, to compare the studied metrics, we simply provide experiment results of two evaluation approaches, one is a validation approach relied on Enrollment Selection (ES) [27] and another is an evaluation method with multiple bins of sorted biometric samples [7].

4.1 Software

In the experiment, we use two matching systems where one is the OpenSource NBIS [26] and another is a commercial fingerprint SDK known as 'id3'. The NIST software contains several modules, among which the MINDTCT is used for generating INCITS 378-2004 standard minutiae template and the matching scores are calculated via Bozorth3. The commercial SDK has 6 options of the existing minutiae template standards and the minutiae templates of ISO/IEC 19794-2:2005 standard [10] have been extracted in the experiment. Similarly, a corresponding matcher has also been implemented with the SDK. By using these two sets of programs, the com-

parative study is accomplished via an interoperate analysis of the selected quality metrics.

4.2 Database et Protocol

In the experiment, one dataset of the 2000 Fingerprint Verification Competition (FVC) test, one of FVC2002, three of FVC2004 and two CASIA¹ datasets are employed. Each of the FVC datasets include 800 images of 100 individuals, 8 samples per individual. The CASIA database contains fingerprint images of 4 fingers of each hand of 500 subjects, where each finger has 5 samples. In this study, we create the two re-organized databases by using samples of the second finger of each hand, and they are respectively denoted as CASL2 and CASR2. Therefore, each sub-database has 2500 images of 500 individual (5 samples per individual).

Table 1: Dataset specification.

DB	Sensor	Dim.	Resolution
00DB2A	Low-cost Capacitive	256×364	500dpi
02DB2A	Optical	296×560	569dpi
04DB1A	Optical	640×480	500dpi
04DB2A	Optical	328×364	500dpi
04DB3A	Thermal	300×480	512dpi
CASL2	Optical	328×356	512dpi
CASR2	Optical	328×356	512dpi

The image size of each dataset is different from one another and the resolution is over 500-dpi. A glance of the datasets is given by several samples in Figure 4. In this study, the experiment includes two parts, one is utility-based evaluation and another is quality-based evaluation. The evaluation approach employed in the experiment is based on the Enrollment Selection (ES) [28].

4.3 Results

4.3.1 ES with Quality

The evaluation task is a comparison between variant frameworks of fingerprint quality metric. We use each group of quality values and two types of matching scores to perform enrollment selection for each dataset. The global EER values obtained by the selected quality metrics are given in Table 2.

¹ <http://biometrics.idealtest.org/detailsDatabase.do?id=3>

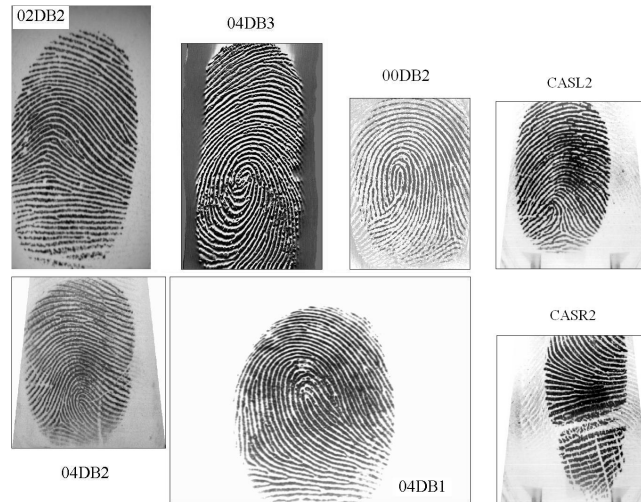


Fig. 4: Illustration of dataset samples.

One can find that the quality metrics providing the lowest global EER are not always ones based on multi-feature, even for an associated vendor such as NBIS matching software of the NFIQ. The quality metric based on a single feature (CWT) also performs well on many datasets. In addition, both the CWT and MSEG demonstrate relatively good generality for the employed matching algorithms, especially when a better matching algorithm is involved.

For instance, MSEG obtains the best results from the last 4 of the 7 employed dataset when performing evaluation with the matching scores of the SDK, while the results obtained from other 3 databases are also not bad. Particularly, MSEG decreases the error rates more than other metrics for the two difficult databases, CASL2 and CASR2. In addition, the CWT also performs well for most of the databases. The QMF and NFIQ do not give dominant results, especially when the NBIS matching scores is used in the experiment because QMF relies on the GMS of the NBIS software, while the NFIQ depends on 11 quality features (or real metrics). The confidence interval (CI) of the global EER values are given in Table 3.

Furthermore, one can observe the effect of matching scores to the knowledge-based metrics: NFIQ and QMF. The NFIQ obtains quite high (bad) EER values from the two CASIA datasets when NBIS matching scores are employed in the evaluation, while it generalizes relatively better results for the two datasets when using the SDK. The QMF obtains better results than NFIQ from 5 (02DB2, 04DB1, 04DB3, CASL2 and CASR2) of the seven databases when using the NBIS matching scores because its training is independently performed for each dataset via the NBIS matching scores, meaning it is appropriate to a specific scenario. However, in comparison with the knowledge-free metrics, both the two metrics do not show a higher performance though they employ different sets of features. Meanwhile, the

Table 2: Global EERs obtained via ES with quality metrics. 'NBIS' and 'SDK' are two sets of matching scores.

DB \ QM	NFIQ	QMF	MQF	CWT	MSEG
00DB2A (N.)	4.97%	6.57%	5.03%	4.93%	4.50%
02DB2A (N.)	13.33%	11.11%	11.18%	11.11%	10.79%
04DB1A (N.)	15.37%	14.72%	14.98%	17.53%	16.54%
04DB2A (N.)	13.32%	16.64%	15.02%	14.16%	14.05%
04DB3A (N.)	7.47%	7.36%	6.87%	7.00%	7.18%
CASL2 (N.)	43.09%	40.64%	40.48%	40.09%	42.30%
CASR2 (N.)	43.51%	41.39%	40.62%	40.45%	43.20%
00DB2A (S.)	0.22%	0.40%	0.76%	0.09%	0.10%
02DB2A (S.)	0.11%	0.30%	0.12%	0.10%	0.20%
04DB1A (S.)	2.66%	1.74%	1.73%	1.91%	1.93%
04DB2A (S.)	3.86%	3.94%	3.43%	3.33%	3.24%
04DB3A (S.)	1.89%	1.66%	1.51%	1.59%	1.51%
CASL2 (S.)	40.92%	42.72%	42.19%	42.35%	38.61%
CASR2 (S.)	38.20%	41.26%	40.94%	39.70%	35.97%

Note: NFIQ and QMF rely on multi-feature and prior-knowledge of GMS; MQF and MSEG are based on segmentation, CWT is a single feature-based metrics.

MQF is a no-image quality metric but the performance is not bad in comparison with the NFIQ and QMF, especially when using the NBIS matching scores because it relies on the minutiae extractor associated to the NBIS software. In this case, one can observe that a good matching algorithm and a relatively good dataset (such as 00DB2, 02DB2 and 04DB3) may blurs the effect of a quality metric, *i.e.* it is easier to approach to a relatively better performance if the matcher is relatively robust. Thus, it is really necessary to perform an off-line biometric test via 'bad' datasets. In addition, it is possible to consider that the implementation of a metric should be independent from the matching performance if we emphasize its 'generality'. The effect of matching performance to quality metrics is further discussed in Section 4.4.

4.3.2 Isometric Bins

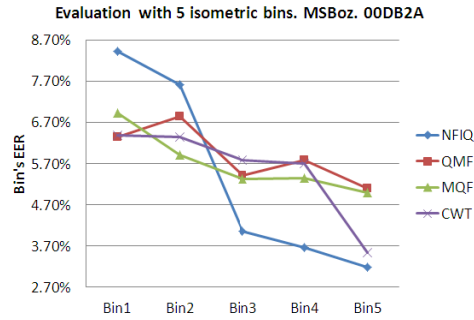
The ES with sample's quality reveals the best of quality metrics' capability in reducing the error rate. In this section, another evaluation is performed by using an approach based on isometric bins of the samples that had been sorted in terms of quality [7]. We don't assert that quality metric is fully able to predict matching performance due to the diversity of matching algorithms. In this case, this kind of evaluation is somehow to demonstrate the linearity between a quality metric and the performance of a matcher. The NFIQ is used as a reference, while the QMF, MQF and CWT represent metrics based on multi-feature fusion, segmentation and

Table 3: The 95% confidence interval of EER of each quality metric.

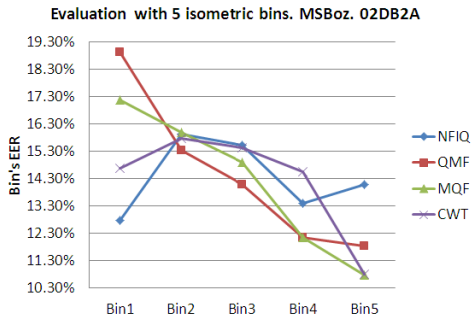
DB \ QM	NFIQ	QMF	MQF	CWT	MSEG
00DB2A (N.)	[0.0492 0.0502]	[0.0651 0.0663]	[0.0497 0.0509]	[0.0488 0.0499]	[0.0450 0.0461]
02DB2A (N.)	[0.1326 0.1340]	[0.1104 0.1118]	[0.1109 0.1128]	[0.1103 0.1119]	[0.1068 0.1084]
04DB1A (N.)	[0.1529 0.1545]	[0.1464 0.1480]	[0.1491 0.1506]	[0.1744 0.1762]	[0.1645 0.1662]
04DB2A (N.)	[0.1321 0.1344]	[0.1651 0.1676]	[0.1489 0.1515]	[0.1407 0.1425]	[0.1396 0.1413]
04DB3A (N.)	[0.0741 0.0752]	[0.0730 0.0742]	[0.0681 0.0693]	[0.0694 0.0706]	[0.0712 0.0723]
CASL2 (N.)	[0.4296 0.4322]	[0.4059 0.4070]	[0.4043 0.4054]	[0.4004 0.4015]	[0.4213 0.4247]
CASR2 (N.)	[0.4337 0.4364]	[0.4134 0.4145]	[0.4057 0.4068]	[0.4039 0.4050]	[0.4307 0.4332]
00DB2A (S.)	[0.0021 0.0023]	[0.0040 0.0043]	[0.0074 0.0078]	[0.0008 0.0009]	[0.0009 0.0011]
02DB2A (S.)	[0.0011 0.0013]	[0.0029 0.0032]	[0.0011 0.0013]	[0.0010 0.0011]	[0.0013 0.0016]
04DB1A (S.)	[0.0268 0.0276]	[0.0172 0.0178]	[0.0171 0.0177]	[0.0188 0.0194]	[0.0189 0.0195]
04DB2A (S.)	[0.0390 0.0402]	[0.0378 0.0389]	[0.0338 0.0349]	[0.0327 0.0338]	[0.0318 0.0328]
04DB3A (S.)	[0.0190 0.0195]	[0.0162 0.0167]	[0.0148 0.0154]	[0.0159 0.0164]	[0.0117 0.0122]
CASL2 (S.)	[0.4087 0.4097]	[0.4266 0.4278]	[0.4213 0.4225]	[0.4229 0.4241]	[0.3856 0.3866]
CASR2 (S.)	[0.3815 0.3825]	[0.4119 0.4132]	[0.4087 0.4102]	[0.3963 0.3977]	[0.3592 0.3603]

single feature, respectively. We do not use all databases and metrics because these results are enough to show the what the quality predicting matching performance is. The results obtained by using two types of matching scores (NBIS and SDK) are given by global EERs' plots in Figure 5 and 6, respectively. One can found that the EER values of the bins obtained by some of the quality metrics are monotonically decreasing, which assert the purpose of proving the validity of a quality metric. Loosely speaking, this kind of property demonstrates so-called quality predicting matching performance. On the other hand, it shows the similarity or linear relationship between the quality scores and GMS. This could be observed with correlation coefficients between the two measurements.

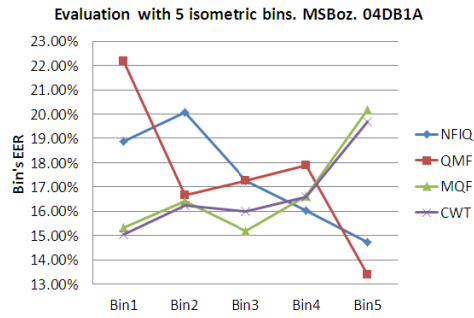
In the experiment, the maximum GMS for each sample is calculated to demonstrate such an observation, see Table 4. For instance, when MSBoz is used, the Pearson correlation coefficients of NFIQ for 00DB2A and QMF for 02DB2 with respect to the maximum GMS are -0.4541 and 0.5127. Similarly, this kind of correlation also could be found for the monotonically decreased cases when MSSDK is employed. Here, we simply gives the result of some opposite cases, where the Pearson coefficients of CWT for 04DB1A, NFIQ for 02DB2A, and MQF for 04DB1A with respect to the maximum GMS of MSSDK are 0.0444, -0.2596 and 0.0585, respectively. These non-correlated values or some negative correlated cases such as the CWT in Figure 5(c) are mostly caused by outliers of either the metric or the matching algorithm. Meanwhile, with the results in Table 2, Figure 5 and 6 together, it reveals that quality predicting matching performance is not always reached linearly, such as the CWT for 04DB2A shown by the 3 sets of results. The global EERs in Table 2 demonstrate that the two metrics perform relatively better for determining the best cases of sample quality, while no linear relationship were found between



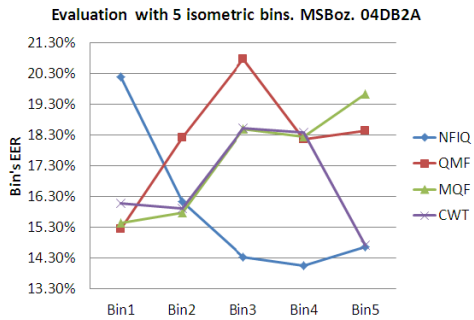
(a) 00DB2A



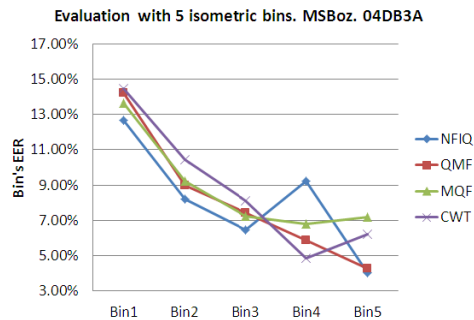
(b) 02DB2A



(c) 04DB1A



(d) 04DB2A



(e) 04DB3A

Fig. 5: Enrollment selection with quality metrics.

Table 4: Pearson correlation between metrics and maximum GMS. 'NBIS' and 'SDK' are two sets of matching scores.

DB \ QM	NFIQ	QMF	TMQ	CWT
00DB2A (N.)	-0.4541	-0.0014	-0.0268	0.2885
02DB2A (N.)	-0.3308	0.5217	0.3940	0.2626
04DB1A (N.)	-0.1579	0.2601	0.0027	0.0122
04DB2A (N.)	-0.3937	-0.0177	0.1450	0.1684
04DB3A (N.)	-0.3063	0.5922	0.3132	0.4604
00DB2A (S.)	-0.4379	-0.0021	0.0402	0.3246
02DB2A (S.)	-0.2596	0.3254	0.3732	0.3230
04DB1A (S.)	-0.1970	0.3734	0.0585	0.0444
04DB2A (S.)	-0.5843	0.0615	0.1309	0.1961
04DB3A (S.)	-0.4131	0.4142	0.4371	0.6121

them and both employed matching algorithms according to Figure 5(d) and 6(d), so is learning-based metric such as Figure 5(d) and 6(b).

4.4 Discussion via sample utility

To validate a biometric quality metric, an objective index [27] is used for representing the quality of a sample. The objective measure is an offline sample EER (SEER) value calculated from a set of intra-class matching scores and a set of inter-class matching scores formulated as $N - 1$ genuine matching scores (GMS)

$$gms_{i,j,k} = R(S_{i,j}, S_{i,k}) \quad j \neq k \quad (3)$$

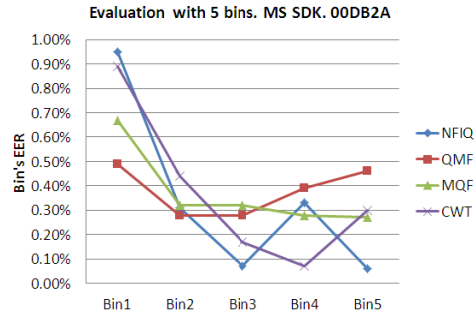
and $N - 1 \times M - 1$ impostor matching scores (IMS)

$$ims_{i,j,l,k} = R(S_{i,j}, S_{l,k}) \quad i \neq l \text{ and } j \neq k, \quad (4)$$

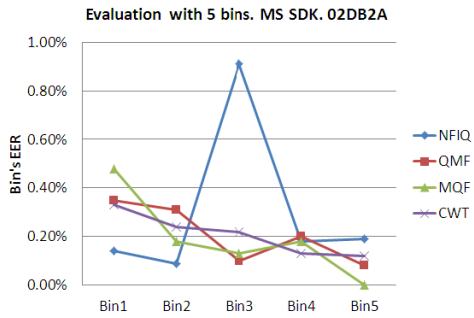
where N and M denote sample number and individual number of a trial dataset, R is a matcher and $S_{i,j}$ indicates the j^{th} sample of the i^{th} individual ($S_{l,k}$ is similar).

Therefore, with a SEER _{i,j} of one sample, one can have a measure of how much the contribution of a sample is within the experimental framework consisted of employed datasets and matching algorithms. The objective measure is denoted as sample's **Utility** throughout the experiments.

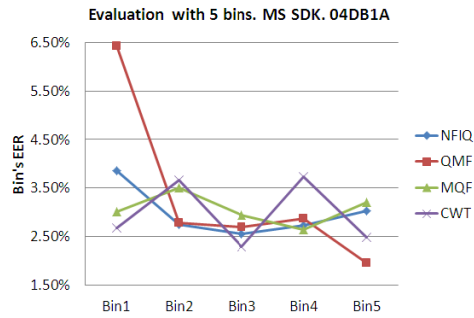
The utility study in this part is actually an ES operation with the objective indexes presented in section 4.4. The objective measure of each sample reflects the behavior of the sample under one matching algorithm of a specific vendor. This kind of measurement is simply used for explaining the limitation of those quality metrics implemented via prior knowledge of matching scores.



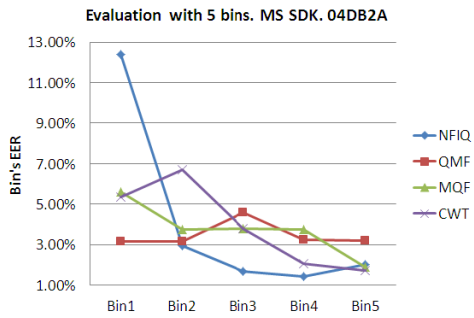
(a) 00DB2A



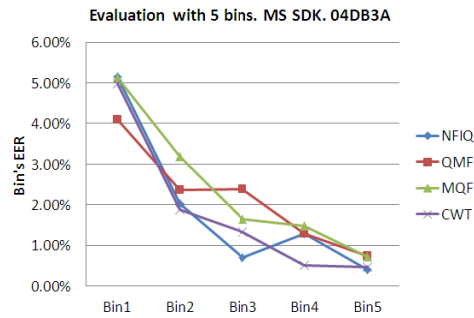
(b) 02DB2A



(c) 04DB1A



(d) 04DB2A

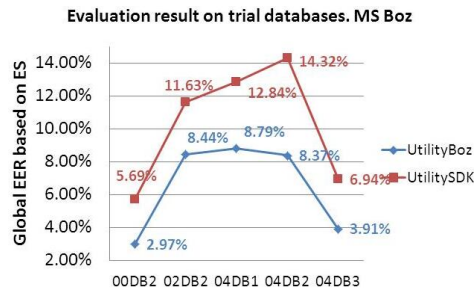


(e) 04DB3A

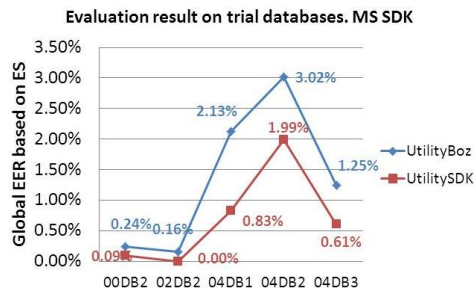
Fig. 6: Enrollment selection with quality metrics.

According to the definition given in 4.4, one can obtain a M -by- N matrix of sample utility for a trial database. The matrix is hence used as a quality result by which the enrollment selection is performed via interoperate matching algorithms, see graphical results in Figure 7.

Figure 7 gives the plots of global EER values obtained by using ES with sample utility values, where Figure 7(a) is the result based on NBIS matching scores (MSBoz) and Figure 7(b) is generated from the SDK's matching scores (MSSDK). In the experiment, first of all, the utility value of each sample ($SEER_{i,j}$) with respect to each matcher is calculated, respectively. In this case, two matrices of the sample utility were figured out and then used for enrollment selection. The utility values correspond to NBIS software and the SDK are denoted as 'UtilityBoz' and 'UtilitySDK', by which the global EER values calculated with ES are plotted in the figure, indicating by blue and red points, respectively.



(a) Matcher of NBIS



(b) Matcher of SDK

Fig. 7: Enrollment selection with objective measures.

The enrollment selection task chooses the best sample of one individual as the enrollment in terms of their utility values. In this case, the best performance of a matching algorithm obtained from a trial dataset cannot go over the global EER value. Apparently, the utility value is mostly dependent on the performance of the matching algorithm which is illustrated by two set of plots. In addition, according to the results, we believe that a quality metric based on a prior knowledge of matching score is not fully able to predict the matching performance in a cross-use. In fact,

one can consider that whether two genuine samples should produce high GMS when one of them is not able to give reliable and sufficient features [29]. Besides, it is not clear that how much the prior knowledge is close to the ground-truth of sample quality.

5 Conclusion

Recent studies of fingerprint quality metrics mainly focus on reducing error rates in terms of utility of the samples. In this study, we make an interoperability analysis to observe the behavior of several representative fingerprint quality metrics from the existing frameworks, and hence reveal the limitations of this issue. Among the experimental study, one can note that it is not very easy to achieve a common good quality metric, even to those with multiple features. For instance, by comparing with the metrics carried out via a single feature, some metrics based on multi-feature do not show the advantage that should have obtained after fusion. Utility-based quality metrics, especially those related to matching scores are more probably affected by the change of matching algorithm, which is clearly brought out with the experiments. Nevertheless, the linear relationship between GMS and quality values is a valid criterion for assessing quality. However, it is not absolutely appropriate for a different matching circumstance. To the end, the offline trials also reveal that quality metrics is not an absolutely predictive measure for matching performance.

References

1. ISO/IEC 29794-1:2009. Information technology ?biometric sample quality ?part 1: Framework. August 2009.
2. Mayank Vatsa A, Richa Singh A, Afzel Noore A, and Max M. Houck B. Quality-augmented fusion of level-2 and level-3 fingerprint information using dsm theory. *International Journal of Approximate Reasoning*, 50(1):pgs. 51–61, 2009.
3. F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, and et al. A comparative study of fingerprint image-quality estimation methods. *Information Forensics and Security, IEEE Transactions on*, 2(4):734–743, 2007.
4. Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP Journal on Image and Video Processing*, 2014(1):1–28, 2014.
5. R. M. Bolle, S. U. Pankanti, and Y. Yao. System and method for determining the quality of fingerprint images, October 5 1999. US Patent 5,963,656.
6. T.P. Chen, X. Jiang, and W.Y. Yau. Fingerprint image quality analysis. In *Image Processing, 2004. ICP '04. 2004 International Conference on*, volume 2, pages 1253–1256 Vol.2, 2004.
7. Yi Chen, Sarat C Dass, and Anil K Jain. Fingerprint quality indices for predicting authentication performance. In *Audio-and Video-Based Biometric Person Authentication*, pages 160–170. Springer, 2005.
8. M. El Abed, A. Ninassi, C. Charrier, and Christophe Rosenberger. Fingerprint quality assessment using a no-reference image quality metric. In *European Signal Processing Conference (EUSIPCO)*, page 6, 2013.
9. J. Fierrez-Aguilar, J. Ortega-Garcia, and et al. Kernel-based multimodal biometric verification using quality signals. In *Defense and Security*, pages 544–554. International Society for Optics and Photonics, 2004.
10. Organization for Standardization. Iso/iec 19794-2:2005: Information technology-biometric data interchange formats-part 2: Finger minutiae data, 2005.
11. Hartwig Fronthaler, Klaus Kollreider, and Joseph Bigun. Automatic image quality assessment with application in biometrics. In *Computer Vision and Pattern Recognition Workshop. CVPRW'06. Conference on*, page 30. IEEE, 2006.
12. Patrick Grother. and Elham Tabassi. Performance of biometric quality measures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):531–543, 2007.
13. R-LV Hsu, Jidnya Shah, and Brian Martin. Quality assessment of facial images. In *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, pages 1–6. IEEE, 2006.
14. B. Lee, J. Moon, and H. Kim. A novel measure of fingerprint image quality using the Fourier spectrum. In A. K. Jain and N. K. Ratha, editors, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, volume 5779, pages 105–112, mar 2005.

15. Guoqiang Li, Bian Yang, and Christoph Busch. Autocorrelation and dct based quality metrics for fingerprint samples generated by smartphones. In *Digital Signal Processing (DSP), 2013 18th International Conference on*, pages 1–5. IEEE, 2013.
16. E. Lim, Xudong Jiang, and WeiYun Yau. Fingerprint quality and validity analysis. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages 1–469–1–472 vol.1, 2002.
17. Loris Nanni and Alessandra Lumini. A hybrid wavelet-based fingerprint matcher. *Pattern Recognition*, 40(11):3146–3151, 2007.
18. M.A. Olsen, E. Tabassi, A. Makarov, and C. Busch. Self-organizing maps for fingerprint image quality assessment. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, pages 138–145, June 2013.
19. Martin Aastrup Olsen, Haiyun Xu, and Christoph Busch. Gabor filters as candidate quality measure for nfiq 2.0. In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 158–163. IEEE, 2012.
20. Norman Poh and Josef Kittler. A unified framework for biometric expert fusion incorporating quality measures. *Pattern Analysis and Machine Intelligence IEEE Transactions on*, 34(1):3–18, 2011.
21. N. K. Ratha, J. H. Connell, and S. Pankanti. Big data approach to biometric-based identity analytics. *Ibm Journal of Research and Development*, 59(2/3):4:1–4:11, 2015.
22. M. Saad, A. C. Bovik, and C. Charrier. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing*, 21(8):3339–3352, 2012.
23. Linlin Shen, Alex Kot, and Waimun Koo. Quality measures of fingerprint images. In *IN: PROC. AVBPA, SPRINGER LNCS-2091*, pages 266–271, 2001.
24. Elham Tabassi, C Wilson, and C Watson. Nist fingerprint image quality. *NIST Res. Rep. NISTIR7151*, 2004.
25. Xunqiang Tao, Xin Yang, Yali Zang, Xiaofei Jia, and Jie Tian. A novel measure of fingerprint image quality using principal component analysis(pca). In *Biometrics (ICB), 2012 5th IAPR International Conference on*, pages 170–175, March 2012.
26. Craig I Watson, Michael D Garris, Elham Tabassi, Charles L Wilson, R Michael McCabe, Stanley Janet, and Kenneth Ko. User’s guide to nist biometric image software (nbis). 2007.
27. Z. Yao, JM. Le Bars, C. Charrier, and C. Rosenberger. Literature review of fingerprint quality assessment and its evaluation. 2016.
28. Z. Yao, J-M. LeBars, C. Charrier, and C. Rosenberger. Fingerprint quality assessment combining blind image quality, texture and minutiae features. In *International Conference on Information Systems Security and Privacy*, Feb. 2015.
29. Z. Yao, J-M. LeBars, C. Charrier, and C. Rosenberger. Quality assessment of fingerprints with minutiae delaunay triangulation. In *International Conference on Information Systems Security and Privacy*, Feb. 2015.
30. C. Charrier C. Rosenberger. Z. Yao, J-M Le Bars. Fingerprint quality assessment with multiple segmentation. In *IEEE International Conference on Cyberworlds (CW)*, Gotland, Sweden, Oct. 2015.