2009 Second International Symposium on Electronic Commerce and Security

Comparative Assessment of Fingerprint Sample Quality Measures Based on Minutiae-based Matching Performance

Changlong Jin, Hakil Kim, Xuenan Cui,Eunsoo Park,Junchul Kim
Department of Information Engineering,
INHA University,
Incheon, Korea
{cljin, hikim, xncui, espark, jckim}@vision.inha.ac.kr

Jinsoo Hwang
Department of Statistics,
INHA University,
Incheon, Korea,
jshwang@inha.ac.kr

Stephen Elliott
Biometric Standards Performance and Assurance Laboratory, Purdue University,
USA,
elliott@purdue.edu

***Abstract*—This Fingerprint sample quality is one of major factors influencing the matching performance of fingerprint recognition systems. The error rates of fingerprint recognition systems can be decreased significantly by removing poor quality fingerprints. The purpose of this paper is to assess the effectiveness of individual sample quality measures on the performance of minutiae-based fingerprint recognition algorithms. Initially, the authors examined the various factors that influenced the matching performance of the minutiae-based fingerprint recognition algorithms. Then, the existing measures for fingerprint sample quality were studied and the more effective quality measures were selected and compared with two image quality software packages, (NFIQ from NIST, and QualityCheck from Aware Inc.) in terms of matching performance of a commercial fingerprint matcher (Verifinger 5.0 from Neurotechnologija). The experimental results over various Fingerprint Verification Competition (FVC) datasets show that even a single sample quality measure can enhance the matching performance effectively.**

Keywords: fingerprint sample quality; matching performance; quality measure; equal error rate

I. INTRODUCTION

Biometrics is a technology of automatically recognizing individuals based on their behavioral or biological characteristics (e.g., fingerprint, face, iris, voice, gait, hand geometry, hand vein). Fingerprint recognition has become the most popular and important biometric modality, primarily because of its permanence, uniqueness, convenience, and high performance [1, 2]. Furthermore, fingerprint recognition can be applied in several practical areas, including (but not limited to) border security control, time and attendance, physical access control, internet authentication.

Although the performance of fingerprint recognition systems has greatly improved [3, 4], it is still influenced by many factors, such as fingerprint sample quality, and common area and deformation between pairs of genuine matching samples. Among these factors, fingerprint sample quality which measures the characteristics of ridge-valley texture, has had the greatest impact on matching performance [4]. There are many fingerprint matching algorithms in the literature [5], most of which follow the scheme of point pattern matching approach, namely, minutiae-based matching algorithm. One of the most important tasks of these matching algorithms is the precise extraction of the ridge characteristics from input fingerprint images, and the performance of this task significantly depends on the quality of the fingerprint images.

According to the on-going ISO/IEC standard on biometric sample quality [6], sample quality can be considered from three aspects: The first is "character" -- that is related with the contributor to the quality of a sample attributable to inherent features of the source. The second is "fidelity" -- that is defined as the degree of the similarity between a biometric sample and its source. The third is the "utility" -- that the observed performance of a sample in a biometric system, or the impact of an individual biometric sample on the overall performance of a biometric system. Most of the sample quality measures refer to the degree to which a biometric sample fulfils its specified requirements for its targeted application with regards to the utility.

Usually, fingerprint sample quality is a scalar quantity that is related monotonically to the performance of the system [4]. However, it is difficult to precisely assess the overall quality of an individual fingerprint. All existing quality methods only assess the quality of ridge-valley texture [7-11], and ignore the common area and deformation which can not be measured for a single sample.

The purpose of this paper is to assess the effectiveness of individual sample quality measures on the performance of minutiae-based fingerprint recognition algorithms. Initially, the authors examined the various factors that influenced the matching performance of the minutiae-based fingerprint recognition algorithms. Next, the existing measures for

This work was supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University, Korea, (No. R112002105080010(2008))

978-0-7695-3643-9/09 $25.00 © 2009 IEEE
DOI 10.1109/ISECS.2009.59

Authorized licensed use limited to: University of Tokyo. Downloaded on March 26,2023 at 14:38:00 UTC from IEEE Xplore. Restrictions apply.

fingerprint sample quality were investigated with respect to their principle and correlation with each other. Then, the most effective quality measures were selected and compared with two image quality software packages, NFIQ from NIST (*NFI*), and QualityCheck from Aware Inc. (*AWA*), and subsequently analyzed in terms of the equal error rate in matching.

This paper is organized as following: absolute and relative fingerprint sample qualities are defined in Section Ⅱ. The existing absolute fingerprint sample quality measures are studied in Section Ⅲ. In Section Ⅳ, the matcher selection, the definition of image segmentation for fingerprint sample quality evaluation, datasets, matching protocol, results and discussion are discussed. Finally, a brief conclusion in Section Ⅴ will summarize the findings of the paper.

II. ABSOLUTE AND RELATIVE SAMPLE QUALITY

The Equal Error Rate (*EER*) denotes the error rate at the threshold *t* for which False Match Rate (*FMR*) and False Non-Match Rate (*FNMR*) are identical: *FMR(t)=FNMR(t)* [12]. In practice, it is usually used as a measure of the performance of biometric systems. For the purposes of this paper, it is used to assess the performance of minutiae-based fingerprint matching systems.

The *EER* is influenced by many factors, such as the Quality of Ridge-Valley Texture (QRVT) which refers to the distinction in ridge-valley contrast, Common Area (CA) and Deformation (DF) between a pair of genuine matching samples, and the performance of the matcher. Since these factors can be considered as independently affecting to the *EER* (For example, a genuine fingerprint matching pair can have a low QRVT, small CA and high DF), the overall *EER* can be described as

$$EER \propto EER_q + EER_a + EER_d + \varepsilon \qquad (1)$$

where EER_q, EER_a, and EER_d represent the error rates mainly caused by QRVT, CA, and DF, respectively, and ε denotes other error factors which are unknown but negligible.

Figure 1 shows the plot of quality score vs. genuine matching score for overall selected datasets - FVC 2000 1a, 2a, 3a, FVC 2002 1a, 3a, 2004 1a, and 2a (refer to Section Ⅳ *A* for continued discussion). The circled region indicates that the genuine matching scores (obtained by Verifinger) can be very low despite of good sample quality, $q = \sqrt{q_1 q_2}$, where q_1, q_2 are Local Clarity Scores (*LCS* refer to Section Ⅲ *C* for definition) of a pair of genuine matching samples.

Typically, for a given dataset, the influences caused by CA and DF are less significant than QRVT. In other words, QRVT is the main factor contributing to *EER*. FVC 2004 DB 1a is a dataset that has a good QRVT as a whole yet varies greatly in position and deformation between genuine pairs. The all-matching experiment after rejecting 12 fingers with severely low CA and high DF out of 100 fingers yielded a 4.88% enhancement in *EER*. Meanwhile, the enhancements in *EER* with the rejection of 5%, 10%, 15% lowest quality samples are 7.76%, 12.83%, 15.11%, respectively. This experiment indicates that sample quality has a greater influence on the performance than common area or deformation. Therefore, in order to investigate the influences caused by QRVT, EER_a and EER_d are taken as constants, and (1) can be simplified as:

$$EER \propto EER_q + C_{ad} + \varepsilon = EER_q + C \qquad (2)$$

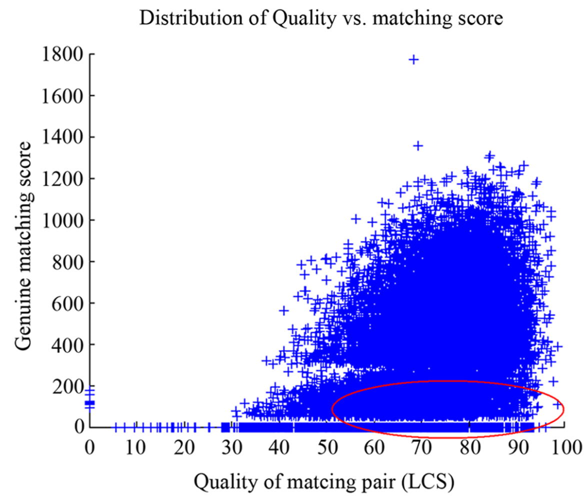

Figure 1. Scatter plot for sample quality (*LCS*) vs. genuine matching score.

In general, the overall fingerprint sample quality in matching is a combination of absoluteness and relativeness. The QRVT of an individual fingerprint can only be assessed by the clarity of ridge and valley. However, when matching is made between a pair of fingerprint samples, the relative factors need to be considered: common area and deformation. These two factors always change with the matching object. As dynamic factors, it is hard to measure these two factors from a single fingerprint image. Hence, the QRVT can be classified as an Absolute Fingerprint Sample Quality (AFSQ), and the score of CA and DF can be defined as Relative Fingerprint Sample Quality (RFSQ). The fusion of these two qualities can be a final sample quality which can then be used to estimate the matching performance within the fingerprint system. However, the challenge is that the enrolled sample (except in the matching phase) is not fixed, and as a result RFSQ is hard to measure. Therefore, fingerprint sample quality methods should always measure the absolute (AFSQ) as opposed to the relative (RFSQ). All existing fingerprint sample quality measures are absolute which only measures the QRVT.

III. ABSOLUTE FINGERPRINT SAMPLE QUALITY MEASURES

A. *Taxonomy of Fingerprint sample quality evaluation*

Since it is impossible to measure the RFSQ for a single fingerprint sample, the existing approaches for fingerprint sample quality produce the AFSQ as quality of a single sample image, and use this AFSQ to evaluate the matching performance of a fingerprint recognition system. The AFSQ in this study measures the clarity and orientation of ridge-valley texture. A summary of existing fingerprint sample quality measures is shown in Table Ⅰ.

Authorized licensed use limited to: University of Tokyo. Downloaded on March 26,2023 at 14:38:00 UTC from IEEE Xplore. Restrictions apply.

TABLE I. SUMMARY OF QUALITY APPROACHES

Classification	Approaches
Ridge orientation strength	Orientation certainty level [10] Gabor filter [11] Spatial coherence [8]
Ridge valley clarity	Local clarity score [7] Contrast [9]
Ridge frequency	Energy concentration [8]
Ridge flow	Local orientation [7] Continuity of the direction field [10]

B. Ridge orientation strength

The rationale behind the ridge orientation strength related approaches is that clear ridge-valley texture has higher orientation concentration thus the difference of eigenvalue of a coherent matrix is large, or the filter results that filtered by Gabor filter has obvious orientation characteristic. Lim et al. [10] proposed a so-called Orientation Certainty Level (*OCL*) method, which measures the energy concentration along the dominant direction of ridges using the intensity gradient. The coherence matrix of blocks is computed in a 32 by 32 block, the size of block is correct for 500 dpi fingerprint. The local *ocl* is computed by the ratio of two eigenvalues of the coherence matrix. Finally, the overall *OCL* is computed by average the local *ocl*. Figure 2 shows the original image and its *OCL* map.

Figure 2. An experiment of *OCL* measure. (Light: good quality; Dark: bad quality).

The advantage of this measure is that it does not require the sinusoid hypothesis while the other measures do and is less time-consuming. However, the *ocl* score tends to become low for regions containing a core or a delta. Further, the OCL value becomes relatively low for dry fingerprints.

Shen et al. [11] proposed a method for measuring fingerprint sample quality by utilizing a Gabor filter bank. The Gabor (*GAB*) filter bank includes eight filters in eight directions that cover equally spaced orientations between 0 and 180 degree. It is used to enhance the R-V region with a dominant direction and depress the disordered R-V region in the fingerprint. Therefore, for fingerprint regions with higher orientation concentration, the variation of the eight directional filtered responses becomes relatively large compared to the disordered R-V regions.

This measure utilizes the characteristic of Gabor filter to distinguish the ridge-like areas and the non-ridge-like areas. Then, the standard deviation of filter responses is calculated for each block. The weakness of this measure is time-consuming and incorrect estimation of ridge frequency in poor quality region.

C. Ridge valley clarity

The approaches in this category include the measure of separating distance between ridge and valley, and the measure of contrast between ridge and valley. Furthermore, they require the hypothesis of sinusoidal wave on fingerprint R-V structures. Chen et al. [7] analyzed the clarity between ridges and valleys in each block by calculating the overlapping region between ridge and valley. After separating the ridges and valleys according to a threshold determined by linear regression, misclassified pixels in both ridge and valley are divided by the total pixels in each region.

The advantage of this method is that the clarity between ridge and valley can be calculated by counting the misclassified pixels while the weakness is that the hypothesis of sinusoidal wave does not hold in the high curvature region such as singularity points.

Hong et al. [9] proposed a fingerprint sample quality measure based on the contrast of ridge and valley (*CNT*), which is defined as

$$CNT = (1/i)\sum_{i}[(\max(V_i) - \min(R_i)) * 100/255] \tag{3}$$

where i is the number of ridge-valley pairs in one block, V_i and R_i stands for grayscale of valley and ridge, respectively. When the region is of good quality, the contrast will be high, and if there is non-ridge-valley structure, the contrast will be low.

D. Ridge frequency

This type of approaches calculates the ridge frequency characteristic in the spectrum domain. Chen et al [8] proposed a method that measures power spectrum energy (*ENG*). The periodic R-V pattern in fingerprint corresponds to a ring pattern in spectrum domain; the strength of the ring pattern corresponds to the ridge strength in the spatial domain. The FFT transforms the spatial image information into the spectrum domain, and then filtered by a Butterworth band-pass filter bank. Afterwards, it computes the entropy for the band powers. Finally, the *ENG* is normalized

The advantage of this measure is that this is a robust global measure while most of the sample quality measures are a local measure. This measure utilizes the characteristic of fingerprint in the spectrum domain. However, this measure cannot distinguish the samples of low quality.

E. Ridge flow

The quality measures in this category quantify the characteristic of ridge flow. Chen et al. [7] proposed a method to compute the average absolute difference of local orientation with the surrounding blocks. A global orientation quality score (GOQS) is finally computed by averaging all of the local orientation quality scores of the image. This measure is not appropriate for the singularity area or the high curvature area.

Authorized licensed use limited to: University of Tokyo. Downloaded on March 26,2023 at 14:38:00 UTC from IEEE Xplore. Restrictions apply.

Lim et al. [10] presented a method by examining the orientation change along each horizontal row and each vertical column of the image blocks. Abrupt direction changes between blocks are accumulated and mapped into a global direction score.

F. Correlations and score distribution analysis

In this paper, the reasonable quality related algorithms were implemented on seven datasets selected from the Fingerprint Verification Competition (FVC) (refer to Section Ⅳ *A*). The quality scores were normalized by global minimum value and maximum value and the sample quality score is within the range from zero which denote low FSQ to 100 which denotes high FSQ. It is assumed that when one measure receives a high FSQ from a certain sample, the other measures have a higher probability to receive a higher score from the same sample, namely, the correct measures should have a high correlation among each other. The high correlation shows information redundancy, in the mean time, it can be used to check the correctness of the method.

The correlations among the group of selected measures and the quality check software were not high, especially for the correlations among the selected measures and NFIQ. This result is because NFIQ is a measure that assesses both the quality of ridge-valley texture and the quality of minutiae. But the selected measures are absolute fingerprint sample quality measures and only assess the characteristic of ridge-valley texture. Correlations among selected measures and the QualityCheck (an alternative image quality software package) are relatively higher than the correlations from NFIQ. Although the correlations among QualityCheck and the selected measures are not so high, it is still an acceptable correlation.

The highest correlation appears between the *LCS* and *OCL* measures. Among all seven selected measures, *OCL*, *GAB*, *LCS*, *ENG*, *CNT* have obviously higher correlations than *COF* and *LOQ*, which are the measures based on ridge flow. It shows that the measure of ridge flow cannot effectively distinguish the FSQ, consequently they are excluded in the next portion-rejection experiments.

IV. Experiments

The aim of this experiment is to assess the individual fingerprint quality measures in terms of matching performance. The quality measures that have high correlations are selected and implemented for portion-rejection experiments.

A. Fingerprint verification matcher AND DatasetS

The Equal Error Rate is influenced by many factors as described previously in Section Ⅱ. One of those factors is the performance of matcher especially without enrollment quality control. In this paper, in order to investigate the influence of AFSQ on matching performance and remove the other factors which influence the EER, a high performance matcher is required. Therefore, the performance caused by fingerprint image quality can be assessed more accurately. Table Ⅱ shows the zero-rejected all-matching EER's for seven selected datasets using two matching algorithms. Therefore, the Verifinger is used in this paper.

TABLE II. All matching performances of two matchers

matcher	FVC Datasets						
	00-1a	00-2a	00-3a	02-1a	02-3a	04-1a	04-2a
Bozorth3	6.86	7.64	13.11	18.66	22.30	21.41	19.24
Verifinger	3.10	1.07	6.76	0.53	1.74	10.85	6.80

For the experiments in this paper, the open datasets of FVC 2000 1a, 2a, 3a, FVC 2002 1a, 3a, 2004 1a, and 2a were selected. All of these datasets comprise of 500 dpi samples. And the results are convenient to compare with the other evaluation methods, especially NFIQ which is tuned for 500 dpi optical fingerprints.

B. Matching protocol

For this experiment, the performance is evaluated by a portion-rejected all-matching experiment. Firstly, a zero-rejected all-matching experiment is implemented. Since there are 100 individuals with 8 impressions for each dataset, the number of genuine matches is $(8\times7)\times100/2=2800$, and the number of impostor matches is $800\times799/2-2800=316800$. The equal error rate for this matching is denoted as EER_0. Then k percent of low quality samples are removed from the dataset based on each quality measure. The rest samples take part in new k%-rejected all-matching experiment. The number of matches is $800\times(1-k/100)\times(800\times(1-k/100)-1)/2$, and the number of genuine matches depends on how many genuine pairs are present. The equal error rate in this case is recorded as EER_k.

The Performance Evaluation Model (*PEM*) of *EER* for the assessment of AFSQ measures can be defined as follow

$$PEM_{EER} = \frac{EER_0 - EER_k}{EER_0}, k = 5,10,\cdots,25 \tag{4}$$

And the performance in each k%-rejected all-matching experiment is denoted by PEM_{EER-k}.

C. Results and discussion

Five selected AFSQ measures are implemented on seven FVC datasets. In general, the Equal Error Rates of this portion-rejection all-matching experiment decreased when the lowest quality samples were rejected. The significant enhancement of *EER* was observed when removing 5%~10% lowest quality samples. After that, there was no significant decrease in *EER*. It means that the dominant factor which influenced the *EER* changed from AFSQ to RFSQ.

The AFSQ of FVC 2002 1a and FVC 2004 1a are best of all engaged datasets (refer to Section Ⅲ *F*). The main factor of influencing the matching performance is RFSQ instead of AFSQ in these two datasets, thus the ratio of performance enhanced is much smaller than the other datasets. No differences between the optical sensor and capacitive sensor were observed. This is because the AFSQ only assesses the

Authorized licensed use limited to: University of Tokyo. Downloaded on March 26,2023 at 14:38:00 UTC from IEEE Xplore. Restrictions apply.

characteristics of ridge-valley, and these characteristics are sensor independent. Figure 3 shows the average performance enhancement for each measure. It shows that *GAB* has the best performance, and *LCS* is the second best measure. However, the *GAB* measure has the weakness of being computationally time consuming, and the performance for each dataset is not even, namely, it is not robust for all datasets. Although the *LCS* is the second best measure, the computation is simple and it is robust for all datasets. Thus, the all-around performance of *LCS* is better than *GAB*. On the other hand, when comparing these two measures with QualityCheck and NFIQ, they are better than the two quality checking software packages. For the *ENG* measure, since it cannot correctly assess the low quality samples, it shows the worst performance in the rejection experiment.

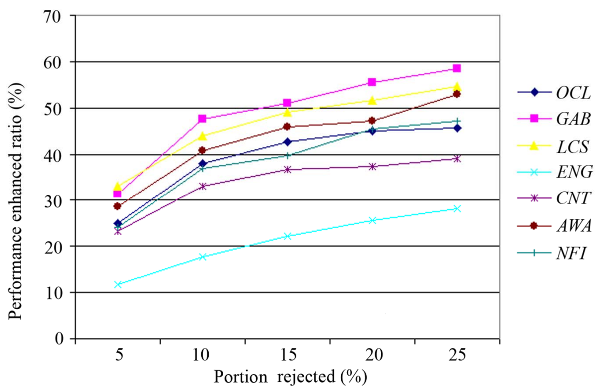

Figure 3. Average performance enhancement ratio for each quality measure

V. CONCLUTIONS

In this paper, factors that influence the minutiae-based matching performance are discussed, such as the quality of ridge-valley texture that refers to the clarity of ridge-valley texture, common area and deformation between a pair of genuine matching samples, and the performance of matcher etc. Then, the absolute and the relative fingerprint sample quality are defined. The existing approaches for fingerprint sample quality are studied and several measures are selected and implemented. High correlation is found between the selected quality measures, which indicate the correctness of each measure.

Five quality measures have been selected and implemented on seven FVC datasets. The results show that these quality measures are sensor independent. Fingerprint sample quality is a key factor in the following datasets: FVC 2000 1a, 2a, 3a, FVC 2002 3a, and FVC 2004 2a. Rejecting the lowest quality samples can significantly improve the matching performance in these datasets. Fingerprint sample quality measures assessed high quality to FVC 2002 1a and FVC 2004 1a, the portion-rejection all-matching experiments show that the quality measures have a slight influence on these two datasets. This means that the factors in these two datasets are common area and deformation. Single quality measures, GAB and *LCS*, have a better capability on minutiae-based matching performance enhancement when compared to the other approaches. *LCS* shows more robust performance and simple computation than GAB. Future work will focus on combination of all these effective quality measures to a single measure and the study of the effects of low-quality images.

REFERENCES

[1] S. Pankanti, S. Prabhakar, and A. K. Jain, "On the individuality of fingerprints," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1010 - 1025, 2008.

[2] Donald E. Maurer and J. P. Baker, "Fusing multimodal biometrics with quality estimates via a Bayesian belief network," *Pattern Recognition*, vol. 41, pp. 821-832, 2008.

[3] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Performance evaluation of fingerprint verification systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, pp. 3 - 18, 2006.

[4] P. Grother and E. Tabassi, "Performance of Biometric Quality Measures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, pp. 531 - 543, 2007.

[5] M. Tico and P. Kuosmanen, "Fingerprint matching using an orientation-based minutia descriptor," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, Issue , Page(s): , pp. 1009 - 1014, 2003.

[6] ISO/IEC FDIS 24794-1 Information Technology - Biometrics - Biometric Sample Quality: Part1 Framework, 2008.

[7] T. P. Chen, X. Jiang, and W. Y. Yau, "Fingerprint image quality analysis," presented at Image Processing, 2004. ICIP '04. 2004 International Conference on, 2004.

[8] Y. Chen, S. C. Dass, and A. K. Jain, "Fingerprint Quality Indices for Predicting Authentication Performance," *Lecture Notes in Computer Science*, vol. 3546, pp. 160 - 170, 2005.

[9] L. Hong, Y. Wan, and A. Jain, "Fingerprint image enhancement: algorithm and performance evaluation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, pp. 777 - 789 1998.

[10] E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis," presented at Image Processing. 2002. Proceedings. 2002 International Conference on, 2002.

[11] L. Shen, A. C. Kot, and W. M. Koo, "Quality Measures of Fingerprint Images," *Lecture Notes in Computer Science* vol. 2091, pp. 266 - 271, 2001.

[12] D. Maltoni, D. Maio, A. k. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*: Springer, 2005.

[13] J. Yin, E. Zhu, X. Yang, G. Zhang, and C. Hu, "Two steps for fingerprint segmentation," *Pattern Recognition*, vol. 48, pp. 5182-5201, 2007.

Authorized licensed use limited to: University of Tokyo. Downloaded on March 26,2023 at 14:38:00 UTC from IEEE Xplore. Restrictions apply.