情報保護學會論文誌
第19卷 第3號, 2009. 6

매칭성능 기반의 지문샘플 품질측정방법에 관한 비교연구*

김 장 룡,[1†] 김 학 일,[1‡] Stephen Elliott[2]
[1]인하대학교, [2]Purdue University

Matching Performance-Based Comparative Study of Fingerprint Sample Quality Measures*

Changlong Jin,[1†] Hakil Kim,[1‡] Stephen Elliott[2]
[1]INHA University, [2]Purdue University

요 약

지문샘플의 품질은 지문인식 시스템에서 매칭 성능에 영향을 주는 중요한 요소이다. 지문인식 시스템에서 낮은 품질의 지문영상을 제거하면 인식 에러율이 현저하게 줄어드는 것을 알 수 있다. 본 논문에서는 minutiae 기반의 지문인식 알고리즘을 이용하여 단일 샘플 품질 측정방법에 대한 유효성을 평가하였다. 우선, minutiae 기반의 지문인식 알고리즘의 매칭성능에 영향을 주는 여러 가지 요소에 대하여 검증하고, 다음, 현재 흔히 사용하는 측정방법에 대해 연구하고, 그 중 효과적인 품질측정방법들을 선택하여 NIST, QualityCheck와 Verifinger 5.0을 이용하여 비교분석 하였다. FVC 데이터베이스로 실험한 결과 단일 측정방법도 매칭성능을 효과적으로 향상함을 알 수 있었다.

ABSTRACT

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, Quality measure, Matching performance, 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

접수일(2009년 4월 22일), 게재확정일(2009년 6월 14일)
* 본 연구는 2002년도 정부(교육과학기술부)의 재원으로 한국과학재단 지정 생체인식연구센터의 지원을 받아 이루어졌습니다. (No. R112002105080010(2009))
† 주저자, cljin@vision.inha.ac.kr
‡ 교신저자, hikim@inha.ac.kr

(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. Most of existing quality methods only assess the quality of ridge-valley texture[7-11], and ignore the common area and deformation, since only the single sample should be evaluated exclude the matching case.

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 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 follows: 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

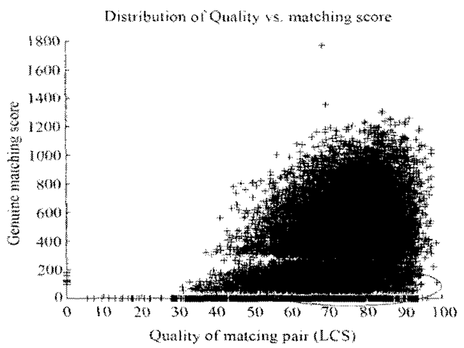
$$EER \propto EER_q + EER_a + EER_d + \epsilon \qquad (1)$$

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

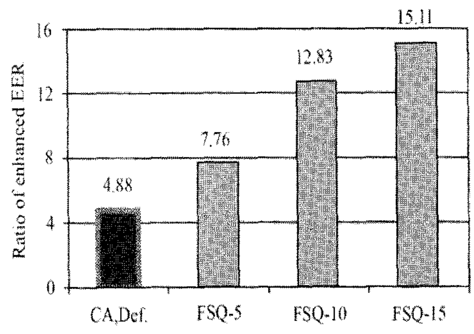
[Fig. 1] shows the plot of quality score vs. genuine matching score for overall selected datasets - FVC 2000 1a, 2a, 3a, FVC 2002 1a, 3a, FVC 2004 1a, and 2a (refer to Section 4.1 for continued discussion). The circled region indicates that the genuine matching scores (obtained by Verifinger) can be very low despite of good sample quality, $q=\sqrt{q_1 q_2}$, where *q1*, *q2* are Local Clarity Scores (refer to Section 3.3 for definition) of a pair of genuine matching samples. The main causes for this type of matching pairs are CA and DF. One example of a low matching score pair with good fingerprint quality is shown in [Fig. 2].

[Fig. 2] Example of a genuine pair with good quality but small common area and high deformation (FVC 2004 DB 1a 6_1 (top), 6_3 (bottom)).

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*. [Fig. 3] shows this assertion: 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



[Fig. 1] Scatter plot for sample quality (*LCS*) vs. genuine matching score.



[Fig. 3] Comparison of influencing factors to *EER*.

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 influence caused by QRVT, EER_a and EER_d are taken as constants, and (1) can be simplified as:

$$EER \propto EER_q + C_{ad} + \epsilon = EER_q + C \quad (2)$$

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

3.1 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 1].

3.2 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

[Table 1] 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]

direction of ridges using the intensity gradient. The coherence matrix of blocks is computed by (3) 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* in (7), where $M \times N$ is the total number of blocks. Similar work has been done by Chen et al. [8], which is measured by (8).

$$Coh = \frac{1}{N}\sum_{N}\left\{\begin{bmatrix} dx \\ dy \end{bmatrix}[dx\ dy]\right\} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{3}$$

$$\lambda_{max} = \frac{(a+b)+\sqrt{(a-b)^2+4c^2}}{2} \tag{4}$$

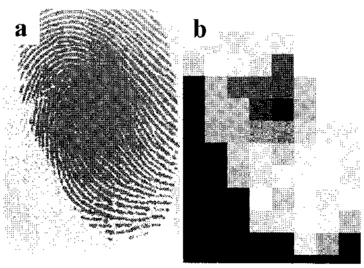
$$\lambda_{min} = \frac{(a+b)-\sqrt{(a-b)^2+4c^2}}{2} \tag{5}$$

$$ocl_1 = \frac{\lambda_{min}}{\lambda_{max}} \tag{6}$$

$$OCL = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} ocl_1(i,j) \tag{7}$$

$$ocl_2 = \frac{(\lambda_{max}-\lambda_{min})^2}{(\lambda_{max}-\lambda_{min})^2} \tag{8}$$

In fact, the original *OCL* measure is designed to find out the edges and corners in digital image. The ridge-valley (R-V) structure of fingerprint, in most cases, contains a lot of noise which causes high frequency components in the spectrum domain. *OCL* measure is very sensitive to this high frequency component. Therefore, the low pass Gaussian filter is used to remove the noise to ensure the correct orientation in ridge-valley structure can be obtained. This is shown in [Fig. 4].



[Fig. 4] 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 oclscore 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. So, 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.

An even type of a Gabor filter in the spatial domain is given by

$$\begin{aligned} & h(x_{\theta_k}, y_{\theta_k}, \theta_k, f) \\ & = \exp\left(-\frac{x_{\theta_k}^2}{2\sigma_x^2} - \frac{y_{\theta_k}^2}{2\sigma_y^2}\right) \times \exp(i2\pi f x_{\theta_k}) \end{aligned} \tag{9}$$

where, for a pixel (x,y) in an image, $(x_{\theta_k}, y_{\theta_k})$ is obtained by

$$\begin{aligned} x_{\theta_k} &= x\cos\theta_k + y\sin\theta_k \\ y_{\theta_k} &= -x\sin\theta_k + y\cos\theta_k \end{aligned} \tag{10}$$

And, θ_k (k=1,...,8) is the angle of the kth Gabor filter, f is the spatial frequency of ridges which equals to the inverse of average interval between parallel ridges, and σ_x and σ_y are the standard deviations of the Gaussian envelope along the x and y

directions, respectively. In the experiment, they are set to the twice of the average ridge interval of a fingerprint.

For each block of size w, the response of the kth Gabor filter is defined as (11), and the standard deviation of the eight responses is obtained as (12), where $\bar{g}(i,j)$ is the average of $g(\cdot)$, $(k=1,\dots,8)$ of (i,j)th block. Then, the final quality score GAB by Gabor filter is the average of the block-wise standard deviations as defined in (13).

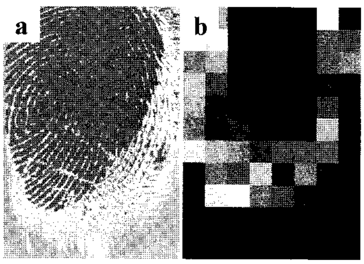
$$g_k(X,Y,\theta_k,f) = \left| \sum_{x=-w/2}^{w/2-1} \sum_{y=-w/2}^{w/2-1} I(X+x,Y+y)h(x,y,\theta_k,f) \right| \quad (11)$$

$$gStd(i,j) = \sqrt{\frac{1}{8}\sum_{k=1}^{8}(g_k(\cdot)-\bar{g}(i,j))^2}, \quad (i=1\dots M, j=1\dots N) \quad (12)$$

$$GAB = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} gStd(i,j) \quad (13)$$

Compared to [8] where each block is classified to either good or bad by thresholding $gStd(i,j)$, GAB in this study produces continuous value to represent the quality of fingerprint image.

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. [Fig. 5] shows the score map for measure GAB. The R-V region of good quality produces a higher score while the R-V region of bad quality does a lower score.



[Fig. 5] Sample image with its block-wise score map for measure GAB
(Light: good quality; Dark: bad quality).

3.3 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 as in (14) and (15). The local clarity score (lcs) is the average value of α and β by (16). Finally, the overall LCS is calculated by (17).

$$\alpha(i,j) = V_B(i,j)/V_T(i,j) \quad (14)$$

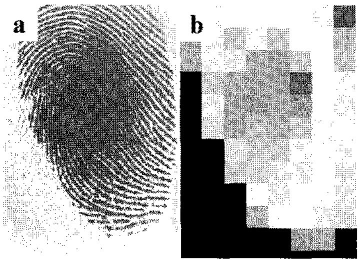
$$\beta(i,j) = R_B(i,j)/R_T(i,j) \quad (15)$$

$$lcs(i,j) = \{\alpha(i,j)+\beta(i,j)\}/2 \quad (16)$$

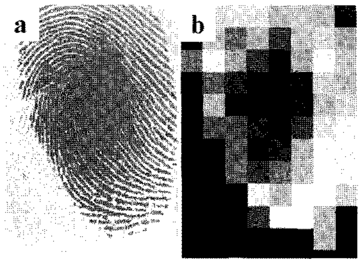
$$LCS = \frac{1}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N} lcs(i,j) \quad (17)$$

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. [Fig. 6] shows the sample image and its lcs score map.

Hong et al. [9] proposed a fingerprint sample quality measure based on the



[Fig. 6] Sample image (a) and *Ics* score map (b) (Light: good quality; Dark: bad quality).



[Fig. 7] An example of *CNT* measure (Light: good quality; Dark: bad quality).

contrast of ridge and valley (*CNT*), which is defined as

$$CNT = \frac{1}{i}\sum_{i}\left[(\max(V_i)-\min(R_i))*100/255\right] \tag{18}$$

where i is the number of ridge-valley pairs in one block, V_i and R_i stands for gray-scale 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. An example is shown in [Fig. 7].

3.4 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 by (22) for the band powers. Finally, the *ENG* is normalized by (23).

$$P(k,l) = |F(k,l)|^2 \tag{19}$$

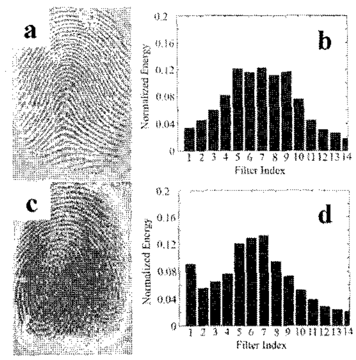
$$E_t = \sum_{k=0}^{N-1}\sum_{l=0}^{M-1} R_t(k,l)P(k,l) \tag{20}$$

$$P_t = \frac{E_t}{\sum_{t=0}^{T-1} E_t} \tag{21}$$

$$E = -\sum_{t=0}^{T-1} P_t \log P_t \tag{22}$$

$$Q = \frac{\log(T-1)-E}{\log(T-1)} \tag{23}$$

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. [Fig. 8] shows two samples of a same low score. The band spectra distributions are obviously different. The pattern of the upper power distribution is more concentrated in the center which represents



[Fig. 8] Two samples with band spectra distribution.

a higher quality. On the contrary, if the band spectra trends to the low frequency, the corresponding fingerprints are worse samples. However, this measure cannot distinguish this case.

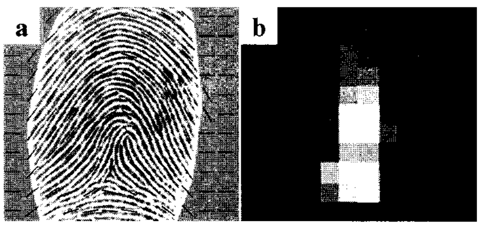
3.5 Ridge flow

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, hence, the Local Orientation Quality (loq) can be computed by (24)

$$loq(i,j) = \frac{1}{8}\sum_{m=-1}^{1}\sum_{n=-1}^{1}|v_4(i,j)-v_4(i-m,j-n)| \quad (24)$$

where v_4 stands for the orientation angle and m, n stands for serial number of 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. An example of local orientation quality computation is shown in [Fig. 9].

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



[Fig. 9] Computation of the local orientation quality. The high curvature area produces a high value.

accumulated and mapped into a global direction score. This method will fail in singular area or high curvature area as shown in [Fig. 9].

3.6 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 4.1). 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. [Table 2] shows the average correlations among each selected measures and two software tools.

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

〔Table 2〕 Correlations among AFSQ measures

Correlation	*OCL*	*GAB*	*LCS*	*ENG*	*CNT*	*LOQ*	*COF*	*AWA*	*NFI*
OCL	1	0.70	0.92	0.66	0.69	0.43	0.22	0.63	0.45
GAB		1	0.74	0.41	0.67	0.05	0.24	0.43	0.32
LCS			1	0.65	0.77	0.37	0.20	0.68	0.45
ENG				1	0.53	0.41	0.15	0.46	0.10
CNT					1	0.21	0.14	0.61	0.34
LOQ						1	0.56	0.39	0.19
COF							1	0.12	-0.03
AWA								1	0.46
NFI									1

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.

〔Fig. 11〕 shows the boxplots for all the implemented measures on all the datasets. *OCL*, *LCS*, *ENG*, and *LOQ* have the similar score distributions on seven datasets. These results show that there are not too much difference among sensors for these measures, namely, these measures are sensor independent. Meanwhile, the rest of the measures and the two quality-check software packages have different quality distributions for different datasets. It shows that these measures have been influenced by sensors.

〔Table 3〕 shows the top three datasets in terms of the mean of the quality score distribution in terms of AFSQ measures. The frequency of measured as good quality datasets are: FVC 2004 1a six times, FVC 2002 1a five times, FVC 2000 3a five times, FVC 2000 1a and FVC 2000 2a one time. Therefore, the FVC 2004 1a is the best dataset from the aspect of AFSQ, and the QualityCheck and NFIQ agree with this result. In other words, all of the quality measures evaluate the FVC 2004 1a as a good dataset. However, the matching performance of this dataset is worst. The reason is that this data has a large influence on CA and DF. For this dataset, the problems of CA and DF have similar influences on the QRVT.

〔Table 3〕 Top three good quality datasets for each measure

Quality measure	Top three datasets
OCL	2000-2a; 2002-1a; 2004-1a
GAB	2004-1a; 2002-1a; 2000-3a
LCS	2002-1a; 2004-1a; 2000-1a
ENG	2000-1a; 2000-3a; 2004-1a
CNT	2002-1a; 2004-1a; 2000-3a
LOQ	2000-3a; 2000-2a; 2000-1a
COF	2004-1a; 2002-1a; 0000-3a

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.

〔Table 4〕 All matching performance of two matchers for the selected datasets

Matcher	FVC 2000 DB 1a(%)	FVC 2000 DB 2a(%)	FVC 2000 DB 3a(%)	FVC 2002 DB 1a(%)	FVC 2002 DB 3a(%)	FVC 2004 DB 1a(%)	FVC 2004 DB 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

〔Table 5〕 The FVC datasets employed in this work

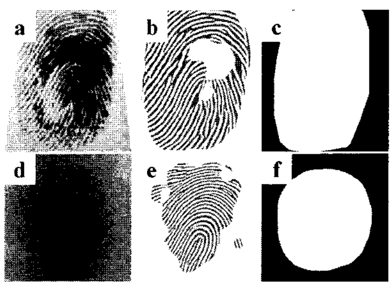
Items	FVC 2000 DB 1a(%)	FVC 2000 DB 2a(%)	FVC 2000 DB 3a(%)	FVC 2002 DB 1a(%)	FVC 2002 DB 3a(%)	FVC 2004 DB 1a(%)	FVC 2004 DB 2a(%)
Sensor type	Optical	Capacitive	Optical	Optical	Capacitive	Optical	Optical
Image size	300x300	256x364	448x478	388x374	300x300	640x480	328x364
Set A (wxd)	100x8	100x8	100x8	100x8	100x8	100x8	100x8
Resolution	500dpi	500dpi	500dpi	500dpi	500dpi	500dpi	500dpi

4.1 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 4〕 shows the zero-rejected all-matching *EER*'s for seven selected datasets using two matching algorithms.

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. More characteristics are described in 〔Table 5〕.

Generally, the foreground in a fingerprint recognition system merely includes the region that contains a valid fingerprint pattern and takes the non-ridge regions and unrecoverable regions as background 〔13〕. This may be a good strategy in feature extraction and can avoid extracting spurious minutiae points. In the evaluation of AFSQ however, the foreground should contain the entire region where the fingertip makes contact with the sensor at one time, in spite of whether it is unrecoverable or not. 〔Fig. 10〕 shows the sample images which have different foreground in AFSQ evaluation. The left column shows the original fingerprint samples, and the middle column shows the segmentation result from a commercial matcher, and the last column showsthe foreground image used in this experiment.



〔Fig. 10〕 The foreground defined in AFSQ. (a),(b),(c): the difference for hole type segmentation. (d), (e), (f): the difference for boundary segmentation.

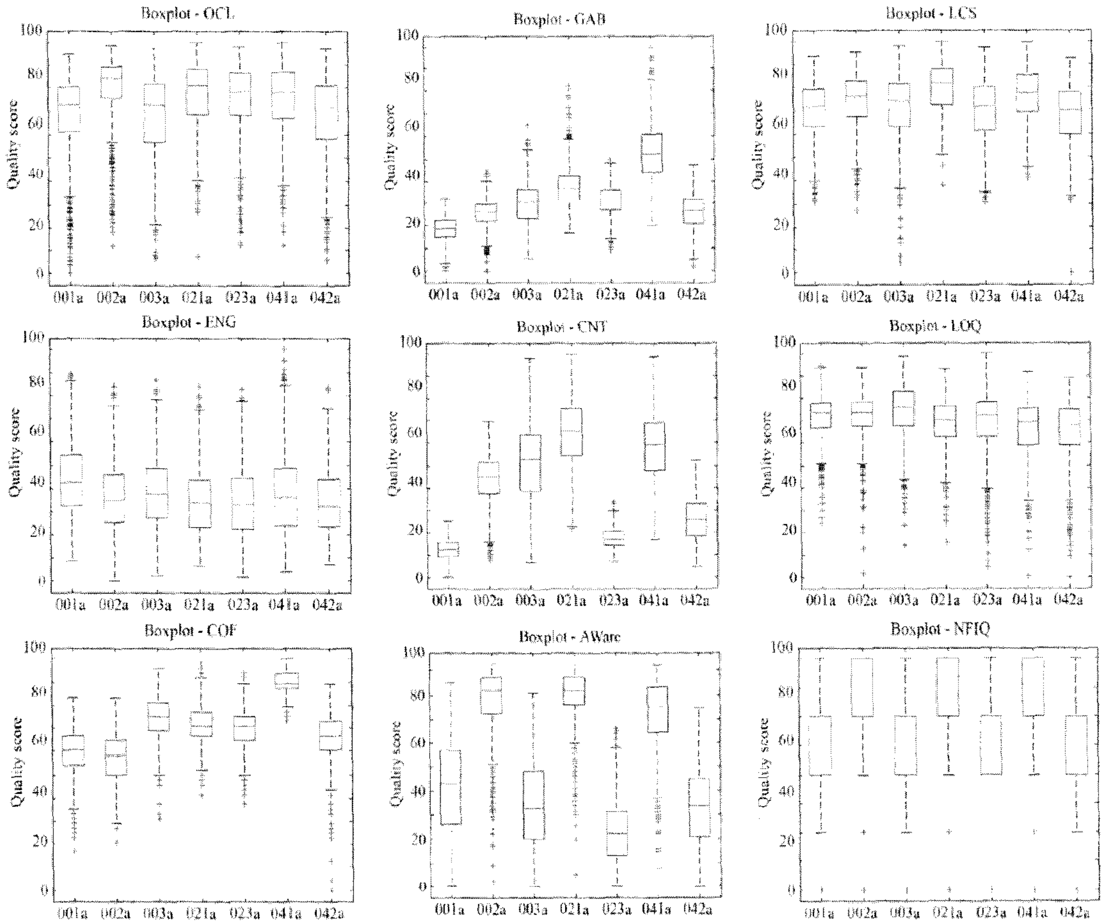
4.2 Matching protocol

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 areremoved 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

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

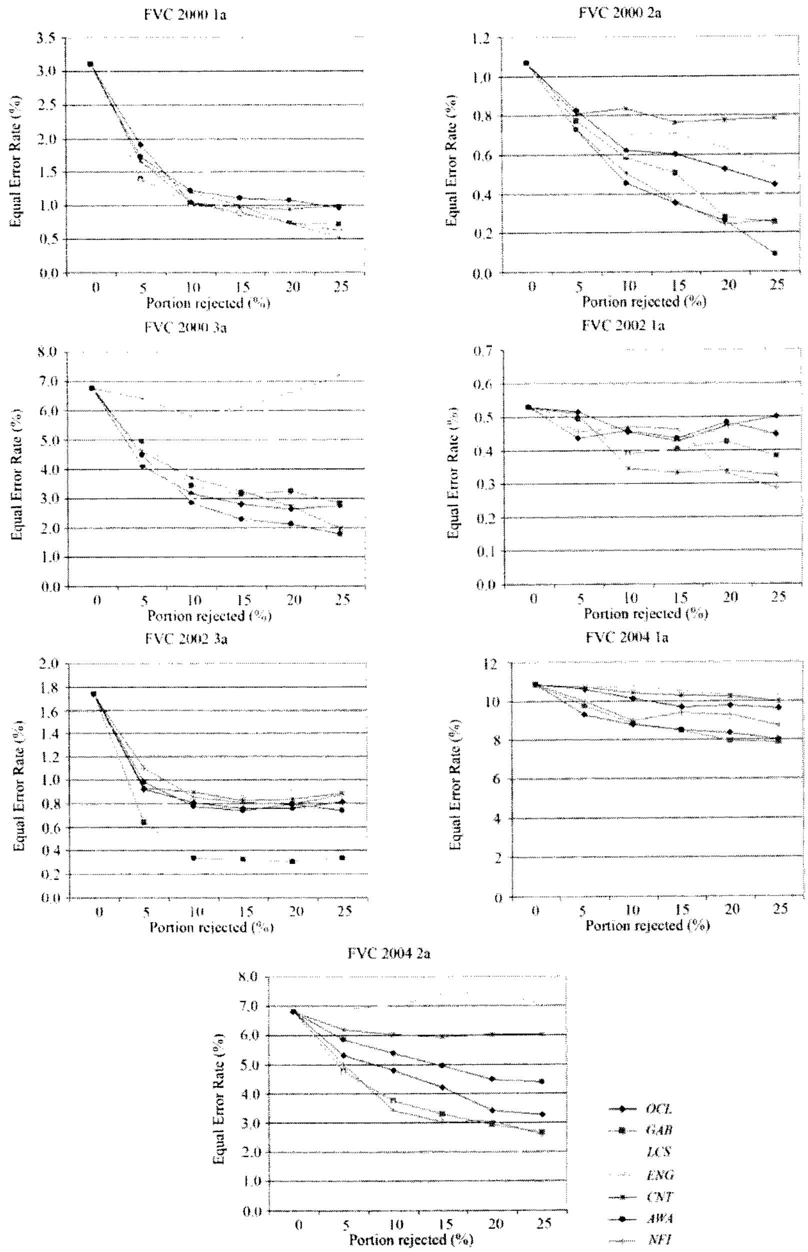
And the performance in each $k\%$-rejected all-matching experiment is denoted by PEM_{EER-k}.



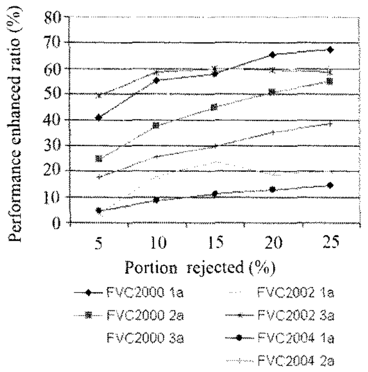
[Fig. 11] Boxplot for each AFSQ measures and two software measures.

4.3 Results and discussion

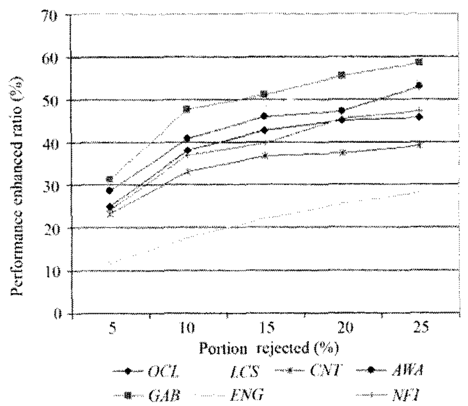
Five selected AFSQ measures are implemented on seven FVC datasets. The results of a portion-rejection all-matching experiment are shown in [Fig. 12]. In general, the Equal Error Rates of this portion-rejection all-matching experiment decreased when the lowest quality samples were rejected. The significant 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



[Fig. 12] Matching performance as samples with the lowest quality value are reject.



[Fig. 13] Average performance enhanced ratio of selected AFSQ measures for all datasets.



[Fig. 14] Average performance enhancement ratio for each quality measure.

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 3.6). 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. The average performance enhanced ratios of the selected AFSQ measures for all dataset are shown in [Fig. 13].

No differences between the optical sensor and capacitive sensor were observed. This is because the AFSQ only assesses the characteristics of ridge-valley, and these characteristics are sensor independent. [Fig. 14] 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.

V. Conclusion

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 indicates 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, no. 8, pp. 1010-1025, Aug. 2008.

[2] D.E. Maurer and J.P. Baker, "Fusing multimodal biometrics with quality estimates via a Bayesian belief network," Pattern Recognition, vol. 41, no. 3, pp. 821-832, Mar. 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, no. 1, pp. 3-18, Jan. 2006.

[4] P. Grother and E. Tabassi, "Performance of Biometric Quality Measures," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, no. 4, pp. 531-543, Apr. 2007.

[5] M. Tico and P. Kuosmanen, "Fingerprint matching using an orientation-based minutia descriptor," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 8, pp. 1009-1014, Aug. 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, pp. 1253-1256, Jan. 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, July 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, no. 8, pp. 777-789, Aug. 1998.

[10] E. Lim, X. Jiang, and W. Yau, "Fingerprint quality and validity analysis," presented at Image Processing. 2002. Proceedings. 2002 International Conference on, pp. 469-472, Sep. 2002.

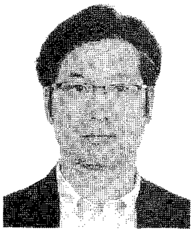
[11] L. Shen, A.C. Kot, and W.M. Koo, "Quality Measures of Fingerprint Images," Lecture Notes in Computer Science vol. 2091, pp. 266-271, June 2001.

[12] D. Maltoni, D. Maio, A.K. Jain, and S. Prabhakar, Handbook of Fingerprint Recognition: Springer, Springer London, 2003.

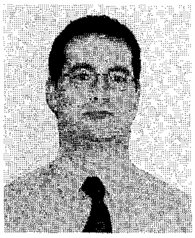
[13] J. Yin, E. Zhu, X. Yang, G. Zhang, and C. Hu, "Two steps for fingerprint segmentation," Image and vision computing, vol. 25, no. 9, pp. 1391-1403, Sep. 2007.

〈著者紹介〉

김 장 룡 (Changlong Jin) 학생회원
1994년 7월: 중국 길림공학원 자동화공학과 학사졸업
2005년 7월: 중국 연변대학교 컴퓨터과학과 석사졸업
2009년 9월~현재: 인하대학교 정보공학과 박사과정
〈관심분야〉 패턴인식, 영상처리, 바이오매트릭스



김 학 일 (Hakil Kim) 종신회원
1983년 2월: 서울대학교 제어계측공학과 학사 졸업
1985년 8월: (미) 퍼듀대학교 전기 컴퓨터공학과 석사 졸업
1990년 8월: (미) 퍼듀대학교 전기 컴퓨터공학과 박사 졸업
1990년 9월~현재: 인하대학교 공과대학 교수
2001년 2월~현재: 한국생체인식포럼 시험평가분과 위원장
2002년 1월~현재: 한국정보보호학회 생체인증연구회위원장
2003년 3월~현재: ISO/IEC JTC1/SC37 (생체인식)WG5(성능평가) apporteur Group
2005년 4월~현재: ITU-T SG17 Q.8 (Telebiometircs)Rapporteur
〈관심분야〉 생체인식, 생체인식 표준화, 정보보호, 패턴인식 및 병렬 영상처리



Stephen Elliott
1996년 8월: (미) Iowa State University, Political Science 학과 학사 졸업
1998년 8월: (미) University of Westminster, Transportation Planning 학과 석사 졸업
2001년 8월: (미) 퍼듀대학교 Curriculum and Instruction 학과 박사 졸업
1992년~현재: 퍼듀대학교 교수
〈관심분야〉 생체인식, 생체인식 표준화, 정보보호 및 패턴인식