Gabor Filters as Candidate Quality Measure for NFIQ 2.0

Martin Aastrup Olsen
Center for Advanced Security Research Darmstadt (CASED)
Mornewegstrasse 32, 64293, Darmstadt, Germany
martin.olsen@cased.de

Haiyun Xu
University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands
h.xu@ewi.utwente.nl

Christoph Busch
Norwegian Information Security Laboratory (NISlab)
Høgskolen i Gjøvik, Teknologivn. 22, 2815 Gjøvik, Norway
christoph.busch@hig.no

Abstract

Quality assessment of biometric fingerprint images is necessary to ensure high biometric performance in biometric recognition systems. We relate the quality of a fingerprint sample to the biometric performance to ensure an objective and performance oriented benchmark. The proposed quality metric is based on Gabor filter responses and is evaluated against eight contemporary quality estimation methods on four datasets using sample utility derived from the separation of genuine and imposter distributions as benchmark. The proposed metric shows performance and consistency approaching that of the composite NFIQ quality assessment algorithm and is thus a candidate for inclusion in a feature vector introducing the NFIQ 2.0 metric.

1. Introduction

In large scale Automated Fingerprint Identification Systems (AFIS), it is important to consider that there will be a certain fraction of individuals who will try to avoid detection. At the border controls of Japan and the United States of America, an individual can be rejected entry if the biometric probe sample captured from the individual matches (has a high comparison score) with a biometric reference already registered in a database, or a watch-list. Consequently, for negative biometric claims a situation can arise where an individual will supply a low quality probe sample on purpose, thus minimizing the chance of detection. Without a method of determining whether the quality of the captured probe sample reaching a sufficient level for recognition purposes, an individual can thus subvert the system. The scenario is substantiated by the findings in [19, 18, 7], where it was established that there is a strong correlation between fingerprint image quality and biometric performance.

Determining the quality of a fingerprint also finds use in other scenarios such as in the context of immigration where a subject can apply for a visa at the embassy or consulate for a given country. In order to verify that the identity of the subject at the border control is indeed the same as the one who received the visa at the embassy, a fingerprint capture is performed at the time of the application. Thus the subject is enroled in the biometric system and can be identified at a subsequent border control. In Europe such a system is implemented for all countries of the Shengen area and is known as the Visa Information System (VIS). In a system such as the VIS the cost of a false reject is high: the subject is either faced with a wasted travel expense and anguish over an unrightful rejection, or the border control will have to employ special procedures to verify the subject identity through other means.

This exemplifies that it is desirable to assess the quality of a fingerprint image before any enrolment transaction is initiated, so that for later comparisons, at least the enrolment sample is of suitable quality. High comparator performance is achieved if a fingerprint's quality is sufficiently good and overall database integrity is improved. This requires poor image samples to be rejected before they are enroled into databases. From an operational perspective, such *Failure To Capture* might trigger a re-capture procedure potentially with more support for the capture subject given by the operator personnel.

Here we present an approach based on Gabor filter responses and evaluate the performance relative to the metrics defined in ISO/IEC TR 29794-4 [10] and NFIQ [18]. The findings presented herein are included in the development considerations of the forthcoming NFIQ 2.0 [3] algorithm, which will be the successor of the widely adopted NFIQ algorithm.

2. Background

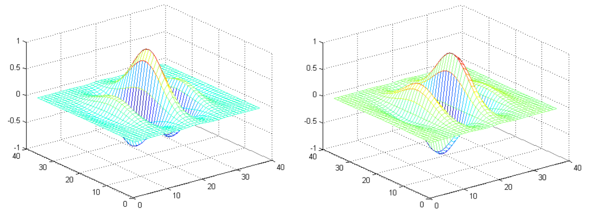
If fingerprint image quality is to be a predictor of the biometric performance, then the quality metric should reflect the signal quality defining both the local and global characteristics in a fingerprint. Locally, the ridge orientation certainty level [12], local orientation [4], or the blockwise similarity to Gabor filter responses [2] can be applied as indicators of quality. In a similar manner, the discrete Fourier transform can be analyzed [13] to give an impression of the global quality. An extensive comparative study on quality metrics was performed by Alonso-Fernandez et al. [1]. Several of these metrics are included in the ISO/IEC technical report on biometric fingerprint sample quality [10].

Fingerprint image quality as a predictor of biometric performance was first published by National Institute of Standards and Technology (NIST) in the technical report Fingerprint Image Quality [18], which documents the development of the NIST fingerprint image quality (NFIQ) algorithm. Here fingerprints are classified into 5 quality levels, ranging from poor to excellent quality, based on several characteristics: ridge orientation flow, ridge orientation certainty, local ridge curvature, and local contrast. Using an interpretation of these characteristics, NFIQ relies on a neural network to perform the classification. In [16], several deficiencies in NFIQ have been identified, among those the important notion that the the definition of the image features used to predict the NFIQ leaves great optimization potential.

Our proposed method and methodology differs from that in [2] as we do not subdivide the input into blocks and we do not use a subjective measure to compare the performance of the algorithm. Instead we work in a pointwise manner and use an objective performance assessment, thus ensuring reproducability of results.

3. Fingerprint Quality Metrics

We propose a fingerprint quality measure based on the Gabor filter responses. In this paper we also evaluate several contemporary quality estimation methods, including *Orientation Certainty Level* (OCL), *Ridge-valley Structure* (LCS), *Ridge-valley Uniformity* (RVU), *Frequency Domain Analysis* (FDA), *Radial Power Spectrum* (POW) and *Orientation Flow* (OF).



(a) (b)

Figure 1. One complex Gabor filter ($f = 0.1, \sigma_x = 6.0, \sigma_y = 6.0, \theta = 1/2\pi$). (a) the real part of the complex Gabor filter; (b) the imaginary part of the complex Gabor filter.

3.1. Quality Metric Based on Gabor Filter Responses

A Gabor filter is a bandpass filter with local support that enhances some spatial frequencies with a certain orientation and attenuates the others [6]. As fingerprints have parallel ridge patterns with well-defined local frequency orientation, the Gabor filter responses of fingerprints reflects the clarity of the fingerprint ridge pattern along a certain direction. Therefore, we choose the responses of a bank of Gabor filters as one quality measure.

The general form of a complex 2D Gabor filter in the spatial domain is given by [8]:

$$h_{\mathrm{Cx}}(x, y; f, \theta, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)\right) \exp\left(j2\pi f x_\theta\right), \tag{1}$$

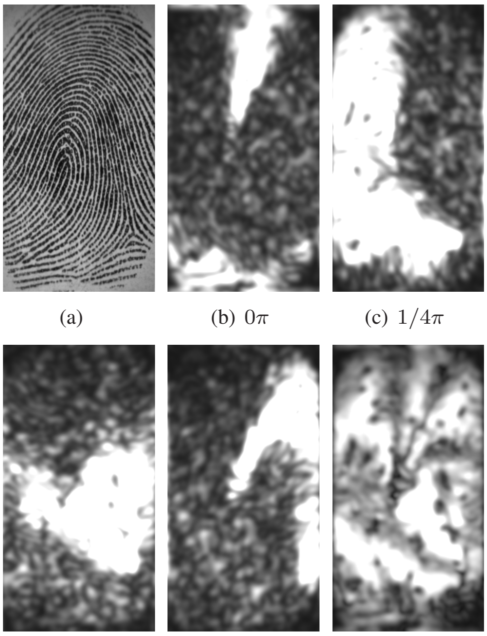
where

$$x_\theta = x \sin\theta + y \cos\theta, \tag{2}$$
$$y_\theta = x \cos\theta - y \sin\theta, \tag{3}$$

f is the frequency of the sinusoidal plane wave along the orientation θ, and σ_x, σ_y are the parameters of the Gaussian window.

We set the filter frequency f as the reciprocal of the average inter-ridge distance. For fingerprint images captured from an adult population with the 500dpi resolution, the average inter-ridge distance is approximately 10 pixels [11]. Therefore, we set $f = 0.1$ in our experiments.

We use $n = 4$ four different orientations, $\theta = (k - 1)/n\pi, k = 1, ..., n$, respectively, for the bank of Gabor filters. Figure 1 presents the real and imaginary part of the Gabor filter with orientation $\theta = 1/2\pi$. The Gaussian parameters are set as $\sigma_x = 6.0$ and $\sigma_y = 6.0$.



(a) (b) 0π (c) $1/4\pi$

(d) $2/4\pi$ (e) $3/4\pi$ (f)

Figure 2. Fingerprint from FVC2002Db2 dataset and its Gabor filter responses. (a) fingerprint image; (b-e) Smoothed absolute value of fingerprint image that have been filtered by Gabor filters with different orientations; (f) the standard deviation value of Gabor responses with different orientations.

We apply the above mentioned bank of Gabor filters to fingerprint images. We take the magnitude of the filter responses and apply a small smoothing filter. A fingerprint and its four Gabor filter responses are illustrated in Figures 2(a)-2(e), respectively.

The Gabor responses reflect the local quality of fingerprints. For the fingerprint areas with clear ridge pattern, the Gabor responses of one or a few orientations should have larger values than other orientations. Whereas for the background or the poor ridge clarity fingerprint areas, the Gabor responses of n orientations will be low and constant. In other words, the variety of the n Gabor responses at a certain area can reflect the clarity of the fingerprint ridge pattern in that area. Therefore, after obtaining the bank of Gabor responses $G_k, k = 1, ..., n$, we take the standard deviation of them, denoted as G_{std} and shown in Figures 2(f). Finally, the Gabor quality is defined as the mean of G_{std}.

In this paper, we also evaluate another Gabor quality measure proposed by Shen et al. [17] in Section 5. The main differences between our method with Shen's are:

- In Shen's method, the fingerprint foreground is segmented into blocks, and each block is marked as "good" or "poor" by comparing the Gabor responses with a fixed threshold T_q. However, for fingerprints obtained by different scanners, optimizing such a fixed threshold T_q is difficult and unreliable. During our evaluation, we noticed such unstable behaviour by setting different T_q. In our method, we obtain the Gabor quality measure without thresholds.
- In Shen's method, another fixed threshold T_b is used to segment fingerprint foreground. However, we notice in our experiments that the size of the fingerprint foreground could influence the recognition performance, i.e., a larger fingerprint foreground could lead to a higher performance. Therefore, in our method, we did not include the fingerprint segmentation part in our evaluation, in order to make the Gabor quality measure also reflect the size of the fingerprint foreground.
- In our method, we evaluated the influence of different Gabor filter orientation number n to the quality measure performance. Instead of using $n = 8$ as proposed by Shen, we found out that setting $n = 4$ will be able to obtain equivalent performances. This setting can increase the calculation speed with a factor of two.

3.2. Other Quality Metrics

The technical report on fingerprint sample quality [10] by the International Organization for Standardization is the most recent overview of recommended fingerprint quality metrics. We use the metrics defined there as a benchmark with which we compare our approach. We use the following metrics:

Orientation Certainty Level (OCL) measures the strength of the energy concentration along the dominant ridge flow orientation within a block by means of computing the blockwise gradient.

Ridge-valley Structure, also called Local Clarity Score (LCS) computes the blockwise clarity of ridge and valleys by applying linear regression to determine a gray-level threshold, classifying pixels as ridge or valley. A ratio of misclassified pixels is determined by comparing with the normalized ridge and valley thickness of that block.

Ridge-Valley Uniformity (RVU) measures the blockwise consistency of the ratio between ridge and valley. A large deviation from the global mean of ridge-valley ratios indicates low quality.

Frequency Domain Analysis (FDA) performs a blockwise Discrete Fourier Transform and measures the energy concentration in a frequency band. Dominance in the very low frequencies indicate low quality.

Radial Power Spectrum (POW) is a measure of maximal signal power in a defined frequency band of the global Radial Fourier spectrum. Ridges can be locally approximated by means of a single sine wave, hence high energy concentration in a narrow frequency band corresponds to consistent ridge structures.

In addition to the quality metrics included in ISO/IEC 29794-4, we also evaluated the Orientation Flow (OF) proposed in [5]. OF is a measure of the rate of change in the blockwise ridge flow across the fingerprint. The quality score decreases as the difference between the dominant ridge orientation of the block and that of its 8 neighboring blocks increases.

4. Quality Assessment

In this paper, we applied the concept of *Biometric Sample Utility* for quality assessment. Biometric Sample Utility is a measure of the biometric performance for a particular sample. The observed utility of the sample d_i^u (the uth presentation of subject i) is computed by means of the genuine and imposter similarity score distributions, e.g. in [9] the utility of a sample is defined as

$$utility_i^u = \frac{\mu_{i,u}^{genuine} - \mu_{i,u}^{imposter}}{\sigma_{i,u}^{genuine} + \sigma_{i,u}^{imposter}}, \tag{4}$$

where μ is the mean and σ is standard deviation. $utility_i^u$ is a measure of the distance between the two distributions: a high utility corresponds to the distributions being well separated and therefore that d_i^u has a low likelihood of being falsely rejected and that the likelihood of an imposter being falsely accepted as d_i^u is also low. Therefore, we say that a high utility reflects a high biometric performance. We will use this definition of utility for our performance comparison.

5. Experiments

The purpose of our experiments is to quantify the predictive performance of the Gabor based quality metric relative to the metrics from ISO/IEC TR 29794-4 and NFIQ. We do so by investigating the Spearman correlation between sample utility as defined in ISO/IEC standard on biometric fingerprint sample quality [9] and the sample quality value for each quality metric. Further we analyse the inter-metric correlations to uncover possible redundant quality metrics.

5.1. Datasets and Protocol

For the experiments we use Db2 (optical) and Db3 (capacitive) of the FVC2002 [14] database and Db1 (optical) and Db2 (optical) of the FVC2004 [15] database. We combine the 8 presentations per subject from the A (100 subjects) and B (10 subjects) sets resulting in 880 samples per dataset.

For each dataset we have computed the utility of each sample according to Eq. 4 using comparison scores obtained using the Neurotechnology Verifinger 6.2 SDK. We investigated 36 parameter permutations of the filter bank size, Gaussian and frequency parameters on the four datasets.

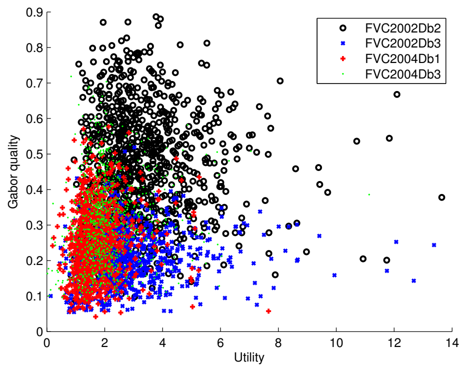


Figure 4. Gabor score plotted against utility for each dataset using $n = 4, f = 0.1, \sigma = 5$.

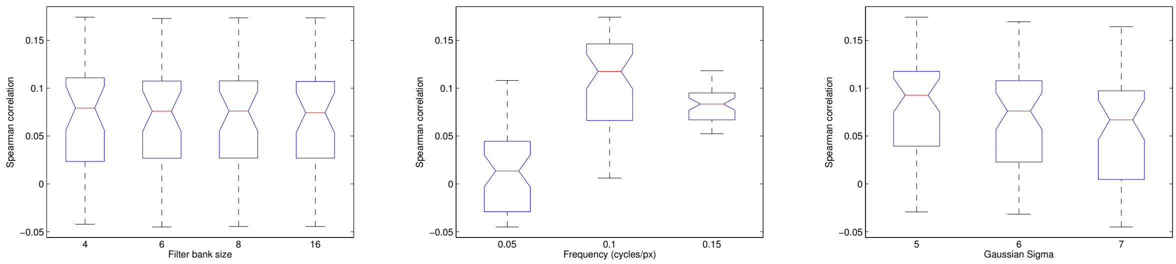
For the Gabor filter bank size we used $n = \{4, 6, 8, 16\}$, the Gaussian parameters σ_x and σ_y were varied with $\sigma = \{4, 5, 6\}$, while the frequency parameter f was varied with $f = \{0.05, 0.1, 0.15\}$.

We compare our approach with NFIQ, OCL, LCS, RVU, FDA, POW, OF and the Gabor filter approach proposed by Shen et al [17]. Except for NFIQ, we apply 36 different parameter configurations on these metrics. In particular, for the blockbased algorithms (OCL, LCS, RVU, FDA) we use blocksizes of $8, 16, 32, 64$.

5.2. Results and Evaluation

The highest Spearman correlation values achieved for each quality metric on each dataset are summarized in table 1. The highest correlation with utility was obtained using the Orientation Flow (OF) approach on the FVC2004Db1 dataset. From the table, we observe that all the investigated metrics have an increased correlation with utility when using the FVC2004Db1 dataset compared to the three other datasets. All metrics have low correlation when using the FVC2002Db2 dataset which we believe is due to the FVC2002Db2 dataset containing particularly difficult to match samples (also called *outliers*). The reason is that outliers will make the estimation of the genuine distributions unreliable, thus influence the utility value. We observe that the proposed Gabor filter approach performs consistently and similarly to the NFIQ metric across the datasets.

The configuration that resulted in the highest correlations for Gabor was obtained with $n = 4, \sigma_x = \sigma_y = 5, f = 0.1$. In our experiments, we observe that the performance of the quality measure does not increase when using values of n higher than 4. In Figure 3 the boxplots (boxes extend to 25th and 75 percentile, with mean marked as a line, and whiskers extending to cover the range excluding outliers) depict the correlation of Gabor quality values across all datasets when fixing respectively the Gabor parameters n 3(a), f 3(b), and



(a) (b) (c)

Figure 3. Spearman correlations between Gabor settings and utility across the parameter space as investigated on the four datasets. (a) for fixed values of n $(4, 6, 8, 16)$; (b) for fixed values of f $(0.05, 0.1, 0.15$; (c) for fixed values of σ $(4, 5, 6)$.

σ. We observe that when viewed across the four dataset these are indeed the settings that show the highest correlations with utility. The influence of the frequency parameter f is dominating the quality score and thus the whiskers extend across a large range in both the filterbank size n and Gaussian σ box plots.

We also investigate the distribution of Gabor quality scores in relation to utility for each of the datasets. We found that the Gabor quality scores are distinct for each dataset and there are some outliers which show high utility and low quality score and vice versa. This is depicted in Figure 4. Neither dataset represents the result of another dataset and hence the consequence is that care must be taken in not extrapolating the results to a general case. The solution is to use datasets containing a larger sample of the population and which contain an even distribution of samples across the entire utility range.

In table 2, we summarize the intermetric correlation to give an overview of which metrics may be redundant. We see almost perfect correlation (0.909) between our approach and the approach presented by Shen et al. This is to be expected as both approaches are based on Gabor filter responses, but we can use this information in combination with the utility correlations in Table 1 to determine that our approach outperforms the Shen approach when seen across all datasets. We also note that there are relatively high correlations with both the OCL and POW metrics, while the correlation between OCL and POW is comparatively lower. This indicates that the Gabor approach may be improved by combining the OCL and POW methods.

6. Conclusions

We introduced a Gabor filter based approach, which correlates with utility on a level comparable or above that of eight contemporary quality metrics. This shows the viability of using the Gabor filter to predict biometric performance on fingerprint images. While we investigated 36 parameter configurations, there is still room for exploring the parameter space for further optimizations.

As the NFIQ metric is a composite metric containing 11 features based on the fingerprint, we conclude that the Gabor filter approach could be included as part of a quality feature vector in order to be able to outperform and replace the NFIQ metric. In relation to the development of NFIQ 2.0, the successor of NFIQ, we have a positive outlook on the inclusion of the presented approach.

Our future research will be improving the Gabor quality approach, focusing on dealing with different types of fingerprints and/or scanners. Another research topic is how to select and combine quality metrics for generating the NFIQ 2.0 metric. Furthermore, the utility measure as defined in Eq. (4) is prone to outliers in the imposter and genuine score distributions, especially the genuine score distributions. This influences the reliability of the utility measure. We can cope with this by using datasets that allow for a larger amount of genuine comparisons, or by using another utility measure, e.g. instead of considering all genuine comparisons, we could consider only the maximum genuine score for a given sample. In our future investigations, we will consider these other options for quality metric performance assessment.

Acknowledgments

This research was supported by the Center for Advanced Security Research Darmstadt (CASED) and the German Federal Office for Information Security (BSI). Further we would like to thank our colleagues from NIST, BSI, secunet, Fraunhofer IGD and BKA for the fruitful discussions which prepared the ground for this paper.

References

[1] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun. A comparative study of fingerprint image-quality estimation methods. 2(4):734–743, 2007. 2

[2] M. S. Altarawneh, W. L. Woo, and S. S. Dlay. Objective fingerprint image quality assessment using Gabor spectrum

Metric	FDA	Gabor	GaborShen	LCS	NFIQ	OCL	OF	POW	RVU
FVC2002Db2	0.072	0.064	0.075	0.016	-0.039	0.072	0.037	-0.047	0.016
FVC2002Db3	0.025	0.132	0.073	0.056	-0.218	0.055	0.061	0.134	-0.107
FVC2004Db1	0.271	0.174	0.190	0.158	-0.213	0.201	0.211	0.119	-0.143
FVC2004Db2	0.070	0.126	0.113	0.125	-0.208	0.152	0.083	0.085	-0.125

Table 1. Spearman correlation between quality metrics and utility for each dataset. The highest correlation found among 36 parameter configurations for each metric is listed. Negative values indicate higher quality score correlate with lower utility (e.g. NFIQ).

Metric	FDA	Gabor	GaborShen	LCS	NFIQ	OCL	OF	POW	RVU
FDA	1.000	0.507	0.464	0.173	-0.068	0.481	0.322	0.364	-0.315
Gabor	0.507	1.000	0.909	0.512	-0.266	0.713	0.449	0.665	0.201
GaborShen	0.464	0.909	1.000	0.522	-0.296	0.709	0.469	0.564	0.188
LCS	0.173	0.512	0.522	1.000	-0.066	0.742	0.441	0.499	0.553
NFIQ	-0.068	-0.266	-0.296	-0.066	1.000	-0.273	-0.319	0.023	0.127
OCL	0.481	0.713	0.709	0.742	-0.273	1.000	0.685	0.482	0.181
OF	0.322	0.449	0.469	0.441	-0.319	0.685	1.000	0.244	-0.006
POW	0.364	0.665	0.564	0.499	0.023	0.482	0.244	1.000	0.284
RVU	-0.315	0.201	0.188	0.553	0.127	0.181	-0.006	0.284	1.000

Table 2. Intermetric Spearman correlations on FVC2004Db1.

approach. In *Proc. 15th Int Digital Signal Processing Conf*, pages 248–251, 2007. 2

[3] O. Bausinger and E. Tabassi. Fingerprint Sample Quality Metric NFIQ 2.0. In *Proceedings of the International Conference of the Special Interest Group on Biometrics (BIOSIG 2011), September 8-9*, pages 167–171, 2011. 2

[4] T. Chen, X. Jiang, and W. Yau. Fingerprint image quality analysis. In *ICIP '04. 2004 International Conference on Image Processing, 2004*, volume 2, pages 1253 – 1256 Vol.2, 2004. 2

[5] T. Chen, X. Jiang, and W. Yau. Fingerprint image quality analysis. In *ICIP '04. 2004 International Conference on Image Processing*, volume 2, pages 1253 – 1256 Vol.2, oct. 2004. 4

[6] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, Jul 1985. 2

[7] J. Fierrez-Aguilar, Y. Chen, J. Ortega-garcia, and A. K. Jain. Incorporating image quality in multi-algorithm fingerprint verification. In *Proc. IAPR Intl. Conf. on Biometrics, ICB, Springer LNCS-3832*, pages 213–220. Springer, 2006. 1

[8] Y. Hamamoto, S. Uchimura, M. Watanabe, T. Yasuda, Y. Mitani, and S. Tomita. A gabor filter-based method for recognizing handwritten numerals. *Pattern Recognition*, 31:395–400, 1998. 2

[9] ISO/IEC. IS 29794-1. Information technology - biometric sample quality - part 1: Framework. Technical report, January 2009. 4

[10] ISO/IEC. TR 29794-4. Information technology - biometric sample quality - part 4: Finger image data. Technical report, 2010. 2, 3

[11] A. Jain, S. Prabhakar, L. Hong, and S. Pankanti. Filterbank-based fingerprint matching. *IEEE Trans. Image Processing*, 9(5):846–859, May 2000. 2

[12] E. Lim, X. Jiang, and W. Yau. Fingerprint quality and validity analysis. In *Proc. Int. Conf. Image Processing 2002*, volume 1, 2002. 2

[13] E. Lim, K.-A. Toh, P. Suganthan, X. Jiang, and W.-Y. Yau. Fingerprint image quality analysis. In *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 2, pages 1241 – 1244 Vol.2, 2004. 2

[14] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2002: Second fingerprint verification competition. In *Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 3 - Volume 3*, ICPR '02, pages 30811–, Washington, DC, USA, 2002. IEEE Computer Society. 4

[15] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2004: Third fingerprint verification competition. In *ICBA*, volume 3072 of *Lecture Notes in Computer Science*, pages 1–7. Springer, 2004. 4

[16] J. Merkle, M. Schwaiger, and M. Breitenstein. Towards improving the NIST fingerprint image quality (NFIQ) algorithm. In *BIOSIG*, volume 164 of *LNI*, pages 29–44. GI, 2010. 2

[17] L. Shen, A. C. Kot, and W. M. Koo. Quality measures of fingerprint images. In *AVBPA*, pages 266–271, 2001. 3, 4

[18] E. Tabassi, C. L. Wilson, and C. I. Watson. NISTIR 7151 - fingerprint image quality. Technical report, NIST, August 2004. 1, 2

[19] C. L. Wilson, M. D. Garris, and C. I. Watson. Matching performance for the US-VISIT IDENT system using flat fingerprints NISTIR 7110. Technical report, NIST, May 2004. 1