Finger image quality assessment features – definitions and evaluation

Martin Aastrup Olsen[1] ✉, Vladimír Šmida[2], Christoph Busch[3]

[1]*Norwegian Biometrics Laboratory, Gjøvik University College, Teknologivegen 22, Gjøvik, Norway*
[2]*Department of Intelligent Systems, Faculty of Information Technology, Brno University of Technology, Antonínská 1, Brno, Czech Republic*
[3]*da/sec – Hochschule Darmstadt, Haardtring 100, Darmstadt, Germany*
✉ *E-mail: martin.olsen@hig.no*

Abstract: Finger image quality assessment is a crucial part of any system where a high biometric performance and user satisfaction is desired. Several algorithms measuring selected aspects of finger image quality have been proposed in the literature, yet only few of them have found their way into quality assessment algorithms used in practice. The authors provide comprehensive algorithm descriptions and make available implementations of adaptations of ten quality assessment algorithms from the literature which operates at the local or the global image level. They evaluate the performance on four datasets in terms of the capability in determining samples causing false non-matches and by their Spearman correlation with sample utility. The authors' evaluation shows that both the capability in rejecting samples causing false non-matches and the correlation between features varies depending on the dataset.

1 Introduction

In large-scale automated fingerprint identification system, there will be a fraction of individuals who are unable to interact with the biometric sensor in the intended way and there will be a certain fraction of individuals who will try to avoid detection. In both cases, quality control in the signal processing subsystem of the biometric system must give an actionable response to the interacting individual or a supervising individual such that errors are addressed immediately.
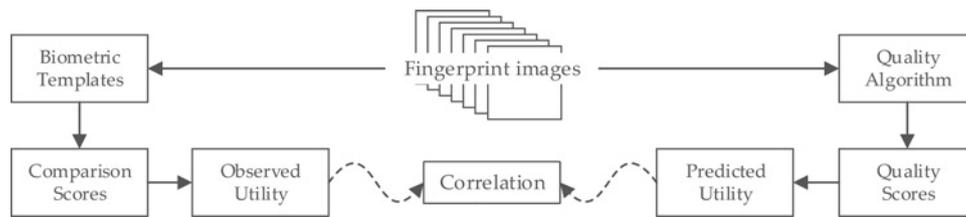
At the border controls of Japan and the United States of America, an individual may be rejected entry if the biometric probe sample captured from the individual matches with a biometric reference already registered in a database, or a watch-list. Consequently, for negative biometric claims a situation can arise where an individual supplies a low quality probe sample on purpose, with the aim of minimising the risk of detection. Thus, without a method of determining whether the quality of the captured probe sample is reaching a sufficient level for recognition purposes, an individual can subvert the system. The scenario is substantiated by the findings in [1–3], where it is established that there is a strong correlation between fingerprint image quality and biometric performance. Determining the quality of a fingerprint finds use in other scenarios such as immigration where an individual can apply for a visa at the embassy or consulate for a given country. To verify that the identity of the individual at the border control is indeed the same as the one who received the visa at the country's consulate, a fingerprint capture is performed at the time of the application. Thus the subject is enrolled in the biometric system and can be identified at a subsequent border crossing.

In Europe, such a system for the exchange of visa data is being implemented for the Schengen area and is known as the Visa Information System (VIS). Its stated purpose is to facilitate the visa application procedure, the fight against fraud and the prevention of visa shopping. Thus the new system shall contribute to the prevention of threats to European countries' internal security. The Council of the European Union (EU) decided to establish the VIS on 8 June 2004 [4] and the European Commission (EC) determined on 30 November 2009 that the VIS should be implemented progressively starting with visa applications first from the North Africa, second the Near East and third the Gulf Region [5].

The VIS has a centralised structure with a central information system, the Central Visa Information System (CS-VIS) and the National Interfaces Visa Information System (NI-VIS). Alphanumeric data as well as biometric data, in particular fingerprints and photographs, are stored in the CS-VIS which can be queried through the NI-VIS by authorised staff of the visa authorities in member states. The standard use case for a query is a Schengen border crossing point, where probe fingerprint images of the traveller are compared against the reference fingerprint images of the visa applicant. The biometric data entering the VIS has to pass certain quality requirements on the local and central levels. The EC has provided member states with a dedicated software kit, known as Kit 4 – Quality Check (USK4) [6], which is used to validate the quality of all fingerprints prior to being inserted in the central system. In the period October 2011 until June 2012 results from consular posts show that around 11% of all visa applications containing fingerprint image failed to meet the local quality assurance [7].

A different approach to biometric border control exists in the United States and is managed under Department of Homeland Security (DHS). Office of Biometric Identity Management (OBIM) which replaced United States Visitor and Immigrant Status Indicator Technology (US-VISIT) in March 2013 is an immigration and border control system used for registering all non-residents entering or exiting the United States of America. Fingerprints are an integral part of the system supporting the goal of establishing an identity to a subject. Having an identity and subject linked allows for checks against watch lists, verification of existing identity documentation such as passports, and discovery of attempted fraudulent identification. In addition to the fingerprint, a face mug shot is also recorded for manual validation purposes.

The cost of a false reject is high in VIS and US-VISIT for the individual, who is applying for entry or a visa, as he is potentially faced with wasted travel expenses and anguish over being unrightfully rejected, or the border control will have to employ special second line procedures to verify the individuals identity claim through other means if at all possible. On the other side, the authorities may fail to identify an individual who is placed in a watch list or is a known visa shopper with potential negative economic or human consequences.

Fig. 1 *Quality scores define predicted utility that should correlate with observed utility derived from comparison scores*

In India, the Unique Identification Authority of India (UIDAI) is managing the Unique Identification number (UID). The 102.5 million UID had been issued by the end of the year 2011 [8] and UIDAI expect 600 million enrolled by 2014 [9]. The goal of the UIDAI is to issue a UID for all Indian residents and thus enable financial inclusion for particularly the rural areas. It is estimated that 40% of Indian citizens in rural areas do not have a bank account and are thus excluded from large parts of society [10]. The UIDAI has faced the challenge of including individuals who work extensively with manual labour that usually leads to worn down and eventually damaged fingerprint patterns of particularly low quality.

For the UID the cost for the individual is high as the success of a bank transaction or any interaction with government officials is dependent on successfully proving an identity claim. Due to the risks exposed by falsely rejecting or accepting a biometric claim, quality control measures must be implemented in any biometric system.

The rest of the paper is organised as follows: Section 2 discusses biometric sample quality, Section 3 contains a detailed overview of finger image quality features with algorithm descriptions, and an evaluation in terms of error-reject curve (ERC), correlation and area under curve (AUC) is found in Section 4. Concluding remarks are found in Section 5.

2 Biometric sample quality

2.1 Components of sample quality

The first component of a biometric system is the capture subsystem, which acquires biometric characteristics. All subsystems operate and depend on the successful acquisition of a biometric sample; hence the acquisition influences all parts of the biometric system. Consequently, it is important to specify requirements for the biometric sample in order to ensure that a sufficient amount of information is available. In the biometric terminology, these requirements are conveyed by the concept of biometric sample quality.

There are different aspects of quality, and one effort to specify those is represented through the International Standard ISO/IEC 29794-1 where three aspects of quality are identified [11]:

- *Character* of a sample is the quality attributable to the inherent features of the source from which the biometric sample is derived. For example a scarred finger has a poor character.
- *Fidelity* of a sample is the quality that describes the degree of similarity between the biometric sample and its source. A sensor with low spatial resolution typically results in a low fidelity sample.
- *Utility* of a sample refers to the predicted impact of an individual sample to the overall performance of the biometric system. It is dependent on both the character as well as the fidelity of the sample.

It is the utility that is of most interest as it is the objective measure for a biometric sample and directly relates the influence of character and fidelity to expected biometric performance. As an objective measure it has the advantage that the utility derived is fully defined by a set of reproducible algorithms and thus it is possible to trace how a particular sample were assigned a given utility.

For the scope of the definition sample quality assessment features and their evaluation it is essential to correlate the predicted with the observed utility. This relationship is depicted in Fig. 1.

Current efforts on standardising fingerprint image quality algorithms are carried out by International Organisation for Standardisation/ International Electrotechnical Commission (ISO/IEC), subcommittee 37, working group 3. The efforts will result in an update to 29794-4 Biometric sample quality – Part 4: Finger image data.

2.2 Influence of capture device technology

There exists a diverse set of capture device technologies each with different properties some of which may negatively or positively affect the recognition error. International Standard ISO/IEC 19794-4:2011 [12] lists a wide range of different sensor technologies, and with each technology exists some limitation with regard to sensor surface area and resolution.

The influence of the capture surface area on recognition accuracy has been addressed in the literature where it has been found that the equal error rate increases as the sensor size decreases [13, 14]. The more general problem of recognition errors due to different sensor technologies used for probe and reference samples is discussed by Ross and Jain [15] who found that there is a strong need for sensor independent algorithms, in particular with regard to matching algorithms.

2.3 Influence of user behaviour and skin condition

User interaction with the acquisition device is another source of recognition error, e.g. placing only the fingertip or side of the finger on the sensor surface is likely to provide insufficient information for the comparator, while pressing the finger too hard on the sensing surface causes elastic deformations, which may exceed compensation capabilities of the feature extraction or matching subsystems. The impact of force on measured image quality and minutia count is addressed by Kukula *et al.* [16].

Degradations in measured quality are expected for certain fingerprint skin conditions [17]. Skin conditions are likely to change due to environmental factors such as temperature and moisture levels, as well as based on users profession (e.g. manual labour or office work). Moreover, it can be observed that specifically in Western cultures there is an increasing number of individuals that suffer from dermatological diseases, that have a direct temporal or persistent impact on the fingerprint pattern. Examples for such diseases are atopic eczema, hyperceratotic eczema and thromboangitis. The skin condition for persons aged 62 and above is generally less moist than that of persons aged 18–25 and the former group generally receives lower quality scores than the latter [18].

2.4 Application areas of quality

Quality assessment can provide real-time actionable feedback at the capture level providing useful information that can lead to recommendations for the re-capture process and thus eventually to a higher quality of captured finger images and consequently increased biometric performance over systems not using quality control.

Further quality survey statistics are helpful in system monitoring to assess the performance of individual capture stations or in the evaluation of capture station operators. Distributed enrolment

systems that operate, e.g. with public–private partnership can be supported by incorporating the quality statistics as one of the criteria, when computing the enrolment fee for the station operator. Measurements over time can also point to systematic or system wide degradation of capture devices as they age and device components such as light emitting diode deteriorate.

Modern biometric system architectures like the UIDAI operate with multiple biometric feature extractors and comparison subsystems and apply load balancing between these background components. With conditional processing of biometric samples of low quality may be sent to a feature extractor that is particularly robust to low quality, but is much slower than the normal feature extractor.

Full-reference and no-reference image quality assessment algorithms have proven useful in the context of fake biometric detection, i.e. the detection of artefacts used in active subversion of a biometric system. An extensive review of the application of this type of image quality measures to perform such detection is addressed in depth by Galbally et al. [19].

2.5 Measuring biometric performance

2.5.1 Overview: Aspects of biometric quality must be expressed in an objective manner to ensure that performance can be measured and compared between different systems.

Tabassi et al. [2] proposed with National Institute of Standards and Technology (NIST) Fingerprint Image Quality (NFIQ) an approach for objective performance assessment based on a measure of the distance between the genuine and imposter comparison score distributions for a given sample; well separated distributions imply that the likelihood of false accept or false reject is low and that it increases with greater overlap between the distributions. This approach is generalised in ISO/IEC 29794-1:2009 [11] which requires that the quality score output by a biometric quality assessment algorithm conveys the predicted utility of the biometric sample. The method suggested for determining the utility of a sample is similar to the one used in NFIQ, i.e. the utility is estimated from the distance between the observed distributions of the sample's comparison scores.

2.5.2 Utility: While no specific method is mandated to derive the utility of a sample Annex A of ISO/IEC 29794-1 [20] suggests that the utility value of a sample may be computed as follows.

Given a set of comparison scores computed using a single comparator the utility of a sample j from subject i is

$$\text{utility}_i^u = \frac{m_{i,u}^{\text{mated}} - m_{i,u}^{\text{non-mated}}}{\sigma_{i,u}^{\text{mated}} + \sigma_{i,u}^{\text{non-mated}}} \tag{1}$$

where $m_{(i,u)}^{\text{mated}}$ and $m_{(i,u)}^{\text{non-mated}}$ are, respectively, the mean of the genuine and imposter comparison score distributions of sample j from subject i and $\sigma_{(i,u)}^{\text{mated}}$ and $\sigma_{(i,u)}^{\text{non-mated}}$ are the standard deviations of, respectively, the genuine and imposter comparison score distributions. Such a utility value can be computed for each comparator using the comparison score provided by that comparator. Methods for fusing utility values exist and a more detailed description of the entire utility computation process can be found in the annex of ISO/IEC 29794-1:2009 [11].

The Spearman's rank correlation coefficient is used in this work as a non-parametric measure of statistical dependence between the quality score and utility.

2.5.3 Error-reject curve: Grother and Tabassi [21] introduced the ERC as a method of assessing the predictive performance of a quality algorithm in terms of false non-match rate (FNMR). We adopt the equations for the one-dimensional case (1D), i.e. where a rejection is driven by the minimum quality q_i of the qualities in a pair of samples. In particular, the combination function H is chosen as the min function

$$q_i = H\left(q_i^{(1)}, q_i^{(2)}\right) = \min\left(q_i^{(1)}, q_i^{(2)}\right) \tag{2}$$

Using the combination function in (2) we form the set $R(u)$ containing the pairwise minima less than u

$$R(u) = \left\{ j{:}H\left(q_j^{(1)}, q_j^{(2)}\right) < u \right\} \tag{3}$$

We then use $R(u)$ to exclude comparison scores starting with the lowest of the pairwise minimums up to some fixed threshold t which corresponds to a FNMR of interest, f. The threshold is obtained using the empirical cumulative distribution function of the comparison scores

$$t = M^{-1}(1 - f) \tag{4}$$

$$\text{FNMR}(t, u) = \frac{\left|\{s_{jj}{:}\, s_{jj} \leq t, j \notin R(u)\}\right|}{\left|\{s_{jj}{:}\, s_{jj} \leq \infty\}\right|} \tag{5}$$

The method models the operational case in which samples are rejected due to low quality. The ERC is determined by progressively excluding a fraction of samples and recalculating the FNMR as the proportion of non-excluded scores which are below the threshold. For good algorithms which output quality scores that are monotonically related to the comparison score the desirable result is that the FNMR decreases quickly with the fraction of samples rejected.

To quantify the decrease in FNMR we propose two metrics which consider the AUC of the ERC with respect to the theoretical best case where the decrease in FNMR equals the fraction of samples rejected due to quality. The first metric, $\eta_{\text{auc}}^{\text{erc}}$ (6) is expressed as the AUC of the ERC subtracted the area under theoretical best

$$\eta_{\text{auc}}^{\text{erc}} = \int_0^1 \text{ERC} - \text{area under theoretical best} \tag{6}$$

The second metric $\eta_{\text{pauc20}}^{\text{erc}}$ (7) is similar to $\eta_{\text{auc}}^{\text{erc}}$ with the modification that only the first 20% the ERC is considered

$$\eta_{\text{pauc20}}^{\text{erc}} = \int_0^{0.2} \text{ERC} - \text{area under theoretical best} \tag{7}$$

2.6 Quality in large-scale biometric systems

When large biometric systems are analysed certain properties will become apparent due to the scale of the operative system. Due to the wide availability of solutions and systems providing a range of possible performance, usability and conformance constellations German Federal Office for Information Security (BSI) have provided technical guidelines for the usage of biometrics in the public sector [22].

The influence of fingerprint quality on biometric performance has been studied in detail by Wilson et al. in their 2004 study on recognition accuracy for US-VISIT [1]. The primary dataset in the study is based on 274,000 right index finger pairs sourced from the Mexican visa program. They show that a proprietary quality measure from Cogent Systems (now 3M [23]) with eight quality bins is a good rank statistic with all eight levels ordered such that the error rate of fingerprints with level 1 is always lower than any other quality level, and the error rate of quality level 2 is always lower than the next six quality levels and so on. The same property is demonstrated for the NFIQ, which categorises finger images into five quality bins. Due to the open source availability of NFIQ this software became widely adopted since its release in 2004. More details on NFIQ will be provided in Section 3.2.

On the two comparators tested they show that the performance degrades consistently with lower quality levels for both the NFIQ and Cogent quality algorithms. An interesting aspect shown in the study is that at quality levels 1–4 the true accept rate is at least 98.2% at a false accept rate of 1.0% for the comparator used in

US-VISIT Automated Biometric Identification System (IDENT). For quality levels 5–8 the true accept rate further degrades from 95.2 to 53.6%. The implication is that half the quality scale is used for finger images that are highly likely to be correctly matched while much less fidelity is given to the remaining more problematic finger images.

Finally, it is concluded that finger image quality is the most critical single factor impacting the performance of fingerprint recognition systems. Wilson *et al.* recommend that image quality should be the first place to improve if the performance of an otherwise reasonably engineered system is poor. One step towards better image quality in biometric systems is to detect when an image is of insufficient quality such that a new capture can be obtained (i.e. actionable quality).
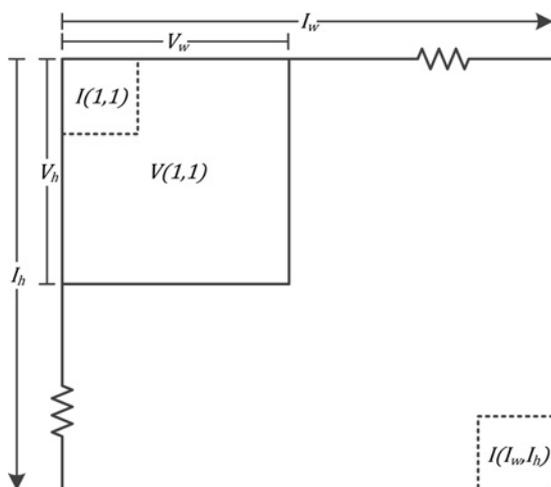
The first study on finger image quality in India was performed by Vatsa *et al.* in a UIDAI case study analysing fingerprints of Indian population using image quality [24]. Using NFIQ, three datasets are analysed with the objective of determining quality aspects of fingerprints obtained from urban and rural population. In their analysis, it is shown that for the ten-print dataset collected among 20,000 individuals in the rural population, 83% of the 200,000 finger images result in a NFIQ score of 1 or 2, indicating good or very good quality. This is unexpected and the authors do show cases where NFIQ has incorrectly assessed the quality of two very poor quality images as being quality scores 1 and 2, respectively. This indicates that NFIQ does misclassify finger image quality classes under some circumstances. Using a much smaller dataset of 1620 images in a prepared dataset with finger images from 27 urban individuals and 81 rural individuals an identification accuracy of 90 and 64%, respectively, was achieved.

It is observed that the finger image quality values reported by NFIQ are highly correlated with comparison scores, which supports the findings by Wilson *et al.* on quality and biometric performance being closely related.

3 Review of finger image quality assessment features

3.1 Overview

In finger image quality analysis, it is important to consider that the finger image can be viewed at a high level (global finger image quality) and a local level (local finger image quality). Analysing the finger image at the global level can provide a fast assessment, but with the tradeoff that local context is neglected. The quality score derived from a global quality feature does reflect the case where, e.g. a part of the fingerprint is blurry or missing, but it does not provide information about the spatial location of the defect. This is in contrast to quality features operating on the local

Fig. 2 *Illustration of block and pixel indexing within an image I with dimensions I_w, I_h. Shown is the pixel $I(1,1)$, the block $V(1,1)$, with dimensions V_w, V_h*

level where spatial location is preserved yielding a more nuanced assessment, and potentially providing important information to the minutiae extractor in the form of a quality vector representing a quality map of the finger image. The common approach to compute local level information is to subdivide the image into blocks as illustrated in Fig. 2. This figure introduces the block and pixel indexing used in the quality feature definitions in Sections 3.4 and 3.5. A taxonomy of finger image quality algorithms has been proposed by Alonso-Fernandez *et al.* [25] where three classes of quality algorithms are described in the context of existing algorithms, namely those which are (a) based on local features; (b) based on global features; and (c) those based on classifiers.

Several fingerprint image quality assessment algorithms have been proposed in the literature, in particular, the classifier-based NFIQ algorithm is used extensively in industry and forms the quality control aspect in operational biometric systems. An extensive review of finger image quality algorithms has been performed by Alonso-Fernandez *et al.* [26], and more recently by Bharadwaj *et al.* [27].

The source code for the quality features specified in Sections 3.4 and 3.5 are made available online [28].

3.2 NFIQ algorithm

The NIST developed in 2004 the NFIQ algorithm and with it a new definition of quality of fingerprint impressions and algorithms for measuring quality was proposed. The work at the time was the first open algorithm for an assessment of finger image quality, which is a predictor of comparator performance [2].

The NFIQ image features are derived from an interpretation of a fingerprint quality map and minutia counts. The localised quality of the fingerprint is measured by computing four quality maps; a direction map, a low contrast map, a low ridge flow map and a high curvature map. For the local analysis, the image is divided into a grid of blocks, where all pixels within a block are assigned the same value derived from all pixels in the block. The direction map represents the local ridge flow as derived through application of the discrete Fourier transform (DFT). The low contrast map separates the background from the fingerprint and detects areas which are blurry. The low flow map is derived from the direction map by assessing if the block has a dominant ridge flow. If no dominant ridge flow exists within a block then minutiae in that block are assumed to be unreliable. The high curvature map is derived from the direction map as the coherence between the directions of one block relative to its eight neighbouring blocks giving an indication of the directional stability in a region. The four maps are fused into one single quality map describing the block wise quality of the finger image. The quality blocks and minutiae from the fingerprint are used to compose a quality feature vector which is used as input for a neural network (i.e. a multi-layer perceptron) which outputs a corresponding quality number representing the expected utility of the given sample.
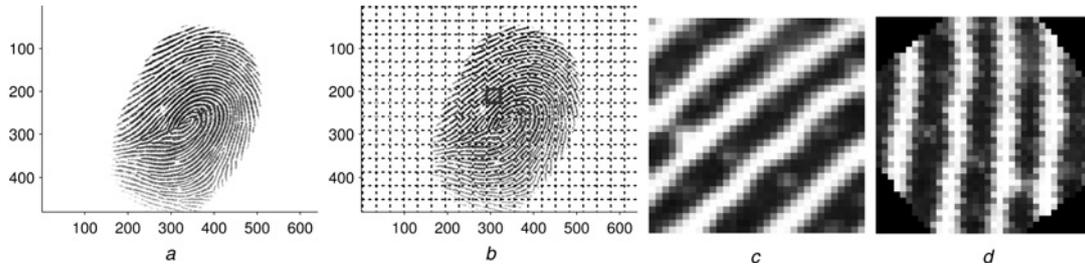
3.3 Improvements to NFIQ

Since its development, NFIQ has been analysed and areas of improvement have been identified by Merkle *et al.* [29]. Herein, a two-step process is proposed where the optimisation of NFIQ is split into classification and prediction parts which improve NFIQ in the points of fingerprint data basis, similarity score statistic, image feature selection, and neural network.

Some of these findings are being applied in the developments of NIST Fingerprint Image Quality 2 (NFIQ 2) [30] which is the evolution of NFIQ initiated by NIST and BSI.

3.4 Local finger image quality

3.4.1 Overview: Local finger image quality assessment algorithms operate on local regions of the image where each local region (i.e. an image block) contains at least two ridge lines. For images with 500 pixels per inch a ridge and valley pair are 8–12

Fig. 3 *Input image used in illustrative examples of the processing of quality*
a Input image
b Subdivision into blocks with example block $\mathbf{V}(10,7)$ marked
c Enlarged view of $\mathbf{V}(10,7)$
d $\mathbf{V}(10,7)$ rotated according to eq. (17)

pixels wide [31] and consequently, to cover two ridge lines a local region must be at least 24 pixels in either dimension.

Valid for all local finger image quality metrics is that they each assign a quality value to each local region, which is aggregated in a specified manner to yield a global quality score for the finger image. It is recommended that the quality algorithm returns a map or histogram of the quality values assigned to each local region. At a later stage, such multiple quality maps can be combined to provide a robust local quality value.

In the remaining of this section state-of-the-art of local finger image quality features are reviewed and detailed algorithm descriptions are provided. Default values provided for the variables are based on an empirical study and shall represent a good balance between computational complexity and accuracy (i.e. utility correlation).

In the following sections, a common input image will be used as an example to illustrate the result of specific points in the specified algorithms. Several of the algorithms perform the same initial steps which are shown in Fig. 3. The input image $\mathbf{I}$ shown in Fig. 3a is image 8_5.bmp from FVC2004 DB1 [32]. For algorithms operating in a block-wise manner, the image is subdivided into blocks according to the overlay grid shown in Fig. 3b. The block $\mathbf{V}(10,7)$ is used as example in local processing and is marked up using a bold line. Fig. 3c shows an enlarged view of $\mathbf{V}(10,7)$ and Fig. 3d shows $\mathbf{V}(10,7)$ rotated according to ridge orientation its dominant ridge orientation as determined using (17).

Computing the block orientation from gradients. From a single block representing a local region of a fingerprint image, the dominant ridge flow orientation is determined by computing the gradient information and then determines the orientation of the principal variation axis.

The numerical gradient of the block is determined using finite central difference for all interior pixels in the x-direction (8) and the y-direction (9)

$$f_x = \frac{I(x+1, y) - I(x-1, y)}{2} \tag{8}$$

$$f_y = \frac{I(x, y+1) - I(x, y-1)}{2} \tag{9}$$

Algorithm 1

Input: Fingerprint image $\mathbf{I}$
Output: FDA quality score Q_{FDA}
1 **for** *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
2   rotate $\mathbf{V}$ such that dominant ridge flow is perpendicular to x-axis
3   crop $\mathbf{V}$ such that no invalid regions are included
4   with $\mathbf{V}$ obtain the ridge-valley signature $T(x)$ (eq. (18))
5   compute the DFT of T to obtain F
6   discard the first component of F
7   determine F_{max} as the index of the term in F with largest magnitude
8   compute $\mathbf{Q}_{FDA}^{local}$ of $\mathbf{V}$ using F (eq. (19))
9 **end**
10 compute Q_{FDA} as the arithmetic mean of all $\mathbf{Q}_{FDA}^{local}$ (eq. (20))

Fig. 4 *FDA algorithm*

With $\boldsymbol{f}_x$ and $\boldsymbol{f}_y$ are the principal axes of variation of $\boldsymbol{V}$ is determined analytically using the sine and cosine doubled angles determined from the arithmetic means of the image gradient covariances (17)

$$a = \overline{\boldsymbol{f}_x^{\,2}} \tag{10}$$

$$b = \overline{\boldsymbol{f}_y^{\,2}} \tag{11}$$

$$c = \overline{\boldsymbol{f}_x \cdot \boldsymbol{f}_y} \tag{12}$$

$$\boldsymbol{C} = \begin{bmatrix} a & c \\ c & b \end{bmatrix} \tag{13}$$

$$d = \sqrt{c^2 + (a - b)^2} + \epsilon \tag{14}$$

$$\sin 2 = \frac{c}{d} \tag{15}$$

$$\cos 2 = \frac{a - b}{d} \tag{16}$$

$$\text{angle}(V) = \frac{a \tan 2(\sin 2, \cos 2)}{2} \tag{17}$$

3.4.2 Frequency domain analysis (FDA): The FDA algorithm (see Fig. 4) operates in a block-wise manner. A 1D signature of the ridge–valley structure is extracted and the DFT is computed on the signature to determine the frequency of the sinusoid following the ridge–valley structure [20, 33].

The value of Q_{FDA}^{local} is undefined if $F_{max} = 1$ or $F_{max} = A(\text{end})$ as both $A(0)$ and $A(\text{end} + 1)$ are not accessing valid indices. Workaround in that case is to set $\boldsymbol{Q}_{FDA}^{local} = 1$. Despite International Organisation for Standardisation (ISO) recommendation [11] of a high quality value indicating a high quality, this is not the case for FDA as specified in ISO/IEC TR 29794-4:2010 [20].
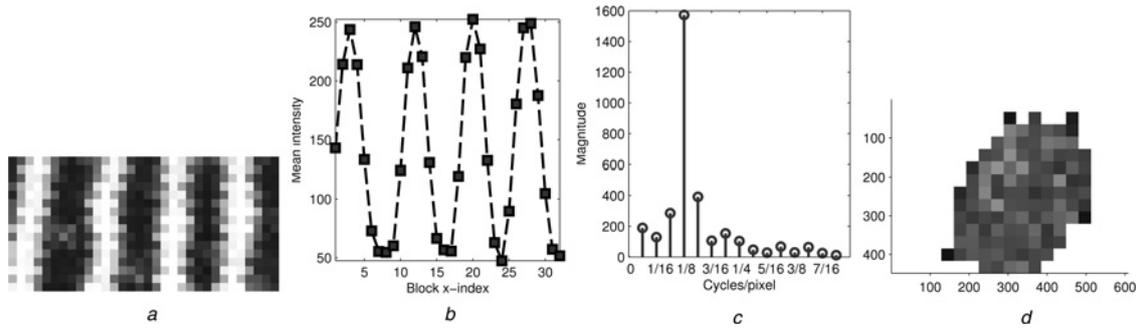
A visual overview of the algorithm outputs is depicted in Fig. 5 where Fig. 5a shows V cropped to contain central area of rotated V; Fig. 5b shows the ridge-valley profile T; Fig. 5c shows the DFT of T after the first component has been removed; and Fig. 5d shows $\boldsymbol{Q}_{FDA}^{local}$ for each V in $\mathbf{I}$.

Ridge–valley signature. The ridge–valley signature is a projection of the mean values of the local region along the y-axis onto a 1D vector. This effectively gives an approximated representation of the fundamental periodicity within the local region. The signature is computed as

$$T(x) = \frac{1}{V_h} \sum_{k=1}^{V_h} I(x, k) \tag{18}$$

Computing the local FDA quality score. The local FDA quality score, $\boldsymbol{Q}_{FDA}^{local}$ is computed as

$$\boldsymbol{Q}_{FDA}^{local} = \frac{A(F_{max}) + C(A(F_{max} - 1) + A(F_{max} + 1))}{\sum_{F=1}^{N/2} A(F)} \tag{19}$$

Fig. 5 *Processing steps of FDA quality algorithm*
a Central area of input block
b Ridge-valley profile
c DFT of ridge-valley profile
d Local qualities

where $C = 0.3$ according to the definition appearing in ISO/IEC TR 29794-4:2010. The effect of the constant is to retain an attenuated amplitude of the frequency bands immediately surrounding F_{max}

Computing the final FDA quality score. The final quality score is computed as the mean of local quality scores

$$Q_{FDA} = \frac{1}{N * M} \sum_{i=1}^{N} \sum_{j=1}^{M} Q_{FDA}^{local}(i,j) \qquad (20)$$

3.4.3 Gabor (GAB) quality:

The GAB quality feature [34] operates on a per-pixel basis by calculating the standard deviation of the Gabor filter bank responses. The size of the filter bank is used to determine a number of filters oriented evenly across the half circle. The strength of the response at a given location corresponds to the agreement between filter orientation and frequency in the location neighbourhood. For areas in the fingerprint image with a regular ridge–valley pattern there will be a high response from one or a few filters likely neighboured orientations. In areas containing background or unclear ridge–valley structure the Gabor response of all orientations will be low and constant.

The Gabor quality feature is resolution dependent (see Fig. 6).

A visual overview of the algorithm outputs is depicted in Fig. 7 where Fig. 7a shows the $\hat{I}$, Figs. 7b–e and f–i depict, respectively, the real and imaginary parts of the Gabor filter at four orientations. The filter response is depicted in Figs. 7j–m and the corresponding Gaussian filtered response is depicted in Figs. 7n–q. Standard deviation of the Gaussian filtered responses is shown in Fig. 7r.

3.4.4 Gabor filter:

The general form of the complex 2D Gabor filter h_{cx} in the spatial domain is given by

$$h_{cx}(x, y; f, \theta, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}\left(\frac{x_\theta^2}{\sigma_x^2} + \frac{y_\theta^2}{\sigma_y^2}\right)\exp\left(j2\pi f x_\theta\right)\right) \quad (21)$$

where

$$x_\theta = x \sin\theta + y \cos\theta \qquad (22)$$

$$y_\theta = x \cos\theta - y \sin\theta \qquad (23)$$

and f is the frequency (cycles/pixel) of the sinusoidal plane wave along the orientation θ. The size of the Gaussian smoothing window is determined by σ_x, σ_y. The filter bank size n is used to compute the differently oriented Gabor filters composing the filter bank. Computing θ given n is done as
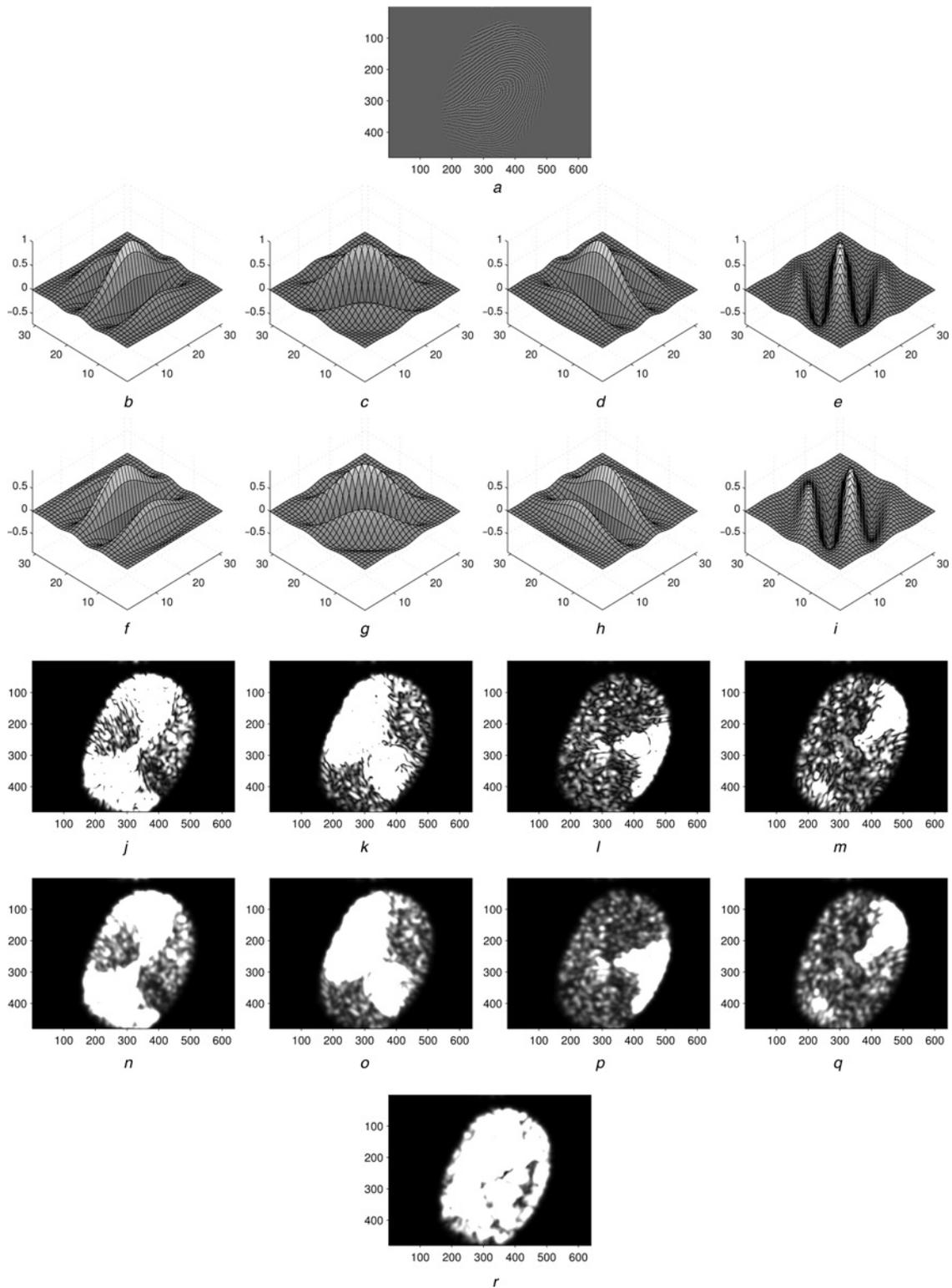
$$\theta = \pi\frac{k-1}{n}, \quad k = 1, \ldots, n \qquad (24)$$

3.4.5 Gabor-Shen (GSH) quality:

GSH [35] is a Gabor-based feature separating blocks into two classes: good and bad. The scalar quality is the ratio between the number of foreground blocks and the number of foreground blocks marked as poor.

The filter response of each Gabor kernel in the filter bank is computed on the pixels in each block and a standard deviation is computed on the responses. Using thresholding each block is determined to be foreground, background, and poor or good quality. The algorithm is outlined in Algorithm 3 (see Fig. 8).

The GSH quality feature is resolution dependent. Shen *et al.* suggest $\sigma_x = \sigma_y = 4$, $f = 0.12$, $n = 8$, $b = 30$ and that T_b and T_q are empirically determined according to dataset. We have found that setting $T_b = 1$ and $T_q = 2$ to yield good results across several datasets.

A visual overview of the algorithm outputs is depicted in Fig. 9 where Figs. 9a–d and e–h depict, respectively, the real and imaginary parts of the Gabor filter at four orientations. The filter response is depicted in Figs. 9i–l and the standard deviation of the responses is shown in Fig. 9m. The mean of standard deviations in each block is shown in Fig. 9n indicating in light grey-scale values those image regions, where good quality is assumed. Blocks marked as foreground are shown in Fig. 9o and blocks marked as poor are shown in Fig. 9p.

Algorithm 2

Input: Fingerprint image **I**
Output: GAB quality score Q_{GAB}

1. Convolve **I** with 2D Gaussian kernel with $\sigma = 1$ and subtract from **I** to yield $\hat{I}$
2. Generate filter bank by computing Gabor filter for each rotation θ (eqs. (21) and (24)); Compute the Gabor response for each pixel in $\hat{I}$ for each filter in the filter bank
3. Convolve the magnitude (complex modulus) of each Gabor response with a 2D Gaussian kernel with $\sigma = 4$
4. Compute the standard deviation of the Gabor magnitude response values at each location yielding a map of standard deviations
5. Sum the map of standard deviations and normalise according to number of sample points (typically size of **I**) to produce the Gabor quality score Q_{GAB}

Fig. 6 *GAB algorithm*

Fig. 7 *Processing steps of GAB algorithm*

a Input image filtered by subtraction of Gaussian convolution

b–e Real part of Gabor filter at $\frac{0\pi}{4}$, $\frac{1\pi}{4}$, $\frac{2\pi}{4}$, $\frac{3\pi}{4}$

f–i Imaginary part of Gabor filter at orientations $\frac{0\pi}{4}$, $\frac{1\pi}{4}$, $\frac{2\pi}{4}$, $\frac{3\pi}{4}$

j–m Magnitude of filter responses at each orientation

n–q Gaussian convolution of magnitudes

r Standard deviation of per pixel magnitude

3.4.6 Local clarity score (LCS): LCS [20, 36] computes the block-wise clarity of ridges and valleys by applying linear regression to determine a grey-level threshold, classifying pixels as ridge or valley. A ratio of misclassified pixels is determined by comparing with the normalised ridge and valley widths of that block (see Fig. 10).

Algorithm 3

Input: Fingerprint image $\mathbf{I}$
Output: GSH quality score Q_{GSH}

1 Generate filter bank of size n by computing Gabor filter for each rotation θ
2 **for** *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
3   Compute the n Gabor responses at each pixel in $\mathbf{V}$
4   Compute the standard deviation G of Gabor reponses for $\mathbf{V}$
5 **end**
6 With G segment $\mathbf{I}$ into foreground and background
7 Determine foreground blocks V_f as those $\mathbf{V}$ where $\mu_i > T_b$
8 Determine poor quality blocks V_p as those $\mathbf{V}$ where $\mu_i > T_b$ and $\mu_i < T_q$
9 Compute quality score Q_{GSH} as $1 - \frac{V_p}{V_f}$

Fig. 8 *GSH algorithm*

Particular regions inherent in a fingerprint will negatively affect Q_{LCS}. For example, ridge endings and bifurcations or areas with high curvature such as those commonly found in the vicinity of core and delta points.

A visual overview of the algorithm outputs is depicted in Fig. 11a crop of the current block shown in Fig. 11a with the average profile depicted in Fig. 11b. The grey-scale values of the profile and the regression line determined using linear regression is shown in Fig. 11c, with the corresponding binarisation into ridge (black) and valley (white) in Fig. 11d. Figs. 11e–g show, respectively, pixels in

the input block determined to be ridge using threshold; grey-level threshold for ridges across the block shown; pixels above threshold marked in white. Similarly, for valleys in Figs. 11h–j. The LCS for each block is visualised in Fig. 11k with white being maximum clarity.

Determining the proportion of misclassified pixels. For a block V there are v_T pixels in the valley region and v_B pixels in the valley region with intensity lower than a threshold DT. Similarly there are r_T pixels in the ridge region and r_B pixels in the ridge region with intensity lower than a threshold DT. α and β are expressions of these ratios

$$\alpha = \frac{v_B}{v_T} \tag{25}$$

$$\beta = \frac{r_B}{r_T} \tag{26}$$

Determining the normalised ridge and valley widths. The normalised valley width $\bar{W}_v$ and the normalised ridge width $\bar{W}_r$ are determined as
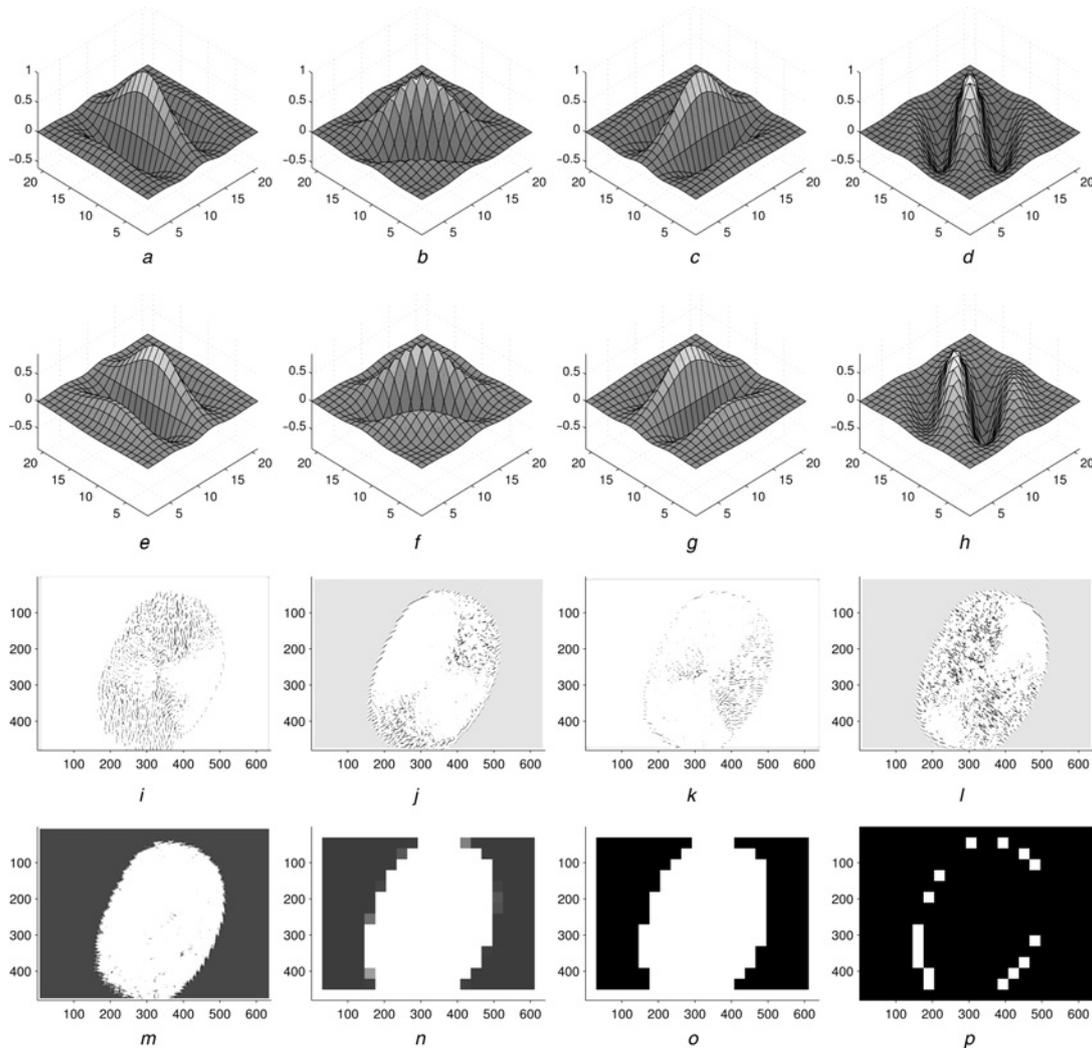
$$\bar{W}_v = \frac{W_v}{(S/125)W^{max}} \tag{27}$$

Fig. 9 *Processing steps of GSH algorithm*

a–d Real part of oriented Gabor filters
e–h Imaginary part of oriented Gabor filters
i–l Magnitude of responses at each orientation
m Standard deviation of filter responses
n Quality level blocks marked on grey scale with white indicating highest quality
o Foreground blocks marked in white
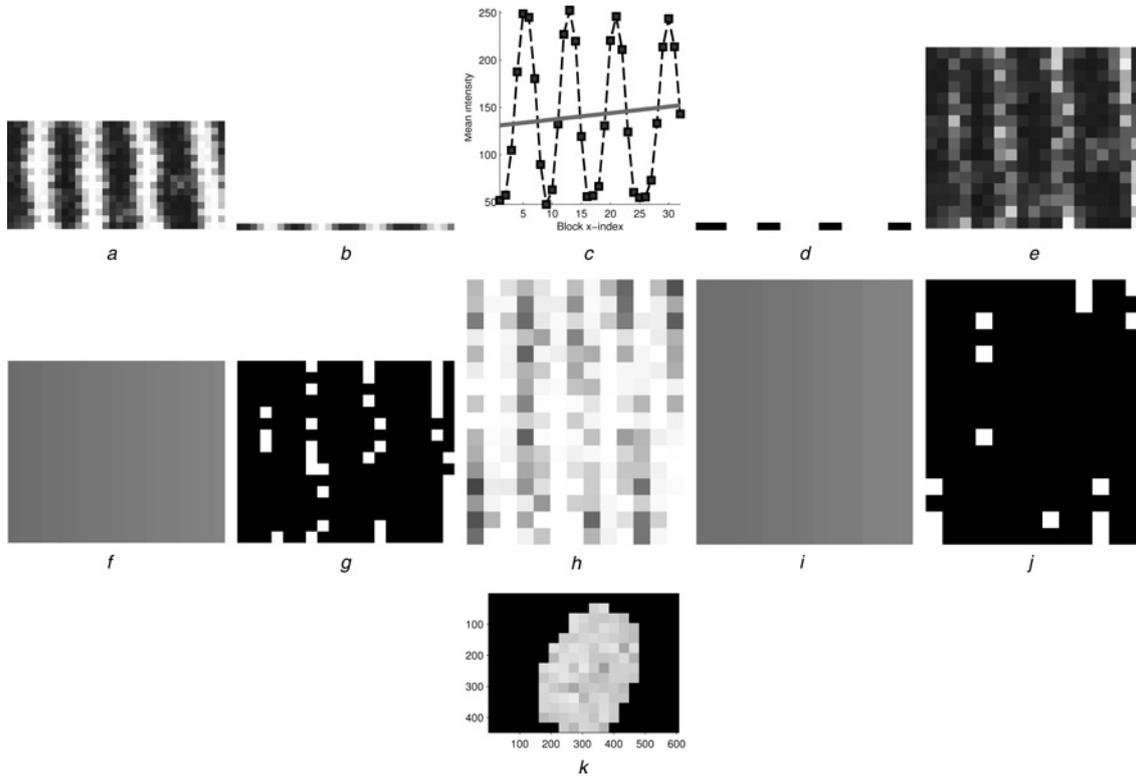*p* Bad quality blocks marked in white

```
Algorithm 4
    Input: Fingerprint image I
    Output: LCS quality score Q_LCS
1 for each block V in I do
2 |    rotate V such that dominant ridge flow is perpendicular to x-axis
3 |    crop V such that no invalid regions are included
4 |    with V obtain the ridge-valley signature T(x) (eq. (18))
5 |    Determine a threshold DT using linear regression on T(x)
6 |    Determine proportion of misclassified pixel β and α in the ridge and valley regions (eqs. (25) and (26))
7 |    Determine normalized ridge width and valley width W̄_r and W̄_v (eqs. (27) and (28))
8 |    Compute the local quality score Q_LCS^local (eq. (31))
9 end
10 Compute quality score Q_LCS (eq. (32))
```

Fig. 10 *LCS algorithm*

Fig. 11 *Processing steps of LCS algorithm*

a Crop of current block
b Average profile of block
c Average block profile with linear regression line
d Binarisation mask with ridge and valley regions based on regression line
e Pixels determined to be ridge based on mask
f Threshold across the region
g Pixels misclassified as valley based on the threshold
h–j same as *e–g* but for valley region
k Local clarity scores

$$\bar{W}_r = \frac{W_r}{(S/125)W^{\max}} \qquad (28)$$

where S is the sensor resolution in dpi, $W^{\max}$ is the estimated ridge or valley width for an image with 125 dpi resolution, and W_v and W_r are

the observed valley and ridge widths. According to [31], $W^{\max} = 5$ is reasonable for 125 dpi resolution. For 500 dpi resolution, (27) and (28) may be expressed as

$$\hat{W}_v = \frac{W_v}{20} \qquad (29)$$

$$\hat{W}_r = \frac{W_r}{20} \qquad (30)$$

Computing the LCS. The final quality score Q_{LCS} is computed using the average values of α and β in valid ridge and valley regions (see (31))

```
Algorithm 5
    Input: Fingerprint image I
    Output: OFL quality score Q_OFL
1 for each block V in I do
2 |    Determine the dominant ridgeflow orientation (eq. (17))
3 |    Compute the absolute orientation difference D(i, j) (eq. (32))
4 |    Compute the local orientation quality score Q_OFL^local(i, j) (eq. (33))
5 end
6 Compute the quality score Q_OFL (eq. (34))
```
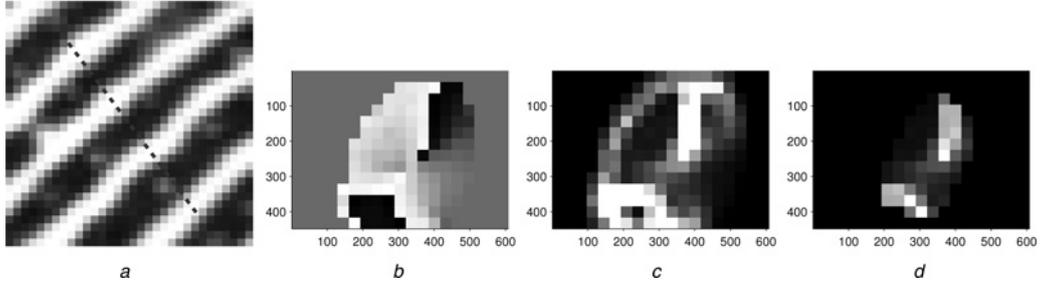
Fig. 12 *OFL algorithm*

where W_r^{nmin} and W_v^{nmin} are the minimum values for the normalised ridge and valley widths, and W_v^{nmax} and W_v^{nmax} are the maximum values for the normalised ridge and valley widths.

Fig. 13 *Processing steps of OFL algorithm*

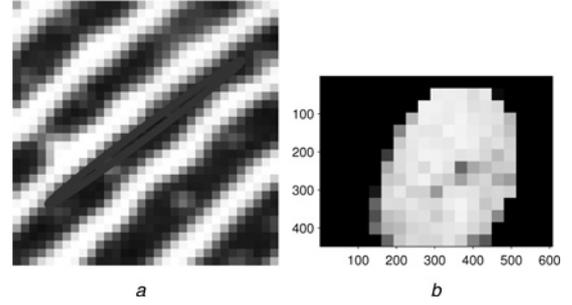*a* Line marking the normal to the ridgeline orientation
b Local orientations
c Orientation differences
d Local quality scores

Fig. 14 *OCL algorithm*

Computing the quality score. With $N*M$ local orientation quality score blocks the global local clarity score is computed as

$$Q_{\text{LCS}} = \frac{1}{N*M} \sum_{i=1}^{N} \sum_{j=1}^{M} \boldsymbol{Q}_{\text{LCS}}^{\text{local}}(i,j) \tag{32}$$

.

3.4.7 Orientation flow (OFL): OFL [20, 36] is a measure of ridge flow continuity which is based on the absolute orientation difference between a block and its neighbouring blocks (see Fig. 12).

In ISO/IEC TR 29794-4:2010 [20], the parameter $\theta_{\min}$ is a constant such that the angular tolerance is $8°$ ($\theta_{\min} = 8°$). The local orientation quality score in this case is assigned such that $\boldsymbol{Q}_{\text{OFL}}^{\text{local}}(i,j) = 100$ when the local quality is the highest. This is the opposite behaviour of the ISO/IEC TR 29794-4:2010 definition where $\boldsymbol{Q}_{\text{OFL}}^{\text{local}}(i,j) = 0$ when the local quality is the highest.

A visual overview of the algorithm outputs is depicted in Fig. 13 where a line marking the normal to the ridge line orientation in Fig. 13*a*; the blockwise dominant orientations are depicted in Fig. 13*b* and the orientation differences are shown in Fig. 13*c*; and the local quality scores are shown in Fig. 13*d*.

Absolute orientation difference. The ridge flow is determined as a measure of the absolute difference between a block and its neighbouring blocks. The absolute difference for block $V(i,j)$ is

$$D(i,j) = \frac{\sum_{m=-1}^{1} \sum_{n=-1}^{1} |V(i,j) - V(i-m, j-n)|}{8} \tag{33}$$

Local orientation quality score. The local orientation quality score $\boldsymbol{Q}_{\text{OFL}}^{\text{local}}(i,j)$ for the block orientation difference $D(i,j)$ is determined as

$$\boldsymbol{Q}_{\text{OFL}}^{\text{local}}(i,j) = \begin{cases} \dfrac{D(i,j) - \theta_{min}}{90\deg - \theta_{min}}, & D(i,j) > \theta_{min} \\ 0 & , \quad \text{otherwise} \end{cases} \tag{34}$$

Fig. 15 *Processing steps of OCL algorithm*

a Current block with ratio between eigen values marked as ellipse
b Local quality scores

where $\theta_{\min}$ is a threshold for the minimum angle difference to be considered as significant.

Global orientation quality score. With $N*M$ local orientation quality score blocks the global orientation quality score is computed as

$$Q_{\text{OFL}} = 1 - \frac{1}{N*M} \sum_{i=1}^{N} \sum_{j=1}^{M} \boldsymbol{Q}_{\text{OFL}}^{\text{local}}(i,j) \tag{35}$$

3.4.8 Orientation certainty level (OCL): OFL quality (OCL) (see Fig. 14) [20, 37] is a measure of the strength of the energy concentration along the dominant ridge flow orientation. The feature operates in a block-wise manner.

The computation of OCL presented here deviates from ISO/IEC 29794-4:2010 [20] in that we subtract the ratio between the eigenvalues from 1 such that $Q_{\text{OCL}} = 0$ reflects the lowest quality and $Q_{\text{OCL}} = 1$ the highest quality.

A visual overview of the algorithm outputs is depicted in Fig. 15 where the ratio of the eigenvalues is shown in Fig. 15*a* as an ellipse and the local qualities are depicted in Fig. 15*b* where higher intensity indicates higher $\boldsymbol{Q}_{\text{OCL}}^{\text{local}}$ score.

Computing the eigenvalues and local orientation certainty. From the covariance matrix $\boldsymbol{C}$ the eigenvalues $\lambda_{\min}$ and $\lambda_{\max}$ are computed as

$$\lambda_{\min} = \frac{a + b - \sqrt{(a-b)^2 + 4c^2}}{2} \tag{36}$$

$$\lambda_{\max} = \frac{a + b + \sqrt{(a-b)^2 + 4c^2}}{2} \tag{37}$$

$$\boldsymbol{Q}_{\text{LCS}}^{\text{local}} = \begin{cases} \left(1 - \dfrac{\alpha + \beta}{2}\right), & \left(W_{\text{v}}^{\text{nmin}} < \bar{W}_{\text{v}} < W_{\text{v}}^{\text{nmax}}\right)\left(W_{\text{r}}^{\text{nmin}} < \bar{W}_{\text{r}} < W_{\text{r}}^{\text{nmax}}\right) \\ 0 & , \quad \text{otherwise} \end{cases} \tag{31}$$

Algorithm 7

Input: Fingerprint image $\mathbf{I}$
Output: Ridge Valley Uniformity (RVU) quality score Q_{RVU}
1 **for** *each block* $\mathbf{V}$ *in* $\mathbf{I}$ **do**
2   Determine dominant ridgeflow orientation $\theta(\mathbf{V})$ according to eq. (17)
3   Rotate $\mathbf{V}$ such that $\theta(\mathbf{V})$ is perpendicular to x-axis
4   crop $\mathbf{V}$ such that no invalid regions are included
5   with $\mathbf{V}$ obtain the ridge-valley signature $T(x)$ (eq. (18))
6   Determine a threshold DT using linear regression on $T(x)$
7   Binarize $T(x)$ using DT
8   Determine transition points in $T(x)$
9   Drop first and last transition from $T(x)$ to remove incomplete ridges or valleys
10   Determine $\mathbf{Q}_{\text{RVU}}^{local}$ from the ratio between remaining ridge and valley widths
11 **end**
12 Compute the quality score Q_{RVU} as the standard deviation of all $\mathbf{Q}_{\text{RVU}}^{local}$

Fig. 16 *RVU algorithm*

this yields a local orientation certainty level $\boldsymbol{Q}_{\text{OCL}}^{\text{local}}$

$$\boldsymbol{Q}_{\text{OCL}}^{\text{local}} = 1 - \frac{\lambda_{\min}}{\lambda_{\max}} \tag{38}$$

which is a ratio in the range 0 to 1, respectively, the lowest and highest orientation certainty levels.

Computing the quality score. The quality score Q_{OCL} is computed as the mean of local scores $\boldsymbol{Q}_{\text{OCL}}^{\text{local}}$ using (38) where a higher Q_{OCL} score indicates higher quality

$$Q_{\text{OCL}} = \frac{1}{N*M} \sum_{i=1}^{N} \sum_{j=1}^{M} \boldsymbol{Q}_{\text{OCL}}^{\text{local}}(i, j) \tag{39}$$

3.4.9 Ridge–valley uniformity (RVU): RVU is a measure of the consistency of the ridge and valley widths (see Fig. 16) [20, 37]. The
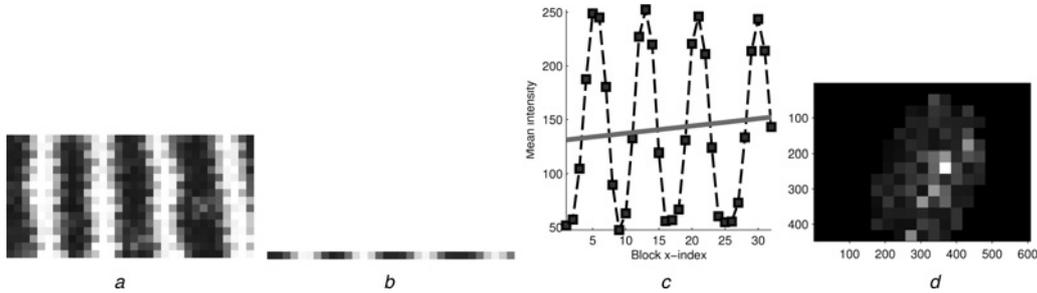

Fig. 17 *Processing steps of RVU algorithm*
a Crop of current block
b Average profile of block
c Average block profile with linear regression line
d Local quality scores as the standard deviation of local ridge to valley ratios
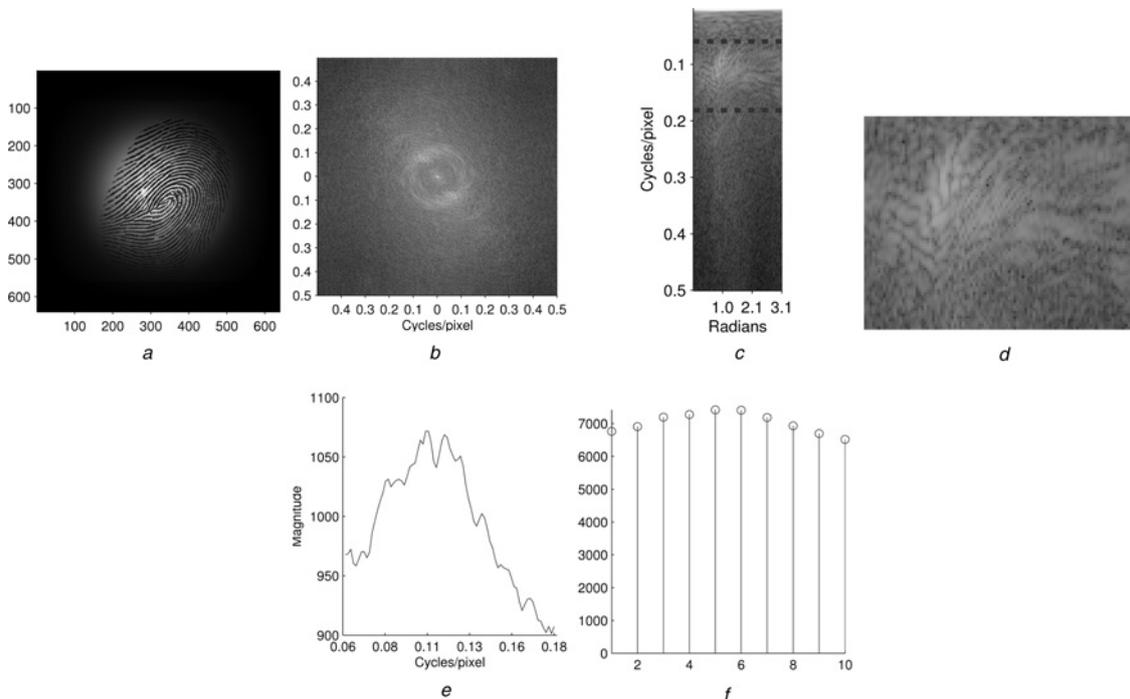

Fig. 18 *Processing steps of RPS algorithm*
a Blackman filtered input image
b Log transformed Fourier spectrum of filtered input
c Transform to polar coordinates with dashed lines indicating frequency band of interest
d Crop of frequency band of interest
e Magnitude in frequency band of interest
f Power in sub bands in the frequency of interest

Fig. 19 *RPS algorithm*

expectation for a finger image with clear ridge and valley separation is that the ratio between ridge and valley widths remains fairly constant and thus the standard deviation of ratios is used as an indication of the sample quality. The RVU quality feature is resolution dependent.

A visual overview of the algorithm outputs is depicted in Fig. 17 where a crop of the valid area of the current block is depicted in Fig. 17a; the ridge–valley signature is shown in Fig. 17b and the result of linear regression on the intensities is depicted in Fig. 17c. The resulting standard deviation of local ratios is shown in Fig. 17d.

3.5 Global finger image quality

3.5.1 Radial power spectrum (RPS): RPS [20, 38] 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.

A visual overview of the algorithm outputs is depicted in Fig. 18 where the input image convolved with a Blackman filter [Algorithm 8 (Fig. 19), line 1] is shown in Fig. 18a. The log transformed spectrum of the DFT is shown in Fig. 18b and the conversion to polar coordinates in Fig. 18c with lines in upper part indicating the frequency band of interest. The extracted band of frequencies is shown in Fig. 18d, and the power is shown in Fig. 18e and the power of binned frequencies is shown in Fig. 18f.

2D Fourier transform. The 2D DFT $F(u, v) = \mathcal{F}\{\mathbf{I}(x, y)\}$ is

$$F(u, v) = \frac{1}{I_h I_w} \sum_{x=0}^{I_w-1} \sum_{y=0}^{I_h-1} \mathbf{I}(x, y) \exp\left(-j2\pi\left(\frac{ux}{I_w} + \frac{vy}{I_h}\right)\right) \quad (40)$$

The magnitude of $F(u, v)$ is computed as

$$|F(u, v)| = \sqrt{\Re(F(u, v))^2 + \Im(F(u, v))^2} \quad (41)$$

Magnitude of frequency bands polar coordinates. The magnitude of the annular band between r and $r + \delta_r$ in the polar Fourier spectrum $F(\alpha, r)$ is computed as

$$J(r) = \sum_{\alpha} F(\alpha, r) \quad (42)$$

Determine quality score. The quality feature Q_{RPS} is found as

$$Q_{\text{RPS}} = \max |J(r)| \quad (43)$$

Fig. 20 *MU algorithm*

3.5.2 Image mean (MU): The MU quality feature is the mean value of the input image (see Fig. 20).

3.5.3 Image standard deviation (SIG): The SIG quality feature is the standard deviation of the input image (see Fig. 21).

4 Performance evaluation

For the performance evaluation, we have chosen to assess the predictive performance of NFIQ and the quality features specified in Sections 3.4 and 3.5.

We base our results on finger image data from Fingerprint Verification Competition 2004 (FVC 2004) [32], Fingerprint Verification Competition 2006 (FVC 2006) [39], MCYT Fingerprint 330 (MCYT-330) [40], and Wet-Dry Dataset 2

Fig. 21 *SIG algorithm*

Table 1 Datasets used for performance evaluation of quality algorithms

Database	FVC2004 DB1 [32]	FVC2006 DB2A [39]	MCYT330 DP [40]	WDSET02 [41]
Sensor	optical (single)	optical (single)	optical (single)	optical (multiple)
Subjects	30		330	33
Fingers	4	140	10	10
Impressions	8	12	12	20
Total images	880	1680	39,600	6600
Genuine	6160	18,480	435,600	125,400
Imposter	17,600	840,000	19,800,000	660,000

Fig. 22 *DET curves for datasets FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, and WDSET02 using comparison scores from providers A, B, and C*

(WDSET02) [41] databases which are summarised in Section 4. For the dataset WDSET02 5 different capture devices were used and the capture protocol specified four finger surface conditions: no treatment with capture as the subject arrives; application of an alcohol based solution to dry the skin surface; application of hand-moisturiser; and finally soaking in water. FVC2004 DB1 and FVC2006 DB2A were captured in a supervised manner were subjects were instructed to perform various actions which result in various degradations such as elastic deformation and uneven pressure. Samples in MCYT330 DP were captured using the three levels finger placement requirements on the sensor: relaxed, but under supervision by operator; placement within specified

rectangle; and high level of control where one core and/or delta must fall within a specified rectangle. The datasets combined provide a total of 48,760 fingerprint image samples captured using optical sensors.

For each dataset, comparison scores are computed using three commercial state-of-the-art minutiae-based comparison systems denoted as A, B, and C as agreements prohibit us from revealing the names. All genuine comparisons were made and imposter scores were determined by randomly selecting a set of probes from other subjects (see Section 4) and compared with the reference template. In the cases where a provider failed to either compute a template or a comparison score we have set the comparison score to 0.

Fig. 23 *Spearman correlation table and scatter plots of quality features and utilities for each dataset*
a FVC2004 DB1
b FVC2006 DB2A
c MCYT330 DP
d WDSET02
The upper half shows the Spearman correlation between two variables, the lower half depicts a scatter plot of the two variables and the diagonal contains the name of the variable in that row and column

We evaluate and rank the quality features according to their ERC characteristics and Spearman correlation with utility values as computed in (1). In the case of ERC, we focus on the decrease in FNMR achieved when rejecting the lowest 20% quality samples as rejecting more than 1 in 5 samples due to quality is unlikely to be acceptable in an operational system (see Section 2.5.3) (see Table 1).

For features Q_{FDA}, Q_{LCS}, Q_{OCL}, Q_{OFL}, Q_{RVU} the quality values were computed on the image foreground as determined by segmentation based on local standard deviation image with a block size of 32 by 32 pixels and a threshold of 0.1 with ridgesegment.m method [42].

4.1 Detection error tradeoff (DET)

To assess the characteristics of the datasets and the biometric feature extractors and comparators a benchmark is performed by computing the DET on each dataset. The DET plots are shown in Fig. 22. It is observed that at a fixed false match rate at 10^{-4} the FNMR shifts over several orders of magnitude depending on the dataset used. It is observed that A is performing best across datasets, while C is the worst performing. Further, providers perform the worst on WDSET02 and the best on FVC2006 DB2A.

Providers perform the worst on WDSET02, one reason for this is that the data collection protocol for that dataset included the capture of soaking wet fingerprints which results in very poor quality samples for some capture devices.

4.2 Feature correlations

Correlation of features has a significance when assessing which features to include in a composite quality algorithm, i.e. one that uses input from multiple quality algorithms to determine a scalar quality value or a vector of multiple selected quality values for a sample. In such case it is advisable to select features which complement each other, i.e. they do not correlate with each other.

Correlation tables for quality values and utility computed on the four datasets are shown in Fig. 23. The upper half shows the Spearman correlation between two variables, the lower half depicts a scatter plot of the two variables and the diagonal contains the name of the variable in that row and column.
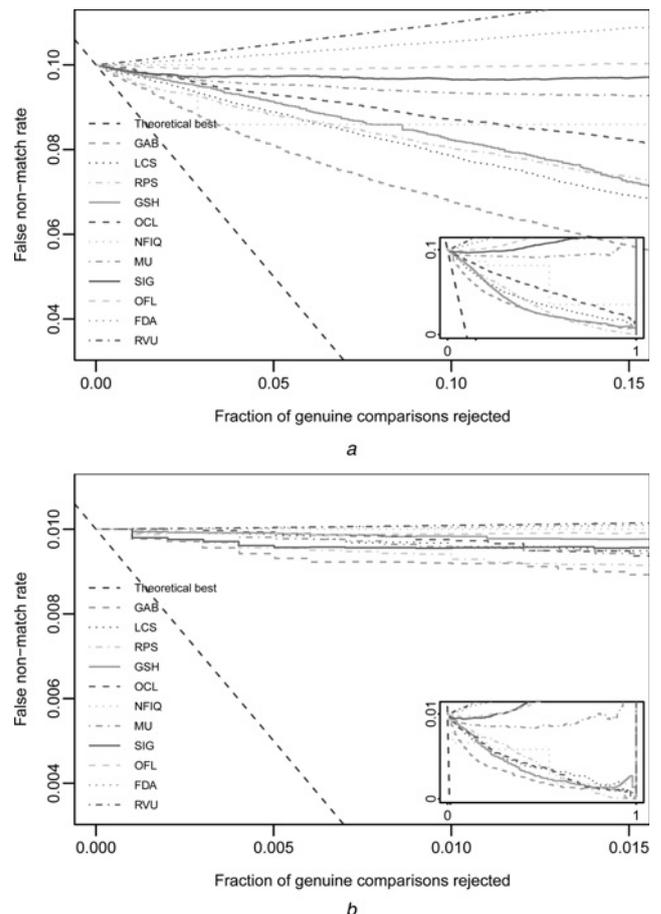
We note that the correlation between features is clearly dependent on the dataset on which they were computed, e.g. consider the correlation between FDA and MU, which is varying significantly by being positively correlated (0.25) in FVC2004 DB1 and WDSET02 (0.32), not correlated (0.00) in FVC2006 DB2A, and negatively correlated (−0.23) in MCYT330 DP. The correlations between the features and utility computed according to (1) for each provider also reveals significant differences across datasets, e. g. the correlations between NFIQ and each of A, B, and C are −0.18, −0.15, and −0.17, respectively, for FVC2004 DB1, while for WDSET02 the correlations with the same providers are −0.65, −0.56, and −0.52. Similarly, for GAB we see major differences, and in the case of B on MCYT330 DP there is a correlation of −0.03, while it is 0.49 on WDSET02.

4.3 Error-reject curves

We determine the behaviour of quality features over the range of quality values in relation to the FNMR using ERC (Section 2.5.3). Ideally, when a sample is removed due to its computed quality value, a corresponding decrease in FNMR occurs, i.e. the idealised quality feature identifies exactly those samples which contribute to FNMR and its behaviour in the ERC is that the resulting reduction in FNMR occurs at the same rate as the rejection due to quality. When a sample which did not cause a false non-match is removed due to its quality level, it will result in an increase in the FNMR which occurs because the FNMR is computed using the genuine comparison scores of those samples remaining after samples have been removed due to their inferior quality as determined be the respective algorithm [see (5)].

For each feature and combination of dataset and comparison score provider we computed the ERC, noting the η_{auc}^{erc} and η_{pauc20}^{erc} as computed according to (6) and (7). The resulting curves for the MCYT330 DP dataset using provider A for comparison scores are displayed in Fig. 24, where we have used two different values of f, one $f = 0.1$ to simulate an operational case where the FNMR is 10%, and one with $f = 0.01$ where the FNMR is 1% [see (4)]. We show, respectively, the first 15 and 1.5% of samples being rejected due to computed quality value. The inlay shows the full ERC from no samples removed to all samples removed. The resulting η_{auc}^{erc} and η_{pauc20}^{erc} computed from ERC for each case of f are summarised in Tables 2 and 3.

All the presented features provide a wide range of possible quality values, while NFIQ by design provides five quality levels. The ERC shows that NFIQ is more robust than individual features as the FNMR in the ERC is decreasing as comparisons containing samples associated with the rejected quality level are removed. Only in a single instance, the combination of WDSET02 and provider B, do we note that NFIQ exhibits unexpected behaviour where the FNMR is increasing. The cost of the relatively small number of distinct quality levels is that one must always choose to reject an entire level of quality, resulting in a rough step function when then the fraction of samples rejected is plotted against the resulting FNMR. This behaviour is visible in Fig. 24a where the FNMR remains at 0.10 until around 3% of the samples are rejected where after the FNMR drops to slightly more than 8% and stays at that level until around 20% of the samples are rejected and the FNMR is further reduced to around 4%.

Fig. 24 *ERC for features on MCYT330 DP using A as basis for FNMR computation with ERC starting, respectively, at*
a 10% FNMR (f = 0.1)
b 1% (f = 0.01). See Tables 2 and 3 for a summary of each instance of f

Table 2 Summary of AUC and PAUC (see (6) and (7)) for ERC plots of quality features on FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, and WDSET02 computed using $f = 0.1$. The values marked in bold indicate the smallest AUC and PAUC for each data set

Feature	FVC2004 DB1		FVC2006 DB2A		MCYT330 DP		WDSET02	
	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}
GAB	0.068	0.014	**0.023**	**0.010**	**0.026**	**0.009**	**0.017**	**0.008**
LCS	0.072	0.014	0.046	**0.010**	0.036	0.011	0.023	0.010
RPS	0.062	0.014	0.051	0.013	0.031	0.011	0.089	0.014
GSH	0.097	0.014	0.028	0.011	0.030	0.011	0.036	0.014
OCL	0.081	0.015	0.039	0.011	0.049	0.013	0.063	0.012
NFIQ	0.068	**0.013**	0.062	0.014	0.045	0.013	0.040	0.015
MU	0.076	0.015	0.062	0.013	0.092	0.014	0.066	0.015
SIG	0.115	0.016	0.077	0.014	0.105	0.014	0.120	0.014
OFL	**0.066**	0.013	0.083	0.012	0.140	0.015	0.048	0.011
FDA	0.121	0.015	0.147	0.015	0.174	0.016	0.181	0.015
RVU	0.138	0.017	0.206	0.017	0.207	0.017	0.177	0.015

Table 3 Summary of AUC and PAUC (see (6) and (7)) for ERC plots of quality features on FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, and WDSET02 computed using $f = 0.01$. The values marked in bold indicate the smallest AUC and PAUC for each data set

Feature	FVC2004 DB1		FVC2006 DB2A		MCYT330 DP		WDSET02	
	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}	η_{auc}^{erc}	η_{pauc20}^{erc}
GAB	0.008	0.002	**0.002**	**0.001**	**0.003**	**0.001**	**0.001**	**0.001**
LCS	0.010	0.002	0.004	**0.001**	0.004	**0.001**	0.002	**0.001**
RPS	**0.005**	0.002	0.004	0.002	0.004	0.002	0.017	0.003
GSH	0.011	0.002	**0.002**	0.002	0.004	0.002	0.004	0.003
OCL	0.007	0.002	0.004	**0.001**	0.004	0.002	0.013	0.002
NFIQ	0.007	0.002	0.005	0.002	**0.003**	**0.001**	0.004	0.002
MU	0.009	0.002	0.007	0.002	0.011	0.002	0.004	0.002
SIG	0.011	0.002	0.006	0.002	0.014	0.002	0.022	0.003
OFL	0.011	0.002	0.008	**0.001**	0.014	0.002	0.002	**0.001**
FDA	0.010	0.002	0.017	0.002	0.021	0.002	0.017	0.002
RVU	0.016	0.002	0.029	0.002	0.026	0.002	0.008	**0.001**

5 Conclusion

Fingerprint sample quality continues to play a large role in biometric systems and the presented review and performance evaluation of existing quality metrics has shown that the algorithms have varying performance across dataset and comparison score vendor. This considerably complicates the task of selecting a subset of features which are suitable for a combined or aggregated quality metric. Inter-feature Spearman correlation give indications as to which features are overlapping with respect to the aspect of the fingerprint they measure.

An important point to address in future research is how to best combine quality features in a way which minimises the dependence on individual comparison score providers, while maintaining a sufficient predictive performance with respect to the biometric performance.

The selection of quality features is further complicated by their non-monotonic behaviour, i.e. the feature might only be indicative of sample utility in a range of the values it produces, and it is not guaranteed that a high quality value indicates higher biometric performance across all ranges within the possible values. Using η_{auc}^{erc} and η_{pauc20}^{erc} as indicators for feature provides insights to the behaviour of quality features throughout the quality values it provides in the context of FNMR, however, it is clear from our results that the ordering of the features using these indicators changes depending on the chosen f parameter in the calculation of ERC. With $f = 0.01$ and using η_{auc}^{erc} as selection criteria, we conclude that GAB, LCS, and GSH provide consistent high performance across the tested datasets, and that RPS, OFL, and OCL appear sensitive to the dataset used.

Our results show that correlation with utility as computed using the definitions by ISO/IEC is not indicative of the behaviour of the quality feature when evaluated using ERC and our recommendation is to apply feature correlations only in the context of identifying potentially overlapping methods of quality measurements.

The quality features we have specified and evaluated are made available to the community [28]. We anticipate that parameters in

the quality features can be tuned to achieve a higher performance on specific datasets or a better generalisability when faced with new datasets.
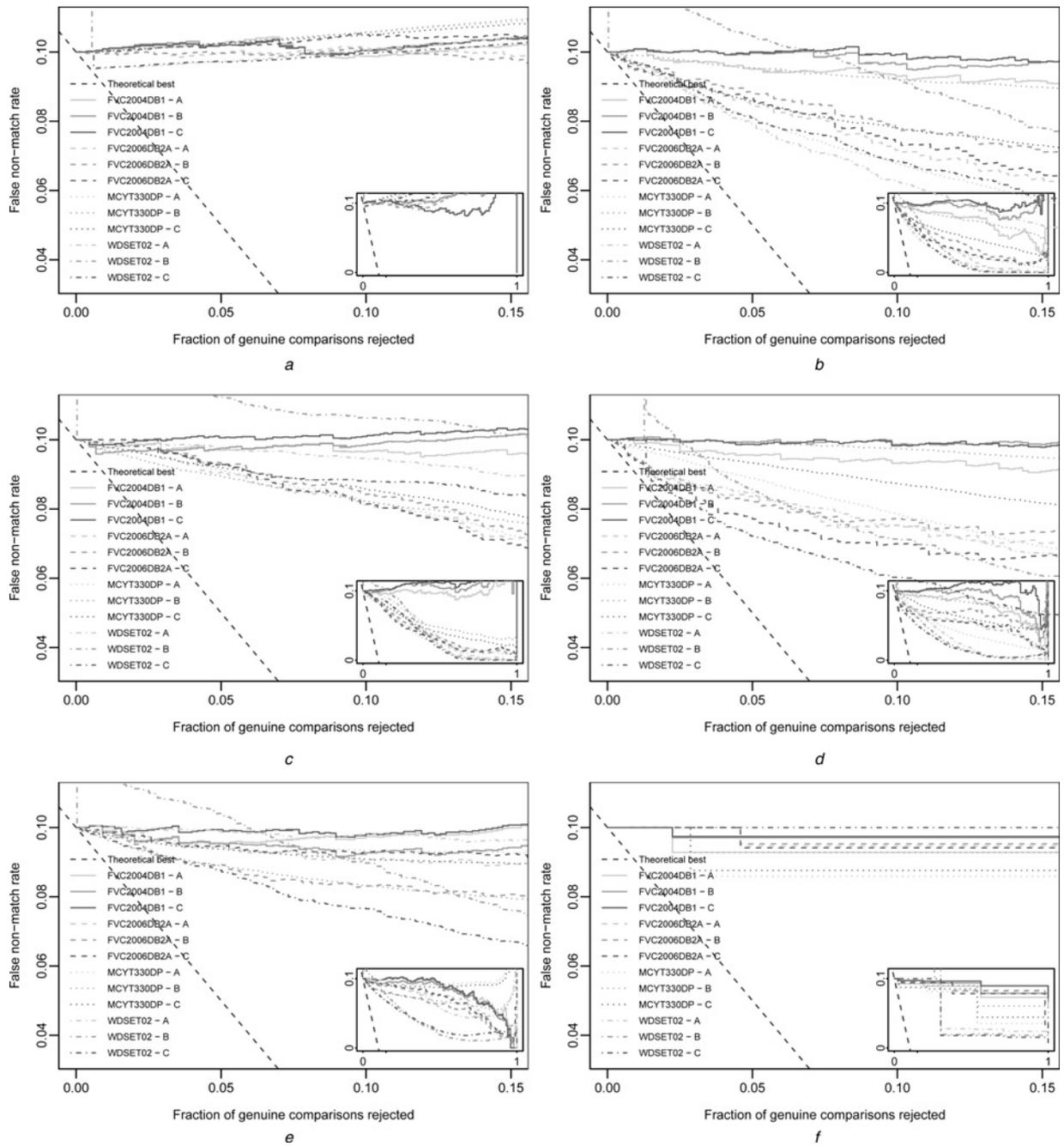
6 Acknowledgments

7 References

1 Wilson, C.L., Garris, M.D., Watson, C.I.: 'Matching performance for the US-VISIT IDENT system using flat fingerprints NISTIR 7110'. Tech. Rep., NIST, May 2004
2 Tabassi, E., Wilson, C.L., Watson, C.I.: 'NISTIR7151 – fingerprint image quality'. Tech. Rep., NIST, August 2004
3 Fierrez-Aguilar, J., Chen, Y., Ortega-garcia, J., *et al.*: 'Incorporating image quality in multi-algorithm fingerprint verification'. Proc. IAPR Int. Conf. on Biometrics, ICB, 2006 (*LNCS*, 3832), pp. 213–220
4 Council E: 'Establishing the visa information system (VIS) (2004/512/EC)' (Official Journal of the European Union, 2004), pp. 5–7
5 Commission E: 'Determining the first regions for the start of operations of the visa information system (VIS), C(2009) 8542)' (Official Journal of the European Union, 2010), vol. 23, pp. 62–64
6 Morpho: 'Secure biometric access – Morpho', 2014. Available at: http://www.morpho.com/identification/secure-biometric-access/ (visited on 2014)
7 Kneidinger, P., Schwaiger, M.: 'Evaluation and monitoring of fingerprint acquisitions for the European Visa Information System experiences from Austria'. Proc. of the Int. Conf. of the Biometrics Special Interest Group (BIOSIG, 2012), 2012, pp. 1–12
8 Unique Identification Authority of India: 'Role of biometric technology in aadhaar enrollment', 2012
9 Unique Identification Authority of India: 'UIDAI strategy overview creating a unique identity number for every resident in India', 2010

10 Unique Identification Authority of India: 'Exclusion to inclusion with micropayments', 2010. Available at: http://www.uidai.gov.in/UID_PDF/Front_Page_Articles/Strategy/Exclusion_to_Inclusion_with_Micropayments.pdf, accessed February 2013

11 ISO/IEC: '29794-1:2009. Information technology – Biometric sample quality – Part 1: Framework'. Tech. Rep., JTC 1/SC 37/WG 3, January 2009

12 ISO/IEC JTC 1/SC 37: 'IS 19794-4:2011. Information technology – Biometric data interchange formats – Part 4: Finger image data', 2011

13 Schneider, J., Richardson, C., Kiefer, F., *et al*.: 'On the correlation of image size to system accuracy in automatic fingerprint identification systems'. Audio- and Video-Based Biometric Person Authentication 2003 (*LNCS*, 2688), pp. 895–902. Available at: http://www.dx.doi.org/10.1007/3-540-44887-X_104

14 Arnold, M., Daum, H., Busch, C.: 'Comparative study on fingerprint recognition systems – project BioFinger'. Proc. BIOSIG, 2003, pp. 33–38

15 Ross, A., Jain, A.K.: 'Biometric sensor interoperability: a case study in fingerprints'. Proc. ECCV Workshop BioAW, 2004, pp. 134–145

16 Kukula, E., Elliott, S., Kim, H., *et al*.: 'The impact of fingerprint force on image quality and the detection of minutiae'. IEEE Int. Conf. on Proc. Electro/Information Technology, 2007, 2007, pp. 432–437

17 Blomeke, C., Modi, S., Elliott, S.: 'Investigating the relationship between fingerprint image quality and skin characteristics'. 42nd Annual IEEE Int. Carnahan Conf. on Proc. Security Technology, 2008, ICCST 2008, 2008, pp. 158–161

18 Sickler, N., Elliott, S.: 'An evaluation of fingerprint image quality across an elderly population vis-a-vis an 18–25 year old population'. 39th Annual 2005 Int. Carnahan Conf. on Proc. Security Technology, 2005, CCST'05, 2005, pp. 68–73

19 Galbally, J., Marcel, S., Fierrez, J.: 'Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition', *IEEE Trans. Image Process.*, 2014, **23**, (2), pp. 710–724

20 ISO/IEC: '29794-4:2010. Information technology – Biometric sample quality – Part 4: Finger image data'. Tech. Rep., JTC 1/SC 37, 2010

21 Grother, P., Tabassi, E.: 'Performance of biometric quality measures', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2007, **29**, (4), pp. 531–543

22 Bundesamt für Sicherheit in der Informationstechnik: 'Technical guideline TR-03121-1, biometrics for public sector applications, part 1: framework, version 2.3'. Proc. 2011

23 3M Cogent: '3M Cogent', 2014. Available at: http://www.solutions.3m.com/wps/portal/3M/en_US/Security/Identity_Management/3M_Cogent/ (visited on 2014)

24 Vatsa, M., Singh, R., Bharadwaj, S., *et al*.: 'Analyzing fingerprints of Indian population using image quality: a UIDAI case study'. Int. Workshop on Proc. Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), 2010, August 2010, pp. 1–5

25 Alonso-Fernandez, F., Fierrez-Aguilar, J., Ortega-Garcia, J.: 'A review of schemes for fingerprint image quality computation'. Proc. of the 3rd COST 275 Workshop, COST-275, Hertfordshire, UK, 2005

26 Alonso-Fernandez, F., Fierrez, J., Ortega-Garcia, J., *et al*.: 'A comparative study of fingerprint image-quality estimation methods', Information Forensics and Security, IEEE Transactions on, 2007, **2**, (4), pp. 734–743

27 Bharadwaj, S., Vatsa, M., Singh, R.: 'Biometric quality: a review of fingerprint, iris, and face', *EURASIP J. Image and Video Process.*, 2014, **2014**, (1), Doi: 10.1186/1687-5281-2014-34

28 Olsen, M.A., Smida, V., Busch, C.: 'Fingerprint quality assessment algorithms', 2015. Available at: http://nislab.no/software/fingerprintquality/

29 Merkle, J., Schwaiger, M., Breitenstein, M.: 'Towards improving the NIST fingerprint image quality (NFIQ) algorithm'. Proc. BIOSIG, 2010, pp. 29–44

30 Commerce, N. US Department of: Development of NFIQ 2.0, Development of NFIQ 2.0, 2014. Available at http://www.nist.gov/itl/iad/ig/development_nfiq_2.cfm (visited on 2014)

31 Maltoni, D., Maio, D., Jain, A.K., *et al*.: 'Handbook of fingerprint recognition' (Springer Publishing Company, Incorporated, 2009, 2nd)

32 Maio, D., Maltoni, D., Cappelli, R., *et al*.: 'FVC2004: third fingerprint verification competition'. Proc. ICBA, 2004, pp. 1–7

33 Lim, E., Toh, K.-A., Suganthan, P., *et al*.: 'Fingerprint image quality analysis'. Int. Conf. on Proc. Image Processing, 2004, ICIP '04, 2004, vol. 2, pp. 1241–1244

34 Olsen, M.A., Xu, H., Busch, C.: 'Gabor filters as candidate quality measure for NFIQ 2.0'. Proc of the 5th IAPR Int. Conf. on Biometrics (ICB 2012), New Delhi, India, 29 March–1 April 2012

35 Shen, L., Kot, A., Koo, W.: 'Quality measures of fingerprint images'. Proc. AVBPA, 2001 (*LNCS* 2091), pp. 266–271

36 Chen, T., Jiang, X., Yau, W.: 'Fingerprint image quality analysis'. Int. Conf. on Proc. Image Processing, 2004, ICIP '04, 2004, vol. 2, pp. 1253–1256

37 Lim, E., Jiang, X., Yau, W.: 'Fingerprint quality and validity analysis'. Proc. Int. Conf. Image Processing 2002, 2002, vol. 1

38 Chen, Y., Dass, S., Jain, A.: 'Fingerprint quality indices for predicting authentication performance'. Proc. AVBPA, 2005 (*LNCS* **3546**), pp. 160–170

39 Cappelli, R., Ferrara, M., Franco, A., *et al*.: 'Fingerprint verification competition 2006', *Biometric Technol. Today*, 2007, **15**, (7–8), pp. 7–9. Available at: http://www.sciencedirect.com/science/article/pii/S0969476507701406

40 Ortega-Garcia, J., Fierrez-Aguilar, J., Simon, D., *et al*.: 'MCYT baseline corpus: a bimodal biometric database', *IEEE Proc., Vis. Image Signal Process.*, 2003, **150**, (6), pp. 395–401

41 Olsen, M.A., Dusio, M., Busch, C.: 'Fingerprint skin moisture impact on biometric performance'. Proc. Int. Workshop on Biometrics and Forensics 2015, 2015

42 Kovesi, P.D.: 'Functions for Computer Vision and Image Processing'. Kovesi-MATLAB and Octave-2005, 2005. Available from: http://www.peterkovesi.com/matlabfns/

8 Appendix: ERCs for quality features

See Fig. 25.

Fig. 25 *ERCs for each feature on each dataset FVC2004 DB1, FVC2006 DB2A, MCYT330 DP, WDSET02 with each providers A, B, C, and f = 0.1*

a FDA
b GAB
c GSH
d LCS
e MU
f NFIQ
g OCL
h OFL
i RPS
j RVU
k SIG

Fig. 25 *Continued*