

Fingerprint Image Quality Estimation and its Application to Multi-Algorithm Verification

Hartwig Fronthaler, Klaus Kollreider, Josef Bigun, *Fellow, IEEE*,

Julian Fierrez-Aguilar, *Student Member, IEEE*,

Fernando Alonso-Fernandez, *Student Member, IEEE*,

Javier Ortega-Garcia, *Member, IEEE*

and Joaquin Gonzalez-Rodriguez, *Member, IEEE*

H. Fronthaler, K. Kollreider and J. Bigun are with Halmstad University, Box 823, SE-30118, Halmstad, Sweden. E-mail: hartwig.fronthaler, klaus.kollreider, josef.bigun@ide.hh.se.

J. Fierrez-Aguilar, F. Alonso-Fernandez, J. Ortega-Garcia and J.Gonzalez-Rodriguez are with ATVS, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Avda. Francisco Tomas y Valiente 11, Campus de Cantoblanco, 28049 Madrid, Spain. E-mail: fernando.alonso, julian.fierrez, javier.ortega, joaquin.gonzalez@uam.es

Abstract

Recently, image quality awareness has been found to increase recognition rates and to support decisions in multimodal authentication systems significantly. Nevertheless, automatic quality assessment is still an open issue, especially with regard to biometric authentication tasks. Here we analyze the orientation tensor of fingerprint images with a set of symmetry descriptors, in order to detect fingerprint image quality impairments like noise, lack of structure, blur, etc. Allowed classes of local shapes are a priori application information for the proposed quality measures, therefore no training or explicit image reference information is required. Our quality assessment method is compared to an existing automatic method and a human opinion in numerous experiments involving several public databases. Once the quality of an image is determined, it can be exploited in several ways, one of which is to adapt fusion parameters in a monomodal multi-algorithm environment, here a number of fingerprint recognition systems. In this work, several trained and non-trained fusion schemes applied to the scores of these matchers are compared. A Bayes-based strategy for combining experts with weights on their past performances, able to readapt to each identity claim based on the input quality is developed and evaluated. To show some of the advantages of quality-driven multi-algorithm fusion, such as boosting recognition rates, increasing computational efficiency, etc., a novel cascade fusion and simple fusion rules are employed in comparison as well.

Index Terms

Structure tensor, orientation fields, biometrics, fingerprints, quality assessment, filtering, symmetry descriptors, simple fusion schemes, cascaded fusion, adaptive fusion, Bayesian statistics.

I. INTRODUCTION

Automatic assessment of image quality by a machine expert is challenging, but useful for a number of tasks: monitor and adjust image quality, optimize algorithms and parameter settings or benchmark image processing systems [1]. Image quality assessment methods can be divided into full/reduced/no-reference approaches, depending on how much prior information is available on how a perfect candidate image should look like. In this work we study quality assessment of the second kind, where images come from a specific application. There exist general quality metrics originally suggested in image compression studies [2], e.g. mean square error (MSE) or peak signal to noise ratio (PSNR). These earlier approaches are excluded here because of their notorious poor performance in recognition applications, which do not have the same objectives as compression applications.

In biometrics, a “universal” quality metric appears to be impossible as one application may use relevant information of an image not useful to another application. For example, a face image contains information not useful to a fingerprint matcher. Ideally a quality model involves features that should be reusable for different applications. In this work symmetry features are used to automatically assess the quality of fingerprint images. We are forced to use generic models when trying to estimate the quality of biometric images, since a high-quality reference image of the same individual is usually not available, i.e. the link to the individual cannot be established in advance.

Once available, the benefits of having an automatic image quality estimate include the following. First, when acquiring biometrics, all samples presented by a person (either for an enrolment or authentication purpose) can be checked automatically to assure quality for stored templates, [3]. Second, in an authentication configuration involving several modalities, e.g. face and fingerprint, the quality of the presented images can be used to adjust the weight given to the respective expert at fusion stage, where a final decision is made. The recognition improvement benefits of quality-aware fusion have been shown previously, although mainly involving quality assessment done by humans [4]–[6]. Third, as the quality of an image might vary across different regions, when measuring the similarity among biometric samples, high quality regions can be favored [7], [8]. Although we address here a specific image domain, even in other computer vision applications which involve visual recognition, e.g. object recognition, image database retrieval, tracking, geometry reconstruction etc. are in need of automatic image quality assessment, which is still an open research issue. As a result of recent fingerprint verification competitions involving particularly low quality impressions, even state-of-the-art systems’ performance decreases remarkably [9]. Recent advances in fingerprint quality assessment include [3], [8], [10]. A taxonomy of fingerprint quality assessment methods is given in [11]. A novelty of the present paper consists in the continuous modeling of all details in a fingerprint allowing them to be used for dual purposes, recognition and quality estimation.

Specializing in a single biometric modality and taking into account several systems’ responses on the present claim, is commonly referred to as the monomodal multi-expert or multi-algorithm type of fusion (in contrast to multimodal fusion which involves different modalities). In this study we are combining fingerprint recognition systems at score level, and we prefer the term “multi-algorithm” for it. Furthermore, we will use “system” or “expert” to address a fingerprint matcher,

whereas we refer to a quality assessment method as “method” or “approach”. Besides, this will become clear from the context. Considering fusion within a modality, in particular fingerprint recognition, [12], [13] showed that combining systems with heterogeneous matching strategies is most desirable, leading to recognition rates even higher than the combination of the best systems relying on common features. This is further developed in [5], where the image quality of the fingerprints is used to adjust two experts’ weights in the final decision using a weighted sum rule. Furthermore, the additional information through quality estimation is exploited as a substitution for training the experts’ weights, due to the experts’ different skills with regard to image quality. When trying to fuse several experts with unknown skills and matching strategies, some sort of training is mandatory to systematically improve the combined performance [14], [15]. The performance can be increased even more, if the trained fusion scheme is adaptive as well, meaning that it takes into account current signal conditions trial by trial. This was also confirmed in [4], although unlike here, for a multimodal configuration and employing quality estimates by humans. For this reason, a trained adaptive fusion scheme, in which Bayes theory is used to calibrate the different experts based on their misclassification history and on the current image quality, has been used in this study. Interestingly enough, some recent works have nevertheless reported comparable performance between fixed and trained combining strategies [16], [17] and a debate has come out investigating the benefits of both strategies [18], [19]. As an example, and within this debate, some researches have shown how to learn user-specific parameters in a trained fusion scheme [20], [21]. As a result, they have reported that the overall verification performance can be improved significantly.

We report quantitative and comparative experimental results of our quality assessment with respect to an existing fingerprint quality estimation method [3], as well as, manually assigned quality labels on the QMCYT database [22], [23], and also on two databases employed in FVC2004 [9]. Three fingerprint recognition systems [3], [7] and the modified version introduced in [24] are employed to *a)* observe how well manually and automatically assigned quality estimates are agreeing, and *b)* carry out the quality-based multi-algorithm fusion. These recognition systems are relying on minutiae, texture/minutiae and texture features respectively, using thus expert diversification in fingerprint recognition.

II. QUALITY ESTIMATION

In this work we employ a novel method to automatically assess the quality of fingerprint images. Although the algorithm is mainly used for fingerprint image quality assessment, its applicability to other biometric modalities was indicated by means of face images [25]. In the first part of this section a more general description of the quality assessment features is given. The ideas are then adapted to fingerprint quality estimation.

A. Quality Assessment Features

The orientation tensor holds edge and texture information, which is exploited in this work to assess the quality of an image. We wish to determine whether this information is structured and generic in some sense, i.e. to distinguish noisy content from relevant non-trivial structures, e.g. minutiae. These relevant structures are for example essential for many recognition algorithms, representing the individuality of a biometric signal. Our method decomposes the orientation tensor of an image into symmetry representations, where the included symmetries are related to the particular definition of quality and encode the a priori content-knowledge about the application (e.g. fingerprints, face images, ...). The resulting quality metric mirrors how well a test image comprises the expected symmetries.

The orientation tensor is given by the equation

$$z = (D_x f + iD_y f)^2, \quad (1)$$

where $D_x f$ and $D_y f$ denote the partial derivatives of the image w.r.t. x - and y -axes. The squared complex notation directly encodes the double angle representation [26]. For the computation of the derivatives, separable Gaussians with a small standard deviation σ_1 are used. Next, the orientation tensor is decomposed into symmetry features of order n , where the n th symmetry is given by $\exp(in\phi + \alpha)$ [26]–[29] representing the argument of (1). The corresponding patterns are shown in figure 1, e.g. straight lines for $n = 0$, parabolic curves and line endings for $n = \pm 1$. Higher orders include circular, spiral and star patterns. In figure 1, the so called class member α , which represents the global orientation of the pattern, is zero. Filters modeling these symmetry descriptions can be obtained by

$$h_n = (x + iy)^n \cdot g \quad \text{for } n \geq 0, \quad (2a)$$

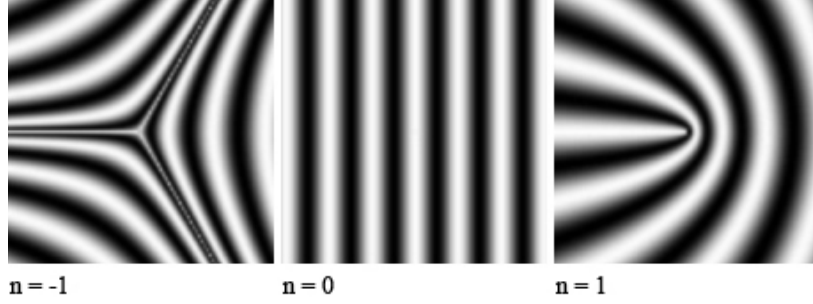


Fig. 1. Patterns with orientation description $z = \exp(in\phi)$: Straight lines for $n = 0$ (linear symmetry); parabolic curves and line endings for $n = \pm 1$ (parabolic/triangular symmetry)

$$h_n = (x - iy)^{|n|} \cdot g \quad \text{for } n < 0, \quad (2b)$$

where g is denotes a 2D Gaussian with standard deviation σ_2 in x and y direction. These features are algebraic invariants of physical operations e.g. translation, rotation and zooming (locally). For a more detailed review of symmetry filters and the symmetry derivatives of Gaussians, we refer to [26]. Decomposing an image into certain symmetries involves calculating $\langle z, h_n \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the 2D scalar product, yielding complex responses $s_n = c \cdot \exp(i\alpha)$, with c representing the certainty of occurrence and α (class member) encodes the direction of symmetry n (for $n \neq 2$). Normalized filter responses are obtained by calculating

$$s_n = \frac{\langle z, h_n \rangle}{\langle |z|, h_0 \rangle}, \quad (3)$$

where the nominator is the total energy of the symmetry (all possible orders) [26]. In this way, $\{s_n\}$ describe the symmetry properties of an image in terms of n orders. $\{n\}$ can be chosen to match the expected symmetries in a candidate image, thus modeling a reference image by a limited number of symmetry features. The definition of quality for a specific application determines the orders and scales (σ) used by this model. Furthermore, we demand $\{s_n\}$ to be well separated over the image plane, in which we look for a high and dominant symmetry at each point. Equation 4 denotes an inhibition scheme [29]

$$s_n^I = s_n \cdot \prod_k (1 - |s_k|), \quad (4)$$

where k refers to the remaining applied orders, to sharpen the spatial extension of filter responses and I is a label that stands for inhibition. Consequently, a high certainty of one symmetry type

requires a reduction of the other types. We calculate the covariance among $\{|s_n^I|\}$ in blocks of size $b \times b$ in order to test if the filter responses have been mutually exclusive. A large negative covariance supports that this is the case and the neighborhood behaves as a high quality local image. On the other hand, positive covariance implies the co-occurrence of mutually exclusive symmetry types in the vicinity of a point, which is an indication of noise or blur. We incorporate this information by weighting the symmetry certainty. We sum $\{s_n^I\}$ over n at each pixel resulting in a total symmetry image

$$s = \sum_n s_n^I \quad (5)$$

The s is further averaged in the blocks (tiles) of size $b \times b$ yielding \bar{s} (we use $\bar{\cdot}$ to denote block wise operating variables). The quality measure \bar{q} for each block is the computed as follows

$$\bar{q} = y(|\bar{r}|) \cdot \chi(-\bar{r}) \cdot \bar{s}, \quad (6)$$

where χ represents the Heavyside function (1 for positive arguments, 0 otherwise) and \bar{r} denotes the block wise correlation coefficient among $\{|s_n^I|\}$. The quantity \bar{r} is calculated as an average of the correlation coefficients between any two involved orders $\bar{r}_{k,l}$, as defined by

$$\bar{r}_{k,l} = \frac{\text{Cov}(|s_k^I|, |s_l^I|)}{\sqrt{\text{Var}(|s_k^I|)\text{Var}(|s_l^I|)}} \quad (7)$$

Note that $\bar{r}_{k,l} = \bar{r}_{l,k}$, and that in case of employing only two orders for the decomposition, e.g. 0 and 1, \bar{r} equals \bar{r}_{01} . The expression $\chi(-\bar{r})$ signals a contribution to \bar{q} if and only if the average covariance \bar{r} is negative. The mapping function m controls the influence sensitivity of \bar{r} and is chosen empirically, e.g. $m(t) = t^2$ makes the method more responsive to quality changes. A quality metric is established by averaging \bar{q} over the “interesting” blocks \bar{i} , which are represented by blocks where $\bar{s} > \tau$, thus having a minimum total symmetry response. The proposed technique is implemented and tested by means of automatic fingerprint image quality estimation.

B. Fingerprint Quality Estimation

By human opinion, the quality of a fingerprint image is usually expressed in terms of the clarity of ridge and valley structures, as well as the extractability of certain points (minutiae, singular points) [8]. We can model the behavior of the orientation tensor of a typical fingerprint entirely with symmetry features. On one hand, a coherent ridge flow has mostly linear symmetry and thus

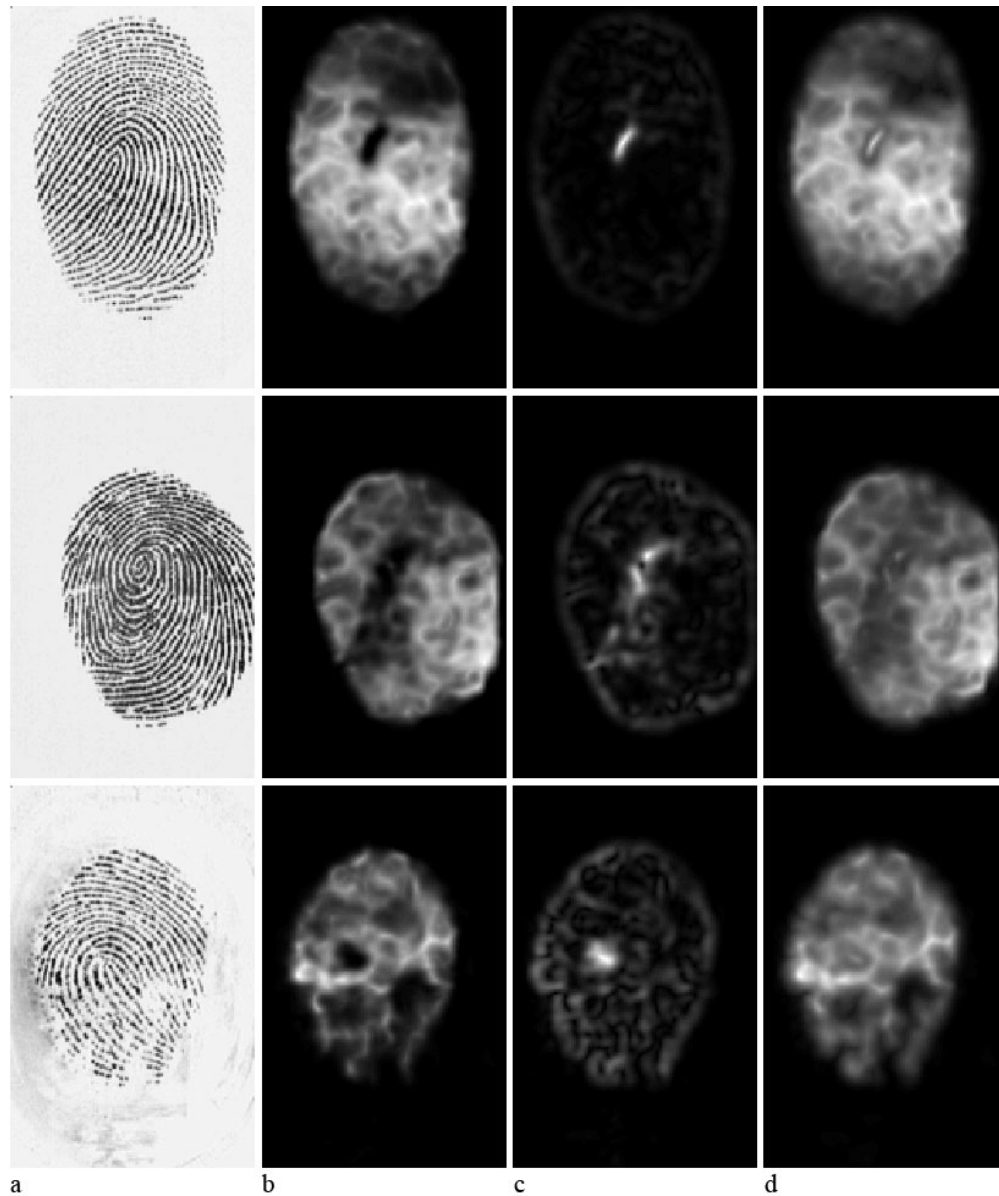


Fig. 2. Decomposition of example fingerprints: a) Original fingerprint; b) linear symmetry magnitude (s_0^1); c) parabolic symmetry magnitude (s_1^1); d) “total symmetry” (summed magnitudes) contains relevant portions (s)

can be modeled by symmetry features of order 0. On the other hand, far fewer minutia points such as ridge bifurcation and ending have parabolic symmetry and can be modeled locally by symmetry features of order 1. An effective local feature extraction method for fingerprint images using these symmetry orders as traits to recognize a fingerprint is presented in [7]. There, linear and parabolic symmetries have been employed to extract minutia points, and spurious minutiae

could successfully be excluded because they would be surrounded by lower linear symmetry compared to authentic ones. This property is encoded in the correlation coefficient here. Other prominent points in fingerprints such as core and delta points can likewise be modeled by symmetry features of order 1 and -1 respectively by using a different scale. Intuitively, features of order $|n| > 1$ are considered not meaningful here and are therefore omitted. Because the pattern of symmetry order -1 contains implicitly subpatterns characterized by order +1, we can model both patterns with an appropriately sized symmetry filter of order +1. Since there might occur difficulties to distinguish some small minutia points from ridge-valley breaks, e.g. scars, we focus on the global ridge structure. This can be achieved by employing large filters. Alternatively, as we are doing, one can downsize the original image by one half before applying the algorithm. The more a fingerprint looks like either of the symmetry patterns $n=0$ and $n=1$ in a small neighborhood (e.g. 20×20 px), and the better classification possible, the higher will the assessed quality be. So, good quality ridge-valley structure exposes smooth transitions between the two (and only these) involved symmetries. Only three scalar products are needed with the orientation tensor, $\langle z, h_0 \rangle$, $\langle |z|, h_0 \rangle$ and $\langle z, h_1 \rangle$. The first two scalar products essentially correspond to averaging the orientation tensor z and its magnitude $|z|$ respectively, whereas the last scalar product corresponds to a derivation of z . All convolutions can be implemented employing 1D Gaussian filters and their derivatives. The used value for block size b is 8 px, the symmetry features use a σ_2 of 3. For the construction of the orientation tensor, a σ_1 of 0.6 is employed. These values were chosen in an optimization search. Slight variations will however not affect the functionality of the method fundamentally.

Figure 2 depicts the results for some example fingerprints of the QMCYT database. As can be seen, s (column d) contains the relevant portion of the image. s_0^I and s_1^I are (depending on the quality) well defined at linear and parabolic symmetry neighborhoods respectively. The quality of the fingerprint shown in the first row in figure 2 is rather good, although appears to be lower in the upper right part. This is also mirrored in the linear and the total symmetry image (columns b and d respectively). The fingerprint depicted in the center row indicates that the core point does not need to be of type “loop” in order to be included in s_1^I as its subpatterns can be modeled by the order 1 filter. The quality of the fingerprint represented by the bottom row of figure 2 worse. The sensor appears to be dirty and the finger is too dry. s_0^I reflects the structural weaknesses by exhibiting fewer points having large magnitudes and the contrast in magnitude

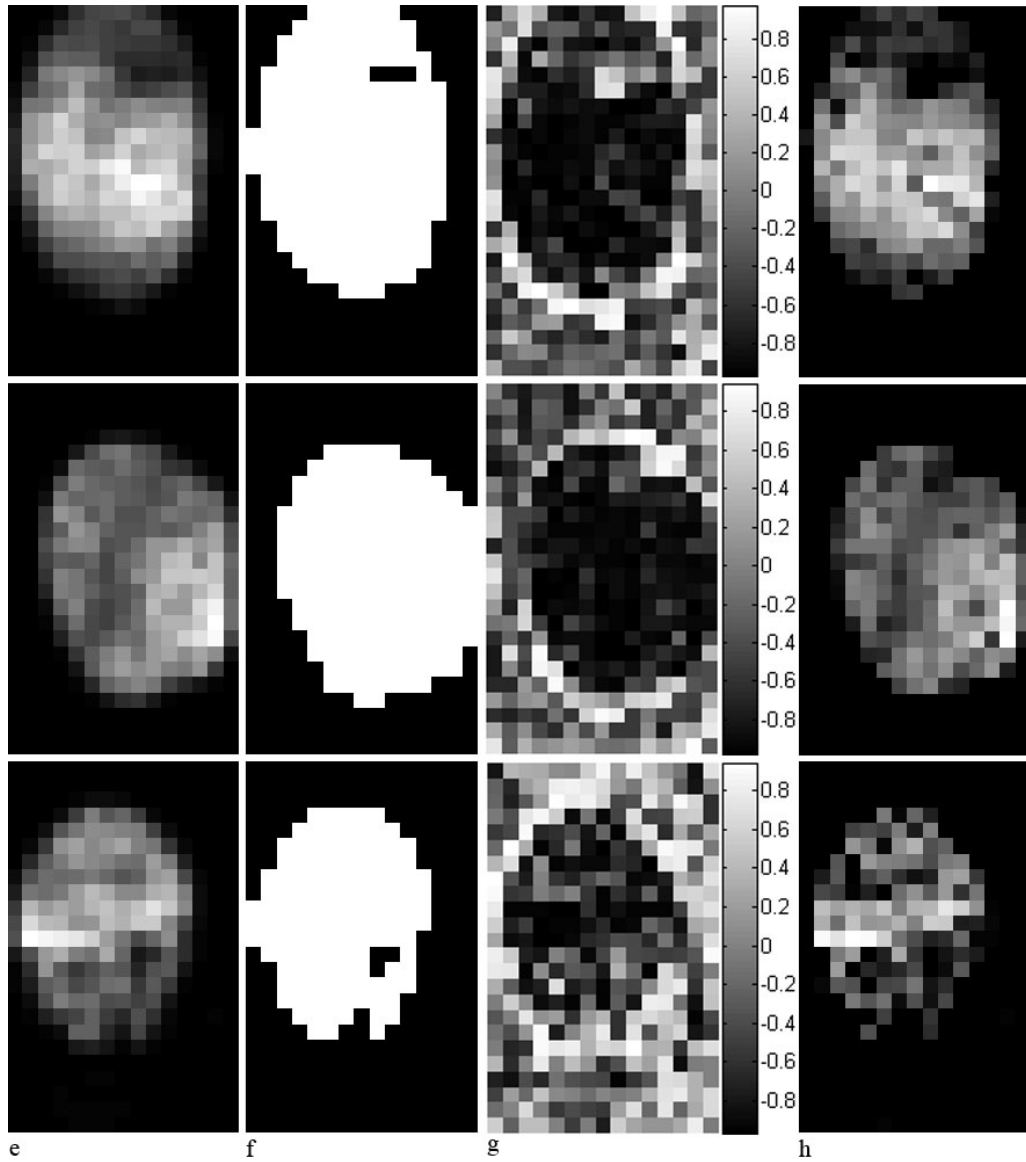


Fig. 3. Intermediate steps in fingerprint quality estimation: e) Tiled (averaged block by block) total symmetry (\bar{s}) and f) thresholded (\bar{i}); g) Correlation coefficient between the tiled parabolic and linear symmetry magnitudes (\bar{r}); h) Tiled quality measure (\bar{q})

between the core point and the rest in s_1^1 is not that clear any more. In this case the correlation coefficient is even in some blocks of the foreground.

The tiled images representing the block wise variables for the example fingerprints are displayed in figure 3. Column f (\bar{i}) indicates that fingerprint segmentation is done implicitly. We

observe, that the covariance is negative in reasonably good-quality regions, whereas it is positive in noisy and bad-quality regions. This separation is not so apparent when considering \bar{s} only. Column h (\bar{q}) visualizes the final quality blocks, with brightness representing good quality. By averaging \bar{q} a overall quality metric Q in $[0, Q_{\max}]$ for some constant $Q_{\max} \leq 1$ is retrieved. Although we have confined our reporting to adapt multi-algorithm fusion, it should be noted that \bar{q} can also be used in other fingerprint processing modules, e.g. to steer a fingerprint enhancement process, to favor robust feature extraction or matching. Note that the symmetry features applied here have already been used for fingerprint alignment as well as matching [7], [30], indicating its added value by dual usage.

Previously published studies on fingerprint quality assessment methods measure spatial coherence of the ridge flow only, by essentially determining or approximating s_0 . Additionally the latter is commonly partitioned into blocks \bar{s}_0 , which are then weighted decreasingly with distance to the fingerprint's centroid when calculating a quality metric. Inspecting, figure 4 reveals that this strategy may not be enough, because important regions such as singular points (e.g. core, delta) are per definition incoherent to the ridge flow, and their strong presence therefore automatically impairs the estimated quality. In figure 4 \bar{s}_0 and our metric are shown on three images from the QMCYT database. Note the different shape of the singular point regions not leading to different results though. Quantitative results with comparisons will be presented further below. To the best of our knowledge, there is no other method to measure quality from both typical and high curvature ridge-valley structure, or to deal with it explicitly, e.g. to preventively disregard high curvature regions. It is often the latter region, which carries the most discriminative information, though.

III. FUSION

In section III-A, Bayes-based training shall help to weight several recognition experts responses for an optimal joint decision. By additionally incorporating a confidence measure modeled by the inverse quality of the fingerprints, the quality sensitivity of a single recognition algorithm is used to adapt (shift) those weights based on the current claim - trained, (quality) adaptive fusion. In section III-B, we propose a cascade type fusion, which exploits a similar effect, but without training and a more aggressive weight adaptation. The purpose here is to save resources (computation time) by executing just one recognition system mostly, while several systems are

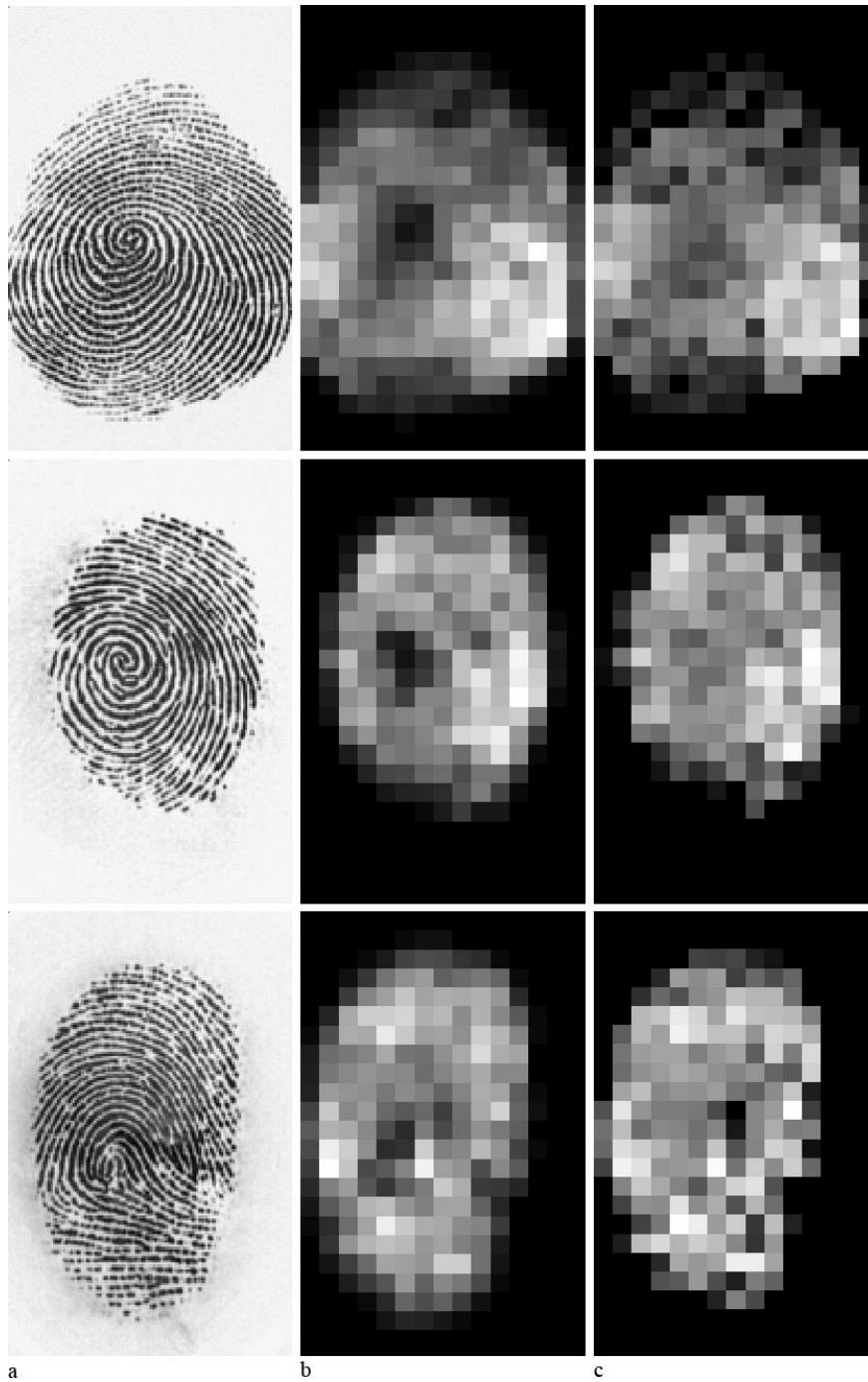


Fig. 4. Illustrating the difference s_0 (b) and \bar{q} (c): Here we can see that the core point is misinterpreted in terms of quality when just averaging s_0

executed if the current fingerprints are of bad quality - untrained, (quality) adaptive fusion. Finally, in section III-C, we list and motivate some simple fusion schemes, where different recognition experts are weighted equally, because we use such schemes for comparison in our experiments - untrained, non-adaptive (to quality) fusion.

A. Bayesian Supervisor

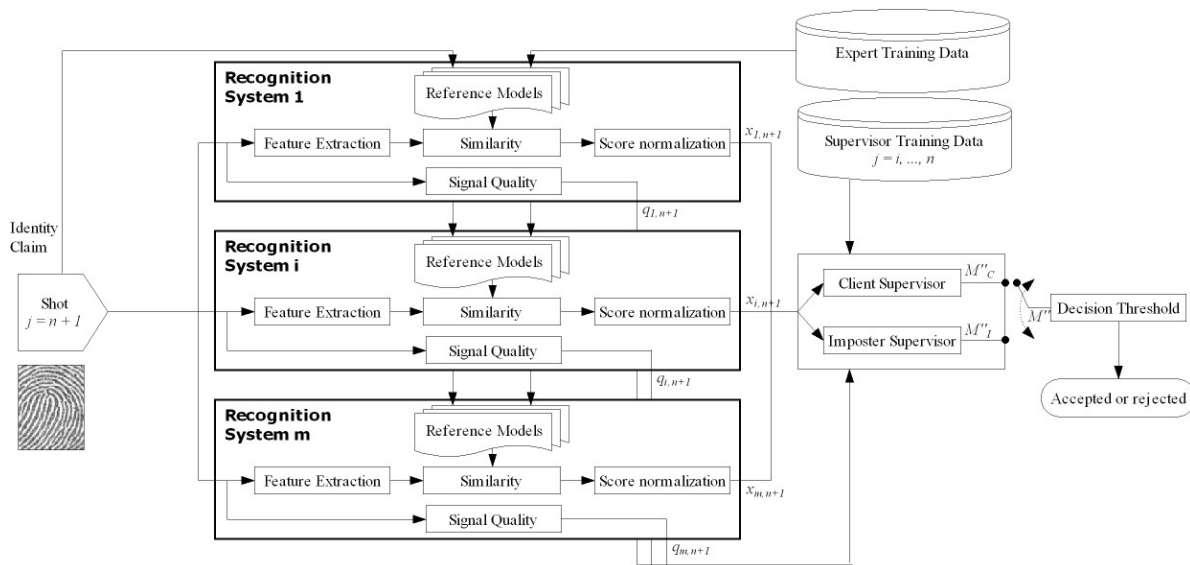


Fig. 5. Multi-algorithm system model: Schematics including all components of the proposed Bayesian supervisor. All experts deliver a certainty in addition to their score, which is estimated as the inverse of the image quality here

This section is devoted to an adaptive fusion scheme using Bayes theory [31]. For a more profound description of the employed model we refer to [4], [14]. Its probabilistic background is further detailed in [32]. As indicated in figure 5 we combine independent fingerprint recognition systems yielding a monomodal multi-algorithm environment. An input fingerprint is referred to as a *shot*. For each shot we get different expert opinions which are delivered to the Bayesian supervisor. The following notation is used when describing the statistical model and the supervisor within this paper:

- i : Index of the experts, $i \in 1 \dots m$
- j : Index of shots, $j \in 1 \dots n, n + 1$. It is the system clock since an expert has one shot per evaluation time.
- x_{ij} : The authenticity score computed by expert i based on shot j
- s_{ij} : The variance of x_{ij} (estimated by expert i).
- y_j : The true authenticity score of shot j
- z_{ij} : The error (mis-identification) score of an expert $z_{ij} = y_j - x_{ij}$

The true authenticity score y_j can only take two numerical values, namely “True” or “False”. So if the values of x_{ij} are between 0 and 1, the values of y_j are chosen to be 0 and 1 respectively. The training of the supervisor is performed on the shots $j \in 1 \dots n$, where x_{ij} and y_j are known. When the supervisor is operational, we consider the shot $j = n + 1$ as a test shot. In this case only $x_{i,n+1}$ is known and the task of the supervisor is to estimate $y_{i,n+1}$. It is assumed that the single experts and the supervisor are trained on different sets.

Note that the experts provide a quality estimate in addition to each score which is modeled to be inversely proportional to s_{ij} . This variance is then used by the supervisor for evaluation.

1) *Statistical Model*: The employed adaptive fusion strategy uses Bayesian statistics and assumes the errors of the single experts to be normally distributed, i.e. z_{ij} is considered to be a sample of the random variable $Z_{ij} \sim N(b_i, \sigma_{ij}^2)$. This does not strictly hold for common audio- and video-based biometric machine experts [14]. Nevertheless it was shown that this problem can be addressed by considering client and impostor distributions separately. Thus, the following two supervisors representing the expert opinions $y_j = 1$ and $y_j = 0$ are constructed:

$$\mathcal{C} = \{x_{ij}, s_{ij} | y_j = 1 \text{ and } 1 \leq j \leq n\} \quad (8)$$

$$\mathcal{I} = \{x_{ij}, s_{ij} | y_j = 0 \text{ and } 1 \leq j \leq n\} \quad (9)$$

The two supervisors will be referred to as *client supervisor* and *impostor supervisor*, respectively.

The task of the client supervisors is to estimate the expected true authenticity score y_j based on its knowledge of client data i.e. computing $M_C'' = E[Y_{n+1} | \mathcal{C}, x_{i,n+1}]$. The prime notation is used to distinguish the 3 different supervisor states. No prime means training, one denotes calibration and two indicate the authentication (operational) phase. The impostor supervisor estimates y_j by computing $M_I'' = E[Y_{n+1} | \mathcal{I}, x_{i,n+1}]$.

The supervisor which comes closer to the ideal case (1 for the client supervisor, 0 for the impostor supervisor) is considered as the final conciliated overall score M'' :

$$M'' = \begin{cases} M_C'' & \text{if } |1 - M_C''| - |0 - M_I''| < 0 \\ M_I'' & \text{otherwise} \end{cases} \quad (10)$$

2) *Supervisor*: Having the experts scores and the quality estimates, the Bayesian supervisor can be summarized as follows:

1) *Training Phase*: In case of the client supervisor, the bias parameters for all experts are estimated as follows:

$$M_{C_i} = \frac{\sum_j \frac{z_{ij}}{\sigma_{ij}^2}}{\sum_j \frac{1}{\sigma_{ij}^2}} \quad \text{and} \quad V_{C_i} = \frac{1}{\sum_j \frac{1}{\sigma_{ij}^2}} \quad (11)$$

here, j is the index of the training set \mathcal{C} . The variances σ_{ij}^2 are calculated by $\bar{\sigma}_{ij}^2 = s_{ij} \cdot \alpha_{C_i}$, where

$$\alpha_{C_i} = \frac{\left(\sum_j \frac{z_{ij}^2}{s_{ij}} - \left(\sum_j \frac{z_{ij}}{s_{ij}} \right)^2 \left(\sum_j \frac{1}{s_{ij}} \right)^{-1} \right)}{n_C - 3} \quad (12)$$

n_C denotes the number of shots in \mathcal{C} . If one or more experts do not provide any quality estimates s_{ij} is set to 1. The bias parameters M_{I_i} and V_{I_i} for the impostor supervisor can be estimated similarly.

2) *Operational Phase*: At this stage authentication on “live” data is performed i.e. the time instant is $n + 1$ and the trained supervisors can access the expert opinions $x_{i,n+1}$ but not to the true authenticity score y_{n+1} . In a first step, the client and impostor supervisor have to be calibrated in order to adapt to their past performance. In case of the client supervisor this calibration is denoted according to

$$M'_{C_i} = x_{i,n+1} + M_{C_i} \quad \text{and} \quad V'_{C_i} = s_{i,n+1} \cdot \alpha_{C_i} + V_{C_i} \quad (13)$$

Having the calibrated experts, they are combined as follows:

$$M_C'' = \frac{\sum_{i=1}^m \frac{M'_{C_i}}{V'_{C_i}}}{\sum_{i=1}^m \frac{1}{V'_{C_i}}} \quad (14)$$

The computations for the impostor case (M'_T , V'_T and M''_T) follow the same pattern. The final supervisor decision is made according to 10.

The procedure described above was successfully applied in risk analysis [32] and in multimodal authentication applications [4], [33]. In these studies verification performance improvements of almost an order magnitude were achieved, compared to the best single modality. In our multi-algorithm framework the quality of a shot is estimated automatically and not by humans, which constitutes one novelty of this work.

3) *Quality adaptive strategy*: As indicated in figure 5, each expert provides a score x_{ij} and a variance s_{ij} for every single authentication assessment. The variance is not an estimation of the general reliability of the expert itself. It is considered as a certainty measure for the current score based on the quality of the input shot. So we propose to calculate s_{ij} using the qualitative knowledge of the experts on the input biometric data they assess. Section II details our approach to extract such a quality estimate from a shot. In the second part of equation 13 the trained supervisor adapts the weights of the experts employing the input signal quality. We define quality index q_{ij} of the score x_{ij} as follows:

$$q_{ij} = \min(Q_{ij}, Q_{i,\text{claim}}) \quad (15)$$

where Q_{ij} is the quality estimate produced by expert i in shot j and $Q_{i,\text{claim}}$ is the average quality of the biometric samples used by expert i for modeling the claimed identity. All quality values are in the range $[0, q_{\text{max}}]$ where $q_{\text{max}} > 1$. In this scale 0 is the poorest quality, 1 is considered as normal quality and q_{max} corresponds to the highest quality. The final variance parameter s_{ij} of the score x_{ij} is obtained by

$$s_{ij} = \frac{1}{q_{ij}^2} \quad (16)$$

Training is the key point with the Bayes-based fusion approach. The biases $M_{Ci}/(M_{Ti})$ and $V_{Ci}/(V_{Ti})$ that represent classification performance of expert i during training are used to weight the experts scores in the joint accept/reject decision. This is done in non-adaptive fusion without incorporating any experts confidences into their scores. In adaptive fusion, these confidences are included with $s_{ij} \neq 1$ to the effect that low confidence in its score for the current claim decreases the experts say in the joint decision. This confidence is modeled inversely by the square of the quality measure for the involved fingerprints, and the according biases are adapted

during training. So a connection between image quality and an experts credibility is established, which is later exploited to continuously shift decision power among experts as a function of both image quality and past classification performance.

B. Cascaded Fusion

One may ask why to use several experts within one modality, instead of using multimodal configurations or simply combining the best capabilities of each expert into a single one? For a variety of reasons, e.g. vendors' conditions, legal confinements, the single experts have to be regarded as black boxes (closed systems) while recognition rates still have to be raised. One can also argue that the time issue is problematic if several systems have to be executed for every single match, i.e. for identification within a large database e.g. U.S.-VISIT. A reasonable way to address this issue is to dynamically include further experts if a single one cannot come up with a clear decision. In such a configuration a minimal number of experts is active most of the time. This is also visualized in figure 6, where we see a series of systems - primary, secondary, etc. system in the following - triggered by certainty thresholds, meaning that system n is utilized if and only if c_{n-1} is below a certain threshold. Afterwards all available scores x_i are fused according to a fusion rule f , which can be chosen simple. In our experiments, the systems are ordered by recognition performance. This configuration is inspired by cascaded classifiers

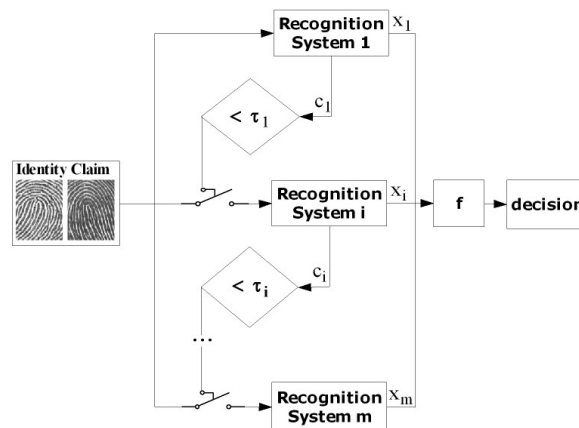


Fig. 6. Cascaded fusion: Experts are triggered on demand and combined only under uncertainty (here: bad quality)

[34], i.e. degenerate decision trees [35], although classifiers could be ordered following different

aspects and scores are not fused there in general. Using scores themselves as certainty thresholds is not recommendable since they are naturally low in most of the cases for identification, and they might be wrong as well. In contrast, image quality is practicable, since the probability of a false acceptance or rejection is higher if the quality of the involved impressions is lower, while multi-algorithm fusion opposes this fact. The image quality used as certainty threshold is relatively independent of the single experts, such that c_i can be shortened to c (compare figure 6). So the number of experts included into the current decision is determined by a single certainty. The primary system can be a single system, although results will probably improve if combinations are already used. But the idea is to utilize as few systems for any match as possible, motivated by faster execution, while still getting the benefit of improved recognition rates. It would be further desirable, if the quality assessment method exploited computational steps necessary by the primary system to save resources. As we use an adaptive, yet untrained, fusion scheme, inferior results compared to the previous strategy (under discussed conditions) are to be expected but its advantages are clearly evident.

C. Simple Schemes

Past experiments indicated that combining systems in simple ways could already lead to relatively good results. Such fusion schemes include, for example, sum and max rule, meaning that the average respectively the maximum of all experts' score is taken as the final score. Because they are non-adaptive we also refer to them as global max, global sum, etc. It has been claimed in several studies that simple schemes are not clearly outperformed by trained (non-adaptive) strategies, for example, support vector machines, in neither monomodal fusion [13] nor multimodal fusion [16]. Simple, yet adaptive schemes have been successfully applied in quality-based multi-algorithm fusion [5]. In the latter study, minutia and texture-based, hence heterogeneous systems have been simply combined but weighted according to the image quality, because the texture based system proved to be more robust against quality impairments. In this work, only non-adaptive simple schemes are used to facilitate comparison.

IV. EXPERIMENTS

As recent studies confirmed, recognition performance is heavily affected by the quality of the input images [3], [22], [36], i.e. reliable feature extraction and image enhancement are

confined by the a priori image quality [8]. An approach to measure quality degradation effects on the recognition performance is to divide the database into several quality groups, and to run recognition tests within these groups only. Likewise, a quality assessment method can be benchmarked against another one, here, estimates by automatic methods versus human grading. We chose NFIQ¹ to compare our method with. NFIQ was introduced as an independent fingerprint quality estimator that is intensely trained to forecast matching performance in [3] and is publicly available as a package of NIST² FIS2³ [37]. The method was tested on 300 different combinations of fingerprint image data and fingerprint recognition systems and found to predict matching performance for all systems and data sets.

In this study, all experiments are conducted on the QMCYT fingerprint database [23], and some on two databases employed in FVC2004 [9]. The former is defining 75×10 fingerprints $\times 12$ impressions, whereas the latter contain 100 fingerprints $\times 8$ impressions per database. For each impression in the QMCYT database a manually annotated quality label is available [38]. We employ a recently developed fingerprint recognition system [7], called system A in the following to validate the quality estimates. To investigate feature independency we also employ the NIST FIS2 fingerprint recognition packages MINDTCT and BOZORTH3 - jointly referred to as system B - in a similar test. Note, that system B is entirely minutia-based whereas system A is exploiting both minutia and texture features for fingerprint alignment and matching respectively. As a third expert, system C represents a non-minutia based recognition system utilizing Gabor features similar to [24], as described in [22]. While there is a need to separate tasks, it is worth noting that both recognition systems A and B use feature vectors which are effectively used by the proposed quality assessment method and NFIQ respectively. The 750 fingerprints of the QMCYT database are split into 5 equally sized partitions of increasing quality. The criteria for a fingerprint to be part of a certain group I-V includes the average quality index for its genuine trials. The latter are chosen to be 150×9 per group, while 150×74 impostor trials are performed, taking fingers of the same type only as impostors (1 impression). We show the EER of system A, B and C for all quality groups, which have been established according to the

¹NIST Fingerprint Image Quality

²National Institute of Standards and Technology

³NIST Fingerprint Image Software 2

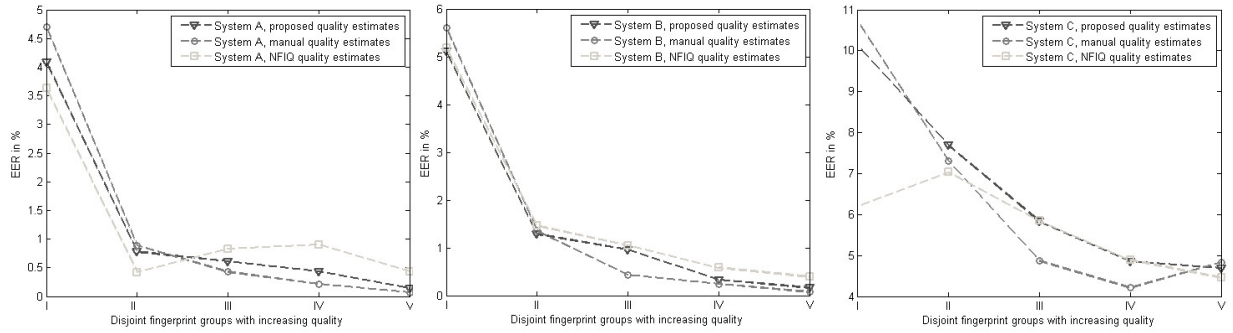


Fig. 7. EER for systems A, B, and C (from left to right) within quality groups I-V from the QMCYT database. The partitions are established by means of different quality assessment methods (see legend).

different quality assessment methods (see figure 7). According to the EER curves we can observe that both automatic quality assessment methods approximate the manual estimates very well, and that the proposed method shows most similar behavior. Note that both recognition systems may make errors independent of the image quality, but the performance trajectories given manually divided quality groups support the view that the quality estimates of our proposal follow the opinion of the human opinion quite well. To further investigate the equivalence of the differently derived quality estimates, we calculate correlation coefficients (ρ) similar to equation 7, but between arrays of quality values. The NFIQ and manual labels have the correlation coefficient $\rho = 0.38$, whereas the proposed estimates and the manual labels correlate with $\rho = 0.47$, thus providing further support for the results of the previous experiment. It is worth mentioning that the grading by the proposed method is continuous in $[0,1]$, whereas it is discrete for NFIQ and the human opinion being in $[1..5]$ and $[0..9]$ respectively. When applicable, the latter two output ranges are normalized into $[0..1]$.

The same experiment is repeated for databases DB2 and DB3 employed in FVC2004. The 100 fingerprints of each database are split into partitions following the rules from above. For each database and per quality group, 20×28 genuine trials and 20×99 impostor trials are performed. We show the EER of system A and B for all quality groups in the top row (DB2) and bottom row (DB3) in figure 8. System C is left out due to the undesirable findings in the previous experiment (lower recognition rates at higher quality). When looking at figure 8 we can observe a generally higher EER level and variance. The correct estimation of the different quality categories has more

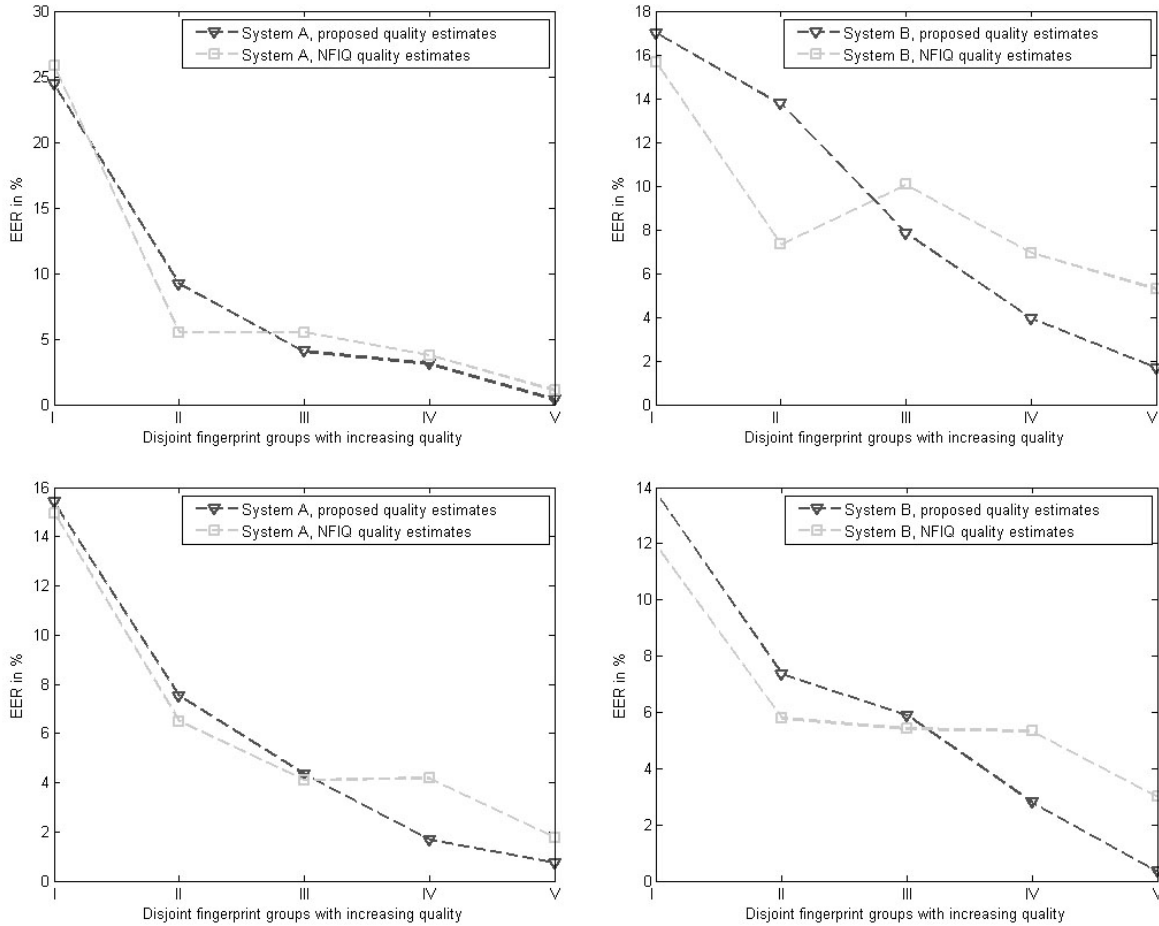


Fig. 8. EER for systems A and B (from left to right) within quality groups I-V from DB2 (top row) and DB3 (bottom row) of FVC2004. The two automatic quality assessment methods are used to establish the partitions (see legend).

impact on recognition rates than before, even in absolute terms (compare figure 7). This lies with the increased difficulty of the FVC2004 databases for recognition systems, which we also consider relevant for quality estimation methods. The extreme conditions (extra low/high finger pressure, fingers dried/moistened by intention) lead to severe image quality impairments, which were obviously detected well by both quality estimators. The quality grading by our method remarkably leads to monotonically decreasing EER curves for all involved recognition systems and databases.

It is worth noting, that the QMCYT and FVC2004-DB2 databases were acquired with the help of optical sensors, whereas a thermal sweep sensor was employed for the acquisition of

FVC2004-DB3. Different sensors produce different ground qualities as well as fingerprint looks, which are obviously handled well by our untrained quality assessment method. The latter also manages to deliver a fine quality grading that is most similar to those of independent humans. We put this down to the usefulness of the employed symmetry features and their energy-independent usage in our algorithm (using normalized filter answers).

	A	B	C		A,B	A,C	B,C,	A,B,C
EER %	1.22	1.9	6.37	SUM	1.06	1.22	1.36	1.56
				MAX	0.75	0.84	1.16	1.16

TABLE I

EER OF SINGLE EXPERTS (SYSTEMS) AND WHEN THEY ARE COMBINED USING SIMPLE FUSION SCHEMES (MAX/SUM)

In table I we state the EER for each recognition system (A, B and C) over the whole QMCYT database, i.e. when the quality division is dissolved again. Note, that system A and B are remarkably better than system C on the test set. In the following, the three systems A-C are combined (at least two experts) using the fusion schemes explained in the section above. A jackknife (leave-one-out) strategy is employed whenever training is involved, meaning that the training set is all users but one (who together with the impostors forms the test set), and all users are tested at some point, giving an averaged EER rate. A number of 4 impressions is used for both client and impostor supervisor training, whereas 9 respectively 74 impressions not belonging to the training set are being tested on. Note that each fingerprint is effectively treated as user and that we take impostors of the same finger type only.

When employing non-trained fusion schemes, the test set is all available users, giving 750×9 genuine and 750×74 impostor trials again. The performance (EER) of expert combinations using simple, non-adaptive schemes is given in table I. We can observe, that combinations involving the best expert (system A) deliver the best results, actually outperforming the best expert almost every time. In this test, fusion applying the MAX rule is better than using SUM, although, the former was favored by shifting the experts to a common operating point. The overall best result using simple schemes involves the two first systems and enables a drop in EER of $\approx 38\%$ compared to the best expert when isolated. It is worth noting that combining all three experts can worsen the joint performance compared to when selecting two (which need

not even be the leading ones). This lies with “simply” fusing severely differently skilled experts without training.

Figure 9 shows the performance of cascaded fusion of system A and B as a function of certainty τ , chosen as the thresholded quality index. Manual quality estimates are taken in case of the dotted gray line to illustrate a best case, while estimates by our method are considered along the path of the dotted black line. Recognition performance of the single systems, further fused by simple schemes - independent of quality though - are indicated as well, with the MAX rule giving the best result (EER of 0.75%). Employing a cascade with system A and B as primary and secondary system respectively, this 0.75% are approached from above with a small remainder, more precisely, with a difference in EER of 0/0.11% when employing manual/automatic quality indices respectively. The big difference is that system A only is utilized in $\approx 84\%$ of all trials. The corresponding choice for the quality threshold is marked by the leftmost arrow in the left-hand part of figure 9, and its “efficiency impact” is marked by the corresponding topmost arrow in the right-hand part of the figure. As illustrated, we (almost) maintain the best error rate for simple fusion of the two systems, but actually need system B every sixth time only. Another interesting “operating point” is indicated by the second arrow in the left-hand part of figure 9, at which the maximum is reached (EER of 0.75%) while both systems are utilized in $\approx 49\%$ only. For these experiments, the MAX rule was employed as cascaded fusion function f .

For the Bayesian-based fusion scheme, indices derived from a quality assessment method are assigned to either one of the systems A-C. This is because we wish to quantify the impact of the image quality on the bayesian supervisor fusion and a certain expert’s ability. The remaining two experts are assigned a quality of 1 (normal) for each trial. The best results in terms of EER are shown in figure 10. It turned out that system A is most suitable to attach certainties based on image quality, which is indicated by q_A instead of A in figure 10. Worth noting, we can observe a drop in EER of $\approx 97/95\%$ when adaptively fusing all experts (q_{A-B-C}) compared to system A in isolation. If manually assigned quality is entering fusion, the EER drops by $\approx 97\%$ and if quality indices derived from our method are used the error rate drops by $\approx 95\%$. Adaptive fusion is able to significantly increase recognition performance independent of the quality assessment method employed, while the improvement using three experts as compared to two is relatively small: By including system C in non-adaptive Bayesian supervisor fusion the EER drops by $\approx 35\%$. This improvement is remarkably better than in case of the simple fusion schemes where the EER

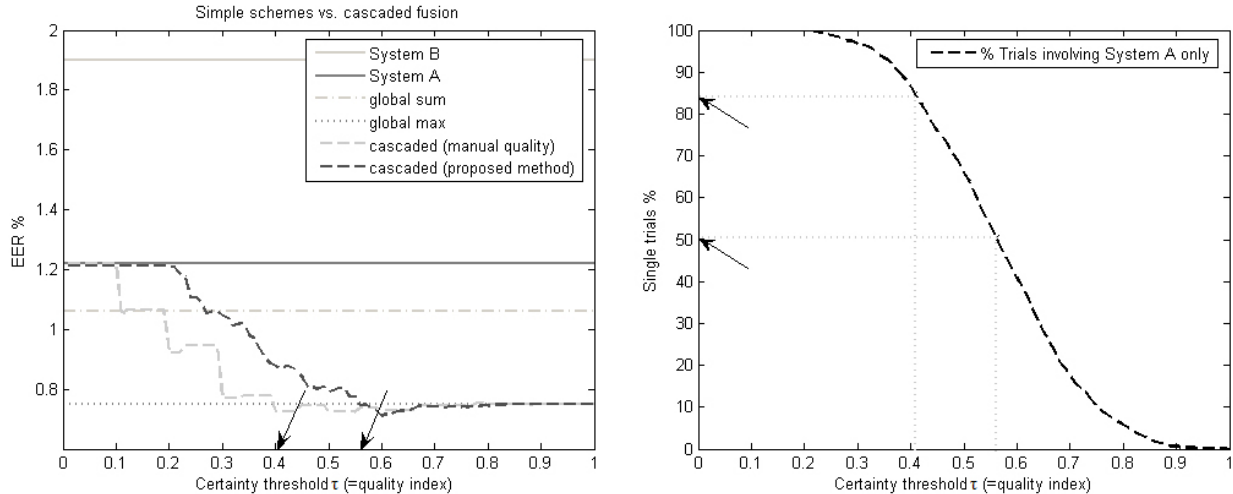


Fig. 9. Left: Results for cascaded fusion compared to simple schemes; The arrows indicate favorable threshold values for the certainty (=image quality) at which the secondary system is triggered. Right: “Efficiency impact” of cascaded fusion; The two arrows are connected to the former arrows through the dotted lines. The function indicates, how many trials (matches) in percent are performed by System A only at a certain trigger threshold (and a certain joint recognition accuracy).

even increase when systems A and B are complemented by system C. This is obviously another effect of training. Previous work has shown that the training of these supervisors is relatively soon satisfied (20 out of 75 users [4]).

It is worth noting that both training and non-training supervisors are important to different applications as demands on computational efficiencies versus decision performance are different. However, in both cases the automatic quality estimates delivered significant benefits as the above experiments indicate. While there have been some studies on how to incorporate quality into training supervisors, the corresponding strategies were largely unstudied for non-training supervisors. The cascade strategy presented above intends to contribute to the latter.

V. CONCLUSION

In this work we proposed a reduced-reference image quality assessment method, together with adaptive monomodal multi-algorithm fusion strategies. We showed how a priori content-knowledge about fingerprints can be encoded and used in quality estimation. The practical benefit is avoidance of expensive training. As the experimental results on fingerprint quality estimation underline, the proposed method competes well with another, yet heavily trained automatic method

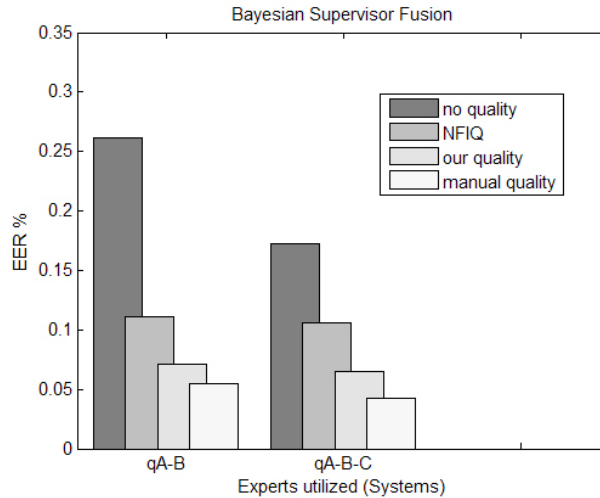


Fig. 10. Best combinations for Bayesian Supervisor Fusion: The quality indices were used to weight system A only (therefore qA)

(NFIQ) on several databases. The introduced method was also behaving closest to human opinion on fingerprint quality, which emerged as excellent in comparison to all computational approaches. When exploited to adapt fusion parameters, levels of agreement studies between human and machine quality assessments have not been reported before, to the best of our knowledge.

We employed three fingerprint verification systems, with varying performance in isolation, to implement multi-algorithm fusion. We wanted to stress the importance of adapting the fusion parameters as a reaction to the image quality of the claim to process. Already by applying a simple fusion scheme (MAX rule) the combined performance exceeded the best expert's (system A with 1.22% EER), leading to an EER of 0.75%, yet could careless simple fusion of several experts also increase the EER. Coming to adaptive fusion, we introduced a non-trained cascaded scheme to dynamically switch on experts in case of uncertainty (low quality), assuming time is the most limited resource. We experimented on two experts in this case, and we could approach the best possible EER, for example, up to 0.11% with the help of our automatic quality indices, while saving to run the second expert 5 out of 6 times. In order to show the full potential of multi-algorithm fusion, we implement a more sophisticated Bayes-based supervisors for both client and impostor expectation estimation. Taking advantage of training and additionally the quality estimates by the proposed method, EERs of 0.17% and 0.07% are achieved, respectively.

It is worth mentioning that similar results can be achieved when the single experts are worse than in this study, as long as they measure complementary information.

REFERENCES

- [1] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Processing*, 13, 2004.
- [2] A. M. Eskicioglu and P. S. Fisher. Image quality measures and their performance. *IEEE Trans. Communications*, 43:2959–2965, December 1995.
- [3] E. Tabassi, C. Wilson, and C. Watson. Fingerprint Image Quality. Technical Report NISTIR7151, Nist, 2004.
- [4] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in mobile communications. In *Proc. of 12'th Int. conf. on image analysis and processing, Mantova, Italy*, pages 2–11. IEEE Computer Society Press, Piscataway, NJ, september 17-19 2003.
- [5] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. Jain. Incorporating Image Quality in Multi-Algorithm Fingerprint Verification. In *Proc. of IAPR Intl. Conf. on Biometrics, ICB, Hong Kong, China*, volume LNCS-3832, pages 213–220. Springer, January 2006.
- [6] J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern Recognition*, 38:777–779, 2005.
- [7] H. Fronthaler, K. Kollreider, and J. Bigun. Local Feature Extraction in Fingerprints by Complex Filtering. In *International Workshop on Biometric Recognition Systems IWBRIS 2005, Beijing, China*, volume 3781, pages 77–84. Springer, 22–23 October 2005.
- [8] Y. Chen, S. Dass, and A. Jain. Fingerprint Quality Indices for Predicting Authentication Performance. In *Audio- and Video-based Biometric Person Authentication (AVBPA) 2005, Rye Brook, New York*, pages 160–170, July 2005.
- [9] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, and A.K. Jain. FVC2004: Third Fingerprint Verification Competition. In *International Conference on Biometric Authentication (ICBA04), Hong Kong*, pages 1–7, July 2004.
- [10] E. Lim, X. Jiang, and W. Yau. Fingerprint quality and validity analysis. In *IEEE International Conference on Image Processing*, volume 1, pages 469–472. IEEE, 2002.
- [11] F. Alonso-Fernandez, J. Fierrez-Aguilar, and J. Ortega-Garcia. A review of schemes for fingerprint image quality computation. In *3rd COST-275 Workshop on Biometrics on the Internet, Hatfield, UK*, pages 3–6. Official Publisher of the European Communities, October 2005.
- [12] A. Ross, A. K. Jain, and J. Reisman. A hybrid fingerprint matcher. *Pattern Recognition*, 36:1661–1673, 2003.
- [13] J. Fierrez-Aguilar, L. Nanni, J. Ortega-Garcia, R. Cappelli, and D. Maltoni. Combining multiple matchers for fingerprint verification: a case study in FVC2004. In *Proc. of 13th IAPR Intl. Conf. on Image Analysis and Processing, Cagliari, Italy*, volume LNCS-3617, pages 1035–1042. Springer, September 2005.
- [14] E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multi modal person authentication systems by Bayesian statistics. In J. Bigun, G. Chollet, and G. Borgfors, editors, *Audio and Video based Person Authentication - AVBPA97*, pages 291–300. Springer, 1997.
- [15] J. Kittler, M. Hatef, R. Duin, and J. Mates. On combining classifiers. *IEEE-PAMI*, 20:226–239, 1998.
- [16] A. Ross and A. K. Jain. Information fusion in Biometrics. *Pattern Recognition Letters*, 24:2115–2125, 2003.
- [17] J. Kittler and K. Messer. Fusion of multiple experts in multimodal biometric personal identity verification systems. In *Proc. of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pages 3–12, 2002.

- [18] R. P. W. Duin. The combining classifier: to train or not to train? In *Proc. of the 16th Intl. Conf. on Pattern Recognition - ICPR 2002*, pages 765–770, 2002.
- [19] F. Roli, G. Fumera, and J. Kittler. Fixed and trained combiners for fusion of imbalanced pattern classifiers. In *Proc. of the 5th Intl. Conf. on Information Fusion - FISOPM 2002*, pages 278–284, 2002.
- [20] A. K. Jain and A. Ross. Learning user-specific parameters in a multibiometric system. In *Proc. of the 2002 IEEE Intl. Conf. on Image Processing - ICIP 2002*, pages 57–60, 2002.
- [21] J. Fierrez-Aguilar, D. Garcia-Romero, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Bayesian adaptation for user-dependent multimodal biometric authentication. *Pattern Recognition*, 38(8):1317–1319, 2005.
- [22] J. Fierrez-Aguilar, L.-M. Munoz-Serrano, F. Alonso-Fernandez, and J. Ortega-Garcia. On the effects of image quality degradation on minutiae- and ridge-based automatic fingerprint recognition. In *IEEE Intl. Carnahan Conf. on Security Technology ICCST, Las Palmas de Gran Canaria, Spain*. IEEE Press, October 2005.
- [23] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. MCYT baseline corpus: A bimodal biometric database. *IEE Proc. VISIP*, 150(6):395–401, 2003.
- [24] A. Ross, J. Reisman, and A. Jain. Fingerprint matching using feature space correlation. In *ECCV 2002 workshop on Biometric Authentication*, LNCS 2359, pages 48–57. Springer, 2002.
- [25] H. Fronthaler, K. Kollreider, and J. Bigun. Automatic Image Quality Assessment with Application in Biometrics. In *IEEE Workshop on Biometrics, in association with CVPR-06, New York*, pages 30–35, June 2006.
- [26] J. Bigun, T. Bigun, and K. Nilsson. Recognition by symmetry derivatives and the generalized structure tensor. *IEEE-PAMI*, 26:1590–1605, 2004.
- [27] J. Bigun. Recognition of Local Symmetries in Gray Value Images by Harmonic Functions. In *Ninth International Conference on Pattern Recognition, Rome*, pages 345–347. IEEE Computer Society Press, November 14–17 1988.
- [28] H. Knutsson, M. Hedlund, and G. H. Granlund. Apparatus for Determining the Degree of Consistency of a Feature in a Region of an Image that is Divided into Discrete Picture Elements. In *US. patent, 4.747.152*, 1988.
- [29] Björn Johansson. *Low Level Operations and Learning in Computer Vision*. PhD thesis, Linköping University, Sweden, Linköping University SE-581 83 Linköping, Sweden, December 2004. Dissertation No. 912, ISBN 91-85295-93-0.
- [30] K. Nilsson and J. Bigun. Localization of corresponding points in fingerprints by complex filtering. *Pattern Recognition Letters*, 24:2135–2144, 2003.
- [31] J. M. Bernardo and M. F. A. Smith. *Bayesian Theory*. Wiley and Son, Chichester, 1994.
- [32] E. S. Bigun. Risk analysis of catastrophes using experts’ judgement: An empirical study on risk analysis of major civil aircraft accidents in europe. *European J. Operational Research*, 87:599–612, 1995.
- [33] J. Bigun, B. Duc, S. Fischer, A. Makarov, and F. Smeraldi. Multi modal person authentication. In H. Wechsler et. al., editor, *Nato-Asi advanced study on face recogniton*, volume F-163, pages 26–50. Springer, 1997.
- [34] P. Simard K. Chellapilla, M. Shilman. Optimally combining a cascade of classifiers. In *Document Recognition and Retrieval XIII*, 15–19 January 2006.
- [35] R. J. Quinlan. *Programs for machine learning*. Morgan Kaufmann Publishers, 1992.
- [36] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance Evaluation of Fingerprint Verification Systems. *IEEE-PAMI*, 28(1):3–18, January 2006.
- [37] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, and S. Janet. Users Guide to Fingerprint Image Software 2 - NFIS2. NIST, 2004.

- [38] D. Simon-Zorita, J. Ortega-Garcia, J. Fierrez-Aguilar, and J. Gonzalez-Rodriguez. Image quality and position variability assessment in minutiae-based fingerprint verification. *IEE Proceedings Vision, Image and Signal Processing, Special Issue on Biometrics on the Internet*, 150(6):402–408, December 2003.