Impact of combining quality measures on biometric sample matching

Krzysztof Kryszczuk, Jonas Richiardi and Andrzej Drygajlo

Abstract— **Biometric matching involves a comparison of two biometric data samples. In practical applications, one or both of the samples may be of degraded quality, in respect to the nominal quality of similar samples acquired in controlled conditions. It has been shown in prior art that in such situations, the integration of quality information into the process of biometric matching can lead to significantly improved classification performance of the biometric matcher. To facilitate such an integration, quality measures originating from both compared biometric samples are usually combined into one quality score. In this paper, we analyze the merit of doing so. We revisit the problem from a pattern classification perspective, and show that using individual quality measures as separate classification features frequently leads to a superior performance of a biometric system in comparison with the system in which quality measures are mapped into one quality score. We provide experimental support of this claim using synthetic data, as well as real biometric database, on the examples of face, fingerprint and multi-modal matching.**

I. Introduction

Biometric sample matching involves a comparison of two biometric samples with the goal of establishing if they originate from the same individuals or from different persons. It constitutes the baseline of biometric identity verification, and is an important concept in forensic science.

Biometric matching systems perform best when the quality of the data samples is comparable and consistent. However, it is a known fact that the conditions of biometric signal acquisition often do not allow such consistency. When external influences, often in conjunction with behavioral factors, change the quality properties of the compared signals, the classification accuracy suffers. The impact of quality degradation on classification performance has been shown to affect all of the prominent biometric modalities. Seeking to address this issue, a number of systems have been proposed, where the quality of the compared signals is explicitly measured, yielding scores known as the biometric quality measures [1], [2], [3].

Biometric sample matching returns only one similarity score. However, each quality measure is derived from one sample, so there are always two quality measures at hand per matching. A prevailing procedure in the literature dictates to combine both individual quality measures into one quality score, which jointly represents the quality of the matched signals. The resulting mapped (combined) quality score has been used in the classification process, using a variety of approaches [4], [5], [6], [7], [8], [9]. These methods have been generalized using a stacking-based framework, *Q-stack* in [10]. Although all methods reported use the mapped quality scores, and not the raw quality measures, remarkably little attention has been paid to the actual mapping function that maps the individual quality measures into one. The problem of the impact of such mapping on classification accuracy also remains unsolved.

In this paper we address the problem of mapping quality measures into one quality score before using them in biometric matching. We argue that mapping of individual quality measures into one quality score comes at a loss of information, potentially detrimental to the classification accuracy. First, we analyze different mapping functions using synthetic data. Then, using real biometric data for face and fingerprint matching, we show that a consistent error reduction is achieved if both individual quality measures appear in the classification process, instead of using only one, combined quality score.

The rest of this paper is structured as follows. In Section II we provide an overview of prior art in mapping quality measures for biometric classification. Section III is devoted to the analysis of the impact of the mapping of two separate quality measures into a joint quality score on class separation. In Section IV we provide experimental support of the claims of this work using real biometric data (face, fingerprint and their fusion). Section V.

II. Mapping quality measures

Biometric sample matching yields a similarity score, used for making a decision whether both samples originate from the same individual or not. Low quality of compared biometric signals is likely to negatively influence the reliability of this decision. In recognition of this fact, quality measures have been used to adjust the score decision threshold [11]. In order to perform such operation, one quality score, which we denote as $\bar{q}$, was used to characterize the compared pair of signals. This single quality measure is a combination of quality estimates q_1 and q_2, derived for both involved biometric samples.

In a multi-biometric system, scores originating from different modalities can be fused to arrive at a final classification decision. Inclusion of the biometric sample quality into the fusion process has been shown to be systematically beneficial to the classification accuracy [1], [8], [7], [10]. Also in these systems, only one quality measure is used per similarity score. That quality measure is either a combination of quality scores originating from both compared samples, or it comes from only one of the compared samples (testing), assuming

K. Kryszczuk is with the IBM Zurich Research Laboratory, Ruschlikon, Switzerland kkr@ibm.zurich.com

J. Richiardi is with PatternLab Sarl, Lausanne, Switzerland jrichiardi@patternlab.ch

A. Drygajlo is with the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland andrzej.drygajlo@epfl.ch

978-1-4244-5020-6/09/$25.00 ©2009 IEEE

that the other one (gallery) has been collected at a high, controlled quality [6].

Combining quality measures from two compared samples into one has become a de-facto standard in biometric matching. Grother and Tabassi [12] combine quality scores for two compared fingerprints using $\bar{q} = min(q_1, q_2)$, arguing that in "positive applications" such as identity verification, high quality of one of the samples can always be assured during the enrolment phase. An assumption that the quality of the gallery image is high may not hold: for instance, certain people are consistently leaving low-quality fingerprint impressions. In the case of a forensic application of biometric sample matching, such an assumption cannot be made at all.

Grother and Tabassi also mention the possibility of using $\bar{q} = \sqrt{q_1 q_2}$, $\bar{q} = |q_1 - q_2|$, $\bar{q} = \frac{1}{2}(q_1 + q_2)$ for combining quality scores. Geometric mean, $\bar{q} = \sqrt{q_1 q_2}$, is a frequently used way of mapping individual quality scores into one, [4], [7], [1], albeit without a sound justification for doing so. The same method has been implicitly used by Chen et al. [2] to weight the importance of bit-wise disagreement between two iris codes.

Nandakumar et al. use two modality-specific schemes of mapping quality measures to one quality score [8]. When matching fingerprints, for each involved quality factor the values pertinent to both matched samples are summed. As such, the final quality measure is an elaborate form of averaging multiple quality measures coming from individual samples. For iris matching, the joint quality score is defined as a correlation coefficient between two quality vectors. So defined, the joint quality measure encodes the quality differences between two irises, and not their individual quality properties.

All the mentioned approaches have in common that they use one joint descriptor of the quality of both matched samples. However, biometric sample matching involves a comparison of two samples, each having its own quality properties. Collapsing two individual quality measures q_1 and q_2 into one $\bar{q}$ must come at a loss of information. We argue that this loss of information is reflected in reduced separability between classes, and consequently in higher error rates than those that can be attained when the quality information is not mapped into one number.

III. The impact of quality measure combination on class separability

Let us consider the matching of two biometric samples, S_1 and S_2, resulting in a similarity score x. Assume we perform independent quality measurements on S_1 and S_2, yielding non-negative quality measures q_1 and q_2, accordingly. These quality measures can be used directly in the classification process. Alternatively, they can be combined into one quality score using some mapping function f: $\bar{q} = f(q_1, q_2)$. The combined quality score $\bar{q}$ is then used in the classification process.

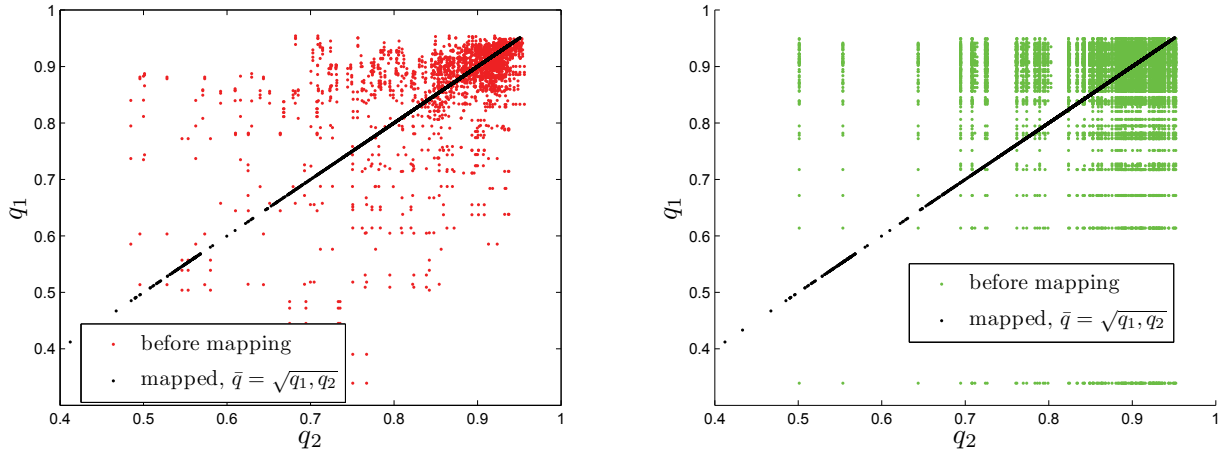
The function $f(q_1, q_2)$ can be of arbitrary nature, but for practical reasons it is usually a symmetric function, $f(q_1, q_2) = f(q_2, q_1)$. As a result, the application of f projects all data points $[x, q_1, q_2]$ onto the same mapping plane that bisects the classification space defined by x, q_1 and q_2. Figure 1 shows an orthogonal projection of the mapping plane onto q_1 and q_2, for $f = \sqrt{q_1 q_2}$. The mapping f reduces the degrees of freedom of the classifier deployed to distinguish between classes. The mapping of $[q_1, q_2] \mapsto \bar{q}$ cannot be reversed in all cases but when $q_1 = q_2$, inadvertently causing an information loss in the system.

Before mapping, the separation E_N between classes A and B can be expressed in terms of Bayes error computed in the evidence space $e_N = [x, q_1, q_2]$. After mapping, the class separation E_C can be computed in the mapped evidence space of $e_C = [x, f(q_1, q_2)]$. We chose to approximate E_N and E_C using numerical simulations. A large number of simulated data samples ensures an arbitrarily close approximation of the true Bayes error. Consequently, the difference $\Delta_E = E_C - E_N$ gives the estimate of the gain that can be expected from using q_1 and q_2 for classification, instead of using $\bar{q} = f(q_1, q_2)$.

In the simulations we have been randomly drawing 10^3 samples per class from class-conditional normal joint distributions of $e_N = [x, q_1, q_2]$. Marginal variances of x, q_1, q_2 were set to unity, and q_1, q_2 were designed as conditionally relevant to x, $p(q_1, q_2|A) = p(q_1, q_2|B)$. As shown in [10], [13], the dependencies between conditionally relevant quality measures and the matching scores x are critical for observing better class separation in the evidence space e_N in respect to that obtained using matching scores alone. We simulate these dependencies by controlling the correlation coefficients ρ_1 and ρ_2, between x and q_1, and between x and q_2, respectively. The difference between the means of marginal class-conditional distributions of x was set to 1.

The data used to compute E_C was derived from the $[x, q_1, q_2]$, by applying the mapping function $\bar{q} = f(q_1, q_2)$. The resulting distributions are given by $e_C = [x, f(q_1, q_2)]$. We have compared the following formulations of $f(q_1, q_2)$: $\bar{q} = \sqrt{q_1 q_2}$, $\bar{q} = |q_1 - q_2|$, $\bar{q} = min(q_1, q_2)$ and $\bar{q} = \frac{1}{2}(q_1 + q_2)$. The results of the simulations in terms of $\Delta_E = E_C - E_N$ are given in Figure 2.

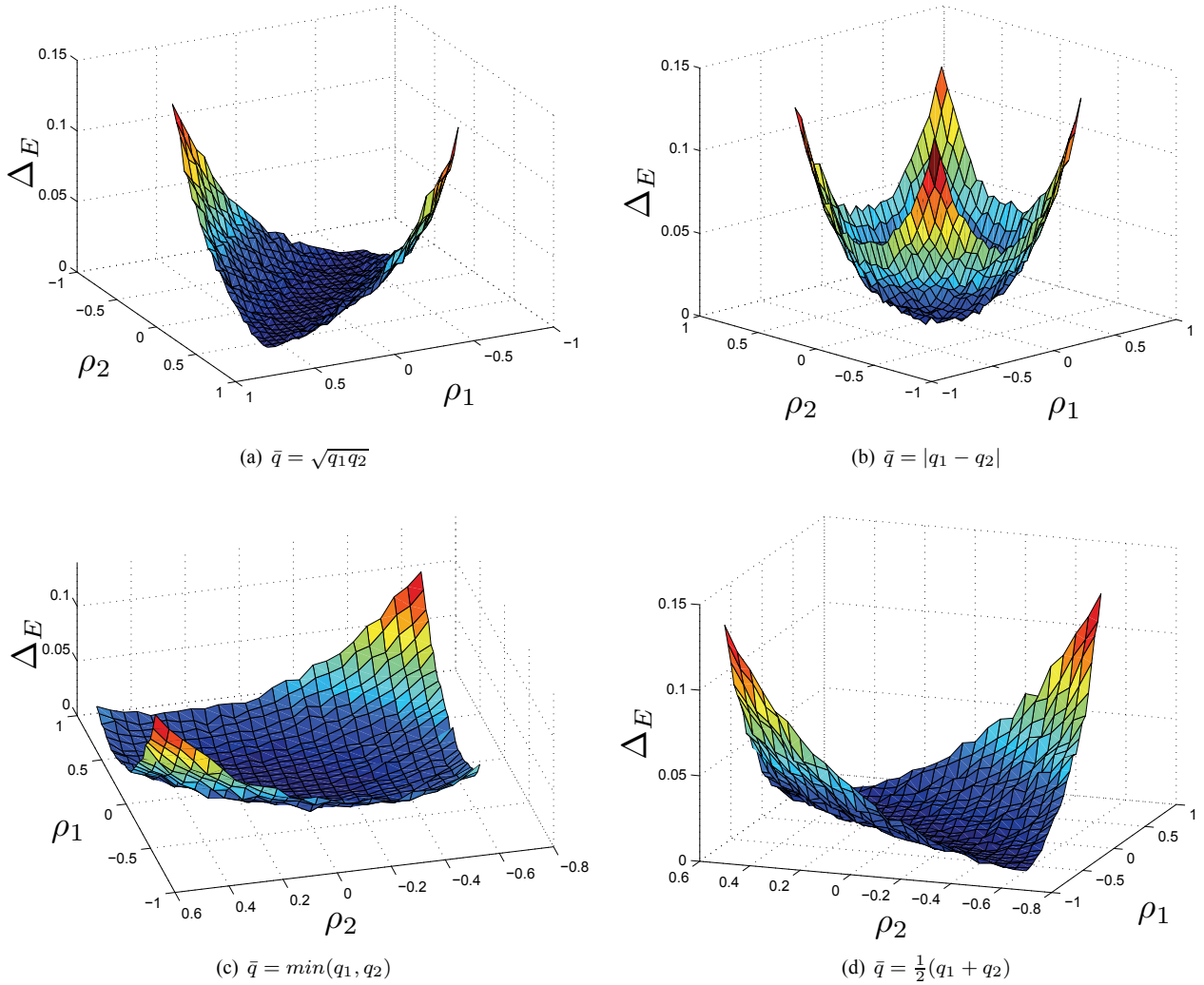
The exclusively non-negative values of Δ_E confirm that no mapping function $f(q_1, q_2)$ resulted in an increase of class separation in respect to that observed for non-mapped quality measures q_1 and q_2. All reported experiments showed $\Delta_E \approx 0$ for $\rho_1 \approx 0 \approx \rho_2$, where no gain is expected from the use of conditionally-relevant quality measures anyway. In the case of $\bar{q} = \sqrt{q_1 q_2}$ and $\bar{q} = \frac{1}{2}(q_1 + q_2)$ we observe small values of Δ_E for $\rho_1 \approx \rho_2$. This result is no surprise - if q_1 and q_2 are highly correlated with x then they are also correlated with each other. Consequently, $q_1 \approx q_2$, and in this case the arithmetical and geometrical mean are both good representations that can replace q_1 and q_2. However, in biometric applications the main difficulty arises when the two matched signals are of different quality. In this case, the simulations show that using q_1 and q_2 explicitly is a better solution than using a mapped quality score $\bar{q}$.



(a) class A (genuine)

(b) class B (imposter)

Fig. 1. Effects of the mapping of quality measures. Effects of multiple re-use of the same samples as imposters clearly visible in (b).



(a) $\bar{q} = \sqrt{q_1 q_2}$

(b) $\bar{q} = |q_1 - q_2|$

(c) $\bar{q} = min(q_1, q_2)$

(d) $\bar{q} = \frac{1}{2}(q_1 + q_2)$

Fig. 2. Comparative error differences between systems using mapped and non-mapped quality measures for classification, as a function of the correlation coefficients ρ_1 and ρ_2 between scores x and quality measures q_1 and q_2. Results obtained using synthetic data.

TABLE I

CLASSIFICATION RESULTS FOR BASELINE SYSTEMS (NO QUALITY MEASURES), ERRORS EXPRESSED IN [%]

	er_A	er_B	$HTER$
face, x_f	14.43	14.38	**14.41**
fingerprint, x_p	0.98	0.94	**0.96**
SVM fusion, $[x_f, x_p]$	0.37	0.40	**0.38**

IV. EXPERIMENTAL EVALUATION

The experimental evaluation involved one-to-one matching of face and fingerprint images from the Biosec database, baseline corpus [14], which contains data collected from 200 subjects. The Biosec database provides 15600 comparisons (matchings), out of which 3200 are genuine, and 12400 are imposter comparisons. The experiments reported here involved face matching using the $DCTmod2$ features, modeled using a Gaussian Mixture Model-based classifier [15]. The fingerprints were processed and matched using the publicly available $NFIS$ system [16]. Since the details of the baseline biometric matchers are not critical for the conclusions of this paper, we decide to skip them for brevity. The performance of the baseline matchers on separating genuine from imposter scores, obtained without the use of quality measures, is given in Table I, in terms of class errors er_A and er_B, and Half-Total Error Rate, $HTER = \frac{1}{2}(er_A + er_B)$. Trained fusion of face and fingerprint scores was performed using a linear SVM classifier.

We have used two face image quality measures, denoted as q^{f1} and q^{f2}. The first face quality measure was the 2D Pearson's correlation coefficient with an average face template [17]. The average face template $\bar{\Gamma}$, measuring $l_x \times l_y$ pixels, was created using PCA reconstruction from images not included in the baseline corpus of the Biosec database. Given a face image Γ_0, the quality measure was computed as

$$q^{f1} = \frac{\Sigma_{i=1}^{l_x l_y - 1}(p_{\Gamma_0} - \mu_{\Gamma_0})(p_{\bar{\Gamma}} - \mu_{\bar{\Gamma}})}{(l_x l_y - 1)\sigma_{\Gamma_0}\sigma_{\bar{\Gamma}}}, \quad (1)$$

where μ_{Γ_0} and σ_{Γ_0} are the mean and variance of all pixels in Γ_0, and $\mu_{\bar{\Gamma}}$ and $\sigma_{\bar{\Gamma}}$ are mean and variance of all pixels in $\bar{\Gamma}$.

As the second quality measure q^{f2}, we used the average image sharpness estimator [17], computed as

$$q^{f2} = \frac{1}{2}\left(\frac{\Sigma_{m=1}^{l_y}\Sigma_{n=1}^{l_x-1}|p_{n,m} - p_{n+1,m}|}{(l_x - 1)l_y} + \frac{\Sigma_{m=1}^{l_y-1}\Sigma_{n=1}^{l_x}|p_{n,m} - p_{n,m+1}|}{l_x(l_y - 1)}\right), \quad (2)$$

where $p_{n,m}$ is the pixel value.

We have employed two fingerprint quality indices, proposed by Chen et al. [3]. The first index measures the energy concentration in the frequency domain as a global feature. The second index measures the spatial coherence in local regions.

A. Experimental protocol and results

The goal of the experiments was to demonstrate that the loss of information associated with the mapping of two distinct quality measures q_1 and q_2 onto one quality score $\bar{q}$ negatively impacts the utility of the quality information in biometric matching. For this purpose, we used the framework of *Q-stack* as a platform for the comparison [10]. For each modality's matcher and quality measure, and their combinations, we ran two concurrent experiments: in the first experiment, a baseline score x was used together with the joint quality score $\bar{q}$, in a similar fashion as in [10]. In the second, concurrent experiment, we have used the same x but this time together with original quality measures q_1 and q_2, without mapping.

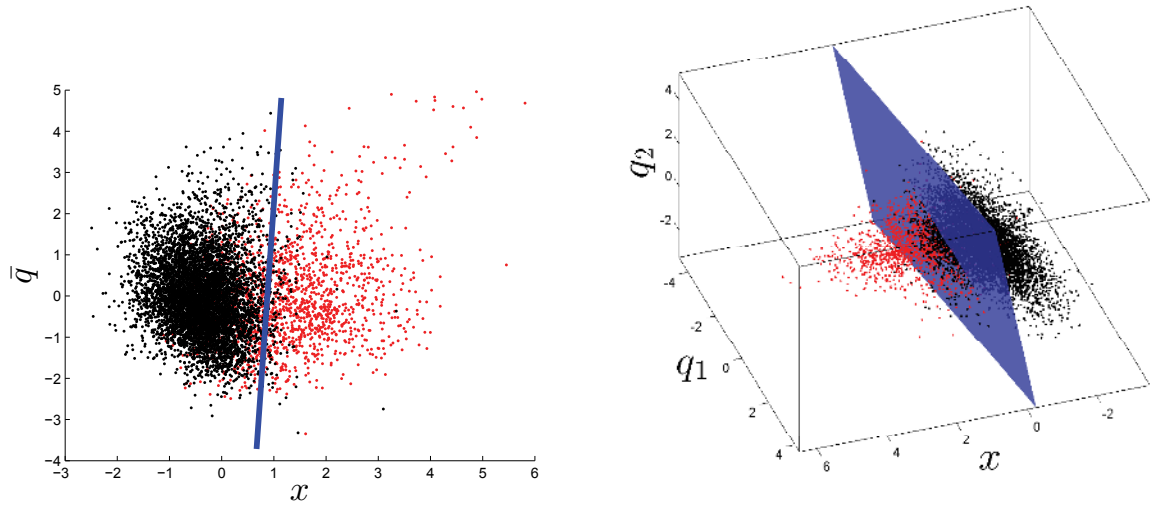
For each comparison we trained one stacked classifier in the evidence space $[x, \bar{q}]$, and another one in the evidence space defined by $[x, q_1, q_2]$. We used $\bar{q} = \sqrt{q_1, q_2}$ as the mapping functions since it is the most commonly used mapping seen in the prior art (Section II), and it also emerges as the best choice of the different mappings considered in Section III. Each trained stacked classifier was a SVM classifier with a linear kernel. A sample of the compared pairs of experiments is shown in Figures 3 and 4. Here, the stacked classifiers are shown in blue color.

In our experiments, we have considered biometric matching using both baseline classifiers alone and multimodal fusion matching using face and fingerprint. The combinations of baseline classifiers and quality measures used in the experiments are summarized in Table II.

Our experiments involved 50-fold cross-validation with random sub-sampling, where 50% of the genuine and imposter matches were used for system training, and the remaining 50% were used for testing. For each run of the experiment, 1600 genuine matches and 6200 imposter matches were used for training, and the same number for testing. The training and testing sets were disjoint at each of the cross-validation runs. A comparison of the mean classification errors after 50-fold cross-validation, for each considered pair of mapped versus non-mapped quality measures, is shown in Table III, in terms of class errors er_A, er_B, and $HTER$.

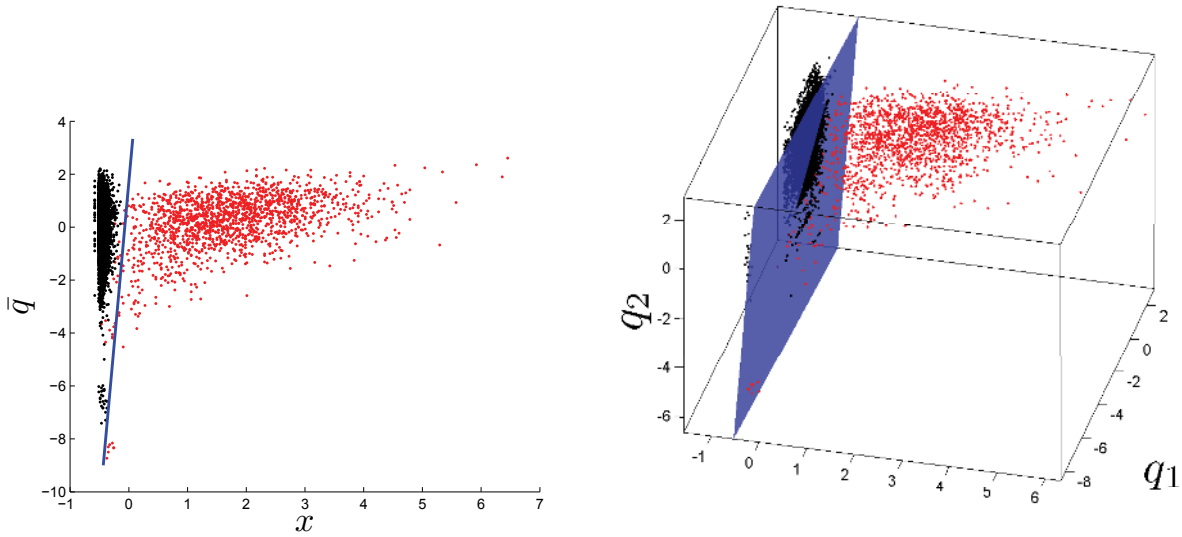
The results obtained using non-mapped quality measures are marked in bold in Table III. The absolute and relative mean differences between $HTER_1$ (mapped $\bar{q}$) and $HTER_2$ (non-mapped q_1, q_2) are shown in columns δ and Δ. The results of a one-sided KS-2 test of the hypothesis that $HTER_1 > HTER_2$, computed over 50 cross-validation runs, are given, together with the corresponding p value.

In all experiments reported, the classification errors were reduced, in terms of class errors and $HTER$, by using original pairs of quality measures q_1 and q_2 instead of a mapped quality score $\bar{q}$, in spite of contrary suggestions from prior art. In all but two experiments, the improvements are statistically significant at $p = 0.05$, according to one-sided KS-2 test. Additionally, the results show that mapping quality measures into one score $\bar{q}$ can render quality measures less, or even not useful for classification. For instance, the baseline



(a) mapped: $x, \bar{q} = \sqrt{q_1, q_2}$ (b) not mapped: x, q_1, q_2

Fig. 3. Comparison of classification using a linear SVM stacked classifier in the evidence spaces containing mapped ($\bar{q} = \sqrt{q_1, q_2}$), versus non-mapped ($q_1$, q_2) quality measures, and baseline classifier scores x, face matching.



(a) mapped: $x, \bar{q} = \sqrt{q_1, q_2}$ (b) not mapped: x, q_1, q_2

Fig. 4. Comparison of classification using a linear SVM stacked classifier in the evidence spaces containing mapped ($\bar{q} = \sqrt{q_1, q_2}$), versus non-mapped ($q_1$, q_2) quality measures, and baseline classifier scores x, fingerprint matching

$HTER = 14.41$ (Table I) is lower than the $HTER$ obtained using x_f and mapped quality scores in Exp. 1 and 3. It is only slightly higher than in Exp. 2. This shows that the mapped quality measures were not useful as classification features. In contrast, in all Exp. 1, 2 and 3 the use of original, non-mapped quality measures resulted in systematic, statistically significant reduction of error rates.

V. Conclusions

In this paper we revisited the common practice of combining quality measures coming from compared biometric signals into one quality score before classification. We provided theoretical and experimental support for the claim that the mapping of quality measures into one quality score inadvertently causes a loss of information and a reduction of the classifier's degrees of freedom, where conditional dependencies between quality measures and baseline scores could be exploited. Using synthetic and real biometric data, we demonstrated that the use of original, non-mapped quality measures systematically leads to significant improvements of classification performance over systems that use mapped quality scores.

TABLE II

EVIDENCE COMBINATIONS USED IN THE EXPERIMENTS. RESULTS OF THE EXPERIMENTS ARE GIVEN IN TABLE III.

Exp.	Evidence: mapped $q_1, q_2 \mapsto \bar{q}$	Evidence: non-mapped q_1, q_2
1	$x_f, \sqrt{q_1^{f1} q_2^{f1}}$	x_f, q_1^{f1}, q_2^{f1}
2	$x_f, \sqrt{q_1^{f2} q_2^{f2}}$	x_f, q_1^{f2}, q_2^{f2}
3	$x_f, \sqrt{q_1^{f1} q_2^{f1}}, \sqrt{q_1^{f2} q_2^{f2}}$	$x_f, q_1^{f1}, q_2^{f1}, q_1^{f2}, q_2^{f2}$
4	$x_p, \sqrt{q_1^{p1} q_2^{p1}}$	x_p, q_1^{p1}, q_2^{p1}
5	$x_p, \sqrt{q_1^{p2} q_2^{p2}}$	x_p, q_1^{p2}, q_2^{p2}
6	$x_f, \sqrt{q_1^{f1} q_2^{f1}}, \sqrt{q_1^{f2} q_2^{f2}}$	$x_f, q_1^{f1}, q_2^{f1}, q_1^{f2}, q_2^{f2}$
7	$x_p, x_f, \sqrt{q_1^{f1} q_2^{f1}}, \sqrt{q_1^{p1} q_2^{p1}}$	$x_p, x_f, q_1^{f1}, q_2^{f1}, q_1^{p1}, q_2^{p1}$
8	$x_p, x_f, \sqrt{q_1^{f1} q_2^{f1}}, \sqrt{q_1^{p2} q_2^{p2}}$	$x_p, x_f, q_1^{f1}, q_2^{f1}, q_1^{p2}, q_2^{p2}$
9	$x_p, x_f, \sqrt{q_1^{f2} q_2^{f2}}, \sqrt{q_1^{p1} q_2^{p1}}$	$x_p, x_f, q_1^{f2}, q_2^{f2}, q_1^{p1}, q_2^{p1}$
10	$x_p, x_f, \sqrt{q_1^{f2} q_2^{f2}}, \sqrt{q_1^{p2} q_2^{p2}}$	$x_p, x_f, q_1^{f2}, q_2^{f2}, q_1^{p2}, q_2^{p2}$
11	$x_p, x_f, \sqrt{q_1^{f1} q_2^{f1}}, \sqrt{q_1^{p1} q_2^{p1}}, \sqrt{q_1^{f2} q_2^{f2}}, \sqrt{q_1^{p2} q_2^{p2}}$	$x_p, x_f, q_1^{f1}, q_2^{f1}, q_1^{p1}, q_2^{p1}, q_1^{f2}, q_2^{f2}, q_1^{p2}, q_2^{p2}$

TABLE III

EXPERIMENTAL RESULTS USING DATA FROM BIOSEC DATABASE, ERRORS EXPRESSED IN [%]. THE EVIDENCE COMBINATION USED IN EACH EXPERIMENT PAIR IS LISTED IN TABLE II. THE COLUMNS TITLED $\delta = HTER_1 - HTER_2$, $\Delta = \frac{\delta}{HTER_1}$, KS-2 IS THE RESULT OF THE ONE-SIDED KOLMOGOROV-SMIRNOV TEST OF THE HYPOTHESIS THAT $HTER_1 > HTER_2$, COMPUTED OVER 50 CROSS-VALIDATION RUNS, AT THE $p = 0.05$ VALUE.

	mapped, $\bar{q} = \sqrt{q_1 q_2}$			non-mapped, q_1, q_2						
Exp.	er_A	er_B	$HTER_1$	er_A	er_B	$HTER_2$	δ	Δ	p	KS-2
face, DCT										
1	14.72	14.70	14.71	12.67	12.72	**12.69**	2.02	13.70	1E-23	1
2	14.12	14.52	14.32	13.6	13.41	**13.5**	0.82	5.69	3E-20	1
3	14.64	14.51	14.58	12.56	12.38	**12.47**	2.11	14.45	1E-23	1
fingerprint										
4	0.363	0.355	0.359	0.351	0.34	**0.345**	0.013	3.75	0.354	0
5	0.495	0.534	0.514	0.38	0.415	**0.397**	0.117	22.72	2E-15	1
6	0.365	0.345	0.355	0.32	0.316	**0.318**	0.037	10.33	9E-3	1
fusion, face+fingerprint										
7	0.203	0.185	0.194	0.189	0.19	**0.19**	0.004	2.21	0.467	0
8	0.172	0.171	0.172	0.14	0.162	**0.151**	0.02	11.84	0.016	1
9	0.235	0.244	0.239	0.214	0.217	**0.216**	0.024	9.93	0.016	1
10	0.216	0.176	0.196	0.182	0.166	**0.174**	0.022	11.19	1E-05	1
11	0.202	0.209	0.206	0.165	0.117	**0.141**	0.064	31.31	5E-09	1

REFERENCES

[1] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, J. Gonzalez-Rodriguez, H. Fronthaler, K. Kollreider, and J. Bigun, "A comparative study of fingerprint image-quality estimation methods," *IEEE Trans. on IFS*, vol. 2, no. 4, 2007.

[2] Y. Chen, S. C. Dass, and A. K. Jain, "Localized iris image quality using 2-D wavelets," in *Proc. of the ICB*, 2006.

[3] Y. Chen, S. Dass, and A. Jain, "Fingerprint quality indices for predicting authentication performance," in *Proc. of AVBPA*, Rye Brook, NY, July 2005, pp. 160–170.

[4] J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez, "Multimodal biometric authentication using quality signals in mobile communications," in *Proc. 12th International Conference on Image Analysis and Processing*, Mantova, Italy, 2003.

[5] K. A. Toh, W.-Y. Yau, E. Lim, L. Chen, and C.-H. Ng, "Fusion of auxiliary information for multi-modal biometrics authentication," in *Proc. of the ICB*, Hong Kong, 2004, pp. 678–685.

[6] J. Fierrez-Aguilar, "Adapted fusion schemes for multimodal biometric authentication," Ph.D. dissertation, Universidad Politecnica de Madrid, 2006.

[7] J. Fierrez-Aguilar, Y. Chen, J. Ortega-Garcia, and A. K. Jain, "Incorporating image quality in multi-algorithm fingerprint verification," in *Proc. of the ICB*, Hong Kong, 2006.

[8] K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain, "Quality-based score level fusion in multibiometric systems," in *Proceedings of ICPR*, vol. 4, Hong Kong, China, August 2006, pp. 473–476.

[9] N. Poh and J. Kittler, "A family of methods for quality-based multimodal biometric fusion using generative classifiers," in *Proc. of the 10th ICARCV*, Hanoi, 2008, pp. 1162–1167.

[10] K. Kryszczuk and A. Drygajlo, "Improving classification with class-independent quality measures: Q-stack in face verification," in *Proc. of the ICB*, Seoul, Korea, 2007.

[11] L. Wein and M. Baveja, "Using fingerprint image quality to improve the identification performance of the U.S. VISIT program," in *Proc. of the National Academy of Sciences, 2005*, 2005.

[12] P. Grother and E. Tabassi, "Performance of biometric quality measures," *IEEE Trans. on PAMI*, vol. 29, no. 4, pp. 531–543, April 2007.

[13] K. Kryszczuk and A. Drygajlo, "Impact of feature correlations on separation between bivariate normal distributions," in *Proc. of the ICPR'08*, 2008.

[14] J. Fierrez, J. Ortega-Garcia, D. Torre-Toledano, and J. Gonzalez-Rodriguez, "BioSec baseline corpus: A multimodal biometric database," *Pattern Recognition*, vol. 40, no. 4, pp. 1389–1392, April 2007.

[15] C. Sanderson, "Automatic person verification using speech and face information," Ph.D. dissertation, Griffith University, Queensland, Australia, 2002.

[16] C. I. Watson, M. D. Garris, E. Tabassi, C. L. Wilson, R. M. McCabe, and S. Janet, "User's guide to NIST fingerprint image software 2," NIST, Tech. Rep. NISTIR 6813, November 2004.

[17] K. Kryszczuk and A. Drygajlo, "On face quality measures," in *Proc. 2nd Workshop on Multimodal User Authentication MMUA'06*, Toulouse, France, May 2006.