
Chapter 1

**Iris Image Quality Metrics with Veto Power and
Nonlinear Importance Tailoring**

John Daugman, Cathryn Downing

Linear combinations of metrics for assessing biometric sample quality are weak, because they lack veto power. For example, a good score for sharp focus of an ocular image would ‘compensate’ in an additive combination for the fact that the eyelids are fully closed; or fully open eyelids would compensate for the image being many diopters out-of-focus. Normalised multiplicative quality factors are better because they are punitive, and thereby confer veto powers. This chapter explains the basis for the product of power functions which underlie the ISO/IEC 29794-6 Iris Image Sample Quality Standard, in particular how the exponents of the power functions allow importance tailoring of each element.

1.1 Introduction

US National Institute of Standards and Technology (NIST) papers by Grother and Tabassi [1], and also by Phillips and Beveridge [2], defined *quality measures* as measurable covariates that are both predictive of biometric recognition performance, and actionable. Quality measures include *subject covariates* that are attributes of a person and which may be transient, such as expression, eyelid occlusion, or the wearing of eyeglasses. Quality measures also include *image covariates* that depend on the sensor and acquisition conditions, such as focus, resolution, and illumination effects. Some subject covariates may not be capable of change or improvement, such as permanent injury or physical deformation; other subject covariates may be improvable but not under voluntary control, such as the degree of pupil dilation. In this chapter we do not explore those distinctions further, but rather our concern is how best to combine multivariate quality measures into one actionable quantity: e.g. to decide whether to enroll a given biometric sample into a database, or to reject it and acquire a new one. In the context of having a plurality of measures, this amounts to computing a single actionable scalar – a *Quality Score QS* – from a vector consisting of several elements. We thus explore a new definition resembling a norm, or rather a semi-norm, of a biometric quality vector. Because of the resemblance, we call it a *pseudonorm*. Its attributes include the ability to treat the different elements of the vector differently and to control their compensatory trade-offs. We construct

2 Iris Image Quality

and demonstrate an algorithmic procedure for determining the parameters of this pseudonorm on quality vectors, using a public test database of iris images.

1.1.1 Related Work

The most comprehensive treatment of *quality-based fusion* – incorporating quality measures together with comparison scores for multi-biometric classifier combination to achieve optimal fusion and decisions within a general Bayesian framework – was by Poh and Kittler [3]. A related proposal for comparator fusion based on quality in multibiometric systems was by Nandakumar *et al.* [4]. They showed that overall performance of multibiometric systems could be improved by dynamically assigning weights to the outputs of individual comparators, using sample quality, in a likelihood-ratio fusion scheme for combining their comparison scores.

Schmid and her co-authors [5] proposed an adaptive iris authentication approach based on quality measures that were selected for their utility in separating genuine from impostor score distributions, in labelled data. A feedforward neural network (two hidden layers, having 16 neurons and 2 neurons) was trained to learn a nonlinear mapping of the measures onto an overall QS, with good results. But it is unknown what actual nonlinear combinations developed in the hidden layers. In other work by Schmid *et al.* [6], the Dempster-Shafer theory of evidence was deployed as a basis for combining quality factors.

Another adaptive use of quality metrics was demonstrated by Li *et al.* [7] for the dynamic selection of the actual recognition strategy. They fused several quality factors into an overall score by products of likelihood ratios, using conditional probabilities of score values given good, versus given poor, quality. The samples may then be processed differently, based on their quality. Belcher and Du [8] used quality measures to calculate a confidence level associated with a comparison score, an idea also present in [5]. Galbally *et al.* [9] proposed that quality related features could be used for liveness detection in an anti-spoofing strategy.

Hofbauer *et al.* [10] even proposed that quality metrics themselves could be used directly for comparing iris images and estimating their similarities. Phillips and Beveridge [2] argued that biometric comparison and quality assessment are actually equivalent, in a formal sense which they call ‘biometric completeness’. Their argument is that a perfect method for quality assessment would predict whenever a one-to-one verification algorithm would give the wrong result. The wrong decision could then be reversed, producing error-free verification.

1.2 Normalisation of Individual Quality Measures

We assume that individual quality measures are non-negative numbers which might have no inherent upper bound, with larger values implying higher quality. We are agnostic here about what they actually measure, but we will illustrate using actual examples. Some measures may already be inherently normalised, such as those that are defined as a percentage. But the goal of our *unity-based normalisation* is not

merely to re-scale, but also to embed a semantics of acceptability and flexibility. We map any measure x onto the $[0, 1]$ unit interval using a sigmoidal function $f(x)$ having two parameters: n controlling the measure’s rigidity or flexibility, and c specifying a ‘set point’ at which the measure is deemed to correspond to 50% quality:

$$f(x) = \frac{x^n}{x^n + c^n} \tag{1.1}$$

Parameter c has the same units as x , whether those be dimensionless (such as the percentage of the iris that is visible between the eyelids), or dimensional (such as square-millimetres of visible iris area). The choice of parameter n should reflect how ‘non-negotiable’ are the demands on x , with larger n meaning greater sensitivity to small differences in the value of x around c . For the measures actually used, we find that a soft $n = 2$ is often a better choice than larger values which begin to effect a rigid ‘brick wall’ threshold. It is easily shown that the slope of $f(x)$, its sensitivity to differences, at the 50% point ($x = c$) is $n/4c$. A family of curves representing such functions for $n \in \{2, 4, 6, 10, 30\}$ are shown in Figure 1.1. Once a quality measure x is normalised onto the $[0, 1]$ unit interval and shaped in this way for appropriate sensitivity to differences, we call it a *quality factor* because it will combine into the overall QS product as a multiplicative factor. (Later we shall tailor the relative importance of quality factors, using empirically tuned power functions of them.)

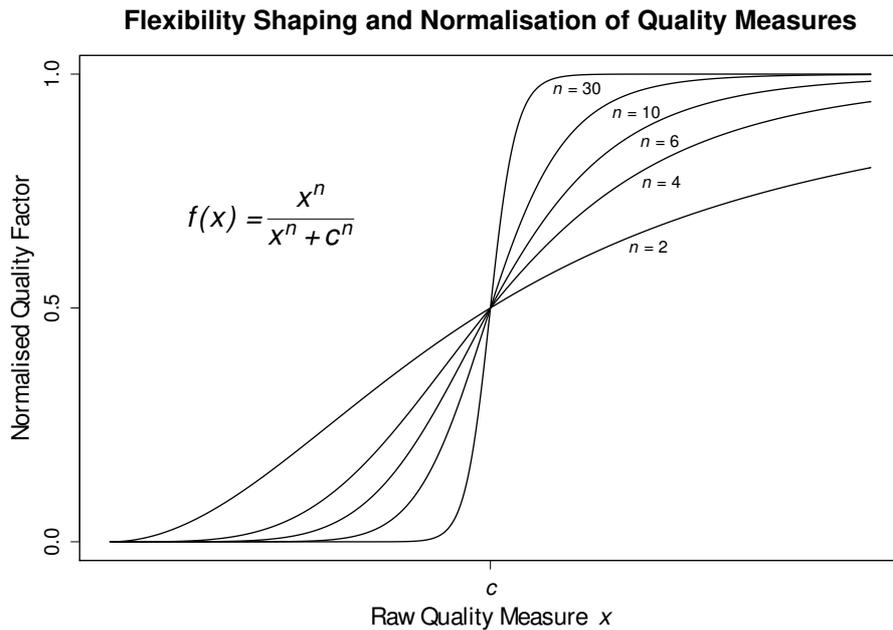


Figure 1.1: Family of functions normalising quality measures into quality factors, and shaping their sensitivity to differences. Larger n means greater rigidity, and the value $x = c$ is the quality factor halfway point.

4 Iris Image Quality

1.3 Effectiveness of Multiplying Quality Factors

We illustrate the nonlinear product approach now by simply multiplying together several quality factors that have predictive value for iris recognition performance. Clearly they each acquire ‘veto power’ in this framework, because if any one of the factors is zero the overall product becomes zero, no matter how good are the others. Conversely, once a quality factor rises enough on its particular sigmoid (Figure 1.1) to approach 1.0, it ceases to matter. Thus, our framework is essentially *punitive*.

Selected quality factors

1. **Sharpness** is computed by 2D Fourier methods as detailed in [11], [12]. This computation is extremely quick, executing much faster than the video frame rate, because it precedes image segmentation. It takes less than a millisecond because convolutions are implemented entirely by pixel additions and subtractions, without kernel multiplications. Thus image focus can be assessed for video frame selection in real-time.
2. **Motion blur** is also computed by Fourier methods on pre-segmented images, within a millisecond. It examines local image structure across the temporal interval between the two fields of a (non-progressive scan) video frame, detecting any interlace shear between even and odd lines.
3. **Texture energy** is a measure of iris contrast, signal-to-noise ratio, and salience, based on the distribution of 2D Gabor wavelet coefficients [11] computed when encoding the iris pattern into an IrisCode. The lowest quartile (in absolute value) of this distribution of coefficients are ignored as being unreliable, when the IrisCodes are compared.
4. **Gaze angle** is estimated as per [13] using Fourier-based trigonometry. Gaze on-axis with the camera’s optical axis is preferred, as the iris is not strictly planar so simple affine transformation to compensate for off-axis gaze is imperfect.
5. **Shaped pupil contrast** is a post-segmentation measure of image focus. Once the actual shape of the pupil has been estimated by a Fourier series expansion [13] of its boundary, a nonlinear Weber contrast is computed using a series of contour integrals that follow various dilations of this shape, ranging from inside to outside the pupil.
6. **Limbus contrast** also integrates various dilations of the outer boundary of the iris once its shape has been estimated by a Fourier series expansion [13]. But unlike the pupil contrast estimate, which is both edge-based and region-based (because the inside of the pupil should be dark), this outer boundary factor uses a linear Michelson contrast measure.

7. **IrisCode entropy** is an estimate of how much useful information the image contains. It takes into account the percentage of iris area that is not occluded by the eyelids, the total number of bits that are unmasked (deemed to be uncorrupted, for example by eyelashes or corneal specular reflections), and it penalises excessive pupil dilation.

For the publically available images in the NIST *Iris Challenge Evaluation* (ICE) database, we computed quality factors using five of these measures and multiplied them together to obtain a provisional quality score for every image. We then studied how these scores predicted Hamming distance for all possible ‘probe’ and ‘gallery’ pairings of images that came from a given eye and thus ought to match. Distances are plotted as a surface function of the quality pairings (after quantisation into ten bins for each) in Figure 1.2, showing a clear bivariate effect. Figure 1.3 shows the effect of these scores on False non-Match Rates, plotted both as a surface and as a contour map, when using a 0.32 Hamming distance match threshold.

Image Quality as a Predictor of Hamming Distance

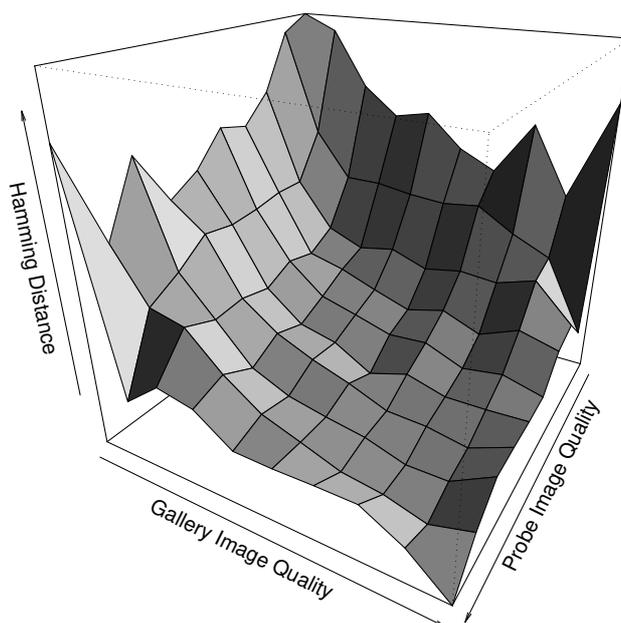


Figure 1.2: Effect of combined image quality factors on same-eye comparison scores. Hamming distance is the fraction of bits that disagree when two IrisCodes are compared [11]. Poor image quality elevates Hamming distances [14].

6 Iris Image Quality

Image Quality as a Predictor of False non-Match Rate

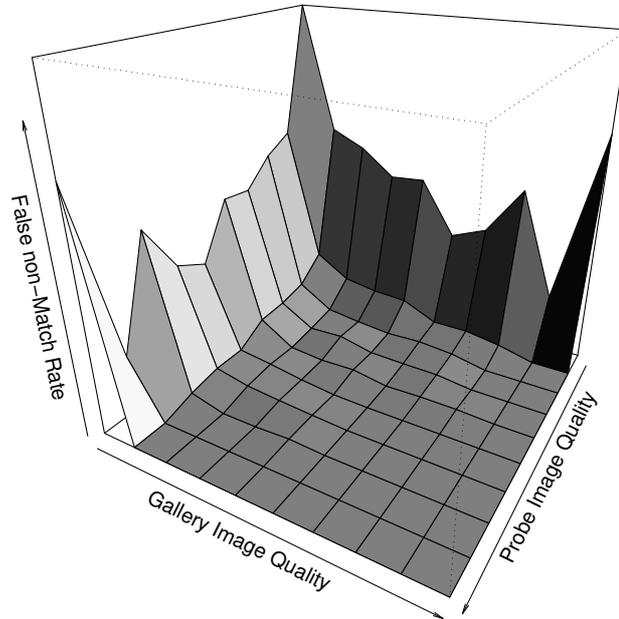


Image Quality as a Predictor of False non-Match Rate

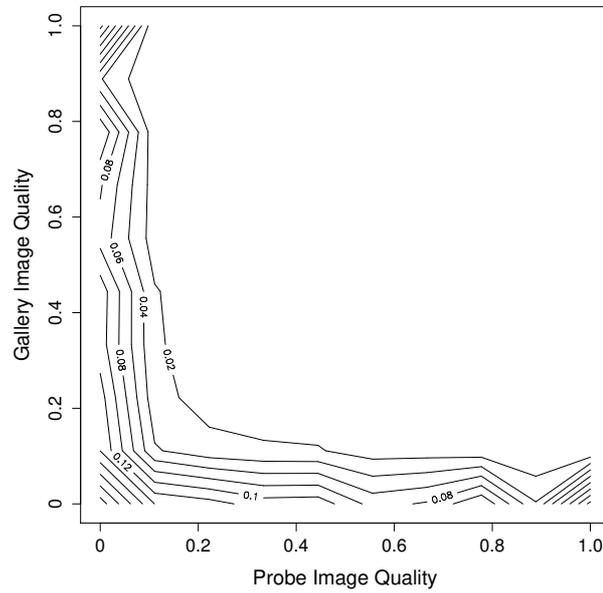


Figure 1.3: Effect of combined image quality factors on False non-Match Rates, plotted as surface and contour maps, at a 0.32 Hamming distance threshold [14].

1.4 Importance Tailoring

We have seen in Figures 1.2 and 1.3 that multiplying quality factors together can produce a strong predictor of match performance. This punitive approach gives veto power to every factor, but intuitively, not every veto should necessarily be equally insistent. Therefore we wish to tailor the relative importance of the quality factors by raising them to different powers, using exponents determined by an empirical multivariate regression procedure.

Using various quality factors X, Y, Z, \dots , we define an overall Quality Score QS as the product of *power functions* of them, with positive exponents $\alpha, \beta, \gamma, \dots$:

$$QS = X^\alpha Y^\beta Z^\gamma \dots \quad \text{for } \{\alpha, \beta, \gamma, \dots\} > 0 \tag{1.2}$$

Component quality factors X, Y, Z, \dots are the normalised functions defined in (1.1). Figure 1.4 shows that exponents larger than $\alpha = 1$ make a quality factor stringent, whereas small exponents make it complacent because it rises to 1.0 very quickly.

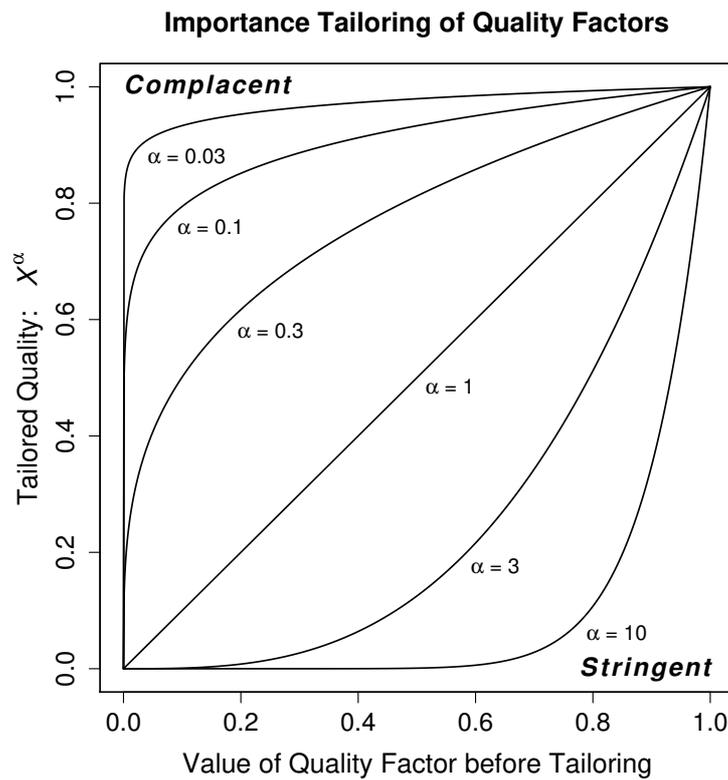


Figure 1.4: Quality factors can be given complacent or stringent exponents, which determine how quickly or how reluctantly they rise to 1.0 and cease to be punitive.

8 Iris Image Quality

Pseudonorm

There exist many alternative definitions for the norm of a vector, using various power functions of its elements, of which the most familiar is the Euclidean norm: $\|(x, y, z, \dots)\| = \sqrt{x^2 + y^2 + z^2 + \dots}$. A subset of norms are semi-norms, which have the property of equaling 0 if any element is 0, thereby providing veto power. Our QS has properties similar to these but it does not actually satisfy the formal properties of norms or semi-norms, so we call it instead a *pseudonorm*, defined on elements in the unit interval $[0, 1]$ and mapping also to the unit interval $[0, 1]$, with parameters $\alpha, \beta, \gamma, \dots$.

In order to select which quality factors to include in the QS and determine the exponents that optimally tailor their relative importance, we submitted Hamming distances (HDs) for a fixed probe-gallery partition of the same-eye images in the ICE 2,953 image database to a series of regression analyses. The goal was to discover which factors X, Y, Z, \dots and corresponding exponents $\alpha, \beta, \gamma, \dots$ in QS make it the best predictor of HD and thus False non-Match Rates. By taking logarithms of both sides of (1.2), the optimisation acquires the form of a linear multivariate regression problem in which the calculated slopes are the exponents sought. In the analyses we used this framework to predict HD, in order to construct empirically an optimal QS product of factors.

Figure 1.5 illustrates the method we used. We first examined the correlations between HD and the log of the smaller of each pair of probe and gallery values for various quality factors (QF). These are shown in the top panel of Figure 1.5, with bar height representing the absolute value of the correlation between the $\log(\text{QF})$ and HD. Many of these candidate factors were experimental variables developed for a NIST trial [15] that launched the ISO/IEC iris quality standardisation project [16]. We have labelled only those that were selected for further analysis and those that correspond either to required (Clause 6.2) or to recommended (Clause 6.3) factors in the ISO/IEC 29794-6 Standard that eventually emerged [16].

The top panel of Figure 1.5 shows that several quality factors were reasonably well correlated with HD. We chose IrisCode entropy, which accounted for 26% of the variability in HD ($R^2 = .26$), as the best single factor and used it in a simple linear regression to predict HD. This quality factor is named and indicated in black in the top panel. Subsequently selected factors for the product sequence are similarly highlighted in the next panels.

The second panel shows the correlations between the $\log(\text{QFs})$ and the residuals from Regression 1 (the difference between observed and predicted HDs). Note that because many of the quality factors were correlated with IrisCode entropy, they have much smaller correlations with the residuals of Regression 1, shown in the second panel, than with HD, shown in the first. Texture energy, however, stands out from the rest in having a substantial correlation that is independent from IrisCode entropy. Therefore, to construct the best two-factor QS, we used IrisCode entropy and Texture

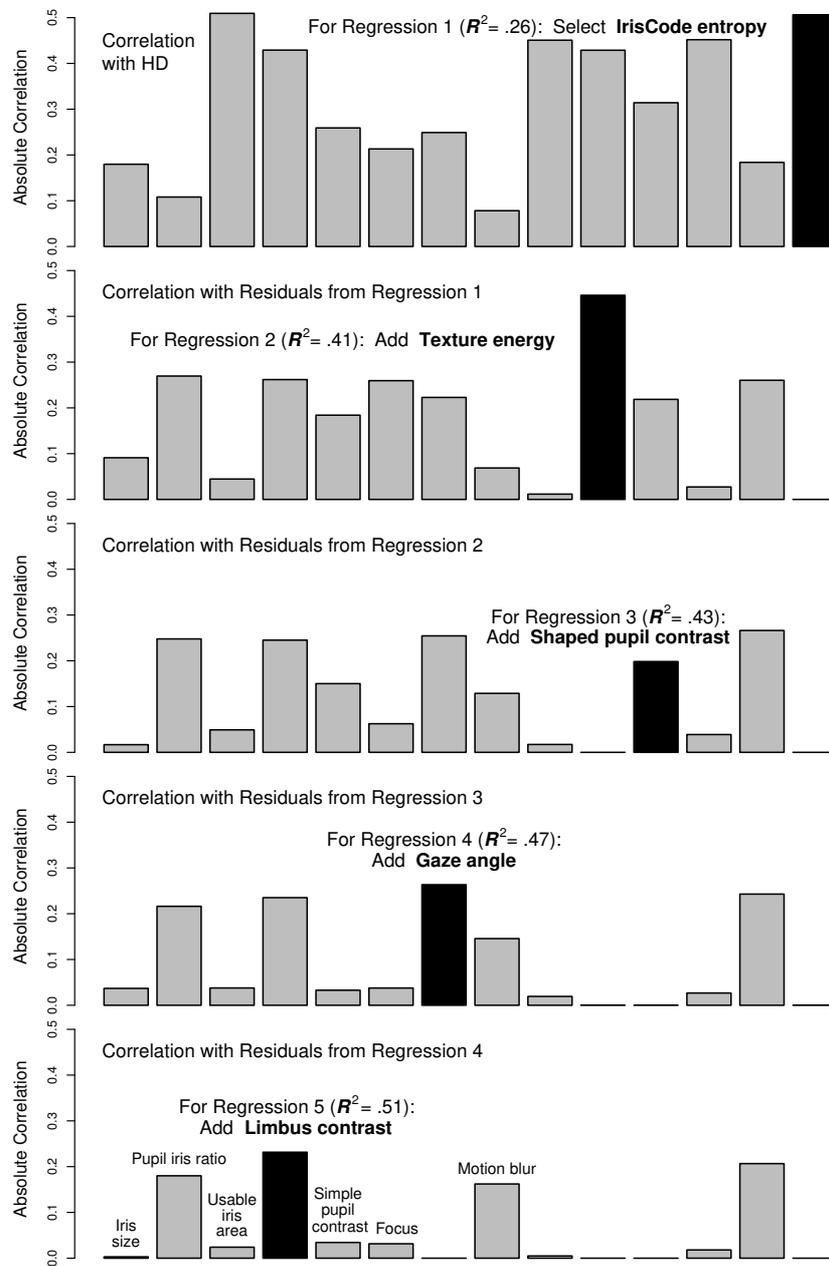


Figure 1.5: Illustration of the correlational method used to select quality factors and their exponents for a multifactorial QS. Bar heights represent absolute correlation. The factors incorporated sequentially into QS are highlighted in black.

10 *Iris Image Quality*

energy in a multiple regression to predict HD. This, Regression 2, accounted for 41% of the variance in HD. We repeated this process of selecting quality factors to add to the regression predicting HD, as shown in the subsequent panels of Figure 1.5, until the increase in R^2 (the proportion of variance accounted for) by adding another variable to the regression equation became nugatory.

Using the five selected quality factors in the multiple regression analysis we can account for almost 51% of the variance in HD, whereas, as shown in Figure 1.5, the best formula with only one or two variables could account for only 26% or 41% of the variance, respectively. Including any one of the remaining factors shown in the fifth panel in a six-variable multiple regression increased R^2 to at most 52%.

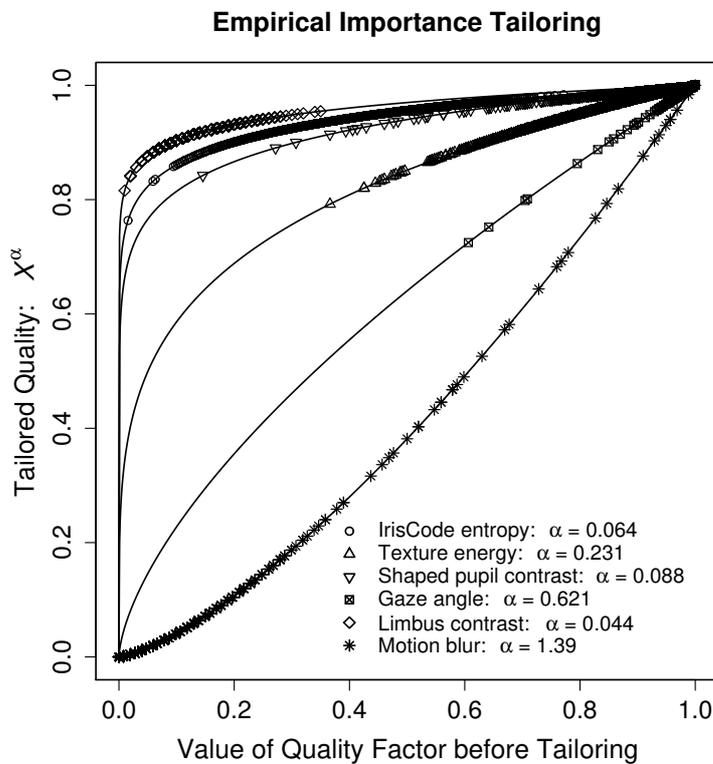


Figure 1.6: Illustration of how several actual quality measures for the images in the NIST Iris Challenge Evaluation database are mapped into tailored quality factors using the exponents derived empirically from multivariate regression analysis.

The regression coefficients obtained for the quality factors in this five-variable formula were then used as the exponents in (1.2) to tailor the factors in the QS. Figure 1.6 lists these exponents in the order in which the factors were included in the QS, and also plots the actual quality factor values for all images in the ICE database, when raised to these powers. The five most predictive quality factors are: IrisCode entropy; Texture energy; Shaped pupil contrast; Gaze angle; and Limbus contrast.

The regression coefficient for Motion blur (when used instead of Limbus contrast) is also given to illustrate the range of empirical fits found. However, its contribution was not sufficient to warrant inclusion in the final QS.

1.5 Error Reject Curves

Figure 1.7 presents the results of our analyses in the form of the Error Reject Curves (ERC) developed by Grother and Tabassi [1]. These rank-order HDs according to the worst of the two image quality measures corresponding to each HD and define successively smaller subsets of HDs by rejecting those where quality falls below a requisite quantile. The residual False non-Match Rate for these shrinking subsets is then plotted as a function of the fraction of HD pairings that have been rejected based on that measure of quality.

The upper panel of Figure 1.7 shows the ERCs for the five individual quality factors selected for our QS. The Figure also shows the correlation of each factor with HD and the exponent fitted for each one within the five-factor QS. The solid line (descending rapidly to 0) is the ideal ERC that would be obtained for a perfect QS ($r = -1$). Again, the factors are listed in the order of their selection for the multiple regression analyses. Over the entire range, among the single factors, Limbus contrast had the best ERC characteristic, followed by Shaped pupil contrast, IrisCode entropy, Texture energy and Gaze angle. However, the ranking of these quality factors varied considerably over different sections of the abscissa.

Generally, our approach recommends a more complacent treatment of those quality factors with rapidly falling ERCs, since we have assigned the smallest exponents to those factors having the best-ranking ERCs: Limbus contrast, Shaped pupil contrast and IrisCode entropy. However, the fitted exponents are only meaningful within the context of a multiple factor QS. The ERCs for tailored and untailored quality factors are identical (except for degenerate cases) since the ordering of the factor is preserved in the tailoring. We must therefore look at ERCs for multi-factor QSs to gain insight into the effects of tailoring.

The lower panel of Figure 1.7 plots ERCs for the best QS from the previous regressions with two, three, four and five factors. These are shown in dark gray. The single factor ERC for IrisCode entropy from the upper panel (in black) is plotted again for reference, as is the ideal ERC. The increase in R^2 reported earlier in the regression analyses is reflected in the notable improvements in the ERC performance as successive factors are included in the QS, with the largest gap between curves corresponding to the largest increase in R^2 . The movement of each successive multiple factor ERC, toward the ideal ERC curve, is clear.

We have also plotted ERCs in light gray for two alternative empirically tailored two-factor QSs. These illustrate the point that the product of the two quality factors with the most impressive single ERCs (Limbus contrast and Shaped pupil contrast) or the two smallest values of α (IrisCode entropy and Limbus contrast) does not necessarily produce the best ERC for a two-factor QS. The choice of IrisCode entropy and Texture energy suggested by the regression analyses is superior to both of these, especially in the initial range of the ERCs where fewer pairings are rejected.

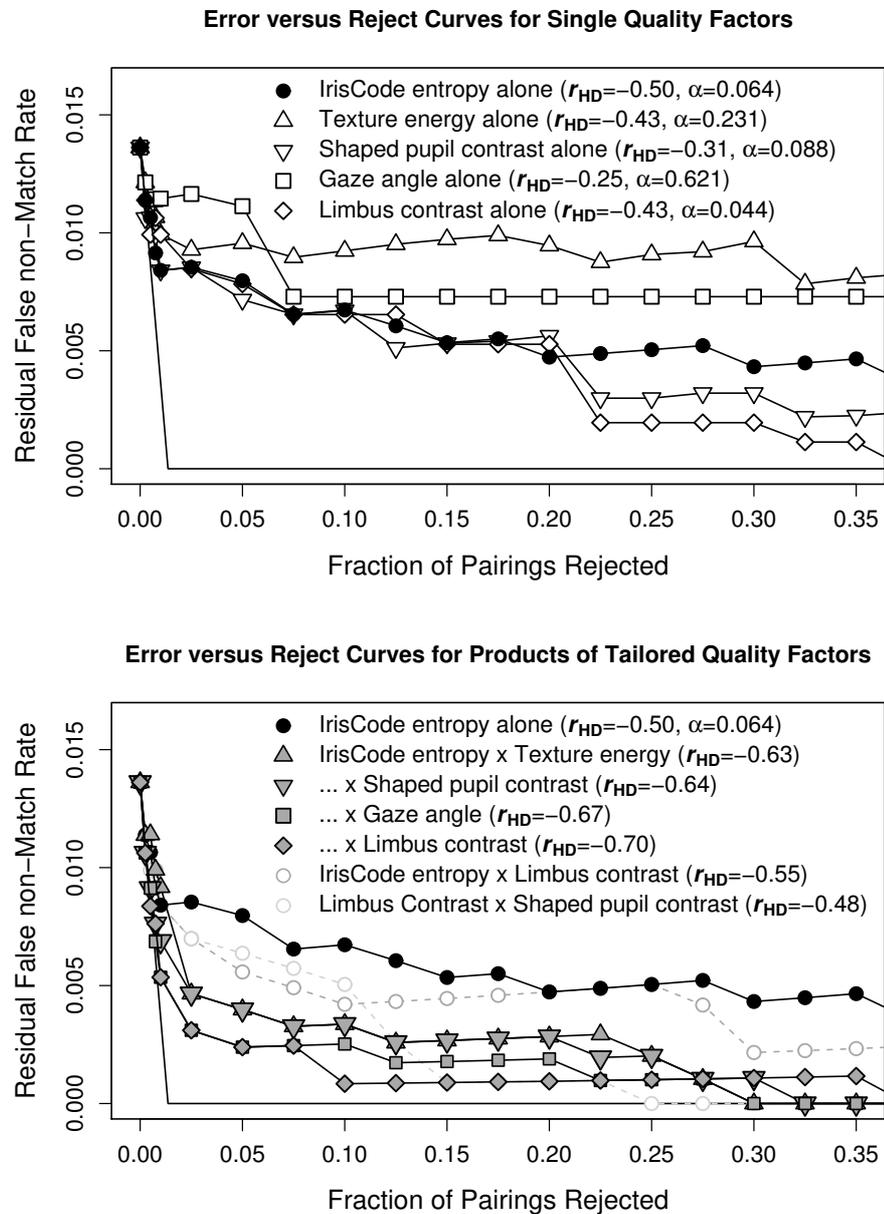


Figure 1.7: Error Reject Curves for single quality factors, upper panel, and multi-factorial QS, lower panel, showing the decline in False non-Match Rate as the worst image pairings are successively dismissed. Ideal performance is indicated by the solid line. Combinations of factors are greatly superior to single factors alone.

1.6 Discipline in Punishment

The rationale for our multiplicative approach to scoring quality was that it confers punitive veto power, which is absent in the additive approach of piece-meal accretion of reward for quality. But discipline must be exercised in the punitive approach, lest the quality factors be either too stringent or too complacent. Empirical tailoring of the exponents based on performance is a way to determine the optimal balance.

Figure 1.5 showed that empirical tailoring of quality factors made five of them more complacent: they acquired exponents smaller than 1.0. In Figure 1.8 we see how both five-factor QSs evolve as successive quality factors are incorporated into the product sequence: X^α ; $X^\alpha Y^\beta$; $X^\alpha Y^\beta Z^\gamma \dots$. In each panel, we show in gray the distribution for QS where all the exponents are 1.0 (the untailed approach), and we show in black the distribution for the corresponding QS where the exponents are tailored for each combination of factors.

By comparing the distributions in gray (untailed) and black (tailed), we can see that the complacent tailoring is mapping all the QSs into significantly higher ranges. This is especially important when the fifth factor, Limbus contrast, is incorporated (bottom panel). For the untailed case, this pushes all images toward very poor quality scores, making this measure far too punitive. But the complacent treatment of Limbus contrast in the tailored QS greatly elevates its values and restores its usefulness. This illustrates the importance of knowing when not to be too punitive.

Clearly, incorporation of successive factors $X_i \in [0, 1)$ in a product sequence can only make the product smaller. However, the use of complacent tailoring generally counteracts this effect, and the judicious assignment of exponents allows the full range of the QS to be used effectively, in addition to fine-tuning the relative contribution of each quality factor.

In Figure 1.8, we have also indicated the positions within each distribution of each of the datapoints linked to the 18 False non-Matches that arose in this database. These are marked by small circles just below the distributions. The QS values below which 2.5% and 5.0% of the scores fall are indicated by the dashed vertical gray lines. We see that for the tailored case, in black, as more factors are incorporated, progressively more False non-Matches are pushed into the long tail to the left of these quantile lines. Whereas for the untailed case, the progression is much less remarkable and the problems caused by the too stringent treatment of Limbus contrast are again highlighted, as the distribution is pushed toward zero and selectivity is lost. This recapitulates the findings of the ERC analyses.

1.7 Predictive Value of the Quality Pseudonorm

In order to determine how much this empirically-derived tailoring improves the QS, we conducted another two linear regressions: one for the product of the five quality factors raised to the empirically-fitted exponents, and a second for the product of the same five factors when all exponents were fixed at unity.

For these analyses, tailored and untailed QSs were computed for each image. The smallest *tailored* five-factor QS for each probe-gallery image pair was entered

14 Iris Image Quality

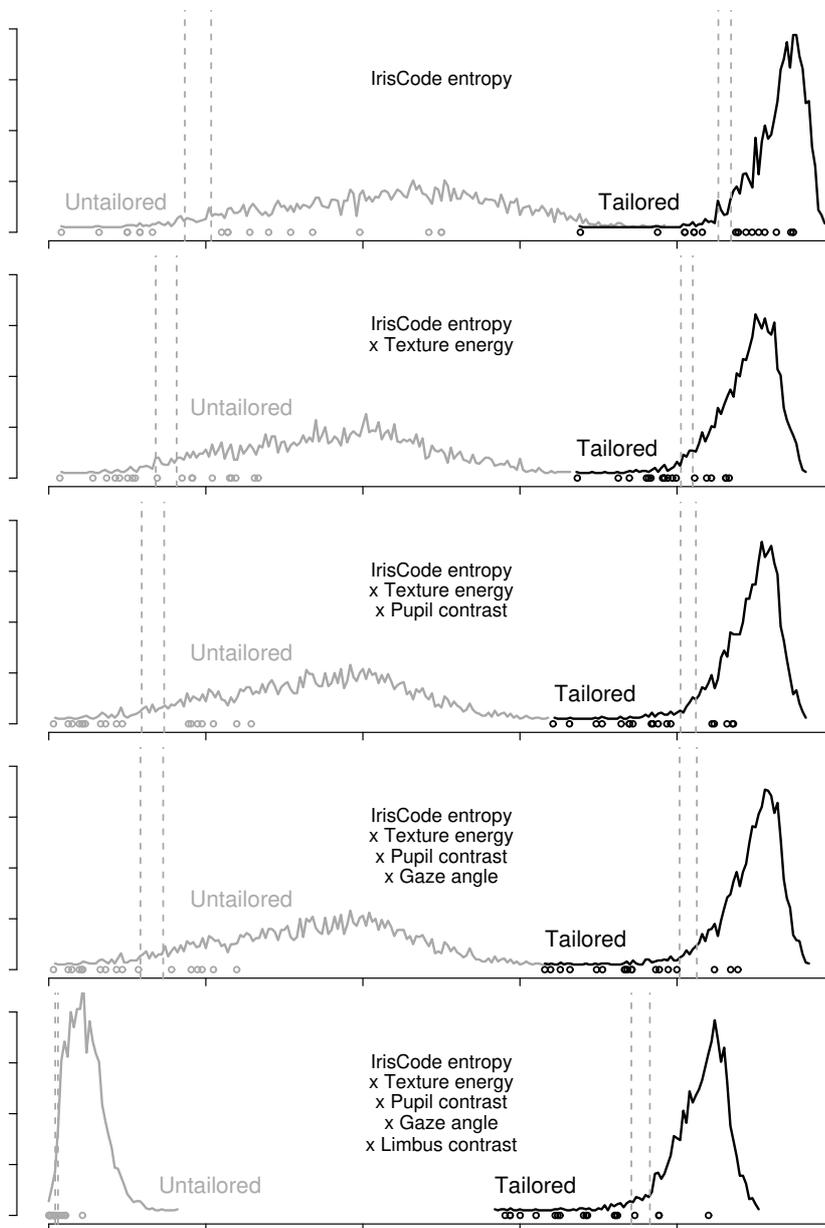


Figure 1.8: Comparison between tailored (black) and untailored (gray) approaches to quality scoring using product sequences with increasing numbers of factors. The small circles indicate the QS value associated with each False non-Match, revealing why such failures are better predicted by the multi-factorial tailored QS.

into a simple linear regression to predict HD. Then the smallest *untailored* five-factor QS for each pair was used in a second simple linear regression on HD.

In Figure 1.9, we plot predicted HD vs observed HD for both of these regressions. The analyses show that by tailoring the QS exponents, we were able to account for 22% more of the variance: R^2 increased from 0.27 for the untailored QS to 0.49 for the empirically tailored QS. The scatterplots reveal a substantial improvement in predictions for the larger HDs, which matter the most for False non-Match Rates.

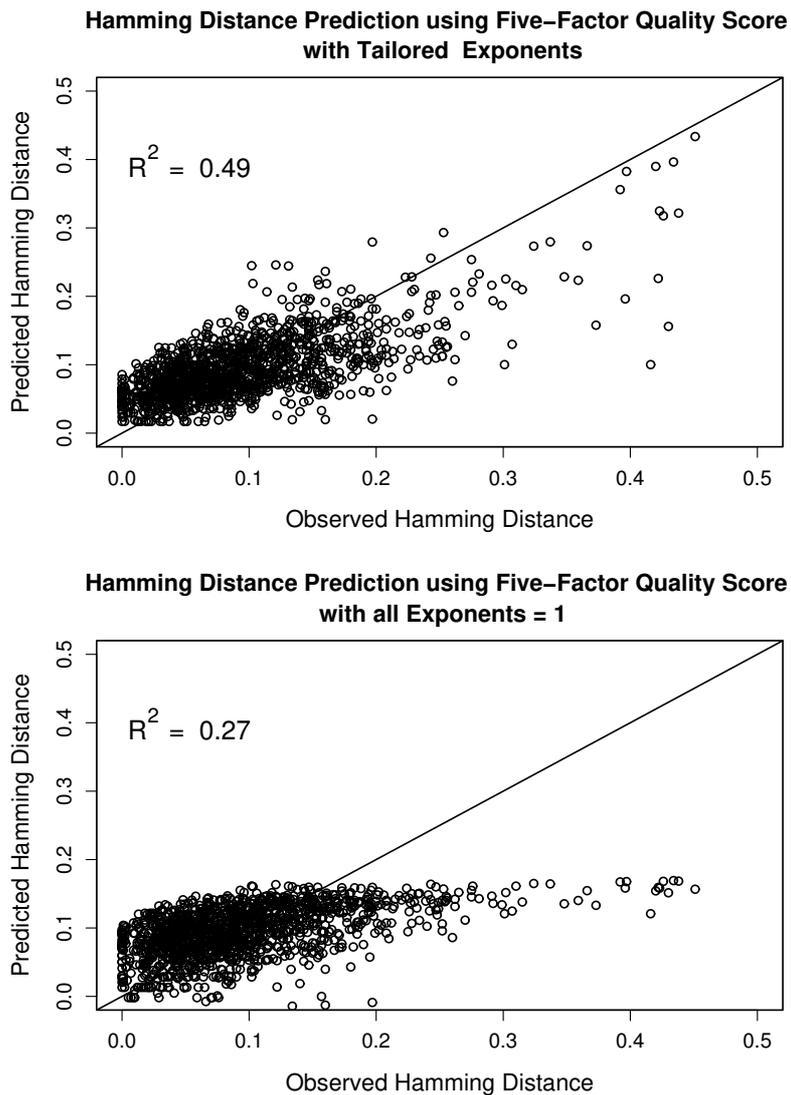


Figure 1.9: Demonstration that importance tailoring of exponents on quality factors improves substantially their ability to predict Hamming distances.

16 *Iris Image Quality*

Because the empirically tailored QS is better able to predict HD than untailored QS, it can be used to flag images that are more likely to produce False non-Matches when same-eye images are compared. On this basis, operational decisions may be taken to repeat image acquisition, or else to proceed with an image. Figure 1.10 shows the consequences of doing so in the ICE database, for each of three QSs considered above. In this database there were 18 same-eye HDs that exceeded 0.32, a threshold that would yield a False non-Match, and the figure indicates (in black) what proportion of these would be excluded as a function of the proportion of images flagged on the basis of poor QS. This is similar to the ERC ‘exclusion’ methodology for evaluating quality assessment algorithms that was originally proposed by Grother and Tabassi [1] and used by NIST in [15], but it is perhaps more straightforward because monotonicity is ensured. ERC curves according to their methodology are also shown in this Figure (in gray) for comparison.

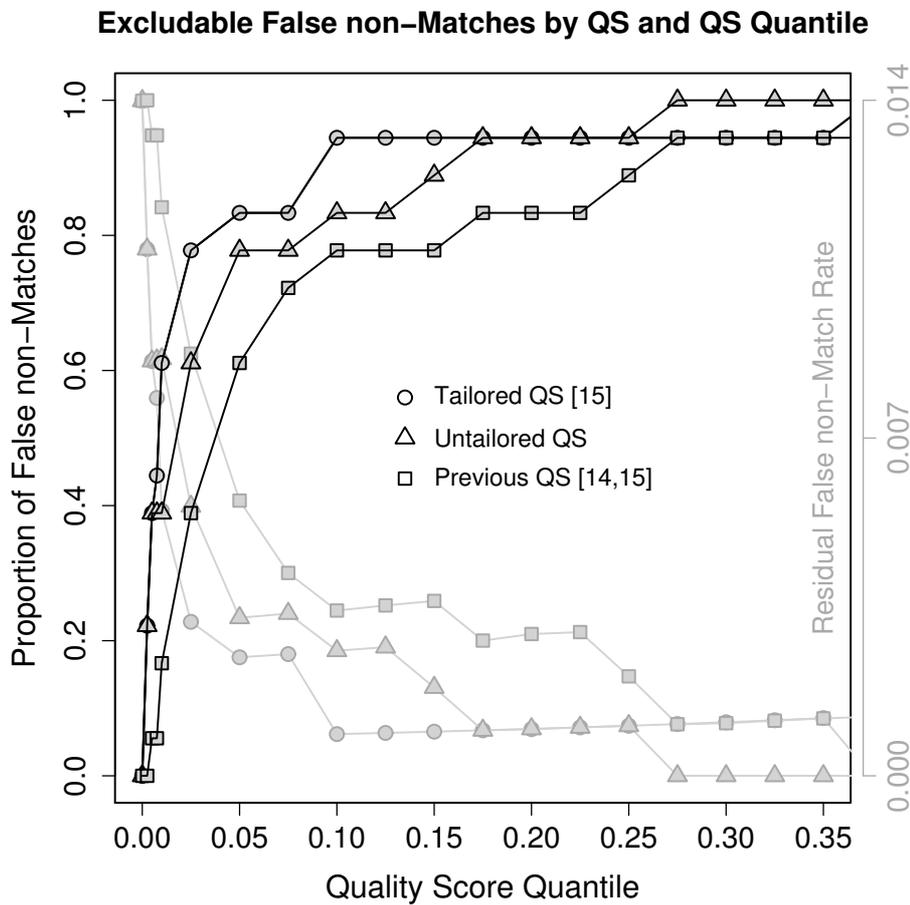


Figure 1.10: Demonstration that importance tailoring of the exponents on quality factors improves substantially their ability to predict recognition performance.

The tailored QS would allow the exclusion of a significantly larger number of False non-Matches than the untailored QS, over the entire quantile range. A t-test confirms significance over the full range but especially up to 12.5% (pairwise t-test (df=8) 3.78, $p = 0.005$), and particularly when considering the poorest 1% to 2% of images. It is also noteworthy that both of the new QSs, tailored and untailored, are substantially better at predicting and excluding False non-Matches than the QS described in [14] and shown in Figures 1.2 and 1.3.

1.8 Testing and Adoption into an ISO/IEC Quality Standard

NIST invited submissions of iris image quality assessment algorithms both from academic and industry providers, and conducted tests of them during 2009 - 2011. The methods described in this chapter were largely developed and refined in response to the NIST initiative, which was called *IQCE: Image Quality Calibration and Evaluation* [15]. The final report on these methods concluded: ‘Implementations from Cambridge University are the most effective in predicting recognition performance of their own native iris recognition algorithm, as well as others’ (p. 6, [15]). It also reported that these quality assessment algorithms ‘are the fastest implementations, with an average 30 milliseconds per image across all four datasets’ (p. 7, [15]).

Consequently an international standardisation project for assessing iris image quality, led by NIST’s Elham Tabassi, incorporated these methods into the ISO/IEC 29794-6 International Standard on iris sample quality [16] during the workprogram of ISO/IEC Joint Technical Committee 1 (SC 37). The pseudonorm QS described here (1.2) became the framework of nonlinear combination of vector elements for generating an actionable scalar, specified now in Clause 6.5.2 of the Standard.

1.9 A Possible Criticism of the Method

It should be noted that the power function form (1.2) defining the QS is monotonic in each individual quality element. Thus the better any one of them is, the better is the overall QS. Our approach may be criticised for this property, because one can imagine interactions of factors for which this ‘collective monotonicity’ may not be optimal. For example, suppose that one quality element is a focus score and another is signal-to-noise ratio, which is degraded by high frequency noise of external origin (not arising in the sensor array itself but present pre-optically, say in a grainy photograph being digitised). If the recognition algorithm is sensitive to such noise, then its performance may actually be enhanced by a degree of defocus, removing the noise. This violates the assumption of collective monotonicity among the elements.

1.10 Acknowledgment

Statistical analyses and generation of figures were performed using the ‘R’ package: <https://cran.r-project.org/>.

1.11 Reference List

- 1 Grother, P., Tabassi, E., ‘Performance of biometric quality measures’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**, (4), pp. 531–543, April 2007
- 2 Phillips, P.J., Beveridge, J.R., ‘An introduction to biometric-completeness: The equivalence of matching and quality’, *IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, Washington, pp. 1–5, Sept. 2009
- 3 Poh, N., Kittler, J., ‘A unified framework for biometric expert fusion incorporating quality measures’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**, (1), pp. 3–18, Jan. 2012
- 4 Nandakumar, K., Chen, Y., Jain, A.K., ‘Quality-based score level fusion in multi-biometric systems’, *18th International Conference on Pattern Recognition (ICPR)*, **4**, pp. 473–476, Hong Kong, 2006
- 5 Schmid, N.A., Zuo, J., Nicolo, F., Wechsler, H., ‘Iris quality metrics for adaptive authentication’, Chapter 4 in *Handbook of Iris Recognition* (Burge and Bowyer, eds.), Springer-Verlag, London, 2013, pp. 67–84
- 6 Kalka, N.D., Zuo, J., Schmid, N.A., Cukic, B., ‘Estimating and fusing quality factors for iris biometric images’, *IEEE Transactions on Systems, Man, and Cybernetics - Part A*, 2010, **40**, (3), pp. 509–524
- 7 Li, X., Sun, Z., Tan, T., ‘Comprehensive assessment of iris image quality’, *18th IEEE International Conference on Image Processing*, pp. 3117–3120, 2011
- 8 Belcher, C., Du, Y., ‘A selective feature information approach for iris image-quality measure’, *IEEE Transactions on Information Forensics and Security*, **3**, (3), pp. 572–577, Sept. 2008
- 9 Galbally, J., Marcel, S., Fierrez, J., ‘Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition’, *IEEE Transactions on Image Processing*, **23**, (2) pp. 710–724, Feb. 2014
- 10 Hofbauer, H., Rathgeb, C., Uhl, A., Wild, P., ‘Iris recognition in image domain: Quality-metric based comparators’, *International Symposium on Visual Computing*, Springer, 2012
- 11 Daugman, J., ‘How iris recognition works’, *IEEE Transactions on Circuits and Systems for Video Technology*, **14**, (1), pp. 21–30, Jan. 2004
- 12 Daugman, J., *US Patent 6,753,919: Fast Focus Assessment System and Method for Imaging*. US Patent Office, 22 June 2004
- 13 Daugman, J., ‘New methods in iris recognition’, *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, **37**, (5), pp. 1167–1175, Oct. 2007

- 14 Daugman, J., Downing, C., ‘Effect of severe image compression on iris recognition performance’, *IEEE Transactions on Information Forensics and Security*, **3**, (1), pp. 52-61, Mar. 2008
- 15 Tabassi, E., Grother, P., Salamon, W., *Iris Quality Calibration and Evaluation: Performance of Iris Image Quality Assessment Algorithms*. NIST Interagency Report **7820**, Bethesda, 30 September 2011.
- 16 International Standard ISO/IEC 29794-6: *Biometric Sample Quality – Part 6: Iris Image*. ISO Copyright Office, Geneva, 2015. See also the overall quality framework document: International Standard ISO/IEC 29794-1: *Biometric Sample Quality – Part 1: Framework*. ISO Copyright Office, Geneva, 2016.