Radial correlations in iris patterns, and mutual information within IrisCodes

John Daugman, Cathryn Downing

[1] *University of Cambridge, Faculty of Computer Science and Technology, Cambridge UK*
* *E-mail: john.daugman@CL.cam.ac.uk*

Abstract: The discriminating powers of biometric patterns derive from their entropy, just as the hardness of cryptographic keys derive from their entropy. The larger the number of independent bits, or the more independent they are, the less chance of collision. We measured the mutual information entailed by radial correlations within each of 632,500 different iris patterns from persons of 152 nationalities. For each iris, we measured how well the sequence of bits in any ring of the IrisCode predicts the sequence of bits in the other rings. Information density is quite non-uniformly distributed across iris patterns radially. Our measurements of mutual information address how much radial resolution is productive to use when encoding an iris, and we show that a non-uniform allocation of encoding resolution radially leads to significant performance improvements by reducing redundancy.

1 Introduction

Iris patterns generally display a radial structure, but this informal impression has not been investigated quantitatively except for an analysis of correlations among pixels by Hu *et al.* [1]. Correlations within random patterns reduce their entropy, which is unfortunate for purposes of discrimination and identification. The main purpose of this paper is to apply the tools of Information Theory to measure that effect, and to document its radial extent both across adjacent and non-adjacent annuli in the iris. We show how these quantitative findings can be exploited in iris encoding to improve recognition performance. When such radial structure is detected and quantified it can also play a useful rôle in iris localisation and confirmation of segmentation, because it is such a distinctive feature in facial images. This paper documents the statistics of radial correlation structure in a large database of iris patterns from persons spanning 152 nationalities, compared with the same metrics in a database of IrisCodes computed from synthetic iris images in which pixels are independent of each other. We investigate correlations not among pixels but among bits within iris templates, and their impact on performance.

Fig. 1: Illustration of radial structure and textural correlations typical of iris patterns, in this case imaged in the visible band.

2 Related literature

Texture analysis has long been an important part of computer vision and image understanding, beginning with terrain classification in aerial or satellite imagery since the early 1970s. Among its many applications are: region segmentation; object or scene classification; inferring 3D surface shape, slant, depth, and material composition; biomedical assessments involving radiological scans or histology; and biometric personal identification in the case of iris recognition. Methods of texture analysis divide broadly into *structural approaches* (best suited for artificial textures with repeated regular patterns), and *statistical approaches* (better suited for natural textures with inexact structure). Statistical metrics and operators include autocorrelation, power spectra and other texture energy measures [2], second moments, entropy, and co-occurrence matrices. Perhaps most well-known is the foundational work of Haralick *et al.* [3] defining textural features by measuring how frequently a given pair of pixel values co-occur in a particular spatial relationship. From these co-occurrence matrices other texture features can be defined including energy, entropy, homogeneity, and correlation [4].

Because texture inherently contains quasi-repetitive or correlated structure for defined scale(s) and direction(s) of propagation, it is natural to characterise texture using directional correlation metrics [5]. In the case of iris patterns, as illustrated in Fig. 1, the dominant structure is invariably radial; indeed the organ and its function are essentially polar. Classical operators for texture analysis are defined in cartesian coordinates, but with the common "unwrapping" of iris patterns into a rectangular picture, polar coordinates are transformed basically into cartesian ones. This coordinate transformation requires a spurious cut to be made for the unwrapping, which fundamentally changes topology and introduces complications that have no benefit, so we avoid doing it. Nonetheless, the only detailed quantitative study yet of iris textural correlations [1] used steering kernels in "unwrapped" coordinates to measure pixel covariance within local neighbourhoods of iris images. They confirmed that the strongest correlations in pixel values were invariably in, or near, the radial direction.

Classical literature in the field of Information Theory [6, 7] provides tools to analyse the randomness and complexity of random variables, their inter-dependencies, their capacity to convey information, and their discriminability. It is therefore surprising that, with rare exceptions [8], Information Theory has not played a rôle in the study of biometric signals and systems other than for encryption. Comprehensive treatment of the concept of *mutual information* between random variables, and of the various entropy metrics to which it is related and on which it is based, can be found in [7].

3 Methods and databases

In this paper we investigate iris textural correlations not among raw pixel values, but rather among radially distributed bits of IrisCodes computed from iris images. We use a database of 632,500 IrisCodes, all arising from different eyes, from persons spanning 152 nationalities who applied for work visas in the United Arab Emirates since 2002 (the "UAE database"). We study these effects in the framework of mutual information and analyse its effects on discrimination power. In order to study how these findings can be exploited for performance benefits, we also use a database containing multiple different images of eyes (the "NIST ICE database" [9]) so that ROC analysis permits trade-offs to be measured between False Match and False non-Match Rates (FMR and FnMR). For these two databases, iris images were acquired in the near-infrared band of illumination (NIR: 700nm – 900nm) and captured at distances of 10cm – 25cm, producing greyscale images such as illustrated in Fig. 2.
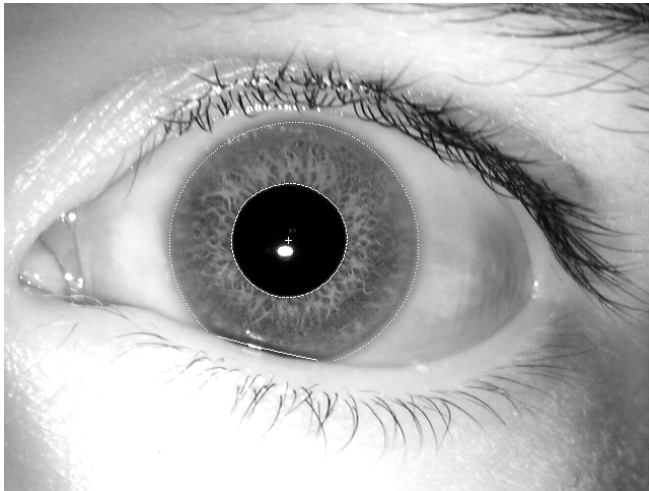

Fig. 2: Example iris image captured in the NIR (700nm–900nm) band for UAE and NIST ICE databases. Overlaid dotted graphics show segmentation output.

The algorithms underlying the computation of IrisCodes have been presented before [10] and will not be rehearsed here. But a crucial point is that IrisCodes are created using 2D wavelets that have bandpass Fourier characteristics (*i.e.* both highpass and lowpass properties in the direction of wavelet modulation), and these properties introduce their own correlations into IrisCode bit streams regardless of the existence of such correlations in an input iris image. The attenuation of higher frequencies causes neighbouring bits to tend to remain the same, while the attenuation of lower frequencies causes bit strings to oscillate, retaining phase coherence across distances reciprocal to the wavelet bandwidth, creating [8] a net "sticky oscillator" effect. Because of these inherent IrisCode properties which must be distinguished from intrinsic iris image properties, throughout this paper all results using human iris patterns are presented with baseline comparisons against results using synthetic "white noise" iris images in which all pixels are independent of each other. A gallery of 500 synthetic iris images was generated by replacing all pixels in an iris with samples drawn from a Gaussian source with $N(\mu, \sigma) = N(128, 30)$. An illustration of such a "white noise" iris image is shown in Fig. 3. The standard IrisCode software was run on all these images to generate their IrisCodes and enroll them into a third database, which we term "IrisCoded noise". Throughout this paper, results using actual human iris patterns are presented graphically in blue, while those using the synthetic noise database for comparison are presented in red.

The quasi-annular zone between the pupil boundary and the outer ("limbus") boundary of an iris receives 64 samples along any ray, and these 64 samples are combined into eight quasi-concentric *rings*. (One consequence of the studies presented here is to show that this
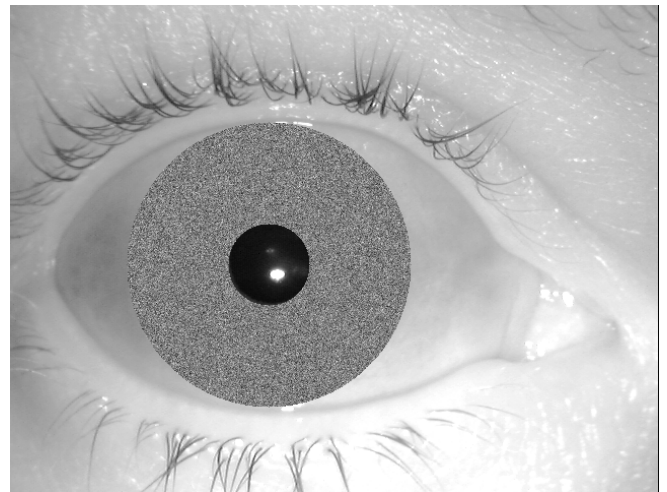

Fig. 3: Example of a synthetic "white noise" image in which all iris pixels are replaced by independent samples from a Gaussian process, in order to distinguish correlation effects within real iris images from those introduced by the IrisCode bandpass wavelet encoding itself.

default grouping is not optimal, and that another is better.) For both the human IrisCodes and the IrisCoded noise, our basic analysis of radial correlation structure was to compute Hamming distances among all possible pairings of the eight rings in an IrisCode. The radial extent of such correlations across "skipped" rings, rather than just between adjacent rings, provides an estimate of radial correlation distance, and it can also reveal how this varies across the iris from the pupil to the limbus. Finally, the distribution of these scores can reveal the variability in these effects across a human population, given the 152 nationalities present in the UAE database.

4 Results

We measured annular correlations between all possible ring pairings among the eight rings ($8 \times 7/2 = 28$ pairings) in every IrisCode. The six panels in Fig. 4 show Hamming distances between adjacent ring pairings, starting from the pupil as Ring 1, using the 632,500 human IrisCodes (blue distributions), and for comparison IrisCodes for the synthetic noise iris patterns (red distributions). A Hamming distance around 0.5 would be expected in the absence of correlations between annuli, as bits that are independent and equiprobable are equally like to agree or to disagree. Values smaller than 0.5 would be observed if such correlations exist. It is clear that strong radial correlations exist between adjacent annuli in the human IrisCodes and that these correlations are larger farther from the pupil, creating greater separation between the iris (blue) and the IrisCoded noise (red) distributions which remain centred around 0.5 and stable. The widths (standard deviations) of all distributions reflect the presence of correlations among nearby bits *within* each ring, due to the bandpass nature of the encoding wavelets, as discussed earlier.

Fig. 5 summarises such distributions of Hamming distance scores between adjacent rings as boxplots. The extent of boxes corresponds to the inter-quartile interval, with median values marked; and the whiskers correspond to 1.5 times the inter-quartile interval above and below the two quartiles. Non-overlap of boxes of the two colours suggests a significant difference. We see that meaningful effects are observed for all six of the adjacent ring comparisons.

The boxplots in Fig. 6 reveal the annular correlations when jumping across rings separated by 1, 2 or 3 skipped rings, thereby providing an estimate of its radial extent. It should be noted that human eyelid occlusion at large radius values should not play a rôle here because IrisCode bits affected by such occlusion are masked. As we saw in Figs. 4 and 5, again here the Hamming distance scores remain centred around 0.5 for the IrisCoded noise (red distributions). But for human IrisCodes (blue distributions), radial correlations exist even between some non-adjacent rings, especially Rings 5 and 7.
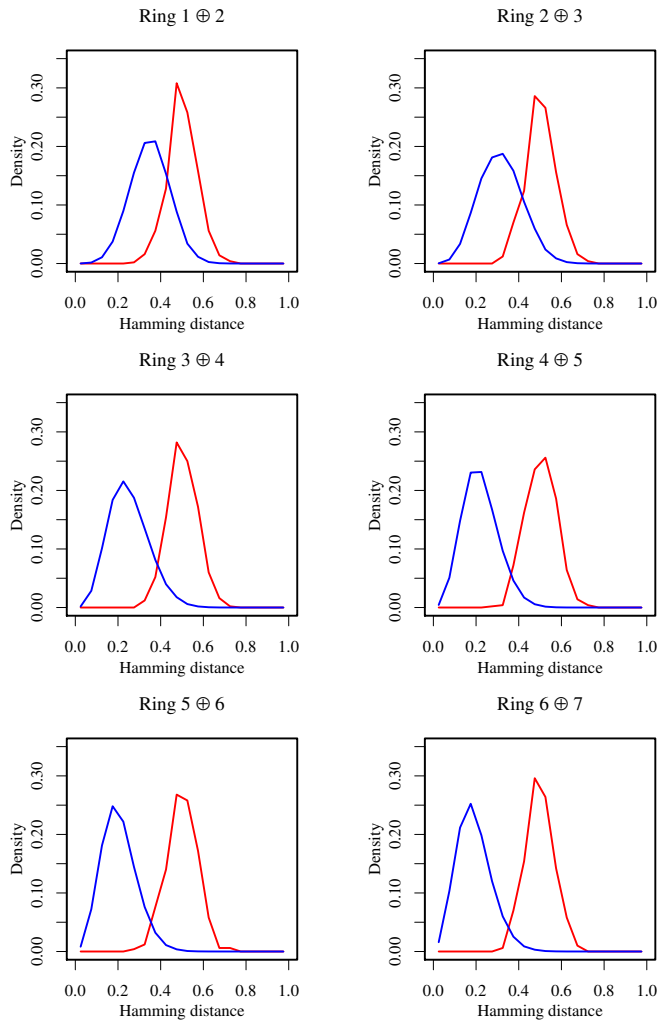
Fig. 4: Distributions of Hamming distance scores when adjacent annuli are compared, both for human IrisCodes (blue) and for IrisCoded noise (red). The ⊕ (Exclusive-OR) operator detects the amount of disagreement. For all adjacent ring comparisons, human IrisCodes are significantly correlated, as shown by their shift from the IrisCoded noise.
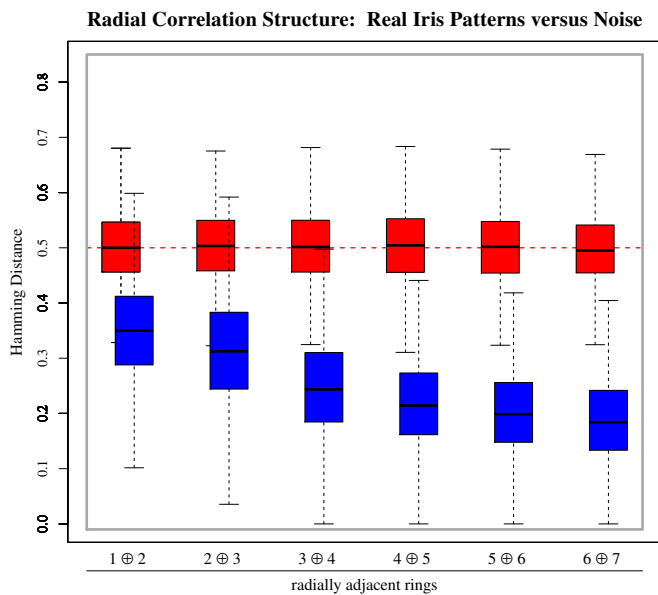

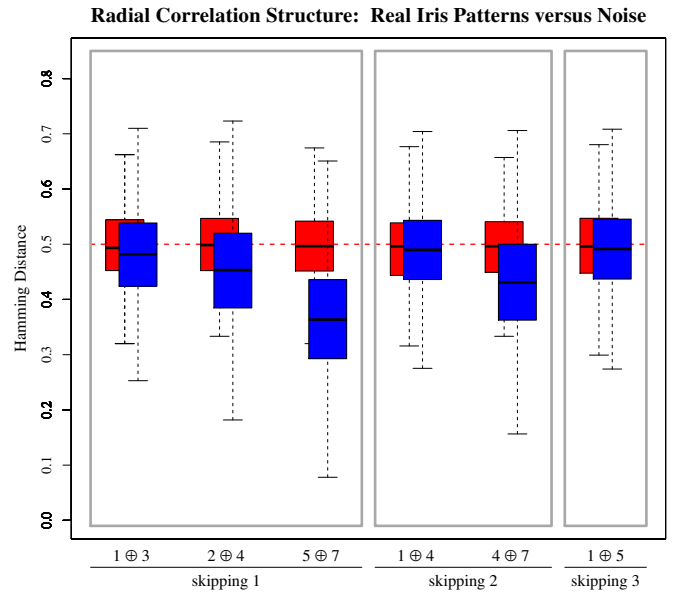**Fig. 5**: Inter-quartile boxplots for the distributions in Fig. 4.

Fig. 6: Inter-quartile boxplots for Hamming distance distributions between annular rings when skipping over 1, 2, or 3 rings.

The complete set of inter-ring comparisons for all the human IrisCodes is plotted in Fig. 7 as a heatmap of their median Hamming distance values. It shows that strong correlations exist between adjacent rings, and even between some nearby but non-adjacent rings, and that these effects are stronger at greater distances from the pupil (the upper-right corner in the heatmap). This confirms that the information density in human iris patterns is greatest near the pupil, and suggests it may be profitable to reduce ring-widths near the pupil, as successive rings there are more different from each other and contain information that is less redundant. We document the performance benefits of doing so in Section 6, while also checking for any cost.
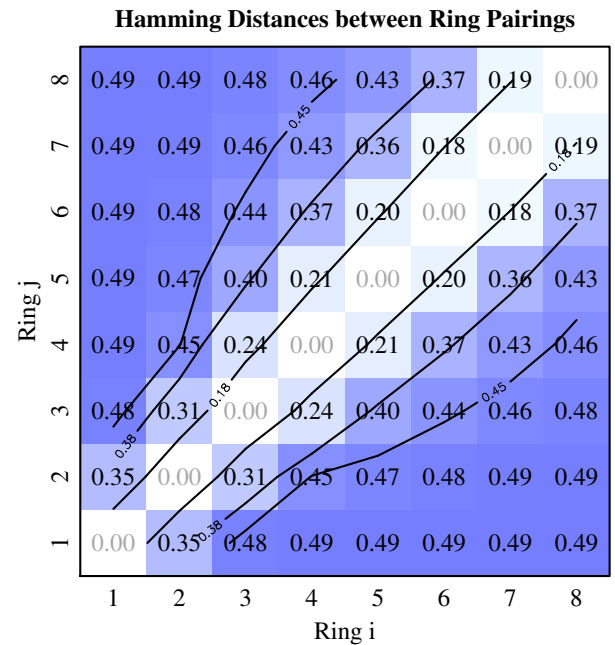

Fig. 7: Heatmap of Hamming distances between all possible pairings of the eight rings in an IrisCode. Median scores across the database of 632,500 human IrisCodes are shown, together with interpolated contour lines for HD values of 0.18, 0.38, and 0.45.

5 Information Theoretic analysis

The discriminating powers of biometric patterns derive from their entropy, just as the "hardness" of cryptographic keys derive from their entropy (which equates directly to key length in bits if the bits are independent). In a biometric template such as an IrisCode, the larger its number of independent bits (or the more independent its bits are from each other), the lower the probability of a collision (False Match) and therefore the higher the tolerance for non-matching bits from the same source can be. Thus both FMR and FnMR can benefit from greater biometric entropy. Correlations, or mutual information, within a particular biometric pattern reduce its entropy and thus its discriminating power. We wish to quantify this negative effect of iris radial correlations; and from this analysis, to reduce their impact.
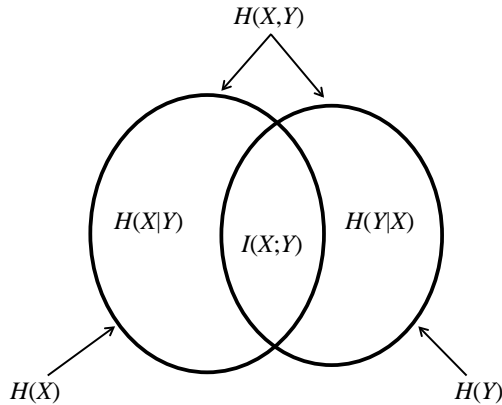

Fig. 8: Classic Venn diagram of mutual information and entropies

Information Theory provides tools for analysing relationships between random variables that are not independent. One such metric is *mutual information* $I(X;Y)$, which measures how much uncertainty about random variable X is removed by knowing random variable Y. These concepts are typically portrayed by a Venn diagram such as Fig. 8 in which the ovals represent the entropies $H(X)$ and $H(Y)$ of the two random variables X and Y, and their union is their joint entropy $H(X,Y)$. Their intersection is their mutual information $I(X;Y)$ showing how much either random variable reveals about the other, while the excluded crescent-shaped regions are the conditional entropies $H(X|Y)$ and $H(Y|X)$ reflecting how much uncertainty remains about X or Y despite knowledge of the other random variable.

Shannon's [6] starting point was to define the information content of an event as the logarithm of its probability p. Thus an observation that was certain anyway ($p = 1$) conveys no information; the lower its probability, the more informative it is; and most importantly, the information gained from *independent* events (joint probability being multiplicative), is *additive*: $\log(p_1 p_2) = \log(p_1) + \log(p_2)$. The entropy H of a random variable X whose states are $\{x_i\}$ is a sum of such terms, each one weighted by the probability $p(x_i)$ of the state, leading to Shannon's foundational formula:

$$H(X) = -\sum_i p(x_i) \log_2 p(x_i) \qquad (1)$$

A pair of random variables X and Y have joint entropy $H(X,Y)$ and conditional entropies $H(X|Y)$ and $H(Y|X)$, defined simply by substituting $p(x_i)$ in (1) with $p(x_i, y_j)$ or $p(x_i|y_j)$ or $p(y_j|x_i)$ where $\{y_j\}$ are the states of Y. Conditionalising over all states of Y requires weighting also by their probabilities $p(y_j)$, leading to the definition of conditional entropy $H(X|Y)$:

$$H(X|Y) = -\sum_j p(y_j) \left[\sum_i p(x_i|y_j) \log_2 p(x_i|y_j) \right] \qquad (2)$$

5.1 Radial correlations as mutual information

There are three equivalent ways to define the mutual information $I(X;Y)$ using the entropies mentioned. The simplest one for our purposes where we regard X as the string of bits in one ring of an IrisCode, and Y as the string of bits in another of its rings, is:

$$I(X;Y) = H(X) - H(X|Y) \qquad (3)$$

where the quantities $H(X)$, $H(X|Y)$, and $I(X;Y)$ are computed in a bitwise manner (measured in bits "per bit of the IrisCode").

The two states $\{x_i\}$ of a bit are equiprobable because the encoding wavelets have zero mean, so projection coefficients are equally likely to be positive or negative; hence $p(x_i) = \frac{1}{2}$. It follows that $H(X) = \frac{1}{2} + \frac{1}{2} = 1$ bit of entropy per bit of the IrisCode, in the absence of any other information. For calculating the conditional entropy $H(X|Y)$ in (2), the conditional probabilities $p(x_i|y_j)$ are captured by Hamming distances HD between rings, because HD measures the relative frequency with which corresponding bits agree or disagree. Thus mutual information (3) evaluates simply to:

$$I(X;Y) = 1 + (\text{HD}) \log_2(\text{HD}) + (1 - \text{HD}) \log_2(1 - \text{HD}) \quad (4)$$

The U-shaped graph in Fig. 9 plots this relationship, with the dependencies between Rings 6 and 7 illustrated both for the human and for the white noise iris patterns (blue and red respectively). The bounded coloured portions correspond to the inter-quartile ranges observed. The mutual information between Rings 6 and 7 of human IrisCodes (having median Hamming distance 0.184) is a substantial 0.311 bits per bit pair, but it is 0.0 bits per bit pair between rings of the noise IrisCodes, as expected. A score of $I(X;Y) = 0$ corresponds to a null intersection in the Venn diagram of Fig. 8, signifying independence of the random variables. Note that if bits always disagreed (HD = 1), their mutual information would be as high (maximal $I(X;Y) = 1$) as if they never disagreed (HD = 0).
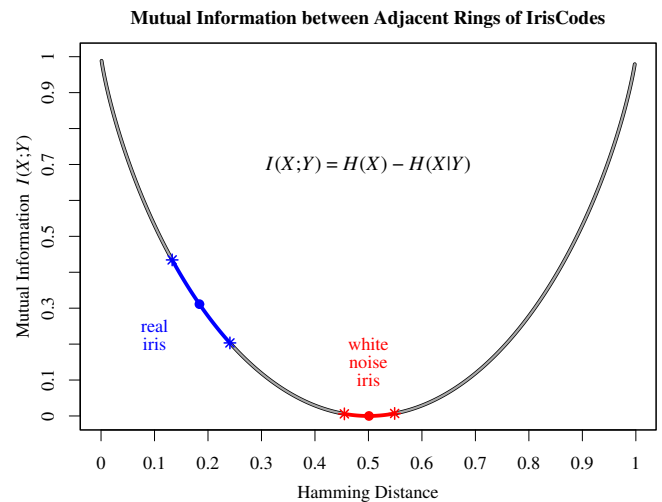

Fig. 9: Plot of (4) mapping Hamming distances between bit strings into their mutual information, highlighting two boxplots from Fig. 5.

The heatmap of Fig. 10 shows the mutual information between every possible pairing of the eight rings, given as the median such $I(X;Y)$ score across all the 632,500 human IrisCodes. Needless to say, every ring has mutual information $I(X;Y) = 1$ with itself, falling off to 0 with distance from this $i = j$ diagonal. It remains significant even across nearby but non-adjacent rings, especially at the larger radius values farther away from the pupil, as indicated by the broadening out towards the upper-right corner of Fig. 10.
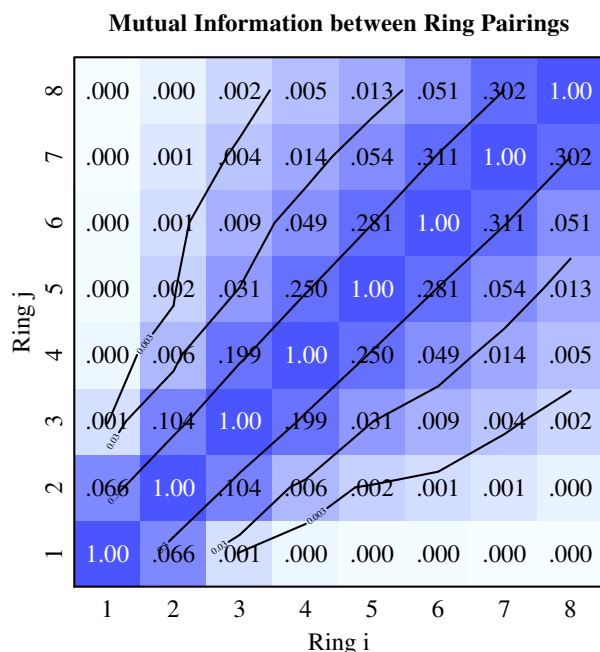
Mutual Information between Ring Pairings

Fig. 10: Heatmap of mutual information between all pairings of the eight rings in IrisCodes, across the database of 632,500 IrisCodes. Interpolated contour lines are shown for 0.3, 0.03, and 0.003 bits per bit. The extent of iris radial correlations is revealed.

6 Exploiting mutual information to improve discrimination performance

Discovering the structure of redundant information in iris patterns can be exploited for performance benefit by appropriately allocating the encoding resolution. The boxplots in Fig. 5 (as well as the heatmaps of Figs. 7 and 10) show that radial correlations between neighbouring rings are progressively greater with farther distance from the pupil, so the encoding budget can be sacrificed at large radii in favour of the inner regions. Varying the annular widths of the sampling rings as a function of radius, in accordance with our observations about mutual information, has the effect of extracting more information, net, from an iris pattern.

We investigated this prediction empirically using the NIST ICE database [9] of about 3,000 images, which include several images of each iris so that it is possible to measure False non-Match Rates (FnMR) and perform an ROC analysis. We compared the standard uniform sequence of ring widths for encoding iris patterns to a sequence of progressively wider sampling annuli that emulates the measured mutual information as a function of radius. Table 1 presents the results for these two ring sequences. (The ring width sequences are in units of 64^{ths} of the normalised radial distance between pupil and limbus; thus both sequences sum to 64 in total.) For same-eye images, there was almost no degradation in HD scores as a result of changing the ring width allocations: the mean HD worsened from 0.114 to 0.118 only. But the improvement in discriminatory power was notable. We measured FnMR at three FMR points, namely 10^{-4}, 10^{-5}, and 10^{-6}, and we observed benefits in all cases. The performance benefit in FnMR is greater at more demanding FMR. We see that if operating at FMR=10^{-6}, the FnMR is reduced by almost 30% in relative terms by adopting the non-uniform ring width sequence shown.

Table 1 ROC analysis: measured False non-Match Rates at various FMR

Width Sequence	@ FMR=10^{-4}	@ FMR=10^{-5}	@ FMR=10^{-6}
8 8 8 8 8 8 8 8	0.0055	0.0079	0.0113
4 5 7 9 9 10 10 10	0.0047	0.0066	0.0080

7 Conclusions

We measured radial correlation structure in a large and diverse iris database. Our main result is to have shown how radial correlations within iris patterns create a non-uniform radial distribution of information. There is greater redundancy between adjacent rings at larger radii, and therefore less contribution to discriminating power than nearer the pupil, where the information density is greatest.

One possibly useful application of detecting and measuring radial structure relates to iris detection and localisation, because radial correlations are a striking feature of human iris patterns. Occasionally, catastrophic segmentation errors can result in some round non-iris object in an image being "detected" as an iris. The boxplots presented in Fig. 5 suggest that Hamming distances larger than (say) 0.45 among most pairs of adjacent rings could be a flag for failed, non-iris, localisation.

The chief findings address the question of how high the radial resolution should be when constructing IrisCodes. The standard algorithm has eight concentric rings, mainly because this efficiently exploits basic computer architecture. Each byte of IrisCode data is one radial traversal of the iris along a particular ray, with the eight bits corresponding to wavelet projections in the eight rings. This facilitates stacking bytes into 16-, 32-, or 64-bit words, and similarly for their corresponding mask bits which indicate occlusion or noisy iris data deemed unreliable, so that the match process executes highly efficiently, on all bits simultaneously within the word length. But some academic research papers in the literature partition the iris into a larger number of concentric rings, often twenty. The measurements presented here of mutual information between rings show that even with only eight rings, there is considerable redundancy. We showed how it can be exploited in the encoding.

Iris recognition has legendary resistance to False Matches. Using databases large enough to make 1.2 trillion different iris comparisons, the National Institute of Standards and Technology confirmed [9] that even if 28% of bits disagree (HD = 0.28) in an accepted match, the FMR is 1 in 40 billion. This is as was predicted from theoretical entropy analysis several years earlier [10], before such mega-databases existed. These results speak eloquently about the importance of biometric entropy. But the present paper has shown that further optimisation can be achieved by considering the mutual information within iris patterns.

Acknowledgment

8 References

1 Hu, Y., Sirlantzis, K., Howells, G.: 'A study on iris textural correlation using steering kernels', *IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2016, pp. 1–8, DOI: 10.1109/BTAS.2016.7791160

2 Wang, L., He, D.C.: 'Texture classification using texture spectrum', *Pattern Recog. Lett.*, 1990, **13**, pp. 905–910

3 R. M. Haralick, K. Shanmugam, I. Dinstein: 'Textural features for image classification', *IEEE Trans. Systems, Man, and Cybernetics*, 1973, **3** (6), pp. 610–621

4 Zucker, S.W., Terzopoulos, D.: 'Finding structure in co-occurrence matrices for texture analysis', *Comput. Graphics and Image Proc.*, 1980, **2**, pp. 286–308

5 Depeursinge, A., Al-Kadi, O., Mitchell, J.: *Biomedical Texture Analysis: Fundamentals, Applications, Tools and Challenges* (Elsevier, 2017)

6 Shannon, C.E.: 'A mathematical theory of communication', *Bell Syst. Tech. J.*, 1948, **27**, pp. 379–423

7 Cover, T., and Thomas, J.: *Elements of Information Theory* (Wiley, 2nd edn. 2006)

8 Daugman, J.G.: 'Information theory and the IrisCode', *IEEE Trans. Info. Foren. Sec.*, 2015, **11** (2), pp. 400–409

9 Grother, P., Quinn, G.W., Matey, J.R., Ngan, M., Salamon, W., Fiumara, G., Watson, C.: 'IREX-III: Performance of Iris Identification Algorithms. NIST Interagency Report 7836' (NIST, Gaithersburg, MD, April 6, 2012)

10 Daugman, J.G.: 'How iris recognition works', *IEEE Trans. Circ. Sys. Vid. Tech.*, 2004, **14** (1), pp. 21–30