THE ROYAL SOCIETY

doi 10.1098/rspb.2001.1696

Epigenetic randomness, complexity and singularity of human iris patterns

John Daugman* **and Cathryn Downing**

University of Cambridge, Computer Laboratory, Cambridge CB2 3QG, UK

We investigated the randomness and uniqueness of human iris patterns by mathematically comparing 2.3 million different pairs of eye images. The phase structure of each iris pattern was extracted by demodulation with quadrature wavelets spanning several scales of analysis. The resulting distribution of phase sequence variation among different eyes was precisely binomial, revealing 244 independent degrees of freedom. This amount of statistical variability corresponds to an entropy (information density) of about 3.2 bits mm^{-2} over the iris. It implies that the probability of two different irides agreeing by chance in more than 70% of their phase sequence is about one in 7 billion. We also compared images of genetically identical irides, from the left and right eyes of 324 persons, and from monozygotic twins. Their relative phase sequence variation generated the same statistical distribution as did unrelated eyes. This indicates that apart from overall form and colour, iris patterns are determined epigenetically by random events in the morphogenesis of this tissue. The resulting diversity, and the combinatorial complexity created by so many dimensions of random variation, mean that the failure of a simple test of statistical independence performed on iris patterns can serve as a reliable rapid basis for automatic personal identification.

Keywords: epigenesis; iris patterns; biometry; wavelets; pattern morphogenesis; statistical independence

1. INTRODUCTION

The iris begins to form during the third month of gestation (Kronfeld 1962). The structures creating its striking patterns are developed by the eighth month (Wolff 1948), although pigment accretion may continue into the first postnatal years. The layers of the iris have both ectodermal and mesodermal origin, consisting of (from back to front): a darkly pigmented epithelium; pupillary dilator and sphincter muscles; a vascularized stroma (connective tissue of interlacing ligaments containing melanocytes); and an anterior layer with a genetically determined (Imesch *et al.* 1997) density of melanin pigment granules. The combined effect (see figure 1) is a visible pattern displaying distinctive features such as arching ligaments, crypts, ridges, and a zigzag collarette. Iris colour is determined primarily by the density of melanin in the anterior layer and stroma, with blue irides resulting from an absence of pigment: long wavelength light penetrates and is absorbed, while shorter wavelengths are reflected and scattered by the stroma (Imesch *et al.* 1997). The heritability and ethnographic diversity of iris colour have long been studied (Davenport & Davenport 1907), but little attention has been paid to the pattern complexity and textural variability of the iris across individuals.

The purpose of our study was to assess the randomness and singularity of iris patterns, and their phenotypic distinctiveness as biometric identifiers, based on video images acquired in public trials of pattern recognition methods proposed earlier (Daugman 1993). Digitized human iris images were acquired over a 3 year period from volunteers at kiosks in field trials at several sites in the USA, UK and Japan. Imaging at arm's length was done by monochrome charge-coupled device (CCD) video cameras, with illumination in the 700–900 nm band from light-emitting diodes or from filtered tungsten sources. Video frames were digitized at a resolution of 640×480 pixels, with iris radius typically in the range 100–140 pixels as seen in figure 1. The iris in each image was automatically localized and isolated by algorithms for finding its inner and outer boundaries (pupil and limbus) as well as any occlusion by upper or lower eyelids, using circular and arcuate edge detectors (Daugman 1993). The isolated iris portion of each image was then submitted to pattern analysis and cross-comparison against all the other irides.

2. PATTERN ANALYSIS AND COMPARISON

Each isolated iris pattern was encoded mathematically by demodulation (Daugman & Downing 1995) to extract its detailed phase information, using complex-valued two-dimensional (2D) Gabor wavelets (Daugman 1980, 1988). This procedure is illustrated in figure 2. It amounts to a local phase sequencing of an iris pattern, by determining in which quadrant of the complex plane the projection coefficients reside when each small image region is projected onto a quadrature pair of 2D Gabor wavelets:

$$h_{\{\mathrm{Re,Im}\}} = \mathrm{sgn}_{\{\mathrm{Re,Im}\}} \int_{\rho} \int_{\phi} I(\rho, \phi) e^{-i\omega(\theta_0 - \phi)} \times e^{-(r_0 - \rho)^2/\alpha^2} e^{-(\theta_0 - \phi)^2/\beta^2} \rho \, d\rho \, d\phi, \quad (2.1)$$

where $h_{\{\mathrm{Re,Im}\}}$ can be regarded as a complex-valued bit whose real and imaginary parts are either unity or zero (sgn) depending on the sign of the 2D integral. $I(\rho, \phi)$ is the raw iris image in a dimensionless polar coordinate system that is size-invariant and translation-invariant. This coordinate system also inherently compensates for pattern elasticity with pupil dilation, because the radial coordinate ρ maps the iris from its inner boundary to its

*Author for correspondence: (john.daugman@CL.cam.ac.uk).

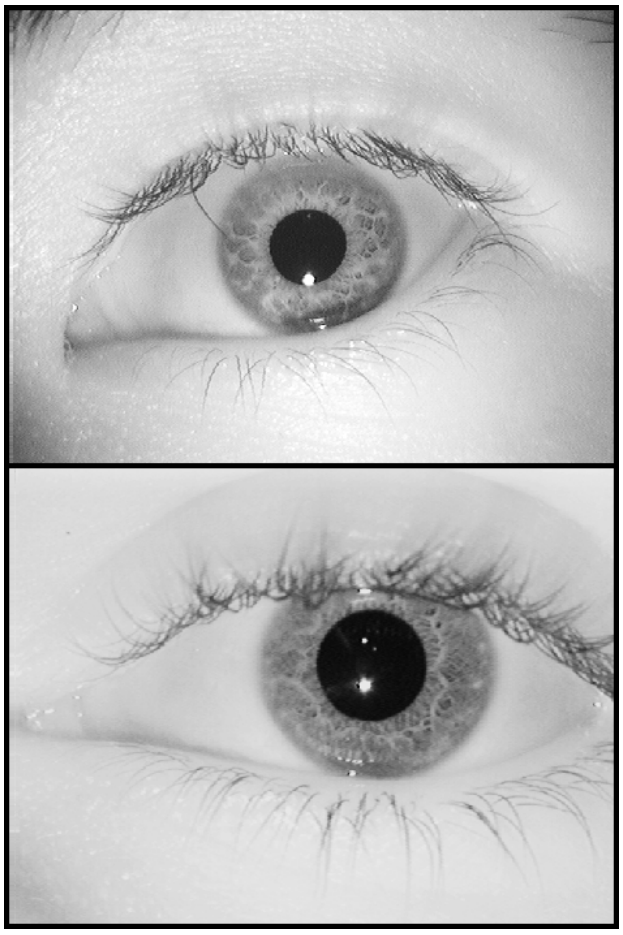

Figure 1. Examples of human iris patterns imaged with monochromatic CCD videocameras at a distance of about 30 cm using light in the 700–900 nm band.

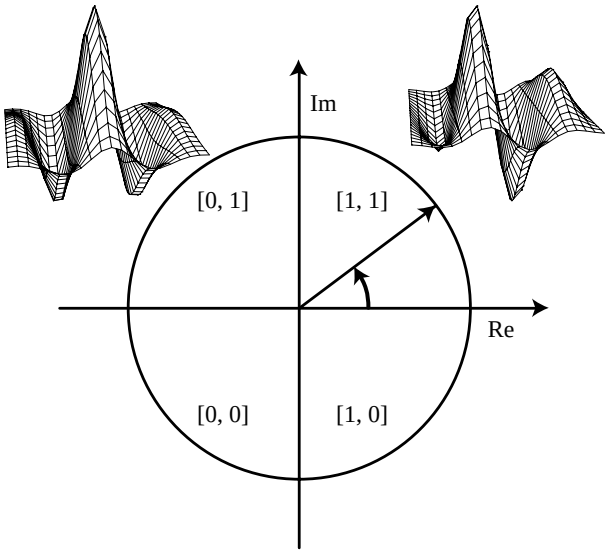

Figure 2. The phase-quadrant demodulation process (Daugman & Downing 1995) used to encode iris patterns. Local regions of an iris are projected (equation 2.1) onto quadrature 2D Gabor wavelets, generating complex-valued projection coefficients whose real and imaginary parts specify the coordinates of a phasor in the complex plane (arrow). The angle of each such phasor is quantized to one of the four quadrants, setting two bits of phase sequence information. This process is repeated all across the iris with many wavelet sizes, frequencies, and orientations, to extract 2048 bits.

outer boundary onto the unit interval [0, 1] and thereby reverses the pattern deformations caused by pupillary action. α and β are the multi-scale 2D wavelet size parameters, spanning an eightfold range from 0.15 mm to 1.2 mm on the iris; ω is wavelet frequency, spanning three octaves in inverse proportion to β; and (r_0, θ_0) represent the polar coordinates of the centre of each region of iris for which the bits $h_{\{\mathrm{Re},\,\mathrm{Im}\}}$ are computed. Altogether 2048 such bits (256 bytes) were computed for each iris, with an equal number of masking bits also computed to signify whether an analysed region was occluded by eyelids, eyelashes, specular reflections, or corrupted by other noise and thus should be discarded from the computations as artefact.

In order to quantify the statistical variability of iris patterns, all irides in our database were compared against all the other ones by computing the Exclusive-OR (XOR) of their 2048 bit demodulation phase sequences. (The XOR operator $\otimes$ detects disagreement between bit pairs; its truth table for the bit pairs 00, 01, 10, 11 is: 0, 1, 1, 0.) The total proportion of disagreeing bits between two encoded iris patterns defines their Hamming distance (HD). If A and B are the 2048 bit phase sequences for two irides, then their dissimilarity is gauged by the total number of '1' bits in the computed bit sequence $A \otimes B$, denoted as vector length by $\| \quad \|$: $\mathrm{HD} = \|A \otimes B\|/2048$. Similar iris patterns would generate small HD scores, and exactly identical iris patterns would have $\mathrm{HD} = 0$. Because any given bit in the phase sequence for an iris is equally likely to be unity or zero, the expected proportion of agreeing bits between two independent sequences would be 50% since the four combinations (00, 01, 10, 11) that are possible when any pair of bits are compared are all equally probable. In summary, the HD metric detects similarities between irides by producing small HD scores, whereas an absence of correlations among irides would produce HD scores that cluster around 0.5.

3. RESULTS

The histogram in figure 3 shows the distribution of HD scores obtained from all 2.3 million comparisons between different iris pairings that were possible in our database. We had 2150 different iris images, including 10 each of one subset of 70 eyes. Excluding those duplicates of (700×9) same-eye comparisons, and not double-counting reversed pairings (e.g. AB and BA), and not comparing any image with itself, the total number of unique pairings between different eye images whose HD scores we computed was $((2150 \times 2149 - 700 \times 9)/2) = 2\,307\,025$. Their observed mean HD was $p = 0.499$ with standard deviation $\sigma = 0.032$. Their entire distribution in figure 3 corresponds almost perfectly to a binomial having $N = p(1-p)/\sigma^2 = 244$ degrees of freedom (d.f.), as shown by the solid curve. The extraordinarily close fit of the theoretical binomial to the observed distribution is a consequence of the fact that each comparison between two bits of phase sequence information from two different irides is essentially a Bernoulli trial, or coin toss. The mean value near 0.5 indicates an absence of correlations

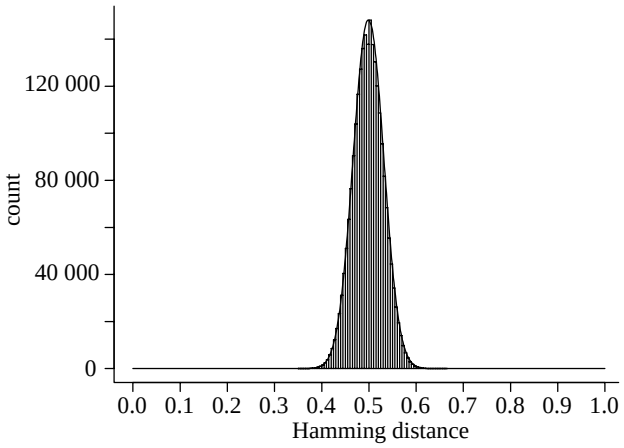

Figure 3. Distribution of Hamming distances (fraction of disagreeing bits) obtained from all 2.3 million possible comparisons between different pairs of irides in our database. The histogram forms a perfect binomial distribution with $p = 0.5$ and $N = 244$ d.f., as shown by the solid curve (equation 3.1). The data imply that it is extremely improbable for two different irides to disagree in less than about one-third of their detailed phase sequences. Mean = 0.499, s.d. = 0.032, minimum = 0.353, maximum = 0.661.

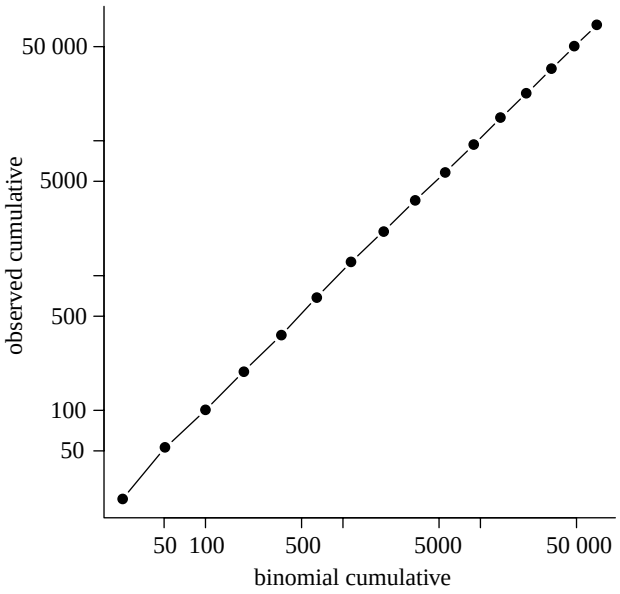

Figure 4. Quantile–quantile plot of the observed cumulatives under the left tail of the histogram in figure 3, versus the predicted cumulatives under a theoretical binomial distribution (244 d.f., $p = 0.5$). Their close agreement over several orders of magnitude strongly confirms the binomial model.

among different iris patterns, and the fact that the smallest HD score observed among the 2.3 million was 0.353 implies that it is extremely improbable for two unrelated irides to agree by chance in even as much as two-thirds of their phase sequence.

In the phase code for any given iris, only small subsets of bits are mutually independent due to the internal correlations, especially radial, within an iris. (If all $N = 2048$ phase bits were independent, then the distribution in figure 3 would be very much sharper, with an expected standard deviation of only $\sqrt{p(1-p)/N} = 0.011$ and so the HD interval between 0.49 and 0.51 would contain most of the distribution.) Bernoulli trials that are correlated (Viveros *et al.* 1984) are still binomially distributed but with a reduction in N, their d.f., and hence an increase in σ. The form and width of the HD distribution in figure 3 tell us that the amount of difference between phase bit sequences for different irides is distributed equivalently to runs of 244 tosses of a fair coin. Expressing this variation as a discrimination entropy (Cover & Thomas 1991), and using typical iris and pupil diameters, the observed amount of statistical variability among different iris patterns corresponds to an information density of about 3.2 bits mm^{-2} on the iris.

The theoretical binomial distribution plotted as the solid curve in figure 3 has the fractional functional form

$$f(x) = \frac{\mathcal{N}!}{m!(\mathcal{N}-m)!}\; p^m (1-p)^{(\mathcal{N}-m)}, \qquad (3.1)$$

where $\mathcal{N} = 244$, $p = 0.5$, and $x = m/\mathcal{N}$ is the outcome fraction of $\mathcal{N}$ Bernoulli trials (or coin tosses that are 'heads' in each run). In our case, x is the HD, the fraction of phase bits that happen to disagree when two different sequences are compared. To validate such a statistical model we must also study the behaviour of the tails, by making quantile–quantile plots of the observed cumulatives versus the theoretically predicted cumulatives from zero up to sequential points in the tail. Such a 'Q–Q' plot is given in figure 4. It reveals very precise agreement between theory and data, over a range of more than three orders of magnitude. The huge factors attenuating the tails of binomial distributions by combinatorics make it extremely improbable that two different irides might have phase sequences which disagree in less than one-third of their bits, as is clear from figure 3. Computing the cumulative of $f(x)$ from zero to 0.333 indicates that the probability of such an event is one in 10 million. The cumulative from zero to just 0.300 (i.e. more than 70% agreement) is about one in 7 billion, and the quarter range cumulative, from zero to 0.25 (i.e. more than 75% agreement), is a vanishing 10^{-15}. Thus, even the observation of a relatively poor degree of match (say 70%, or HD = 0.30) between the phase sequences for two different iris images would still provide extraordinarily compelling evidence of identity, because the test of statistical independence between them is still failed so convincingly.

4. GENETICALLY IDENTICAL IRIDES

Finally, we compared genetically identical eyes in the same manner, to discover the degree to which their textural patterns were correlated and hence genetically determined. A convenient source of genetically identical irides are the right and left pair from any given person; they have the same genetic relationship as the four irides of monozygotic twins or the prospective $2\mathcal{N}$ irides of $\mathcal{N}$ clones. Although eye colour is, of course, strongly determined genetically, as is overall iris appearance, we found the pattern details of genetically identical eyes to be as uncorrelated as they are among unrelated eyes. This is

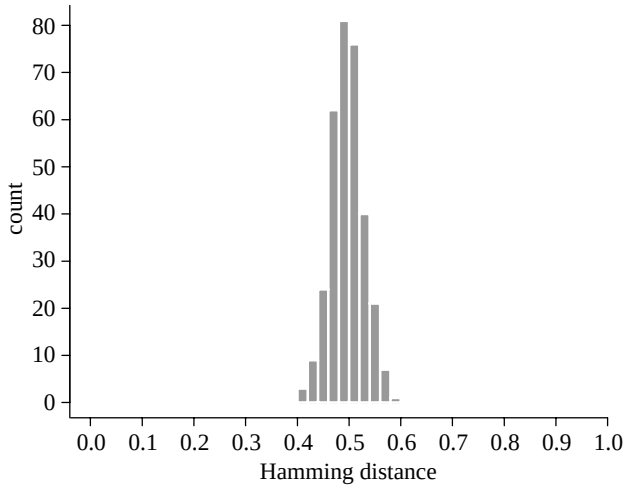

Figure 5. Distribution of Hamming distances between genetically identical irides, in 648 paired eyes from 324 persons. The data are statistically indistinguishable from those shown in figure 3 comparing unrelated irides. Unlike eye colour, the phase structure of iris patterns therefore appears to be epigenetic, arising from random events and circumstances in the morphogenesis of this tissue. Mean = 0.497, s.d. = 0.031.

shown in figure 5, comparing 648 right/left iris pairs from 324 persons. Their observed mean HD was 0.497 with standard deviation (s.d.) 0.031, a result that is statistically indistinguishable from the distribution shown in figure 3 for unrelated eyes. A set of six pairwise comparisons among the eyes of actual monozygotic twins also yielded a result (mean HD = 0.507) expected for unrelated eyes.

5. DISCUSSION

The estimate of 244 d.f. differentiating the patterns of different irides should be regarded as a lower bound. It is possible that computational encoding methods other than those described here would extract still further independent feature dimensions, thereby inferring a larger number of degrees of freedom and a higher discrimination entropy. In any case, the 95% confidence interval for our estimate extends from 227 to 262 d.f.

Notwithstanding this diversity among iris patterns and their apparent singularity because of so many dimensions of random variation, their utility as a basis for automatic personal identification would depend upon their relative stability over time. There is a popular belief that the iris systematically reflects one's health or personality, and even that its detailed features reveal the state of individual organs ('iridology'), but such claims have been discredited (Berggren 1985) as medical fraud. Measuring the repeatability of iris pattern encoding in 7070 comparisons among same-eye images, our mean HD was 0.110 ± 0.065, due to factors such as imaging variation, motion blur, reflections, and CCD noise. In any case, as shown by figure 3, the recognition principle described is intrinsically tolerant of a large proportion of the iris information being corrupted without significantly impairing the inference of personal identity, even in very large database searches, by this simple test of statistical independence.

REFERENCES

Berggren, L. 1985 Iridology: a critical review. *Acta Ophthalmol.* **63**, 1–8.

Cover, T. and Thomas, J. 1991 *Elements of information theory.* New York: Wiley.

Daugman, J. G. 1980 Two-dimensional spectral analysis of cortical receptive field profiles. *Vis. Res.* **20**, 847–856.

Daugman, J. G. 1988 Complete discrete 2D Gabor transforms by neural networks for image analysis and compression. *IEEE Trans. Acoustics Speech Signal Processing* **36**, 1169–1179.

Daugman, J. G. 1993 High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Machine Intelligence* **15**, 1148–1161.

Daugman, J. G. & Downing, C. J. 1995 Demodulation, predictive coding, and spatial vision. *J. Opt. Soc. Am. A* **12**, 641–660.

Davenport, G. & Davenport, C. 1907 Heredity of eye-color in man. *Science* **26**, 589–592.

Imesch, P. D., Wallow, I. & Albert, D. M. 1997 The color of the human eye: a review of morphologic correlates and of some conditions that affect iridial pigmentation. *Surv. Ophthalmol.* **41** (suppl 2), 117–123.

Kronfeld, P. C. 1962 Gross anatomy and embryology of the eye. In: *The eye*, vol. 1 (ed. H. Davson). London: Academic Press.

Viveros, R., Balasubramanian, K. & Balakrishnan, N. 1984 Binomial and negative binomial analogues under correlated Bernoulli trials. *Am. Stat.* **48**, 243–247.

Wolff, E. 1948 *The anatomy of the eye and orbit.* London: H. K. Lewis.