Technical Report

UCAM-CL-TR-482
ISSN 1476-2986

Number 482

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Biometric decision landscapes

John Daugman

January 2000

15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom
phone +44 1223 763500

http://www.cl.cam.ac.uk/

Technical reports published by the University of Cambridge
Computer Laboratory are freely available via the Internet:

http://www.cl.cam.ac.uk/TechReports/

Series editor: Markus Kuhn

ISSN 1476-2986

Biometric Decision Landscapes

John Daugman University of Cambridge The Computer Laboratory[1]

Abstract

This report investigates the "decision landscapes" that characterize several forms of biometric decision making. The issues discussed include: (i) Estimating the degrees-of-freedom associated with different biometrics, as a way of measuring the randomness and complexity (and therefore the uniqueness) of their templates. (ii) The consequences of combining more than one biometric test to arrive at a decision. (iii) The requirements for performing identification by large-scale exhaustive database search, as opposed to mere verification by comparison against a single template. (iv) Scenarios for Biometric Key Cryptography (the use of biometrics for encryption of messages). These issues are considered here in abstract form, but where appropriate, the particular example of iris recognition is used as an illustration. A unifying theme of all four sets of issues is the role of combinatorial complexity, and its measurement, in determining the potential decisiveness of biometric decision making.

Keywords – Statistical decision theory, pattern recognition, biometric identification, combinatorial complexity, iris recognition, Biometric Key Cryptography.

1 Yes/No Decisions

Biometric identification fits squarely in the classical framework of statistical decision theory. This formalism emerged from work on statistical hypothesis testing[1] in the 1920s - 1930s and on radar signal detection analysis[2] in World War II, and its key elements are briefly summarized here in Figures 1 and 2. For decision problems in which prior probabilities are not known, error costs are not fixed, but posterior distributions are known, the formalism of Neyman and Pearson[1] provides not only a mechanism for making decisions, but also for assigning confidence levels to such decisions and for measuring the overall "decidability" of the task.

Yes/No pattern recognition decisions have four possible outcomes: either a given pattern is, or is not, in fact the target; and in either case, the decision made by the recognition algorithm may be either the correct or the incorrect one. In a biometric decision context the four possible outcomes are normally called False Accept (*FA*), Correct Accept (*CA*), False Reject (*FR*), and Correct Reject (*CR*). Obviously the first and third outcomes are errors (called Type I and Type II respectively), whilst the second and fourth outcomes are the ones sought. By manipulating the decision criteria, the relative probabilities of these four outcomes can be adjusted in a way that reflects their associated costs and benefits. These may be very different in different applications. In a customer context the cost of a *FR* error may exceed the cost of a *FA* error, whereas just the opposite may be true in a military context.

It is important to note immediately the uselessness of either error rate statistic alone in characterizing performance. Any arbitrary system can achieve a *FA* rate of 0 (just by rejecting all candidates). Similarly it can achieve a *FR* rate of 0 (just by accepting all candidates). The notion of "decision landscape" is intended to portray the degree to which any improvement in one error rate must be paid for by a worsening in the other. This concept facilitates the definition of metrics quantifying the intrinsic decidability of a recognition problem, and this can be useful for comparing different biometric approaches and understanding their potential.

[1]Cambridge CB2 3QG, England. tel +44 1223 334501 fax +44 1223 334679 john.daugman@CL.cam.ac.uk

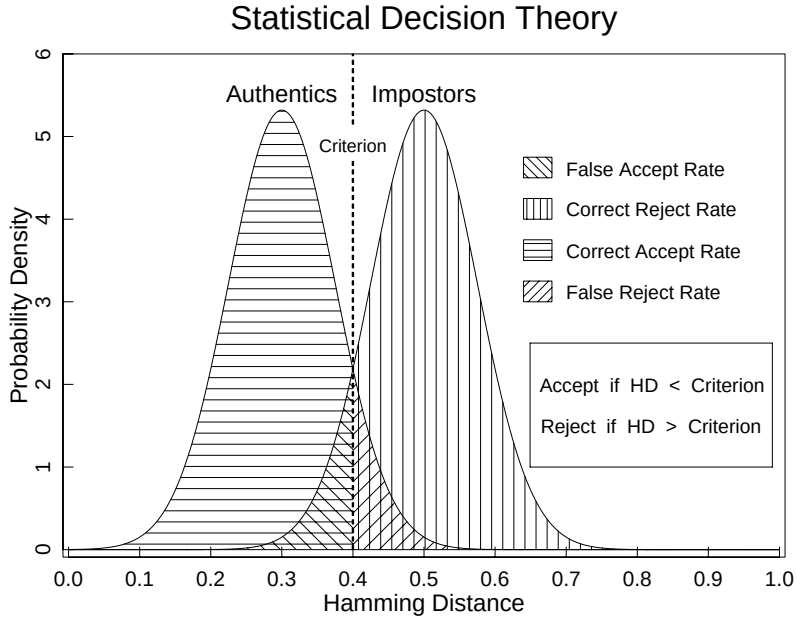

Figure 1: Decision landscape: general formalism for biometric decision making.

Figure 1 illustrates the idea of the decision landscape. The two distributions represent the two states of the world, which are imperfectly separated. The abscissa is any metric of similarity or dissimilarity; in this case it happens to be Hamming Distance, which is the fraction of bits that differ between two binary strings. A decision about whether they are instances of the same pattern (albeit somewhat corrupted), or completely different patterns, is made by imposing some decision criterion for similarity as indicated by the dotted line. Similarity up to some Hamming Distance (0.4 in this case) is deemed sufficient for regarding the patterns as the same, but beyond that point, the patterns are declared to be different.

The likelihoods that these are correct decisions, or not, correspond to the four stippled areas that lie under the two probability distributions on either side of the decision criterion. It is clear that moving the decision criterion to the right or left (becoming more liberal or more conservative) will change the relative likelihoods of the four outcomes. It is also clear that the "decidabiity" of a Yes/No decision problem is determined by how much overlap there is between the two distributions. The problem becomes more decidable if their means are further apart, or if their variances are smaller. One measure of decidability, although not the only possible one, is d' (*d-prime*), defined as follows if the means of the two distributions are μ_1 and μ_2 and their two standard deviations are σ_1 and σ_2:

$$d' = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{1}{2}(\sigma_1^2 + \sigma_2^2)}} \tag{1}$$

(Note that d' has the units of Z-score: distances are marked off in units of a conjoint standard deviation.) A shortcoming of the d' statistic is that it ignores moments higher than second-order, and it becomes less informative if distributions depart significantly from modal form. Nevertheless, it can be a useful gauge for assessing different decision landscapes. It has the virtue of quantifying, in a single number, the intrinsic decidability of a decision task in a way that is independent of the chosen decision criterion. It assesses the degree of inevitable trade-off between the two error rates. Because it measures the separation between the two distributions defining the decision landscape, the higher it is, the better. In the schematic of Figure 1, $d' = 2$.

Let us name the two distributions $P_{Im}(x)$ and $P_{Au}(x)$, denoting respectively the probability densities of any measured dissimilarity x (such as a Hamming Distance) arising from two *different* biometric sources ("Impostor"), or from the *same* source ("Authentic"). Then the probabilities of each of the four possible decision outcomes *FA, CR, CA,* and *FR* are equal to the areas under these

two probability distributions on either side of the chosen decision criterion C:

$$P(FA) = \int_0^C P_{Im}(x)dx \tag{2}$$

$$P(CR) = \int_C^1 P_{Im}(x)dx \tag{3}$$

$$P(CA) = \int_0^C P_{Au}(x)dx \tag{4}$$

$$P(FR) = \int_C^1 P_{Au}(x)dx \tag{5}$$

It is clear that these four probabilities separate into two pairs that must sum to unity, and two pairs that are governed by inequalities:

$$P(CA) + P(FR) = 1 \tag{6}$$

$$P(FA) + P(CR) = 1 \tag{7}$$

$$P(CA) > P(FA) \tag{8}$$

$$P(CR) > P(FR) \tag{9}$$

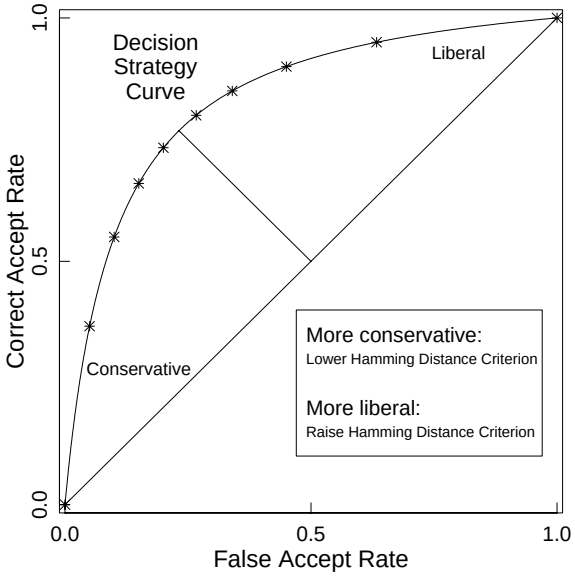

Figure 2: The Neyman-Pearson (ROC) decision strategy curve.

Manipulation of the decision criterion C in the integrals (2) - (5) in order to implement different decision strategies appropriate for the costs of either type of error in a given application, is illustrated schematically in Figure 2. Such a decision strategy diagram, sometimes called a Receiver Operating Characteristic or Neyman-Pearson curve, plots $P(CA)$ from (4) against $P(FA)$ from (2) as a locus of points. Each point on such a curve represents a different decision strategy as specified by a different choice for the operating criterion C, as was indicated schematically in Figure 1.

Inequality (8) states that the Neyman-Pearson strategy curve shown in Figure 2 will always lie above the diagonal line. Clearly, strategies that are excessively liberal or conservative correspond to sliding along the curve towards either of its extremes. Irrespective of where the decision criterion is placed along this continuum (hence how liberal or conservative one wishes to be in a particular application), the overall power of a pattern recognition method may be gauged by how bowed the ROC curve is. The length of the short line segment in Figure 2 is monotonically related to the

quantity d' defined in (1). Ideally one would like a biometric that generates a decision landscape whose ROC curve is extremely bowed, reaching as far as possible into the upper left corner of Figure 2, since reaching that limit corresponds to achieving a Correct Accept rate of 100% whilst keeping the False Accept rate at 0%.

The decision landscape concept that has been illustrated in Figures 1 and 2 and quantified by equations (1) - (9) can be applied to any biometric, regardless of its chosen template or its measure of similarity. Such dual histogram plots should always be provided for a biometric, together with some measure of "separability" or "decidability" such as d' in order to describe the decision landscape more fully than merely citing a *FA* and *FR* rate, either of which can always be brought to 0. It is unfortunate that such fuller portrayals of a biometric's decision landscape are rarely provided.

2 Decisions from Combined Biometrics

In this section we investigate the consequences of combining two or more biometric tests into an enhanced test. There is a common and intuitive assumption that the combination of different tests must improve performance, because "surely more information is better than less information." On the other hand, a different intuition suggests that if a strong test is combined with a weaker test, the resulting decision landscape is in a sense averaged, and the combined performance will lie somewhere between that of the two tests conducted individually (and hence will be degraded from the performance that would be obtained by relying solely on the stronger test).

There is truth in both intuitions. The key to the apparent paradox is that when two tests are combined, one of the resulting new error rates (*FA* or *FR* depending on the combination rule used) becomes *better than that of the stronger of the two tests,* while the other error rate becomes *worse even than that of the weaker of the tests.* If the two biometric tests differ significantly in their power, and each operates at its own cross-over point where $P(FA) = P(FR)$, then combining them actually results in significantly *worse* performance than relying solely on the one, stronger, biometric.

Notation: Two hypothetical independent biometric tests will be considered, named **1** and **2**. For example, **1** might be voice-based verification, and **2** might be fingerprint verification. Each biometric test is characterized by its own pair of error rates at a given operating point, denoted as the error probabilities $P_1(FA)$, $P_1(FR)$, $P_2(FA)$, and $P_2(FR)$:

$P_1(FA)$ = probability of a False Accept using Biometric **1** alone.

$P_1(FR)$ = probability of a False Reject using Biometric **1** alone.

$P_2(FA)$ = probability of a False Accept using Biometric **2** alone.

$P_2(FR)$ = probability of a False Reject using Biometric **2** alone.

There are two possible ways to combine the outcomes of the two biometric tests when forming the conjoint ("enhanced") decision: the Subject may be required to pass both of the biometric tests, or he may be accepted if he can pass at least one of the two tests. These two cases define the disjunctive and conjunctive rules:

Rule A: Disjunction ("OR" Rule) – Accept if either test **1** or test **2** is passed.

Rule B: Conjunction ("AND" Rule) – Accept only if both tests **1** and **2** are passed.

We can now calculate False Accept and False Reject error rates of the combined biometric, both for disjunctive (**Rule A**) and conjunctive (**Rule B**) combinations of the two tests. These new error probabilities will be denoted: $P_A(FA)$, $P_A(FR)$, $P_B(FA)$, and $P_B(FR)$.

Rule A: Disjunction – Accept if either test **1** or test **2** is passed.

If Rule **A** (the "OR" Rule) is used to combine the two tests **1** and **2**, a False Reject can only occur if both tests **1** and **2** produce a False Reject. Thus the combined probability of a False Reject, $P_A(FR)$, is the product of its two probabilities for the individual tests:

$$P_A(FR) = P_1(FR)P_2(FR) \tag{10}$$

(clearly a lower probability than for either test alone). But the probability of a False Accept when using this Rule, which can be expressed as the complement of the probability that neither test **1** nor **2** produces a False Accept, is higher than it was for either test alone:

$$\begin{aligned} P_A(FA) &= 1 - [1 - P_1(FA)]\,[1 - P_2(FA)] &(11)\\ &= P_1(FA) + P_2(FA) - P_1(FA)P_2(FA) &(12)\end{aligned}$$

Rule B: Conjunction – Accept only if both tests **1** and **2** are passed.

If Rule **B** (the "AND" Rule) is used to combine the two tests **1** and **2**, a False Accept can only occur if both tests **1** and **2** produce a False Accept. Thus the combined probability of a False Accept, $P_B(FA)$, is the product of its two probabilities for the individual tests:

$$P_B(FA) = P_1(FA)P_2(FA) \tag{13}$$

(clearly a lower probability than for either test alone). But the probability of a False Reject when using this Rule, which can be expressed as the complement of the probability that neither test **1** nor **2** produces a False Reject, is higher than it was for either test alone:

$$\begin{aligned} P_B(FR) &= 1 - [1 - P_1(FR)]\,[1 - P_2(FR)] &(14)\\ &= P_1(FR) + P_2(FR) - P_1(FR)P_2(FR) &(15)\end{aligned}$$

Example: Combination of two hypothetical biometric tests, one stronger than the other.

Suppose weaker Biometric **1** operates with both of its error rates equal to 1 in 100, and suppose stronger Biometric **2** operates with both of its error rates equal to 1 in 1,000. Thus if 100,000 verification tests are conducted with impostors and another 100,000 verification tests are conducted with authentics, Biometric **1** alone should make a total of 2,000 errors, whereas Biometric **2** alone should make a total of only 200 errors. But what happens if the two biometrics are combined to make an "enhanced" test?

If the "OR" Rule were followed in this same batch of tests, the combined biometric should make 1,099 False Accepts and 1 False Reject, for a total of 1,100 errors. If instead the "AND" Rule is followed, the combined biometric should make 1,099 False Rejects and 1 False Accept, thus again producing a total of 1,100 errors. Either method of combining the two biometric tests produces 5.5 times more errors than if the stronger of the two tests had been used alone.

Conclusion: A strong biometric is better used alone than in combination with a weaker one... when both are operating at their cross-over points. To reap any benefit from the combination, equations (10) - (15) show that the operating point of the weaker biometric must be shifted to satisfy the following criteria: *If the "OR" Rule is to be used, the False Accept rate of the weaker test must be made smaller than twice the cross-over error rate of the stronger test. If the "AND" Rule is to be used, the False Reject rate of the weaker test must be made smaller than twice the cross-over error rate of the stronger test.*

3 Identification Decisions

We now compare the requirements of performing a *verification* (a one-to-one comparison against a single stored template), versus performing an *identification* (a one-to-many comparison against all the enrolled templates in some database containing N impostors). If the presenting template is in fact one of the enrolled templates in the database, the probability of a False Reject when it comes up is obviously the same as in the verification trial. But we are interested now in how much more strenuous the demands against getting a single False Accept need to be, in the case of identification trials involving N other templates.

Notation: Let P_1 = probability of a False Accept in a verification trial. Let P_N = probability of a False Accept in identification trials after an exhaustive search through a database of N unrelated templates. We wish to calculate this.

Clearly the probability of *not* getting a False Accept in any given comparison is $(1 - P_1)$. This must happen N independent times, and so the probability of it not occurring in any of those N comparisons is $(1 - P_1)^N$. Thus the probability of making *at least* one False Accept among those N comparisons is one minus that probability. This is an extremely demanding relationship:

$$\boxed{P_N = 1 - (1 - P_1)^N} \tag{16}$$

Example: Consider a biometric verifier that achieves a 99.9% Correct Rejection performance in one-to-one verification trials. Thus $P_1 = 0.001$, as per equation (7). How will it perform when searching through a database of unrelated templates?

Using equation (16), we see that for the following sizes of databases with N unrelated templates, these will be the probabilities P_N that this biometric makes at least one False Accept:

Database Size	*False Accept probability*
N = 200	$P_N = 18\%$
N = 2,000	$P_N = 86\%$
N = 10,000	$P_N = 99.995\%$

Once the enrolled database size reaches only about 7,000 persons, this biometric actually becomes more likely (99.91%) to produce a False Accept in identification trials than it is to produce a Correct Reject (99.9%) in verification trials.

Conclusion: Identification is vastly more demanding than verification, and even for moderate database sizes, merely "good" verifiers are of no use as identifiers. Observing the approximation that $P_N \approx NP_1$ for small $P_1 << \frac{1}{N} << 1$, when searching a database of size N an identifier needs to be roughly N times better than a verifier to achieve comparable odds against a False Accept.

It has been suggested that this fundamental problem might be overcome by "binning" or "filtering" the database into smaller subsets, thereby reducing the problematical exponent N in (16). For example, fingerprints might be pre-classified into three standard types as whorls, loops, or arches, and then each search could be restricted just to databases of about one-third the full size. But this strategem creates its own problems:

1. Binning is equivalent to "combining multiple tests," as analyzed in the previous section. The additional test here is the classification operation, which itself has its own $P(FR)$, $P(FA)$, ROC curve, and d'. For a Subject to be correctly identified, he must pass both the biometric matching test, and any pre-sorting test which classified his prints. Proponents of such combined schemes often ignore the effect on the *other* error rate when contemplating the benefits

for just *one* of the error rates. In this case, the benefit of reducing the *FA* probability (by reducing the database search size N) is paid for by an increase in the *FR* probability, since any misclassification of a print into the wrong bin must cause a failure to match. It has been reported[3] in tests of four AFIS systems that fingerprint binning classification error rates are in the range of 1% to 5%. Such large increases in failures to match are unacceptable.

2. Even if there were no binning errors, the argument for its benefits requires that $P(FA)$ within a given class (e.g. whorl prints) is no higher than across mixed classes. If instead, as would certainly be the case for matching methods such as optical Moire-pattern analysis, loop prints were only confused with other loops, and whorls with whorls, then there would be no performance benefit from binning. The size of each search database would be smaller, but the probability of false matches within each class would be proportionately higher, since all of the potential false matching prints remain present in the (now smaller) bin.

3. Even if there were no binning errors, and even if the within-bin false match rate were no higher than the across-bin rate, the potential benefit of binning is limited by the number of bins. Their number determines the factor by which the search database size can be reduced. As the numerical examples illustrate, a reduction factor such as 3 or 10 is utterly insufficient.

The only real key to surviving equation (16) when performing large-scale database searches, in which perhaps many millions of templates must be compared in seeking to make an identification while avoiding false matches, is to ensure that P_1 is sufficiently small. For a search database size of N, the single-trial *FA* probability must be significantly smaller than $1/N$. For example, when being compared against a database of 10 million different templates, an "innocent" Subject can only feel 99% certain that he won't be falsely matched with any of them if the single-comparison false match probability is no greater than $P_1 = 10^{-9}$: 1 in a billion.

4 The Degrees-of-Freedom in Biometric Feature Sets

Achieving such demanding confidence levels against false matches when attempting identification decisions against large search databases is only possible if the combinatorics of the biometric are sufficiently vast. The combinatorial complexity of a biometric test can be gauged by its number of degrees-of-freedom, which is essentially the number of independent dimensions of variation, or the number of independent yes/no questions that the biometric decision is based upon. For biometrics that do not compare lists of distinct features but rather use a simple analogue measure such as correlation, the number of degrees-of-freedom is the number of independent data that can be resolved, as limited by (among other things) the imaging quality and the intrinsic auto-correlation within each pattern. This metric of *data size divided by its mean correlation length* corresponds to Hartley's[4] classic definition of the number of degrees-of-freedom in a signal, and to the number of resolvable cells in the Information Diagram proposed by Gabor[5].

Clearly, the larger the number of underlying features that are compared, the less chance that two unrelated templates might just by chance agree in them all. But it also becomes less likely that even their source will be able to produce a perfect match to them all. Therefore it is the combinatorial question of how likely it is that some *proportion* of the features will be matched by chance by different people, and some proportion will fail to be matched even by the same person, that really determines the shape of the decision landscape. The goal of biometric feature encoding is to maximize the number of degrees-of-freedom that will belong to the "impostors" distribution in Figure 1, while minimizing the number that will belong to the "authentics" distribution. Thus for example, in face recognition, one would like to choose feature dimensions that vary the most among different individuals, but that do not vary when a given individual changes expression, pose angle, hairstyle, age, etc.

The number of degrees-of-freedom contained in a wide range of biometrics can be estimated by counting the number of yes/no questions that they ask (e.g. "is this minutia from fingerprint A

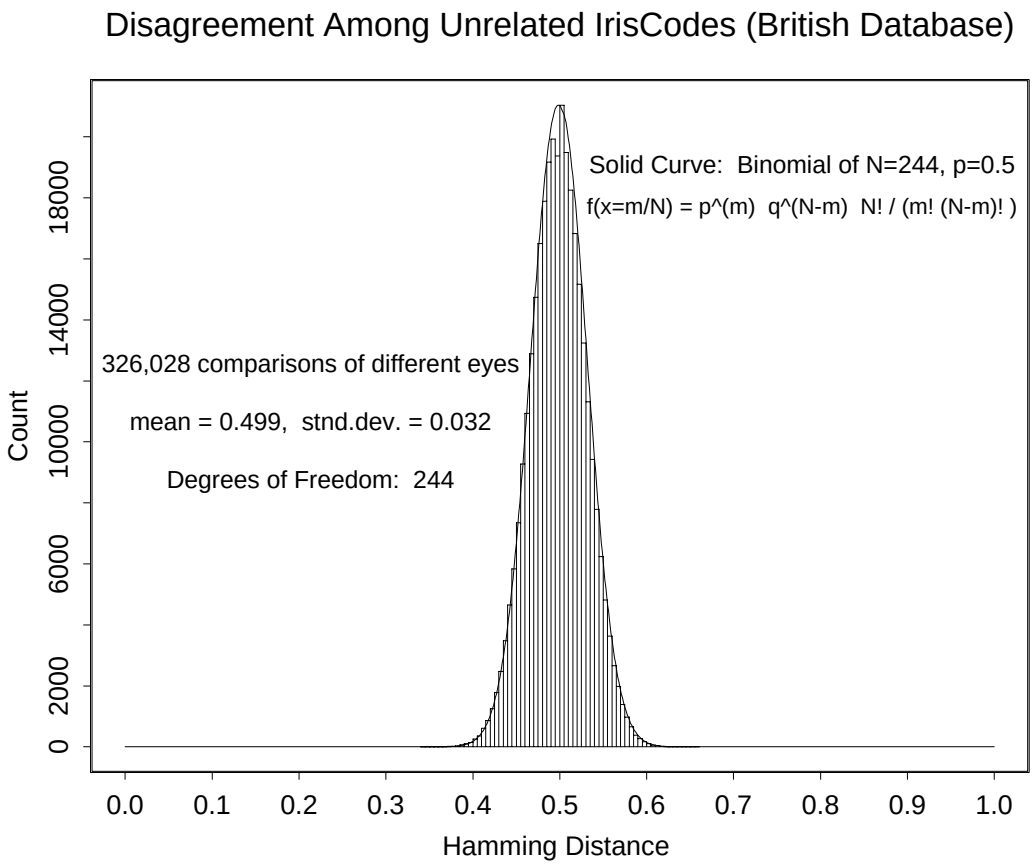

Figure 3: Disagreement among unrelated iris patterns in a British database, providing an estimate of the number of degrees-of-freedom in IrisCodes.

also in B?", or, "does the phase of this local patch in iris A agree with the local phase in B?"), and then measuring the dimension of the subspace of these questions that are independent.

If the thresholds for such yes/no questions can be adjusted so that each has the same probability p of being true, then the proportion of them that are true when unrelated templates are compared will be binomially distributed with parameters (p, N), where N is the size of the subset of questions posed that are independent. (If all the yes/no questions were independent, then N would equal their number.) The probability that some fraction $x = m/N$ of the N independent questions may agree in their answers just by chance between unrelated templates will be distributed according to the density function $f(x)$:

$$f(x = m/N) = \frac{N!}{m!(N-m)!}\, p^m (1-p)^{(N-m)} \tag{17}$$

The mean of such a distribution is $\mu = p$, and its standard deviation is $\sigma = \sqrt{p(1-p)/N}$. In general it is asymmetrical unless it happens that $p = 0.5$, in which case it resembles a Gaussian apart from being discrete rather than continuous, and having strictly finite support (the unit interval for $x = m/N$ in this case.) Although the classic example of a random variable that is binomially distributed is a series of independent Bernoulli trials (N independent coin-tosses with fixed probability p), Viveros[6] *et al.* have pointed out that correlated Bernoulli trials are also distributed as such families of functions but with reduction in N.

An illustration of this principle is seen in Figure 3, which shows the data from 326,028 exhaustive pairwise comparisons among 808 different iris patterns. Each IrisCode contains 2,048 bits of phase data, but these are strongly correlated because of iris radial structure. The fraction of disagreeing bits, or Hamming Distance, between every pair of different IrisCodes is plotted in the histogram, and the solid curve is equation (17) with parameters p and $N = p(1-p)/\sigma^2$ measured from the mean and standard deviation of the data. The quality of the fit between the data and equation (17)

is extraordinary. The calculated value of N is 244, based upon the measurement that $\mu = 0.499$ and $\sigma = 0.032$, and this tells us that among the 2,048 bits computed in an IrisCode, effectively only 244 of them are independent. Hence the number of degrees-of-freedom in IrisCodes acquired with this particular camera focus quality and imaging resolution was 244. The solid curve was computed using Stirling's formula to calculate the large factorials in equation (17):

$$N! \approx e^{N \ln(N) - N + \frac{1}{2} \ln(2\pi N)} \tag{18}$$

which errs by less than 1% in estimating $N!$ for $N \geq 9$. The presence of large factorials causes the binomial probability density to attenuate extremely rapidly in its tails, which in turn means that the cumulative in equation (2) for False Accept probability becomes infinitesimally small when N, the number of degrees-of-freedom in the biometric, is sufficiently large.

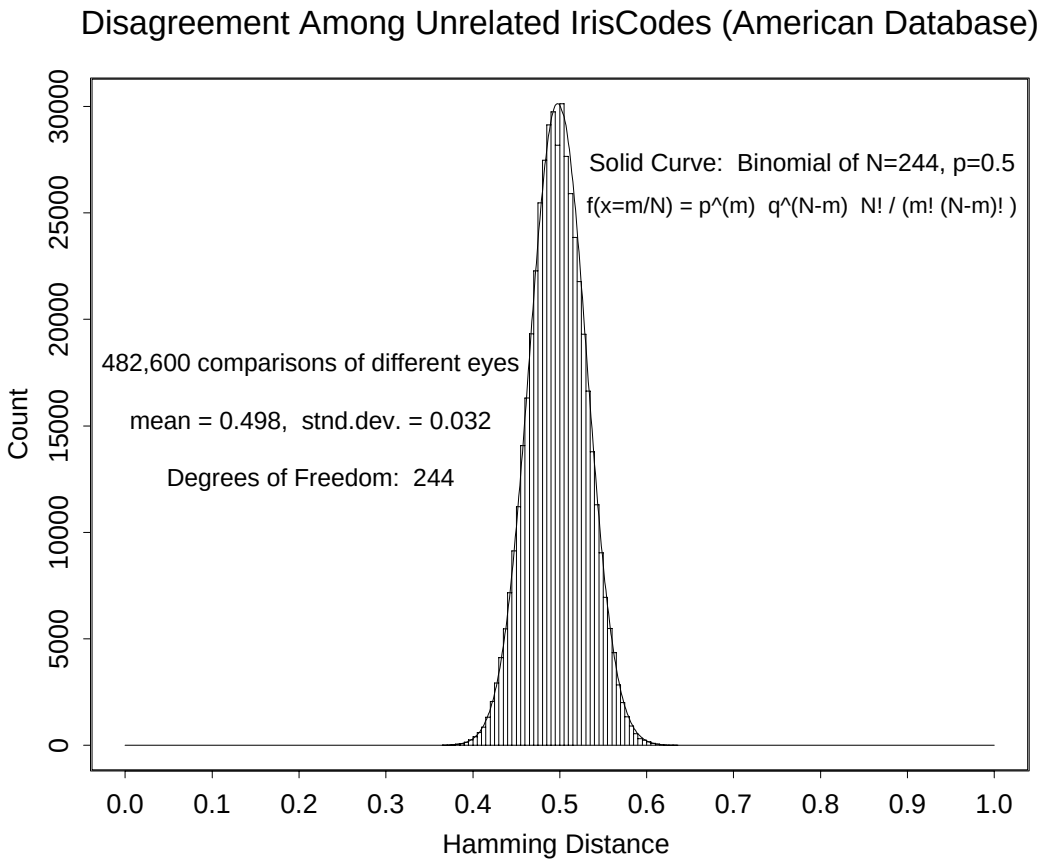

Figure 4: Disagreement among unrelated iris patterns in an American database, providing an estimate of the number of degrees-of-freedom in IrisCodes.

Whereas Figure 3 presented data from 326,028 comparisons among iris patterns acquired in Britain, Figure 4 presents 482,600 such comparisons among an entirely different gallery of iris images acquired on a different optical platform in America. It is once again perfectly described by a binomial, as the solid curve illustrates. But because the optical systems used were quite different, it is only coincidental that the measured number of degrees-of-freedom is again 244, exactly the same as for the data in Figure 3. The actual number can be lower or higher than this (as many as 266 have been reported in one study[7] using the author's algorithms[8]), because variations in imaging focus and resolution affect the auto-correlation within each image.

It would be possible to estimate and compare the number of degrees-of-freedom for different biometrics, using this mapping of biometric feature set comparisons into an equivalent number of Bernoulli trials as described above. Moreover, through this expression of the decision processes as an ensemble of elementary yes/no questions, the different biometrics would all acquire the benefits

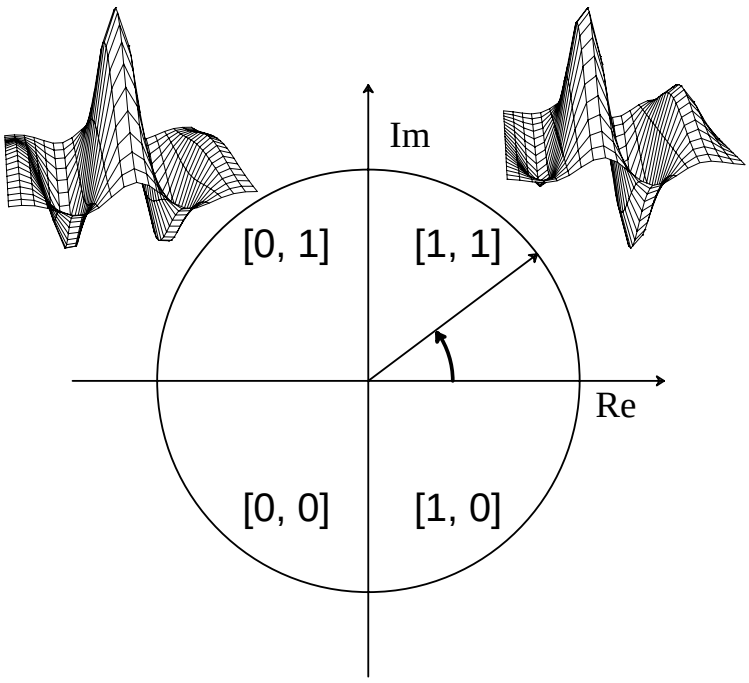

Figure 5: Iris encoding by phase demodulation with complex-valued 2D Gabor wavelets.

of having factorial (binomial-class) tails, i.e. rapidly attenuating densities, instead of exponential or flatter ones. This in turn would facilitate the execution of decision making by a simple test of statistical independence[8]: a Subject becomes highly likely to pass a test of statistical independence when compared with any other person's biometric template, but to fail this same test of statistical independence against his own.

5 Decision Landscape for Iris Recognition

Iris patterns are encoded into "IrisCodes" by a process of phase demodulation[8] employing a two-dimensional generalization[9] of complex Gabor wavelets. These can represent[10] a textured pattern by an ensemble of phasors in the complex plane. In an IrisCode, each phasor angle (Figure 5) is quantized into just the complex quadrant in which it lies for each local patch (r_0, θ_0) of the iris, and this operation is repeated all across the iris, at many different scales (α, β, ω) of analysis. Such local phase quantization is performed by evaluating the real and the imaginary parts of the following integral expression:

$$\int_\rho \int_\phi e^{-i\omega(\theta_0-\phi)} e^{-(r_0-\rho)^2/\alpha^2} e^{-(\theta_0-\phi)^2/\beta^2} I(\rho,\phi)\rho d\rho d\phi \tag{19}$$

where the raw image data is given in a pseudo-polar coordinate system $I(\rho, \phi)$ following spatial localization of an iris using methods described in earlier papers[8]. Bits in an IrisCode are set on the basis of whether the real and imaginary parts of (19) are positive or negative. These form the elementary "yes/no questions" that underlie decision making with iris patterns. All currently available systems for iris recognition are based upon the author's algorithms as described here, in licensed executable code.

The ability of phase demodulation to extract enough degrees-of-freedom from iris patterns to permit their reliable recognition is summarized in Figure 6. This is based upon the same set of 482,600 iris comparisons as shown in Figure 4 but after rotations to each of 7 angles (to correct for head tilt and for imaging through a pan/tilt mirror), keeping only the best value. This selection of the lowest Hamming Distance among 7 rotated comparisons skews the impostors distribution to

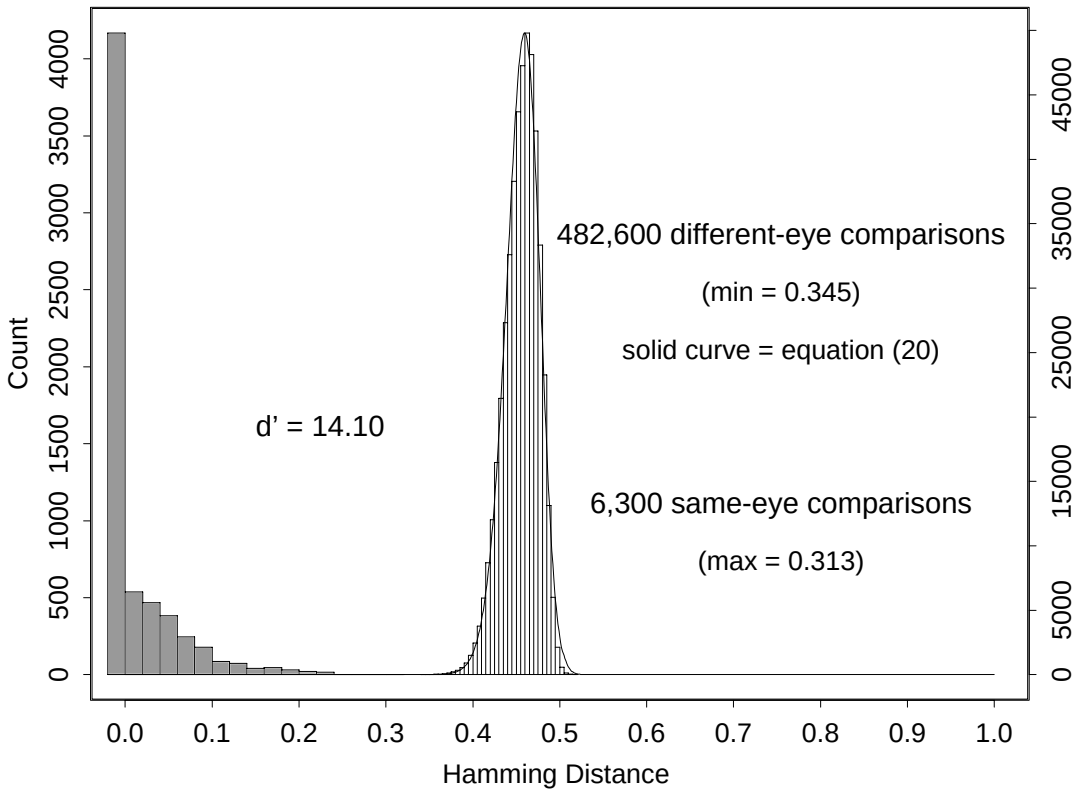

Figure 6: Decision landscape for iris recognition using the same (American) gallery as used in Figure 4, but allowing several image rotations to find each closest fit. The histograms show 482,600 different-eye comparisons, and 6,300 same-eye comparisons. The solid curve is equation (20).

the left, but its functional form remains perfectly described by the appropriate density expression (plotted as the solid curve) which is

$$f_n(x) = nf(x)\,[1 - F(x)]^{n-1} \tag{20}$$

where $f(x)$ is the simple binomial density function given in equation (17), $F(x)$ is its cumulative from 0 to x, and $n = 7$ rotations. The overal decidability for iris recognition in this decision landscape is measured at $d' = 14.1$ as defined by equation (1).

We can now compute $P(FA)$, the probability of False Accept in single comparisons, using the impostors distribution on the right side of Figure 6 as $P_{Im}(x)$ in equation (2). The lowest Hamming Distance actually observed in this empirical series of 482,600 comparisons after 7 rotations each was 0.345. But because of the principled reasons making $f(x)$ in Figure 4 a binomial (17), we may just integrate the expression in (20) for $f_n(x)$ (plotted as the solid curve in Figure 6) from $x = 0$ up to various possible decision criteria C as per (2) in order to calculate the theoretical $P(FA)$ for still lower decision criteria. Such cumulatives are presented in the following table:

HD Decision Criterion C	*False Accept probability*
0.35	1 in 105,000
0.333	1 in 1.7 million
0.30	1 in 985 million
0.25	1 in 10^{14}

We see now that the extremely rapid attenuation of the cumulatives in (2) when there are enough degrees-of-freedom (244 in the case of this data set) allows huge confidence against False Matches, while at the same time tolerating a 25% to 30% disagreement within acceptable biometric matches.

In short, this table shows how biometric identification decisions can survive equation (16). The essence of biometric decision making is combinatorics, and the fundamental strength of a biometric is its combinatorial complexity.

6 Biometric Key Cryptography

We turn finally to a novel potential application of these principles, which is the use of biometrics as encryption keys. We coin the name "Biometric Key Cryptography," or BKC, for this subject. Its purpose is to encrypt data so that only the possessor of (say the encrypting IrisCode) can decrypt it again and thereby "see" it; and also to find means for sending biometrically-secured communications over non-secure channels. Security has many goals besides user authentication, including: integrity certification (non-corruption of data or a message); non-repudiation (e.g. of an e-commerce transaction[11]); and encryption that does not rely on secret keys (which can be stolen, broken, or lost). Cryptography and security are very mature subjects, with elaborate protocols already in place that are deeply understood with formal models. However, these protocols have not yet seriously incorporated biometric keys.

A biometric signature could play a role in the key-generation process, or even serve as the encryption key itself, with the advantage that the user doesn't have to remember it but can always produce it. Let us consider the following trivial example, in which Alice wishes to send Bob a secret message M over a non-secure two-way channel. The operator $\bigotimes$ signifies bitwise Exclusive-OR, and A is Alice's IrisCode while B is Bob's:

message from Alice to Bob	*message from Bob to Alice*
$A \bigotimes M$	
	$B \bigotimes [A \bigotimes M]$
$A \bigotimes (B \bigotimes [A \bigotimes M]) = B \bigotimes M$	

In a final step, Bob now is able to decode Alice's message M by yet once again Exclusive-OR'ing her transmission to him with his own IrisCode B, since $[B \bigotimes (B \bigotimes M)] = M$.

As presented above, this protocol is fatally flawed if no hashes are used because an attacker who eavesdrops on all three communications can extract A, B, and M, simply by Exclusive-OR'ing the messages from Alice to Bob, with that from Bob to Alice. But if the probabilistic character of biometric signatures (the fact that each bit has some error probability associated with it) is exploited, together with the selection of random subsets of A and B for the $\bigotimes$ operations, it is possible to contemplate schemes which cloak these communications yet allow decryption and error-correction by the recipient.

Classically, with few exceptions[12] cryptographic security protocols require that every single bit of a cipher key be correct, or else the ciphertext remains completely unintelligible. But the individual bits in biometric templates are notoriously unreliable; we saw in Figure 6 that up to 30% of the bits might be wrong in an authentic template, due to factors such as sensor noise, physiological variability, or poor imaging focus. Yet as we also saw earlier (Figure 6 and its associated *FA* Table), the combinatorics of IrisCodes have the consequence that there is a "ball" in 2,048-dimensional Boolean space, whose radius is 0.30, centered on any given IrisCode, within which the probability of an intrusion by any other IrisCode is about 10^{-9}. This extreme sparsity in the populating of IrisCode space invites efficient vector-quantization that would both probabilistically ensure key uniqueness (no two different IrisCodes can occupy the same ball with $p > 10^{-9}$), yet at the same time allow the unreliability of the individual IrisCode bits to be exploited positively. The synthesis of existing encryption protocols with these probabilistic properties of biometric templates and their associated decision landscapes, opens new avenues for research.

References

1. J. Neyman and E.S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Phil. Trans. Roy. Soc. London,* Series A, Vol. 231, 1933, pp. 289-337.

2. W.W. Peterson, T.G. Birdsall, and W.C. Fox, "The theory of signal detectability," *Trans. I.R.E.* PGIT-4, 1954, pp. 171-212.

3. J.L. Wayman, "Technical testing and evaluation of biometric identification devices," in *Biometrics: Personal Identification in Networked Society* (A. Jain, R. Bolle, S. Pankanti, eds), Kluwer, Dordrecht, 1999, pp. 345-368.

4. R.V.L. Hartley, "Transmission of information," *Bell Syst. Tech. J.* Vol. 7, 1928, pp. 535-563.

5. D. Gabor, "Theory of communication," *J. Inst. Electr. Eng.* Vol. 93, 1946, pp. 429-457.

6. R. Viveros, K. Balasubramanian, and N. Balakrishnan, "Binomial and negative binomial analogues under correlated Bernoulli trials," *Am. Stat.* Vol. 48, No. 3, 1984, pp. 243-247.

7. C. Seal, M. Gifford, and D. McCartney, "Iris recognition for user validation," *British Telecommunications Engineering Journal* Vol. 16, No. 7, 1997, pp. 113-117.

8. J. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Analysis and Machine Intelligence* Vol. 15, No. 11, 1993, pp. 1148-1161.

9. J. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Am. A* Vol. 2, No. 7, 1985, pp. 1160-1169.

10. J. Daugman, "Complete discrete 2D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoustics, Speech, and Signal Processing* Vol. 36, No. 7, 1988, pp. 1169-1179.

11. M.M. Gifford, D.J. McCartney, and C.H. Seal, "Networked biometrics systems – requirements based on iris recognition," *British Telecommunications Engineering Journal*, Vol. 17, No. 2, 1999, pp. 163-169.

12. S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comp. Sys. Sci.* Vol. 28, 1984, pp. 270-299.