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

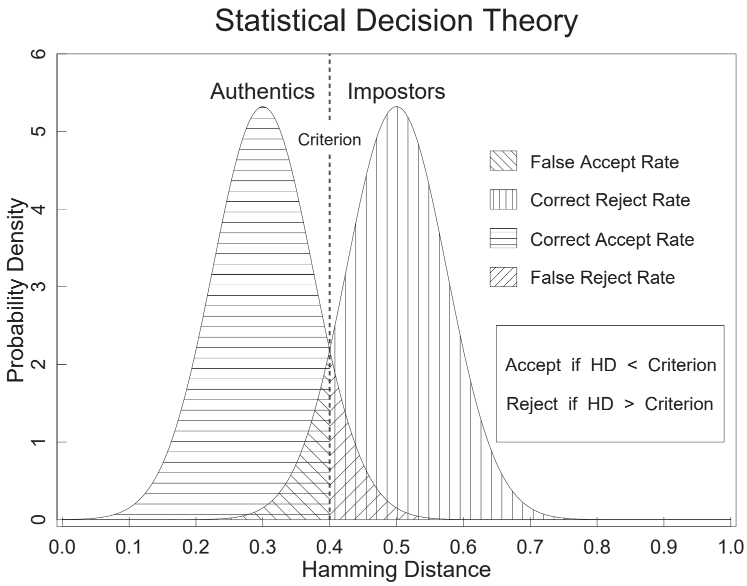

Figure 1: Decision landscape: general formalism for biometric decision making.

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:

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

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

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

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

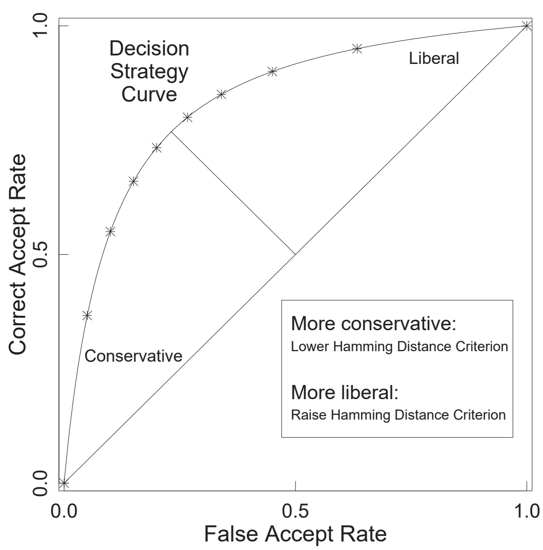

Figure 2: The Neyman-Pearson (ROC) decision strategy curve.

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

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming inter-face) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**

fastcase® Smarter legal research.