Technical Report

UCAM-CL-TR-635
ISSN 1476-2986

Number 635

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Results from 200 billion iris cross-comparisons

John Daugman

June 2005

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/

ISSN 1476-2986

Results from 200 billion iris cross-comparisons

John Daugman

Abstract

Statistical results are presented for biometric recognition of persons by their iris patterns, based on 200 billion cross-comparisons between different eyes. The database consisted of 632,500 iris images acquired in the Middle East, in a national border-crossing protection programme that uses the Daugman algorithms for iris recognition. A total of 152 different nationalities were represented in this database. The set of exhaustive cross-comparisons between all possible pairings of irises in the database shows that with reasonable acceptance thresholds, the False Match rate is less than 1 in 200 billion. Recommendations are given for the numerical decision threshold policy that would enable reliable identification performance on a national scale in the UK.

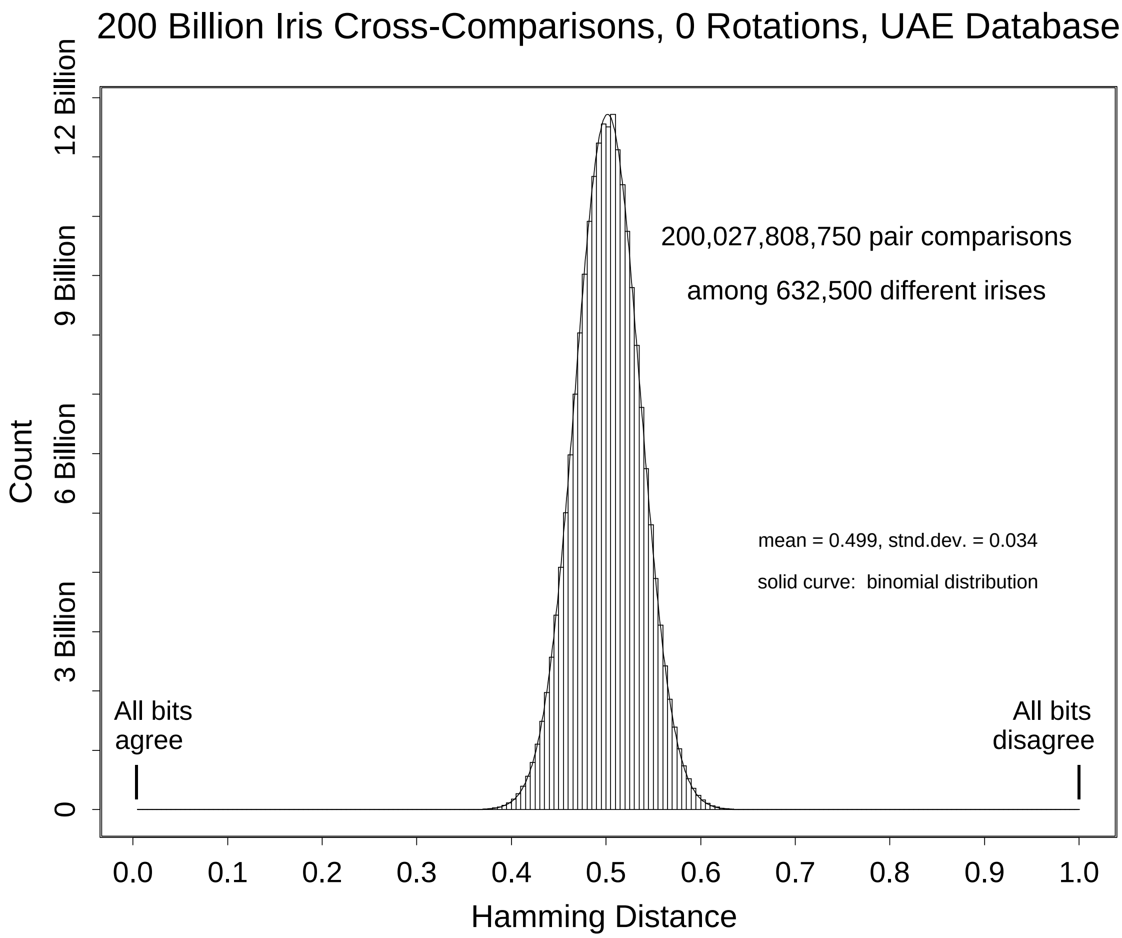
1 Introduction

In 2001 the United Arab Emirates (UAE) Ministry of Interior launched a national border-crossing security programme that is today deployed at all 17 of the UAE's air, land, and sea ports. It is based upon mathematical analysis of the random patterns visible in the iris of a person's eye (*iris recognition*), using algorithms developed by the author of this report and cameras made by LG. All deployments of iris recognition worldwide use these same Daugman algorithms, and most use the same LG cameras; but a valid criticism that has often been raised in discussions of other possible national-scale deployments is the absence of large-scale statistical test reports. For example, although the algorithms have been deployed successfully at Schiphol Airport since 2001 for biometrically controlled entry into The Netherlands without passport inspection, no statistical analysis of performance has been published. Several smaller-scale tests of the iris recognition algorithms have consistently shown them to produce no False Matches (a bibliography of reports is available at: `http://www.CL.cam.ac.uk/users/jgd1000/iristests.pdf`), but those tests involved relatively small databases, and the iris images were often acquired in laboratory rather than real-world conditions.

In the UAE border-crossing deployment, nearly 2 trillion (2 million-million) iris comparisons have been performed to date, as all foreign nationals visiting the Emirates have their irises compared against all the *IrisCodes* (mathematical descriptions of registered iris patterns) stored in a central database. Some 40,000 persons have thereby been caught trying to re-enter the UAE with false travel documents since this deployment began. The Abu Dhabi Directorate of Police report that so far there have been no False Matches; yet it was desired to exploit the large enrollment database to understand better the statistical powers of iris recognition technology. For example, by computing the similarities between all possible pairings of different irises in the database, much could be learned about the robustness of the algorithms against making any False Matches when there are such vast numbers of opportunities for making False Matches. This would help to illuminate the technology's potential for even larger-scale national deployments. UAE Interior Minister Sheikh Saif Bin Zayyed therefore graciously consented to make the enrollment database of IrisCodes available to the University of Cambridge for detailed analysis. This report is the result of that analysis. The iris recognition algorithms have been described previously in detailed aspects in Daugman(1993, 2001, 2003, 2004) and will not be reviewed here.

2 Cross-comparison results

Statistical analysis of the decision-making power of iris identification requires examining the distribution of similarity scores obtained from comparing different irises, and likewise the scores obtained from comparing different images of same irises. Ideally these two distributions should be well-separated, with rapidly decaying "tails," since it is the overlap of the distributions' tails that determines error rates.



The above distribution shows all cross-comparison similarity scores obtained from making all possible pair comparisons amongst 632,500 different irises. N different objects can generate a total of $N \cdot (N-1)/2$ different pairings, which for this database means

200,027,808,750, or about 200 billion different pair comparisons. The similarity metric is *Hamming Distance,* which is simply the fraction of bits that disagree between any two IrisCodes. Thus if two IrisCodes happened to agree in every bit, their Hamming Distance would be 0 and they would score a count at the extreme left in the above distribution. If they disagreed in every bit, they would score a count at 1.0 on the extreme right. The vast majority of IrisCodes from different eyes disagreed in about 50% of their bits, as expected since the bits are equiprobable and uncorrelated between different eyes. Very few pairings of IrisCodes could disagree in fewer than 35% or more than 65% of their bits. The smallest and largest Hamming Distances found in this set of 200 billion simple raw comparisons of different IrisCodes were 0.265 and 0.750 respectively.

The solid curve that fits the data extremely closely in the above plot is a binomial probability density function. This theoretical form was chosen because comparisons between bits from different IrisCodes are Bernoulli trials, or conceptually "coin tosses," and Bernoulli trials generate binomial distributions. If one tossed a coin whose probability of "heads" is p (typically but not necessarily $p = 0.5$) in a series of N independent tosses and counted the number m of "heads" outcomes, and if one tallied this fraction $x = m/N$ in a large number of such repeated runs of N tosses, then the resulting distribution of x would follow the solid curve in the figure above and be described by this equation:

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

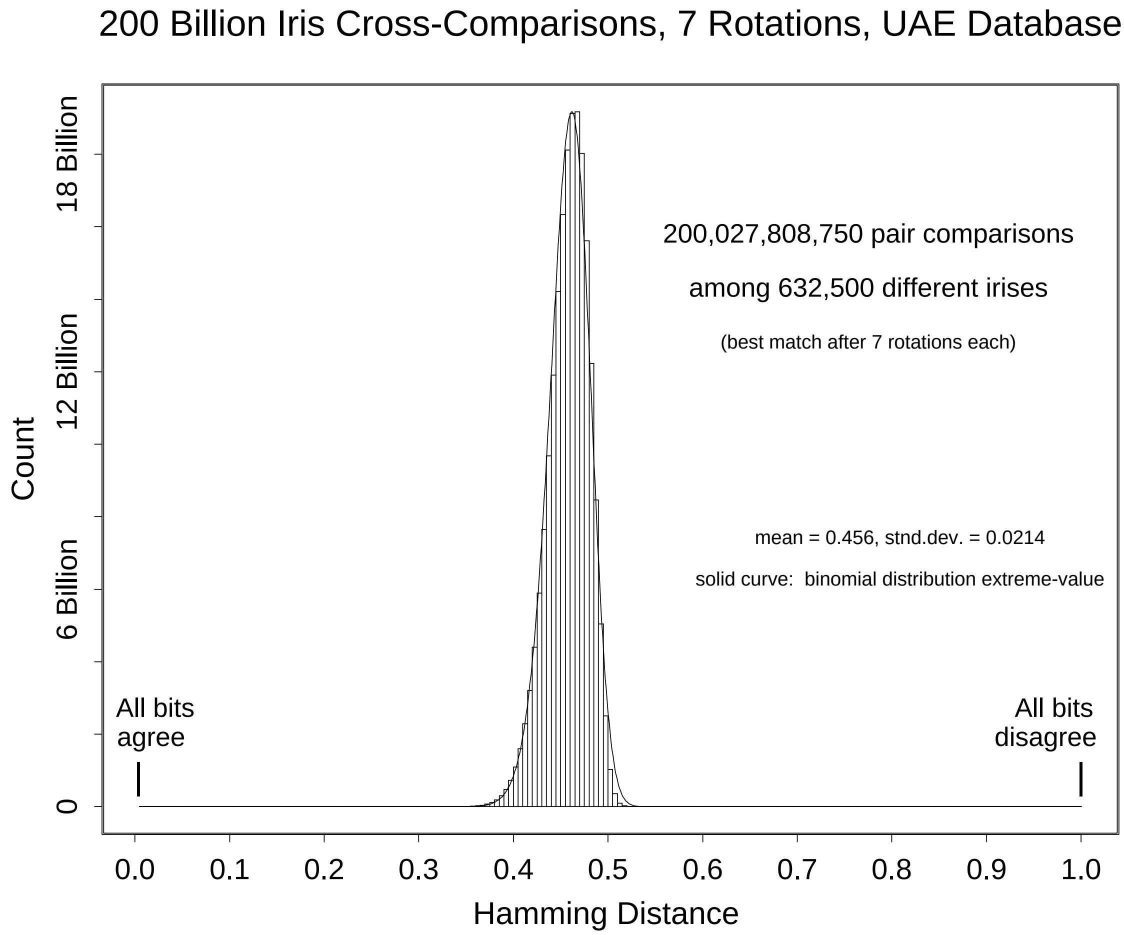
The analogy between tossing coins and comparing bits between different IrisCodes is deep but imperfect, because any given IrisCode has internal correlations arising from iris features, especially in the radial direction. Their effect is to reduce the value of N to a number significantly smaller than the number of bits that are actually compared between two IrisCodes; N becomes the number of effectively independent bit comparisons. The value of p is very close to 0.5 (empirically 0.499 for this database), because the states of each bit are equiprobable *a priori*, and so any pair of bits from different IrisCodes are equally likely to agree or disagree.

Different people expose different amounts of iris between their eyelids, and the amount visible depends also on circumstances. Therefore the number of bits available for comparison between two different IrisCodes is quite variable. A close match (say a Hamming Distance of $HD = 0.100$) based on only few compared bits is much less indicative of identity than an apparently poorer match (say $HD = 0.200$) based on a large number of compared bits. In the coin-tossing analogy, obtaining very imbalanced outcomes just by chance is much more likely in a short series of tosses than in a long series. This requires a renormalisation of any observed raw Hamming Distance score HD_{raw} into one HD_{norm} whose deviation from 0.5 has been re-scaled for statistical significance, based on the number of bits n that were actually compared between the two IrisCodes:

$$HD_{\text{norm}} = 0.5 - (0.5 - HD_{\text{raw}})\sqrt{\frac{n}{911}} \tag{2}$$

The parameters in the above equation influence the standard deviation of the distribution of normalised Hamming Distance scores, and they give the distribution a stable form which permits a stable decision criterion to be used.

When IrisCodes are compared in a search for a match, it cannot be known precisely what was the amount of head tilt, camera tilt, or eye rotation when the IrisCodes were obtained. Therefore it is necessary to make comparisons over a reasonable range of relative tilts (rotations) between every pair of IrisCodes, keeping the best match as their similarity score. This generates a new distribution that is skewed towards lower Hamming Distances even between unrelated irises.



The new distribution after k rotations of IrisCodes in the search process still has a simple analytic form that can be derived theoretically. Let $f_0(x)$ be the raw density distribution obtained for the *HD* scores between different irises after comparing them only in a single relative orientation; for example, $f_0(x)$ might be the binomial defined in Eqn (1). Then $F_0(x) = \int_0^x f_0(x)dx$, the cumulative of $f_0(x)$ from 0 to x, becomes the probability of getting a False Match in such a test when using *HD* acceptance criterion x. Clearly, then, the probability of *not* making a False Match when using decision criterion x is $1 - F_0(x)$ after a single test, and it is $[1 - F_0(x)]^k$ after carrying out k such tests independently at k different relative orientations. It follows that the probability of a False Match after a "best of k" test of agreement, when using *HD* criterion x, regardless of the actual form of the raw unrotated distribution $f_0(x)$, is $F_k(x) = 1 - [1 - F_0(x)]^k$, and the expected density $f_k(x)$ associated with this cumulative is:

$$f_k(x) = \frac{d}{dx}F_k(x) = kf_0(x)\,[1 - F_0(x)]^{k-1} \tag{3}$$

Equation (3) is the solid curve in the above figure, fitting the distribution of the same set of 200 billion IrisCode comparisons after 7 relative rotations of each pair.

3 Observed False Match Rates

The cumulative scores under the left tail of the preceding distribution, up to various Hamming Distance thresholds, reveal the False Match Rates among the 200 billion iris comparisons if the acceptance decision policy used those thresholds. These rates are provided in the following Table. No such matches were found with Hamming Distances below about 0.260, but the Table has been extended down to 0.220 using Equation (3) for extreme value samples of the binomial (plotted as the solid curve in the preceding figure) to extrapolate the theoretically expected False Match Rates for such decision policies.

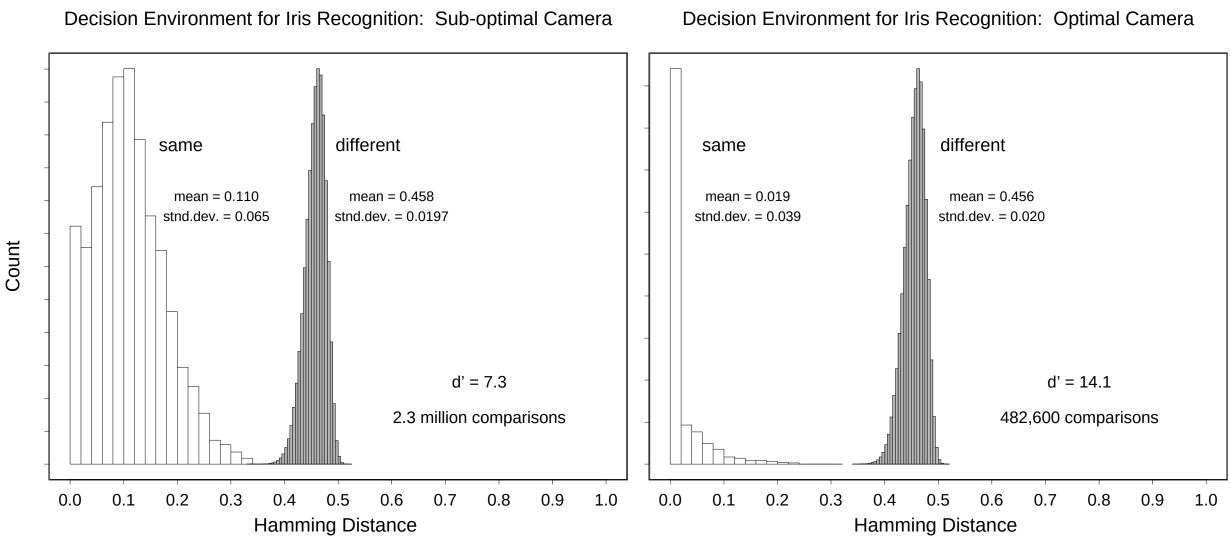
HD Criterion	*Observed False Match Rate*
0.220	0 (*theor: 1 in* 5×10^{15})
0.225	0 (*theor: 1 in* 1×10^{15})
0.230	0 (*theor: 1 in* 3×10^{14})
0.235	0 (*theor: 1 in* 9×10^{13})
0.240	0 (*theor: 1 in* 3×10^{13})
0.245	0 (*theor: 1 in* 8×10^{12})
0.250	0 (*theor: 1 in* 2×10^{12})
0.255	0 (*theor: 1 in* 7×10^{11})
0.262	1 in 200 billion
0.267	1 in 50 billion
0.272	1 in 13 billion
0.277	1 in 2.7 billion
0.282	1 in 284 million
0.287	1 in 96 million
0.292	1 in 40 million
0.297	1 in 18 million
0.302	1 in 8 million
0.307	1 in 4 million
0.312	1 in 2 million
0.317	1 in 1 million

The **US Department of Homeland Security** recently sponsored independent testing of the same Daugman algorithms. In a total of 1,707,061,393 (1.7 billion) cross-comparisons between different irises, the smallest Hamming Distance observed was in the range of 0.280, consistent with the above Table. (Source: International Biometric Group, *Independent Testing of Iris Recognition Technology,* May 2005.)

In the UK with a national population of about 60 million, an "all-against-all" comparison of IrisCodes to detect any multiple identities could be reliably performed using a decision criterion of about 0.220 without making False Matches. Employing such a decision policy in off-line exhaustive comparisons would allow the detection of multiple identities even when up to about 22% of the bits in their IrisCodes disagreed, due to corrupting factors in image acquisition such as poor focus, motion blur, limited cooperation, and so forth. In everyday transactions in which an identity is first asserted, matches could be accepted with a Hamming Distance as high as perhaps 0.33, allowing 33% of the bits to be incorrect, thereby tolerating unfavourable image capture conditions.

4 Camera properties and imaging conditions

Different cameras deliver iris images of differing quality and resolution, in part depending on the sophistication of their optics (e.g. autofocus and autozoom), and in part because they employ different wavelengths of the near-infrared spectrum. The following two graphs from earlier studies (Daugman 2003, 2004) illustrate that although the distribution of similarity scores for different-eye comparisons (shown in black) is stable and similar to that presented in this document for 200 billion UAE iris comparisons, the distribution of similarity scores for same-eye comparisons (shown in white on the left of each graph) is dependent on camera quality and imaging conditions.



5 Conclusion and recommendations

Iris recognition can be reliably used on a national basis in an Identity Cards scheme, including the capability for exhaustive IrisCode comparisons to detect multiple identities, if the decision policy employs a threshold criterion of about 0.22 Hamming Distance for off-line exhaustive comparisons and a criterion near 0.33 for everyday one-to-one matches.

References

Daugman, J.G. (1993) High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15** (11): 1148–1161.

Daugman J.G. (2001) Statistical richness of visual phase information: Update on recognising persons by their iris patterns. *International Journal of Computer Vision,* **45** (1): 25–38.

Daugman J.G. and Downing C.J. (2001) Epigenetic randomness, complexity, and singularity of human iris patterns. *Proceedings of the Royal Society (London): B. Biological Sciences,* **268**: 1737–1740.

Daugman J.G. (2003) The importance of being random: Statistical principles of iris recognition. *Pattern Recognition,* **36**: 279–291.

Daugman J.G. (2004) How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology,* **14** (1): 21–30.