Quick Primer: Iris Recognition based on Hamming Distances

1. Iris patterns are encoded into a sequence of bits (1 or 0), called an IrisCode. These resemble the outcomes of coin-tosses, with both `Heads' and `Tails' having equal probability. IrisCode sequences are aligned, so that they can be directly compared for different iris patterns. (Eyelid occlusions are masked.)

2. Randomness among different irises makes it equally probable that any given pair of aligned bits will agree or disagree. Here is an example, for just 8 bits:

IrisCode bit sequence for Iris `A':	1	1	0	0	0	1	0	0
IrisCode bit sequence for Iris `B':	0	1	1	0	1	1	1	0
Did they disagree? (`yes'=1, `no'=0)	1	0	1	0	1	0	1	0

3. In this simple example, aligned bits for IrisCodes A and B disagreed in 4 cases out of 8, which determines their Hamming Distance: 4/8, or a **0.5 HD score**.

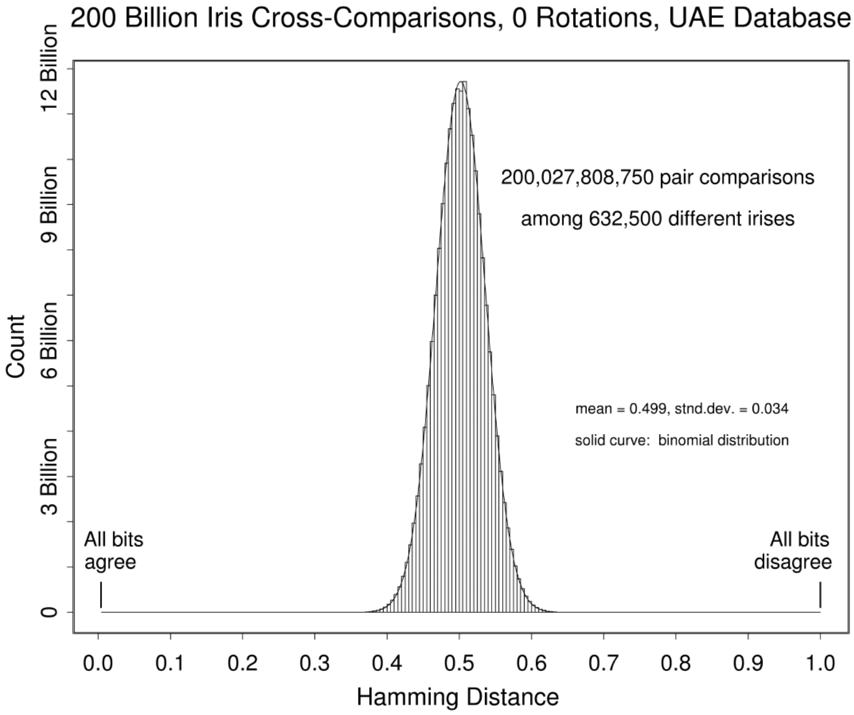
4. When much larger numbers of bits from unrelated IrisCodes are compared, it becomes very improbable to deviate far from a 0.5 HD score. This is true for exactly the same reason that when a coin is tossed many times, the fraction of `Heads' outcomes will remain close to 50%. (Although all possible outcome sequences have the same probability, there are **vastly more sequences** that have a nearly equal mix of `Heads' and `Tails', than a very unbalanced mix.)

5. Each bit in an IrisCode does not represent a point in an iris. It represents a region. These local regions have varying sizes, and they overlap each other. The value of the bit (1 or 0) is a summary of the local phase structure, which is a type of feature that is difficult for human observers to describe or name.

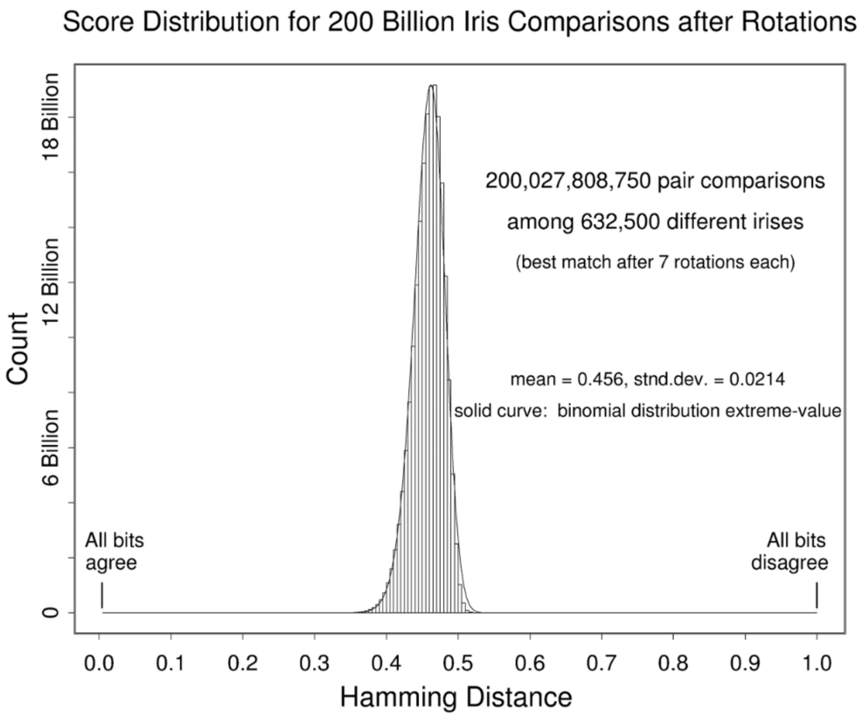
6. Each eye is attached to three pairs of muscles. In addition to eye movements that might be described as "up/down" or "right/left", each eye can rotate by almost $\pm 15^{o}$, which plays a role in stereo alignment. There can also be head tilt and camera tilt by unknown degrees, so IrisCode comparisons must allow for "scrolling" a bit sequence first for alignment at different rotation angles:

IrisCode bit sequence for Iris `A':	1	1	0	0	0	1	0	0
IrisCode bit sequence for Iris `B'-:	1	1	0	1	1	1	0	0
Did they disagree? (`yes'=1, `no'=0)	0	0	0	1	1	0	0	0

7. The distribution of Hamming Distance fractions (or `scores') after comparison of all the bits in IrisCodes for two different eyes, but only in any single relative rotation alignment, obeys the following distribution, with a mean near 0.5:



8. But when several relative rotations are tried and only the `best match' is kept (which is called `extreme value sampling' because we keep only the smallest of all the HD scores), then we get the following distribution, with a slightly lower mean. The distribution has shifted a little to the left, and it is sharper:



9. The solid curves fitted to both of the preceding distributions have precise mathematical forms, expressible in equations, because IrisCode comparisons are well-defined and deeply understood mathematically. They enable exact and testable predictions about the probability of a False Match, expressed as the likelihood that two different eyes could generate, by chance, a HD score smaller than any given value. When the Count distributions shown are taken to be probability distributions (i.e. with total area beneath them equal to 1.0), then the probability that two different eyes could produce an HD score in any given interval (e.g. ≤ 0.3) equals the portion of the distribution lying there.

10. Some normalisations are needed to ensure this good predictable behaviour regardless of how many bits were available for comparison (not masked by eyelids or corneal reflections), and taking into account how many IrisCodes were searched in the Gallery (the "Russian Roulette" problem that the longer you play the game, the more likely that you will eventually become unlucky):

Logic for computing raw Hamming Distance scores, incorporating masks:

$$HD_{\text{raw}} = \frac{\|(codeA \otimes codeB) \cap maskA \cap maskB\|}{\|maskA \cap maskB\|}$$

where $\otimes$ is Exclusive-OR, $\cap$ is AND, and $\|\quad\|$ is the count of 'set' bits.

Score re-normalisation to compensate for number of bits compared:

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

Decision Criterion normalisation by database size and query rate:

$$HD_{\text{Crit}} \sim 0.32 - 0.012\ \log_{10}(N \times M)$$

where N is the search database size, M is the number of queries to be compared against the full database, while requiring nil False Matches

---

11. The overall framework for iris recognition based on Hamming Distances, as summarised in this short document, has been extraordinarily well-confirmed by US NIST in independent tests. The following table shows predictions made about False Match Rates (as a function of HD criterion), long before databases large enough to test them even existed (2nd column). NIST validated these predictions in a Trillion iris comparisons (3rd column), by crossing databases.

HD criterion	*FMR predicted in* [9]	*NIST* [12] [13] *measured FMR*
0.36	1 in 24,000	1 in 25,000
0.35	1 in 110,000	1 in 71,000
0.34	1 in 556,000	1 in 476,000
0.33	1 in 3.1 million	1 in 3.4 million
0.32	1 in 20 million	1 in 24 million
0.31	1 in 137 million	1 in 165 million
0.30	1 in 1.1 billion	1 in 2 billion
0.29	1 in 9 billion	(not measured)
0.28	1 in 92 billion	1 in 40 billion

12. Some remarkable conclusions emerge from such cross-comparisons across large databases: (i) False Match Rates as a function of HD decision criterion are well predicted, from strong theory; (ii) tiny reductions in the HD criterion (e.g. in steps of 0.01, from 0.30 to 0.29 to 0.28) produce massive reductions in FMR; (iii) even when (say) 30% of the bits disagree between two IrisCodes, so HD score = 0.30, declaring a match in such cases has an associated probability of only 1-in-a-billion that it will be a False Match; (iv) if it is reasonably agreed that the distribution below is "narrow", in the sense that 99% of its area lies between 0.4 and 0.5, out of the total range 0.0 to 1.0 of possible HD scores, then: ***"all <u>different</u> eyes are roughly equidistant from each other."***