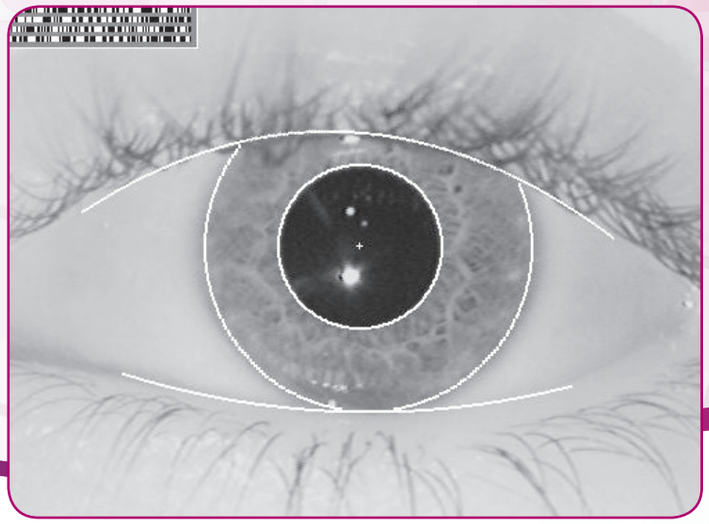
MATHEMATICS MATTERS

Statistically-Secure Identities



It is essential that the biometric systems we depend on for national security can reliably identify individuals and aren't easily deceived. A range of mathematical techniques ensure these systems work as intended, helping to keep the UK safe.

Concerns about criminal behaviour and national security mean that the use of biometric information to identify people is on the rise in the UK. All biometric systems, whether they are based on irises, faces or fingerprints, rely on mathematical methods to convert images into data that can be digitally stored and compared. Research shows that some biometric systems are more reliable than others, and a robust statistical foundation is essential to ensuring they are properly secure.

Iris recognition, invented by John Daugman at the University of Cambridge, is one of the most secure forms of biometric identification, and the methods he introduced in 1993 for encoding and recognising iris patterns form the basis of all iris recognition systems in use today. These systems use a number of mathematical techniques to detect, store and compare iris patterns in a fraction of a second.

Given a video image of an eye, the system first identifies the location of the iris by marking its boundaries with the pupil and sclera (white part) of the eye. This is achieved with Fourier analysis, a mathematical technique often used in image processing. Fourier analysis can also ignore eyelashes or reflections that might be covering the iris, revealing its true location without any of these image artefacts.

Each person has a unique iris pattern that serves as their biometric marker. Once the iris location has been properly identified, the system creates a digital record of the pattern that uniquely identifies its owner. Changes in position or lighting might mean that different images of the same iris produce different records, so it is important that two records of the same irises are similar enough to be matched, while still ensuring that two different irises aren't incorrectly paired.

> "The chance of two different irises matching is one in 10 billion - roughly the number of human eyes on the planet."

Daugman's clever solution encodes each iris pattern as a stream of 1s and 0s, or bits, using 2048 bits in total. When someone registers with the system for the first time, their iris pattern is encoded and stored in a database. When they later present their iris to the system for identification, a new bit stream is calculated and compared to all of the examples in the database by examining each pair of bits in turn. Comparing two bit streams takes around a billionth of a second on a modern computer, making it feasible to search an entire iris database in a short amount of time.

The system is not looking for exact matches because even the same iris in two different images will have slightly different bit streams - only a certain threshold of bits need to agree. It turns out that anything more than 70% of the bit stream agreeing will confirm a person's identity, because the chance of two different irises matching so well is one in 10 billion, or roughly the number of human eyes on the planet. This statistically-proven security has made iris recognition a great success, but Daugman still continues to refine the technique with new mathematical methods, such as a recent improvement that enables the system to adjust for eyes looking away from the camera.

Even with these new developments, iris recognition is not suitable for use in all situations. Imaging a person's iris requires a detailed close-up shot, making it useless for identifying people in crowds or at a distance. While alternative identification methods for these scenarios do exist, it is important that they live up to the same standard as iris recognition.

Nick Fieller at the University of Sheffield has discovered that this may not be the case with

Institute of mathematics & its applications
EPSRC Pioneering research and skills

facial recognition systems. This result may seem surprising, as such systems are found in everything from surveillance equipment to consumer digital cameras, but there is an important distinction between facial recognition and facial detection. The latter merely involves recognising the presence of any face, but the former is the more difficult problem of matching an image of a face to a particular person's face, and quantifying the extent of the match with probability.

Photographs of faces can vary much more than those of irises, from drivers' license-style head shots to blurry CCTV footage, so creating a standardised description is essential. Fieller's method uses established facial landmarks, such as the edges of the eyes or lips, to reduce photographs of faces to around 30 key points. This technique allows for direct comparisons between different individuals' faces, or different images of the same face, by manipulating the landmark points from two images until they align.

Fieller gathered 3,000 different faces to examine their statistical variation. He found that unlike irises, in most cases the statistical differences between faces are too small for them to be distinguished. It is like trying to identify someone by their height – while there are some extremely short or tall individuals, most people are of average height. In the same way, most people have average faces, falling in the middle of a statistical distribution.

Although it seems that faces may not be suitable for secure identification systems, Fieller now plans to apply a similar method to fingerprint recognition. Fingerprints also have landmark points that can be statistically analysed, such the top of arches or the centres of whirls, which will allow him to study their distributions. Current fingerprint recognition methods can be subjective, but Fieller's method would quantify the extent to which two fingerprints agree.

In addition to using these distributions for recognition, Fieller hopes they could also be used to improve incomplete fingerprints found at crime scenes. By analysing the incomplete print and looking at its statistical properties, Fieller could generate a range of possible full fingerprints that might help provide police with a match.

The many different biometric systems in existence all share one aim: to reliably and uniquely identify individuals. This aim can only be realised if biometrics are backed up by the kind of robust statistical evidence provided by statisticians and computer scientists like Fieller and Daugman. Their work guarantees the effectiveness of biometrics, and helps keep the UK safe.

TECHNICAL SUPPLEMENT

Encoding iris patterns

Iris recognition requires that each sample be compared to the entire database, so iris patterns must be stored in a low-memory and easily-compared format. It turns out that mathematical functions known as Gabor filters are very good at detecting iris patterns, and can fulfil both of these requirements. John Daugman's method applies a Gabor filter to one small region of the iris at a time, and the result has two components that are stored as either a 1 or 0. The entire iris can be encoded with just 1024 of these pairs, making a total of 2048 bits.

New irises are matched against the database by computing the difference between the bit streams, known as the fractional Hamming distance (HD). This is done by examining each bit position in turn, counting the number of positions that differ and dividing by the length of the bit stream. For example, comparing the bit streams 1101 and 1010 gives a HD of 0.75, because the last three bits have changed. All HDs range from 0 to 1, with the HD of two randomly selected bit streams being an average 0.5. The probability distribution of iris HDs is a very steep binomial distribution, so the majority fall between 0.4 and 0.6, and the probability of two different irises having a HD below 0.3 is one in 10 billion, proving the statistical security of iris recognition.

Measuring facial similarity

Nick Fieller's method for measuring facial similarity uses the locations of 32 anthropometric landmarks in three dimensions, giving a total of 96 parameters. Fieller uses multivariate techniques to produce statistical distributions of these parameters, giving an indication of the variability between faces within the population. Even with extremely detailed measurements of the facial landmarks, Fieller found it was not possible to clearly distinguish between faces, as most people have "average" faces that sit in the middle of the distribution.

References

Daugman, J. G. (1993) High confidence visual recognition of persons by a test of statistical independence. ***IEEE Transactions Pattern Analysis and Machine Intelligence,*** 15(11), 1148-1161. DOI: 10.1109/34.244676

Daugman, J. G. (2003) The importance of being random: statistical principles of iris recognition. ***Pattern Recognition,*** 36(2), 279-291. DOI: 10.1016/S0031-3203(02)00030-4

Daugman, J. G. (2007) New Methods in Iris Recognition. ***IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,*** 37(5), 1167-1175. DOI: 10.1109/TSMCB.2007.903540

Evison, M., Dryden, I., Fieller, N., Mallett, X., Morecroft, L., Schofield, D. & Bruegge, R. V. (2010) Key Parameters of Face Shape Variation in 3D in a Large Sample. ***Journal of Forensic Sciences,*** 55(1), 1556-4029. DOI: 10.1111/j.1556-4029.2009.01213.x

The IMA would like to thank Professor John Daugman, University of Cambridge and Dr. Nick Fieller, University of Sheffield for their help in the preparation of this document