A Path Forward for Multi-biometrics

James L. Wayman
San Jose State University
San Jose, CA, USA
jlwayman@aol.com

Abstract

In this talk, we define "multi-biometrics" and "multi-modal" biometrics and discuss the reasons why "multi-modal" biometrics has failed to achieve its promise, despite 30 years of research. We give examples of how "multi-biometric" systems have been successfully used in large-scale applications, such as the Australian "SmartGate", US-VISIT and U.K. IRIS systems, and present a Bayesian-based, parameter-free approach to combining multi-biometric data.

Some Newly Proposed Definitions

A biometric *mode* is the combination of a body part, an algorithm and a sensor type. A *multi-modal* biometric system is one in which, for two or more of these components, more than one type of the component is present. *Multi-instance, multi-algorithmic and multi-sensor* biometrics would be terms used for systems which have only single components of a mode (body parts, algorithms or sensors) in multiple. Systems using multiple images over time of a single body part can be called *multi-presentation*. Systems using multiple, but separable, frequency bands for imaging can be called *multi-spectral*. The term *multi-biometrics* can be used as a general descriptor for any of these systems. These definitions have been proposed by the international standards committee ISO/IEC JTC1 SC37 Working Group 1 (vocabulary harmonization) to resolve some definitional difficulties.

For example, under these definitions, facial recognition using high resolution images in which the iris was visible would only be multi-modal if the iris portion of the image were handled by an algorithm different from that used for the facial recognition. Multi-spectral facial imaging would be single modal if the spectral bands were all handled with the same sensors or algorithms. However, facial imaging combined with facial thermography (based on passive IR collection) would be multi-modal if different collection hardware and processing algorithms were used for images of each type. Facial imaging would be single modal, but multi-algorithmic, if both image decomposition and local correlation algorithms were being used simultaneously.

At this writing, the SC37 Working Group 2 (Biometric Technical Interfaces) is preparing a "Technical Report on Multi-Modal and other Multi-Biometric Fusion", to be known as ISO 24722.

The Era of Multi-modal Biometrics: The 1970s

In the 1970s, decision-level biometric fusion based on combining results from fingerprint, handwriting, voice, or hand geometry systems was widely seen as the pathway toward lowering biometric error rates [1-7].

In 1974, the U.S. Air Force, Electronic Systems Division, announced an ambitious program for a world-wide, military-wide, multi-modal biometric identification system as part of a "Base and Installation Security System" (BISS) [4,5]. The Air Force considered methods for hardening the system against forgery attacks and for combining biometric methods to reduce error rates. The Mitre Corporation was contracted to gather over a thousand signature, fingerprint and voice samples on over 200 volunteers in a formal BISS test program [6,7]. Assuming operational independence of error rates, they hypothetically "fused" the test data to look at error rate improvement under both "AND" and "OR" fusion logic for all combinations of two modes and included "2 out of 3" fusion for all three modes. Although scheduled for deployment in 1981, the biometric components of the BISS system were ultimately never deployed for reasons now lost in history. However, the U.S. military does to this day use single-modal biometrics for access control on a local basis throughout the world.

Correlations

All multi-biometric measures (multi-modal or not) from a single person are by necessity correlated. In 1908, Sir Francis Galton made the following statement about the multi-modal Bertillion system

> "The incorrectness (of the Bertillion system) lay in treating the measures of different dimensions of the same person as if they were *independent* variables, which they are not. For example, a tall man is much more likely to have a long arm, foot, or finger, than a short one." [8].

Further, even single images from single sensors contain correlated data within the signal, a fact with allows good compression algorithms to be created [9-11] and leads to claims of "173 degrees of freedom" from 2048-bit "iris codes" [12]. Consequently, all biometric systems whether multi-modal, multi-presentation, multi-instance or multi-sensor, contain correlated data. For improving error rates, ***the emphasis in biometrics should be on maximizing the information content of the total processed signals within the constraints of cost, human interaction time, and system complexity***. At the current state of biometric system development, information content can be most easily increased within the practical constraints by collecting more images (multi-presentation) from more body parts (multi-instance) or by viewing body parts from multiple angles (multi-sensor), and existing information can be better exploited by looking at it in more ways (multi-algorithmic). Adding sensor types, such as combining facial thermography with optical facial imaging or facial imaging with fingerprinting, has not proven

1-4244-0469-X/06/$20.00 ©2006 IEEE

practicable in the history of biometric deployments. Even current large-scale national systems collecting multiple types of biometric data, such as the US-VISIT system or the Hong Kong National ID system [13], do not combine biometric measures across modes.

There are inherent problems limiting the deployment of multi-modal biometrics beyond the cost and complexity of the added sensors, the problems of controlling the acquisition environment simultaneously for several modes, and the added user interface time. Testing of the inherently correlated multi-modal measures has been inhibited by the privacy implications of release of such comprehensive personal data on volunteers. The U.S. National Institute for Standards and Technology (NIST) has been planning a "Multi-modal Biometric Assessment Research Kiosk" program for the last few years, but have announced that the research image data sets developed will only be released as single modes to protect the privacy of the volunteers.

Alternatives to Multi-modal Systems

Although multi-modal systems have not yet become a reality after 30 years of research, multi-presentation and multi-instance systems have been successfully used throughout that time – particularly in forensic applications using images of multiple fingers [14]. A number of tests and deployments have shown the practical benefit of multi-instance, multi-sensor biometrics in face recognition [15-17, 36], multi-presentation biometrics for fingerprint [18,19] and iris recognition [20], and multi-instance (in the sense of longer segments) and multi-algorithmic approaches to text-independent speaker recognition [21].

Fingerprints from the same person are known to be correlated at pattern level [22] but have only modest correlation in errors for automated matching [19] or pattern classification [23] algorithms. NIST reports [18] in the Fingerprint Vendor Technology Evaluation (FpVTE),

> "Thus the major conclusion is that each doubling of the number of fingers produces a fixed factor reduction in false rejection errors… approximately a factor of five as the number of fingers is doubled…
>
> The variables that had the largest effect on system accuracy were the number of fingers used and fingerprint quality
>
> - Additional fingers greatly improve accuracy
> - Poor quality fingerprints greatly reduce accuracy"

The NIST report makes a strong case for improving fingerprint system accuracy by spending available time and money resources in collecting more and better images of the same type (multi-instance/presentation biometrics). Using a "slap" approach, 4 fingers can be collected simultaneously on a single piece of hardware in the same time as required for collection of a single print. This makes it possible, in the case of fingerprinting, to collect multi-instance biometrics with the same basic sensor type and time requirements as single instance biometrics. A great deal of current U.S. government funding for research, development and testing is going into "slap" fingerprint collection [24, 25]. The U.S. government has announced that the US-VISIT border crossing system will transition from the collection of two index fingers to the collection of 10 fingerprints [26]. US-VISIT collects facial images, using them only for human inspection, not as part of a multi-modal biometric system [27].

Facial recognition performance has been shown to improve through multi-presentation, multi-sensor and multi-algorithmic biometrics. NIST reports [17] that, "Fusion of four scores from five images per person cuts verification error rates by about half" and "Fusion of two leading FRVT systems also reduces verification errors by about 50%". The Australian Customs border crossing "SmartGate" found similar facial recognition improvement using 5 enrollment images taken simultaneously at slightly different angles: centered, above, below, left and right. Recent work by McLindin [36] has quantified the improvement in the 5 image approach. The verification image used by SmartGate is the best chosen from the image streams of 3 cameras at different heights (multi-presentation/multi-sensor data) [15, 16].

The single-modal U.K. Iris Recognition Immigration System (IRIS), used at Heathrow Airport "ports of entry", images both irises simultaneously, taking the best images from a data stream [20]. Multiple cameras accommodate users of different heights, making this a multi-presentation/instance/sensor system. Fingerprints are required at enrollment, but are used only for background checks.

Multi-biometrics with Correlated Measures

So the task before us is the efficient usage of correlated data from multi-biometrics of any type to reduce error rates. One classical approach used in the literature for biometric data fusion [28-30] seeks to optimize a cost function, given knowledge of Baysian priors and the cost of both false non-match and false match errors.

Suppose that each user has N (generally correlated) scores, $\vec{x} = (x_1, x_2, \ldots x_N)$. There are two associated probability densities: $f(\vec{x}|G)$ and $f(\vec{x}|I)$, for genuine (truly matching) and impostor (truly non-matching) comparisons, respectively. These densities will be determined empirically and are not assumed to have any particular analytic form. A region R of $\Re^N$ is a non-match decision region and its complement, $R^C = \Re^N \setminus R$, is the region of a match decision. A comparison will be considered non-matching if the scores, $\vec{x} \in R$ and matching if $\vec{x} \in R^C$. Given are the costs of wrong decisions: C_{FM} for falsely matching and C_{FNM} for falsely non-matching,. The prior probability that a comparison is truly non-matching is P(I) and truly matching is P(G) = 1- P(I).

So the penalty function, E(R), for making a wrong decision becomes

$$E(R) = C_{FM} P(I) \int_{R^C} f(\vec{x}|I)\, d\vec{x} + C_{FNM} P(G) \int_{R} f(\vec{x}|G)\, d\vec{x} \qquad (1)$$

Wright and DeVito [31] show that the optimality of choosing the non-matching region,

$$A = \left\{ \vec{x} \in \Re^N \,|\, \frac{f(\vec{x}|I)}{f(\vec{x}|G)} \geq \frac{C_{FNM} P(G)}{C_{FM} P(I)} \right\}, \qquad (2)$$

does not depend on the N elements of $\vec{x}$ (user scores over N systems or measures) being uncorrelated. Of course, adequate data must be available to estimate the surfaces of the multi-dimensional probability densities, $f(\vec{x}|G)$ and $f(\vec{x}|I)$. In the case of N=1, equation (2) reduces to the well-known optimum region for a single-modal decision.

Genuine and impostor distributions for all biometric methods are strongly dependent upon user demographics, user attitudes, and the specifics of the application environment, level of supervision and training and frequency of use. They may differ strongly between applications and change within an application as users become more habituated or the total environment evolves. These distributions can rarely, if ever, be characterized by analytic functions of few parameters, but rather require extensive empirical data from the application of interest for characterization [32, 33]. (As Karl Pearson noted, "I can only recognize the occurrence of the normal curve…as a very abnormal phenomenon" [34]). Equations (1) and (2) are "non-parametric" in that they do not analytically model the distributions.

Although equation (2) does not require statistical independence of measures, it is nonetheless difficult to apply in any more than a heuristic fashion to real systems as the costs of an error, the evolving genuine/impostor distributions, and especially the priors, may be difficult to guess with any certainty. The perceived accuracy of the biometric system may strongly impact the prior probability of deceptive behaviors which, depending upon the application, can be the probability, P(I), that comparison will be truly non-matching (impostors in access control systems) or P(G), truly matching (duplicate enrollments in ID systems or people on a watchlist). For most applications, the probability of deceptive behavior, whether P(G) or P(I), will be very close to 0, but may not be guessable to within even an order or two of magnitude.

Of greater practical value to real biometric systems is the ROC curve, which is inherently empirical and "non-parametric". Real systems generally start out with the hope of establishing a security policy-driven "false acceptance rate", which will be a false non-match rate for negative claim applications, such as ID systems, and a false match rate for access control. A target false acceptance rate of "1 in 10,000" is often used because of the mistaken belief that such a system will have security qualities similar to 4-digit PINs. In actual practice, however, thresholds are tuned in application so that the "false rejection rate" does not become inconvenient. For example, one government system sets the false rejection rate for its biometrics-based ID card to yield no more than 50 false matches per month because staff is available only to investigate that many problems – true rejections for multiple enrollments being rare. A good rule of thumb is: if false rejections exceed true rejections by much more than a single order of magnitude, the security of the exception handling mechanism will break down.

So having established decision thresholds based on the dictates of the false rejection rate, it would be helpful to know the corresponding false acceptance rate, which brings us to consider the ROC curve, which compares false acceptance rates to false rejection rates (or false match rates to false non-match rates) as a function of decision threshold. Comparisons of ROCs can give a ranking of different systems, allowing error rate improvements to be assessed against the added cost and complexity of multi-biometric data, without requiring advanced estimation of the costs of errors and the probabilities of deceptive behavior.

Likelihood Ratio Techniques for Multi-biometric ROC Estimation

The inequality in equation (2) compares a likelihood ratio against a threshold. If the conditional probabilities of $f(\vec{x}|G)$ and $f(\vec{x}|I)$ are known, then the ROC can be computed as a curve parametric with this threshold. Dass, *et al*[32] use correlated data which is multi-instance, multi-algorithmic and multi-modal, consisting of scores from images of two different fingers processed with the same algorithm and one face image processed with two different algorithms [35]. They partition the data, and using the covariance matrix computed from the scores in one partition, develop a parametric model of the conditional probability "couplings" of the non-parametric distributions of the individual scores, $f(x_i|G)$ and $f(x_i|G)$, i=1,…N, which is used to compute the ROC from data in the other partition. Not surprisingly, combining the data from the two face algorithms shows great improvement over either of the face algorithms alone, and using all of the multi-modal/instance/algorithmic data shows the best ROC performance of all.

If only the N scores from each user were uncorrelated, then the model of the conditional probabilities would not be needed and an ROC for the multi-biometric system could be developed simply using the product rule, as

$$Gain = \prod_{i=1}^{N} \frac{f(x_i|I)}{f(x_i|G)} \tag{3}$$

Dass recomputed the ROC by (3) and shows that it is remarkably similar to that computed from the correlation model. This does not mean that other more useful models for correlated data cannot be developed, nor does it mean that ROCs would not be improved if the actual conditional probability density surfaces were known. It does, however, establish that the approximation of (3), may be as useful as the more complicated Dass model, and implies that (3) might be a good, workable estimate of multi-biometric performance at the level of correlation and instability encountered in real biometric data.

Conclusions

All multi-biometric data are correlated, but the history of biometrics has shown that for improving error rates, more is better, regardless of correlation. In practice, the easiest way to obtain additional data has been through single-modal multi-biometric methods, such as multi-presentation (many face images), multi-instance (several fingerprints), multi-sensor (several cameras) or mult-algorithmic. Although multi-modal biometric approaches are theoretically fascinating, the practical path forward in multi-biometrics is in first fully exploiting the time, cost, and complexity economies of multi-presentation/instance/sensor/algorithmic data.

Acknowledgements

Support for this work was received from the Technical Support Working Group under Contract N41756-02-C-4651.

References

[1] D. Raphael and J. Young, "Automated Personal Identification", (SRI, International, Palo Alto, CA, 1974)

[2] U.S. National Bureau of Standards, "Guidelines on the evaluation of techniques for automated personal identification", Federal Information Processing Standard Pub. 48, April, 1977

[3] W.K. Messner, G.A. Cleciwa, G.O. Kibbler and W.L. Parlee, "Research and Development of Personal Identify Verification Systems", Carnahan and International Crime Countermeasures Conference, University of Kentucky, April 16-19, 1974

[4] W. Haberman and A. Fejfar, "Automatic Identification of Personnel through Speaker and Signature Verification – System Description and Testing", Carnahan Conference on Crime Countermeasures, University of Kentucky, May, 1976

[5] A. Fejfar and P. Benson, "Test Results Advanced Development Models of BISS Identify Verification Equipment, Part V – Miscellaneous", Mitre Corp. Technical Report MTR-3442 , Bedford, MA, 16 September, 1977

[6] A. Fejfar and J. Myers "The testing of three automatic identity verification techniques", Proc. Int. Conf. On Crime Countermeasures, Oxford, July, 1977

[7] A. Fejfar, "Combining Techniques to Improve Security in Automated Entry Control", 1978 Carnahan Conf. On Crime Countermeasures,, Mitre Corp. MTP-191, May 1978

[8] F. Galton, Memories of My Life (Methuen, London, 1908)

[9] R. Cox, "Three New Speech Coders from the ITU Cover a Range of Applications." IEEE Communications Magazine 35, no. 9, September 1997, pp. 40 - 47.

[10] "Wavelet Scalar Quantization (WSQ) Gray-Scale Fingerprint Image Compression Specification", Criminal Justice Information Services, Federal Bureau of Investigation, IAFIS-IC-0110v2, February 16, 1993.

[11] "Information Technology – Digital Compression and Coding of Continuous-tone still images: Requirements and guidelines", CCITT Recommendation T.81, ISO/IEC – 10918; 1993(E), available online at www.w3.org/Graphics/JPEG/itu-t81.pdf

[12] J. Daugman, "Biometric Personal Identification System Based on Iris Analysis", US Patent 5,291,560, 1994

[13] R. Wong, "Application of Biometric Technology in HKSAR Immigration Department", International Conference on Biometric Authentication, Hong Kong Polytechnic Univ., July 15-17, 2004

[14] R. Allen, P. Sankar, and S. Prabhkar, "Fingerprint Identification Technology" in J. Wayman, *et al* (eds.) Biometric Systems: Technology, Design, and Performance Evaluation (Springer, London, 2005)

[15] F. Fraser, "Exploring the Use of Face Recognition Technology for Border Control Applications – Australia's Experience", Proceedings of Biometric Consortium Conference 2003, September 22-24, 2003, Washington, D.C.

[16] Australian Customs Service, "Overview of the SmartGate Trial", February, 2004, available at http://www.customs.gov.au /webdata/resources/files/FS_overview_smartgate0406.pdf

[17] P. Grother, "Facial Recognition Vendor Test Supplemental Report", NISTIR 7083, 2 February, 2003, online at ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7083.pdf

[18] C. Wilson, *et al*, "Fingerprint Vendor Technology Evaluation", National Institute of Standards and Technology Report NISTIR 7123, 2003, on-line at http://fpvte.nist.gov

[19] A. Hicklin, C. Watson, and B. Ulery, "The Myth of Goats: How many people have fingerprints that are hard to match?", NIST, NISTIR 7271, Sept. 2005, on-line at www. itl.nist.gov/iad/894.03/pact/ir_7271.pdf

[20] P. Abrahamsen, "Case History: IRIS Gets Off the Ground", Biometrics 2005, London, Oct. 21-23,2005

[21] A. Martin, M. Przybocki and J. Campbell, "The NIST Speaker Recognition Evaluation Program", in J. Wayman, *et al* (eds.) Biometric Systems: Technology, Design, and Performance Evaluation (Springer, London, 2005)

[22] D. Maltoni, D. Maio, A. Jain and S. Prabhakar, Handbook of Fingerprint Recognition (Springer-Verlag, New York, 2003)

[23] J. Wayman, "Multifinger Penetration Rate and ROC Variability for Automatic Fingerprint Identification Systems", in N. Ratha and R. Bolle (eds.) Automatic Fingerprint Recognition Systems (Springer-Verlag, New York, 2004)

[24] B. Ulery, A. Hicklin, C. Watson, M. Indovina, K. Kwong, "Slap Fingerprint Segmentation Evaluation 2004 Analysis Report" , NIST Image Group, Information Access Division, 8 March, 2005, available on-line at http://fingerprint.nist.gov/slapseg04/ir_7209_Summary.pdf

[25] C. Miles, "Biometrics R&D Portfolio at the National Institute of Justice", presentation at the Biometrics Consortium Conference, Crystal City, VA, 21 Sept., 2005

[26] Statement of DHS Sec. Michael Chertoff to the U.S. Senate Committee On Commerce, Science and Transportation, Washington, D.C., 19 July, 2005, available on-line at http://www.dhs.gov/dhspublic/display?theme=45&content=4643

[27] Department of State, 22 CFR Part 51, Public Notice 4993, RIN 1400-AB93, Electronic Passport. *Federal Register*, 70(33), 18 Feb. 2005, available at http://a257.g.akamaitech.net /7/257/ 2422/01jan20051800/edocket.access.gpo.gov/2005/05-3080.htm

[28] P. Griffin, "Optimal Biometric Fusion for Identity Verification" *Technical Report RDNJ-03-0064*, Identix, 2003

[29] L. Hong and A. Jain, "Multimodal Biometrics", in A. Jain, etal (eds), Biometrics: Personal Identification in a Networked Society, (Boston, Kluwer Academic Press, 1999)

[30] L. Hong, A. Jain and S. Pankanti, "Can Multibiometrics Improve Performance", Auto*ID'99*, Summit, NJ, Oct 1999, pp. 59-64, on-line at www.cse.msu.edu/cgi-user/web/tech/document?NUM=99-39

[31] L. Wright and C. De Vito, "Higher Order Biometric Systems", University of Arizona Department of Mathematics internal report, 1999, written under funding from the U.S. National Biometric Test Center

[32] S. C. Dass, K. Nandakumar, A.K. Jain, "A Principled Approach to Score Level Fusion in Multimodal Biometric Systems", Proc. AVBPA 2005

[33] J. C. Wu and C. Wilson "Nonparametric Analysis of Fingerprint Data", NIST, NISTIR 7226, May, 2005 , available on-line at http://www.itl.nist.gov/iad/894.03/pact/ir_7226.pdf

[34] K. Pearson, "On some applications of the theory of chance to racial differentiation" (1901), as quoted in S. Stigler, Statistics on the Table, (Harvard University Press, Cambridge, MA, 1999)

[35] NIST Biometric Scores Set – Release 1 (2004), available on-line at http://www.itl.nist.gov/iad/894.03/biometricscore

[36] B. McLindin, "Improving the Performance of Two Dimensional Facial Recognition Systems: The Development of a Generic Model for Biometric Technology Variables in Operational Environments", Ph.D. thesis, University of South Australia, March 2005, available on-line at http://www.library.unisa.edu.au/adt-root/public/adt-SUSA-31102005-074958