HOLISTIC EVALUATION OF MULTI-BIOMETRIC SYSTEMS

Nick Bartlow, Don Waymire
{nbartlow, dwaymire}@brtrc.com
BRTRC
8260 Willow Oaks Corporate Dr.
Fairfax, VA 20171

Gregory Zektser
zektser_gregory@bah.com
Booz Allen Hamilton
13200 Woodland Park Rd
Herndon, VA 20171

Abstract

As large scale biometric systems have been deployed throughout the landscape of the federal government, a new set of challenges and expectations have arisen. The ability to address increasing database sizes, demands on throughput, and privacy / legal issues are just a few examples of challenges at the forefront. Despite these challenges, the large scale systems are expected to continue to improve in terms of accuracy. In response to such challenges and expectations, systems relying on a single source of biometric input are transitioning to multi-biometric implementations. Traditional motivations behind the application of multi-biometric systems have centered on improving recognition performance, increasing population coverage, deterring spoof attacks, and reducing failure-to-enroll rates. The potential for improvement associated with these motivations is typically thought to come at the expense of increased processing time and computational complexity. In this paper, we provide supporting arguments for expanding the basis of multi-biometric system evaluation to reflect a holistic scope. Namely, the best multi-biometrics systems will maximize the degree in which requirements corresponding to the characteristics of universality, uniqueness, permanence, measurability, performance, acceptability, and circumvention are met. Additionally, we will demonstrate how evaluating the requirements of large scale federal multi-biometric systems necessitate an increased level of granularity beyond the traditional seven characteristics. Finally, we provide an example prioritization scheme which demonstrates how the overall measure of success need not uniformly consider all characteristics.

Introduction

A recent trend in biometrics involves a shift from uni-biometric to multi-biometric systems. Uni-biometric systems make use of a single source of biometric information to perform identity determinations (verification, identification, negative recognition, etc.). Both theoretical research and empirical observation of fielded systems reveals that uni-modal systems are subject to a variety of shortfalls. Ceilings on performance accuracy, poor subject population coverage, relatively high failure-to-enroll rates, and ease of circumvention are classic examples of such shortfalls [RNJ08, RNJ06]. Multi-biometric systems which rely on more than one source of biometric input can be used to alleviate such shortfalls. As can be seen in Figure 1, multi-biometric systems can incorporate information from multiple modalities, instances, sensors, samples, or any combination of the five. Arguably, such systems may also include other sources of information including biographic, travel document-based, etc.

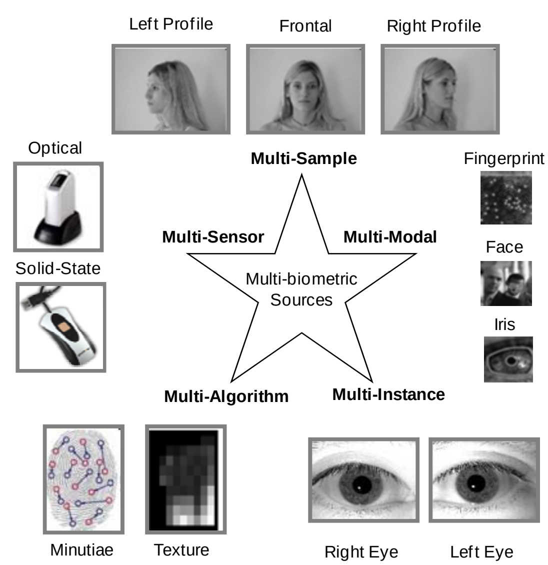

Figure 1 – Multi-Biometric Sources of Input [RNJ06]

The trend toward multi-biometric systems has been particularly prevalent in large-scale U.S. government systems. DoD ABIS, DHS IDENT, and FBI Next Generation IAFIS are all examples of systems which are currently multi-biometric in nature [nga09, ngi09, fsd07]. Furthermore, all three systems are increasing the number of biometric sources which can be leveraged.

Traditionally, it is thought that the additional opportunities (particularly related to performance) associated with multi-biometric systems must come at the expense of increased processing time and computational complexity. However, through careful application of emerging technological advances, multi-biometric systems may not have such negative side effects. One can argue that the best multi-biometric systems will maximize the degree in which requirements corresponding to the characteristics of universality, uniqueness, permanence, measurability, performance, acceptability, and circumvention are met. In remainder of this paper, we provide supporting arguments for expanding the basis of multi-biometric system evaluation to reflect a holistic scope. Additionally, we will demonstrate how evaluating the requirements of large scale federal multi-biometric systems necessitate an increased level of granularity beyond the traditional seven characteristics. Finally, we provide an example prioritization scheme which demonstrates how the overall measure of success need not uniformly consider all characteristics.

Why the Basis of Multi-Biometric System Evaluation Can Be Expanded

As previously mentioned, the traditional role of multi-biometric fusion has been to increase system accuracy, increase the coverage across the population base, decrease instances of failures to acquire / failures to enroll, and increase the difficulty associated with circumvention. These four motivations directly relate to a subset of the characteristics typically used to evaluate a biometric modality.

Table 1 shows the characteristics proposed by Jain [JR08]. We demonstrate a typical high level evaluation through an example multi-biometric system. We consider the example of transitioning from a single print fingerprint biometric system to a multi-modal system which utilizes both the fingerprint and face (basic 2D frontal images). Many years of research have demonstrated that multi-biometric fusion, the process of consolidating multiple sources of biometric information can significantly improve system accuracy over uni-biometric systems [RNJ08, RNJ06]. In the proposed example, the combination of fingerprint and face biometric information can greatly increase a system's ability to accurately discriminate between inter-class (non-matching) and intra-class (matching) samples. This relates directly to the characteristic of performance. Typically this aspect of performance is measured by rates such as False Non-Match Rate (FNMR), False Match Rate (FMR), False Accept Rate (FAR), False Reject Rate (FRR), True Positive Identification Rate (TPIR), False Negative Identification Rate (FNIR), False Positive Identification Rate (FPIR), and Equal Error Rate (EER). Relationships between such rates are commonly visualized in many forms of Receiver Operator Characteristic (ROC) curves. However, as we will see, these measures and visualizations only apply to a specific subset of the ideas encompassed by the characteristic of performance

Population coverage corresponds directly to the characteristic of universality. In the proposed multi-biometric system, a greater percentage of the population will have suitable biometric samples as compared to the uni-modal fingerprint system. It is widely known that some individuals have severely damaged fingerprints or missing digits which may prevent

Table 1 – Characteristics of a Biometric / Biometric System [JR08]

Characteristic	Description
1. Universality	Every individual accessing the
2. Uniqueness	The given trait should be sufficiently different across individuals comprising the population.
3. Permanence	The biometric trait of an individual should be sufficiently invariant over a period of time with respect to the matching algorithm. A trait that changes significantly over time is not a useful biometric.
4. Measurability	It should be possible to acquire and digitize the biometric trait using suitable devices that do not cause undue inconvenience to the individual. Furthermore, the acquired data should be amenable to processing in order to extract representative feature sets.
5. Performance	The recognition accuracy and the resources required to achieve that accuracy should meet the constraints imposed by the application.
6. Acceptability	Individuals in the target population that will use the application should be willing to present their biometric trait to the system.
7. Circumvention	This refers to the ease with which the

fingerprint acquisition / processing. However, it is much less likely that individuals will lack both suitable fingerprint and face samples. Therefore, the proposed multi-biometric system meets the requirements of universality more so than the original uni-modal system. Occasionally this requirement is measured by the Failure to Enroll (FTE) rate.

The ability to decrease observed failures to acquire (FTA) and failures to enroll (FTE) can relate to the characteristics of measurability and universality. By incorporating both fingerprint and face modalities one can tolerate various types of noise that may be observed in either modality. This tolerance will improve failure to acquire rates. As previously indicated the multi-biometric system will increase population coverage. The combination of increased tolerance of noise and increased population coverage will decrease failure to enroll rates as compared to the original uni-biometric system. It is important to note that a multi-biometric system need not be multi-modal in order to observe such improvements. Multi-instance, multi-sample, multi-algorithm, and multi-sensor systems can also deliver such results.

Finally, the example multi-biometric system will likely be more difficult to circumvent than the original uni-biometric system. Considering spoofing and masquerade attacks, one would have to present false samples for multiple biometric sources. In the fingerprint and face system both biometrics would have to be forged and potentially presented simultaneously. From this perspective, circumvention is much more difficult [RLG09, RNJ08, RNJ06]. At present, there is no widespread methodology for quantitatively measure the difficultly of circumventing a system.

To this point, we have exhausted the typical scope of analysis when evaluating a biometric system. However, only four of the seven biometric characteristics have been addressed. This begs the question, can multi-biometric systems be used to increase the degree the requirements associated with the remaining three characteristics (uniqueness, permanence, acceptability) are met? If so, these should likely be incorporated into the evaluation process which determines the success of a multi-biometric system.

If a multi-biometric system is resulting in improved accuracy, it is likely that the uniqueness of the overall feature set in question is more unique than the uni-biometric system in which it is being compared to. Although exact measures of uniqueness for fingerprint and face are often argued, few would argue against the idea that the combination of features from both sources of information will result in a collective set which is more unique than either single source of information. Therefore, it seems that the degree in requirements are met associated with this characteristic can be improved.

Permanence can also be affected when transitioning from a uni-biometric to a multi-biometric system. It is well known that certain biometric modalities such as face change over time. This leads to the concept of template aging. Modifying the example only slightly, if the original uni-biometric system was based on face, a multi-biometric system incorporating both face and fingerprint would certainly improve the permanence of the collective biometric feature set as fingerprints are less susceptible to change over time than faces.

Finally, we consider the notion of acceptability. User friendliness, privacy, and legal issues require that much attention is given to this characteristic. However, rarely do evaluations consider all these ideas collectively. Privacy impact assessments (PIA) do just that; assess the impact of a proposed system on an individual's privacy. Research on the legality of systems is limited to statutes, and precedence. On the other hand, evaluations of user friendliness typically focus on the usability of a system without giving much consideration to privacy and legal issues. Regardless of these issues, one can envision a multi-biometric system which may improve user friendliness, improve subject privacy, and fall within the scope of legal precedence.

As a result, a holistic process for evaluating multi-biometric systems seems possible. This does not indicate that all multi-biometric systems will result in improvements to all characteristics. Rather it highlights the need to consider all characteristics and how they may be positively or negatively impacted.

An Increased Level of Evaluation Granularity

The previous analysis of the uni-biometric to multi-biometric transition hinted at a problem with evaluation across the seven traditional characteristics of a biometric / biometric system. Namely, some characteristics entail distinct aspects which are measured in different ways. For instance, considering the characteristic of performance, the definition rightly calls out the difference between "accuracy" and "resource constraints." While accuracy can easily be measured through rates such as FAR, FRR, FNMR, FMR, etc., these rates do not account for processing time, or financial costs of supporting hardware / software. It goes without saying that these additional aspects are of utmost importance, especially when dealing with large-scale federal multi-biometric systems. For instance, a multi-biometric system capable of a 0.00001% EER is of little use to a DHS border crossing application if it requires 24 hours to retrieve an end to end result. Likewise, a system capable of collecting a multiple sources of biometric input at 1,000 meters is of little use to DoD if it costs ten trillion dollars to produce. At a minimum, aspects such as these must be broken out as a part of a holistic evaluation of a multi-biometric system [Gro09]. However, one could potentially go further. Table 2 reflects a proposed scope of characteristics which are tailored more to the evaluation of a multi-biometric system as opposed to single a biometric modality (as originally proposed by Jain).

While it is beyond the scope of this work to provide detailed examples of each type of characteristic, it is fairly straightforward to envision how the characteristics are particularly important when evaluating large scale multi-biometric systems (such as those observed in the U.S. federal government). Notions behind characteristics such as cost and efficiency are quite obvious and require little justification. However availability, interoperability, mobility, etc. may not be as obvious. For instance, technical advances may provide the ability to capture iris images at a distance beyond 100 meters. However, a multi-biometric system must be able to deal with the constraints of the intended operational environment. Various types of subject occlusion, movement, and other factors may make multi-biometric acquisition impossible in various environments.

The characteristic of interoperability is increasingly important in modern day large scale biometric systems. For instance, a multi-biometric system which collects many modalities and a number of representations may not be able to leverage the information contained in the databases of other large scale systems if the information collected does not follow agreed upon standard formats. The need for considering mobility may also be less than obvious. In many scenarios, it may be necessary to collect, transmit, and or match multi-biometric information in remote environments. The ability of a multi-biometric system to account for such considerations represents an important part of a holistic evaluation process.

A Proposed Methodology for Evaluation

There are a number of situations where it would be beneficial to have a holistic approach to evaluating a multi-biometric system. The example used to this point involved transitioning from a uni-biometric to a multi-biometric system. Such an evaluation scheme would also afford the opportunity to assess multi-biometric systems which are currently in place. Additionally multi-biometric systems often incrementally add sources of input across their life cycles. An evaluation methodology of this nature would allow one to carefully consider whether or not it is a good decision to incorporate one or more new sources of biometric information.

Understanding the importance of having access to a holistic multi-biometric system evaluation methodology, we now consider details of how such a methodology could work. Ideally, a methodology would start by developing a set of quantifiable measurements for each proposed characteristics in Table 2. As indicated previously, many of the traditional seven characteristics of a biometric have associated mechanisms for quantification such as performance, and measurability. Furthermore, these measures have served as the primary basis for large scale evaluations [PG09, GTQS09]. Some of these measures will directly translate into the proposed set of characteristics

Table 2 – Proposed Characteristics for Evaluating a Multi-biometric System

Characteristic	Description
1. Acceptability	The degree in which the subject population in question agrees to use the biometric technology.
2. Accuracy	The accuracy of the multi-biometric system as measured by FAR, FRR, FNMR, FMR, etc.
3. Availability	The availability of the multi-biometric information in the intended operational environment
4. Circumvention	The ease in which an individual can spoof, masquerade, or mask identity causing incorrect identity decisions.
5. Convenience	The ease of use or user friendliness of the multi-biometric system.
6. Cost	The financial cost of the multi-biometric system.
7. Efficiency	The speed of any or all blocks of the multi-biometric system (acquisition, feature extraction, matching, searching, end-to-end processing)
8. Interoperability	The degree in which multi-biometric data conforms to standards and can be utilized by other systems.
9. Legality	The degree in which the multi-biometric system sets or abides by legal precedence.
10. Measurability	The ability to acquire and digitize the multi-biometric traits (assuming availability).
11. Mobility	The degree in which the multi-biometric
12. Permanence	The invariance over time of the collective multi-biometric traits in question.
13. Privacy	The degree in which the biometric system lawfully infringes on personal privacy.
14. Scalability	The ease in which the system can handle and or adapt to growing work volume (i.e. subject population, matching throughput, etc.)
15. Uniqueness	The difference of the collective multi- of
16. Universality	The possession of suitable samples of the collective multi-biometric trait across the population in question.

for evaluating a multi-biometric system as well. Characteristics such as cost, efficiency, and interoperability could also easily be quantified. Other characteristics do not have highly widely accepted measures but may have a new bed of research that can lead to such metrics. For instance, various forms of capacity analysis research shed light into the uniqueness of biometric input based on feature sets and or template structure [SN08]. There are however, characteristics that cannot be easily quantified such as convenience, legality, and privacy.

In lieu of straightforward quantification opportunities, one could introduce an evaluation methodology which which relies on ordinal relativity. For instance, it is difficult to quantify the characteristic of circumvention. But one could simply assign values for increasing levels of difficulty such as easy, moderate, difficult, etc. Such approaches have been adopted before in the form of meta-analyses and often provide valuable information [LBD+09]. Alternatively, one could employ comparative methods which juxtapose one system against another. Therefore, values could be assigned according to impact (i.e. negative impact, neutral, positive impact).

Regardless of the mechanisms, if quantification schemes are developed for each of the proposed characteristics and an objective function can be introduced, evaluation of the multi-biometric system(s) in question simplifies to a multi-dimensional maximization problem. A simple example of such an objective function is presented in Equation 1.

Equation 1 – Example Objective Function for Multi-Biometric System Evaluation

$$\max_{S \in \Re} \left(\sum_{i=1}^{n} C_i * W_i \right)$$

Where
S = Evaluation Score
$\Re$ = Multi-dimensional Input Space
n = Number of Characteristics
C_i = Score for Characteristic i
W_i = Weight of Characterstic i $\quad \sum_{i=1}^{n} W_i = 1$

In the simple objective function, quantitative values for each characteristic are multiplied by the corresponding characteristic weights and subsequently summed. Note, weighting schemes could be devised to uniformly distribute importance across all characteristics, selectively highlight certain characteristics, or completely eliminate one or more characteristics from consideration. At that point, evaluation is simply a matter of determining the appropriate objective function and inserting appropriate values for each system to be evaluated / compared. Such a methodology would allow an evaluation to account for multiple characteristics holistically.

Conclusion

In this paper, we provided supporting arguments for expanding the basis of multi-biometric system evaluation to reflect a holistic scope. Instead of focusing merely on motivations associated with performance, universality, measurability, and circumvention, we showed how all seven traditional characteristics of a biometric can be incorporated in multi-biometric system evaluation. Additionally, we proposed a new set of characteristics which are appropriate for evaluating large-scale federal multi-biometric systems. Having proposed this set of evaluation characteristics, we presented a simple, yet highly adaptable methodology for quantifying and evaluating multi-biometric systems holistically. Finally, we demonstrated how the methodology could be prioritized such that the overall measure of success need not uniformly consider all characteristics.

Bibliography

[fsd07] 10-Fingerprint Scanners to Deploy at all Ports of Entry. Website, Nov 2007. http://www.dhs.gov/xtrvlsec/programs/gc_1194553866460.shtm.

[Gro09] P. Grother. The Need for Large-Scale Biometric Testing. In *Biometrics Consortium Conference, 2009. BCC 2009.*, Sept. 2009. Presentation.

[GTQS09] P. Grother, E. Tabassi, G.W. Quinn, and W. Salamon. IREX I: Performance of Iris Recognition Algorithms on Standard Images. Technical Report 7629, NIST, 2009.

[JR08] A. Jain and A. Ross. Introduction to biometrics. In A.K. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, chapter 1, pages 1–22. Springer, New York, NY, 2008.

[LBD+09] Y.M. Lui, D. Bolme, B.A. Draper, J.R. Beveridge, G. Givens, and P.J. Phillips. A Meta-Analysis of Face Recognition Covariates. In *Biometrics: Theory, Applications, and Systems, 2009. BTAS 2009. Third IEEE International Conference on*, Sept. 2009.

[nga09] Next Generation ABIS Goes Operational, Now Referred To as DoD ABIS. Website, Jan 2009. http://www.biometrics.dod.mil/Newsletter/Issues/2009/Apr/v5issue2_a1.html.

[ngi09] Next Generation Identification. Website, Jun 2009. http://www.fbi.gov/hq/cjisd/ngi.htm.

[PG09] B. Wing P. Grother. MBTE: Assessing Capability of Stand-Off Iris Capture for Border Crossing Evaluation Summary. In *Biometrics Consortium Conference, 2009. BCC 2009.*, Sept. 2009. Presentation.

[RLG09] R. N. Rodrigues, L. L. Ling, and V. Govindaraju. Robustness of Multimodal Biometric Fusion Methods Against Spoof Attacks. *Journal of Visual Languages & Computing*, 20(3):169 – 179, 2009. Advances in Multimodal Biometric Systems - Multimodal Biometrics.

[RNJ06] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics*. Springer-Verlag New York, Inc., 2006.

[RNJ08] A. Ross, K. Nandakumar, and A.K. Jain. Introduction to Multibiometrics. In A.K. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, chapter 14, pages 271–292. Springer, New York, NY, 2008.

[SN08] N.A. Schmid and F. Nicolo. On Empirical Recognition Capacity of Biometric Systems Under Global PCA and ICA Encoding. *IEEE Transactions on Information Forensics and Security*, 3(3):512–528, 2008.