Appeared in Proc. of 12th European Signal Processing Conference (EUSIPCO), (Vienna, Austria), pp. 1221-1224, September 2004.

MULTIMODAL BIOMETRICS: AN OVERVIEW

Arun Ross and Anil K. Jain

West Virginia University
Morgantown, WV 26506 USA
ross@csee.wvu.edu

Michigan State University
East Lansing, MI 48823 USA
jain@cse.msu.edu

ABSTRACT

Unimodal biometric systems have to contend with a variety of problems such as noisy data, intra-class variations, restricted degrees of freedom, non-universality, spoof attacks, and unacceptable error rates. Some of these limitations can be addressed by deploying *multimodal biometric systems* that integrate the evidence presented by multiple sources of information. This paper discusses the various scenarios that are possible in multimodal biometric systems, the levels of fusion that are plausible and the integration strategies that can be adopted to consolidate information. We also present several examples of multimodal systems that have been described in the literature.

1. INTRODUCTION

Establishing the identity of a person is becoming critical in our vastly interconnected society. Questions like "Is she really who she claims to be?", "Is this person authorized to use this facility?" or "Is he in the watchlist posted by the government?" are routinely being posed in a variety of scenarios ranging from issuing a driver's licence to gaining entry into a country. The need for reliable user authentication techniques has increased in the wake of heightened concerns about security and rapid advancements in networking, communication and mobility. Biometrics, described as the science of recognizing an individual based on her physiological or behavioral traits, is beginning to gain acceptance as a legitimate method for determining an individual's identity. Biometric systems have now been deployed in various commercial, civilian and forensic applications as a means of establishing identity. These systems rely on the evidence of fingerprints, hand geometry, iris, retina, face, hand vein, facial thermogram, signature, voice, etc. to either validate or determine an identity [1].

Most biometric systems deployed in real-world applications are unimodal, i.e., they rely on the evidence of a single source of information for authentication (e.g., single fingerprint *or* face). These systems have to contend with a variety of problems such as: (a) Noise in sensed data: A fingerprint image with a scar, or a voice sample altered by cold are examples of noisy data. Noisy data could also result from defective or improperly maintained sensors (e.g., accumulation of dirt on a fingerprint sensor) or unfavorable ambient conditions (e.g., poor illumination of a user's face in a face recognition system). (b) Intra-class variations: These variations are typically caused by a user who is incorrectly interacting with the sensor (e.g., incorrect facial pose), or when the characteristics of a sensor are modified during authentication (e.g., optical versus solid-state fingerprint sensors). (c) Inter-class similarities: In a biometric system comprising of a large number of users, there may be inter-class similarities (overlap) in the feature space of multiple users. Golfarelli et al. [2] state that the number of distinguishable patterns in two of the most commonly used representations of hand geometry and face are only of the order of 10^5 and 10^3, respectively. (d) Non-universality: The biometric system may not be able to acquire meaningful biometric data from a subset of users. A fingerprint biometric system, for example, may extract incorrect minutiae features from the fingerprints of certain individuals, due to the poor quality of the ridges. (e) Spoof attacks: This type of attack is especially relevant when behavioral traits such as signature or voice are used. However, physical traits such as fingerprints are also susceptible to spoof attacks.

Some of the limitations imposed by unimodal biometric systems can be overcome by including multiple sources of information for establishing identity [3]. Such systems, known as *multimodal biometric systems*, are expected to be more reliable due to the presence of multiple, (fairly) independent pieces of evidence [4]. These systems are able to meet the stringent performance requirements imposed by various applications. They address the problem of non-universality, since multiple traits ensure sufficient population coverage. They also deter spoofing since it would be difficult for an impostor to spoof multiple biometric traits of a genuine user simultaneously. Furthermore, they can facilitate a challenge-response type of mechanism by requesting the user to present a random subset of biometric traits thereby ensuring that a 'live' user is indeed present at the point of data acquisition.

In this paper we examine the levels of fusion that are plausible in a multimodal biometric system, the various scenarios that are possible, the different modes of operation, the integration strategies that can be adopted and the issues related to the design and deployment of these systems.

2. LEVELS OF FUSION

A generic biometric system has 4 important modules: (a) the sensor module which captures the *trait* in the form of raw biometric *data*; (b) the feature extraction module which processes the data to extract a *feature set* that is a compact representation of the trait; (c) the matching module which employs a classifier to compare the extracted feature set with the templates residing in the database to generate matching scores; (d) the decision module which uses the matching scores to either determine an identity or validate a claimed identity. In a multimodal biometric system information reconciliation can occur in any of the aforementioned modules (see Figure 1). (a) Fusion at the data or feature level: Either the data itself or the feature sets originating from multiple sensors/sources are fused. (b) Fusion at the match score level: The scores generated by multiple classifiers pertaining to different modalities are combined. (c) Fusion at the decision level: The final out-

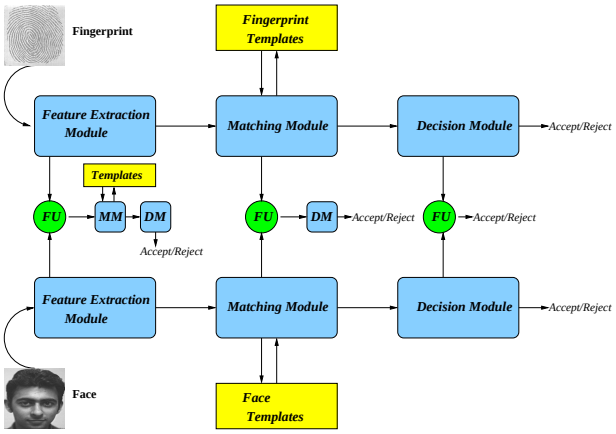

Figure 1: Levels of fusion in a bimodal biometric system; FU: Fusion Module, MM: Matching Module, DM: Decision Module.

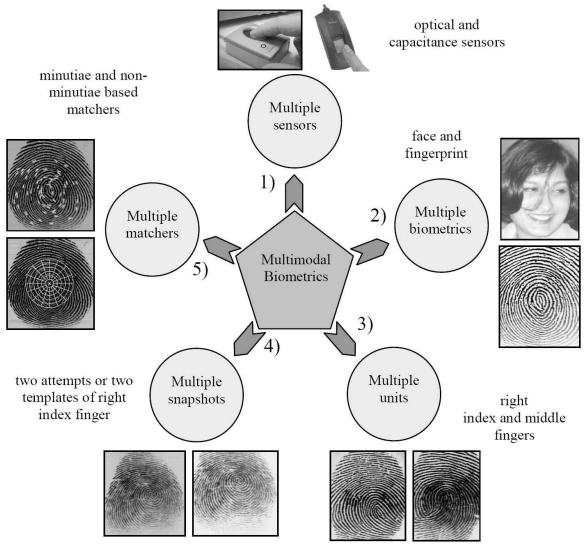

Figure 2: Scenarios in a multimodal biometric system [6].

put[1] of multiple classifiers are consolidated via techniques such as majority voting [5]. Biometric systems that integrate information at an early stage of processing are believed to be more effective than those systems which perform integration at a later stage. Since the feature set contains richer information about the input biometric data than the matching score or the output decision of a matcher, fusion at the feature level is expected to provide better recognition results. However, fusion at this level is difficult to achieve in practice because (i) the feature sets of the various modalities may not be compatible (e.g., eigen-coefficients of face and minutiae set of finger), and (ii) most commercial biometric systems do not provide access to the feature sets (nor the raw data) which they use in their products. Fusion at the decision level is considered to be rigid due to the availability of limited information. Thus, fusion at the match score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities.

3. FUSION SCENARIOS

Depending on the number of traits, sensors, and feature sets used, a variety of scenarios are possible in a multimodal biometric system (Figure 2).

1. Single biometric trait, multiple sensors: Multiple sensors record the same biometric trait. Thus, raw biometric data pertaining to different sensors are obtained. Chang et al. [7] acquire both 2D and 3D images of the face and combine them at the data level as well as the match score level to improve the performance of a face recognition system. Kumar et al. [8] describe a hand-based verification system that combines the geometric features of the hand with palmprints at the feature and match score levels. Interestingly, in their experiments, fusion at the match score level results in better performance than fusion at the feature level. This could be due to the high-dimensionality of the fused feature set (the curse-of-dimensionality problem) and, therefore, the application of a feature reduction technique may have been appropriate (see Section 5).
2. Single biometric trait, multiple classifiers: Unlike the previous scenario, only a single sensor is employed to obtain raw data; this data is then used by multiple classifiers. Each of these classifiers either operate on the same feature set extracted from the data or generate their own feature sets. Jain et al. [9] use the logistic function to integrate the matching scores obtained from three different fingerprint matchers operating on the same minutiae sets (also see [10]). Ross et al. [11] combine the matching score of a minutiae-based fingerprint matcher with that of a texture-based matcher to improve matching performance. Lu et al. extract three different types of feature sets from the face image of a subject (using PCA, LDA and ICA) and integrate the output of the corresponding classifiers at the match score level [12].
3. Single biometric trait, multiple units: In the case of fingerprints (or iris), it is possible to integrate information presented by 2 or more fingers (or both the irises) of a single user. This is an inexpensive way of improving system performance since this does not entail deploying multiple sensors nor incorporating additional feature extraction and/or matching modules.
4. Multiple biometric traits: Here, multiple biometric traits of an individual are used to establish the identity. Such systems employ multiple sensors to acquire data pertaining to different traits. The independence of the traits ensures that a significant improvement in performance is obtained. Brunelli et al. [13] use the face and voice traits of an individual for identification. A HyperBF network is used to combine the normalized scores of five different classifiers operating on the voice and face feature sets. Bigun et al. develop a statistical framework based on Bayesian statistics to integrate the speech (text-dependent) and face data of a user [14]. The estimated biases of each classifier is taken into account during the fusion process. Hong and Jain associate different confidence measures with the individual matchers when integrating the face and fingerprint traits of a user [15]. They

[1] In a verification system the output is an "Accept" or a "Reject" while in an identification system the output is the identity label of an enrolled user.

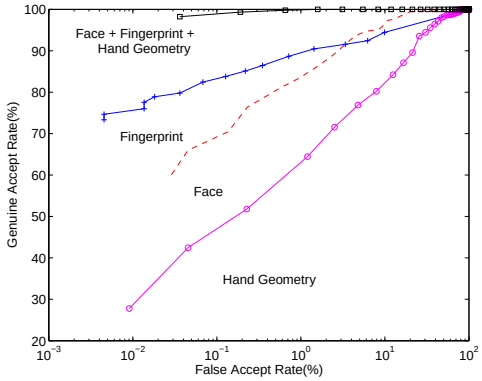

Figure 3: Performance gain using the sum rule [3].

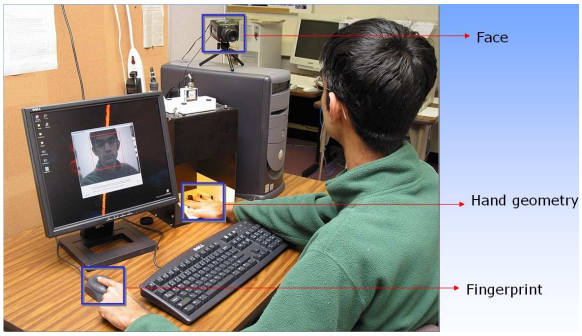

Figure 4: A prototype multimodal biometric login system.

also suggest an indexing mechanism wherein face information is used to retrieve a set of possible identities and the fingerprint information is then used to select a single identity. A commercial product called BioID [16] uses the voice, lip motion and face features of a user to verify identity.

4. MODES OF OPERATION

A multimodal system can operate in one of three different modes: serial mode, parallel mode, or hierarchical mode. In the serial mode of operation, the output of one modality is typically used to narrow down the number of possible identities before the next modality is used [15]. Therefore, multiple sources of information (e.g., multiple traits) do *not* have to be acquired simultaneously. Further, a decision could be made before acquiring *all* the traits. This can reduce the overall recognition time. In the parallel mode of operation, the information from multiple modalities are used simultaneously in order to perform recognition. In the hierarchical scheme, individual classifiers are combined in a treelike structure. This mode is relevant when the number of classifiers is large.

5. INTEGRATION STRATEGIES

The strategy adopted for integration depends on the level at which fusion is performed. Fusion at the feature level can be accomplished by concatenating two *compatible* feature sets. Feature selection/reduction techniques may be employed to handle the curse-of-dimensionality problem. Fusion at the match score level has been well studied in the literature [17, 18]. Robust and efficient normalization techniques are necessary to transform the scores of multiple matchers into a common domain prior to consolidating them [19]. In the context of verification, two distinct strategies exist for fusion at this level. In the first approach the fusion is viewed as a classification problem where a feature vector is constructed using the matching scores output by the individual matchers; this feature vector is then classified into one of two classes: Accept (genuine user) or Reject (impostor) [20]. In the second approach the fusion is viewed as a combination problem where the individual matching scores are combined to generate a single scalar score which is then used to make the final decision [21, 22]. General strategies for combining multiple classifiers have been suggested in [23] and [24]. Ross and Jain have shown [3] that the simple sum rule is sufficient to obtain a significant improvement in the matching performance of a multimodal biometric system (Figure 3). They also suggest a technique to incorporate user-specific weights to further improve the system performance [25]. Fusion strategies at the decision level include majority voting [5], behavior knowledge space method [26], weighted voting based on the Dempster-Shafer theory of evidence [27], AND/OR rules [28], etc.

6. DESIGN ISSUES

A variety of factors should be considered when designing a multimodal biometric system. These include (a) the choice and number of biometric traits, (b) the level in the biometric system at which information provided by multiple traits should be integrated, (c) the methodology adopted to integrate the information, and (d) the cost versus matching performance trade off. The choice and number of biometric traits is largely driven by the nature of the application, the overhead introduced by multiple traits (computational demands and cost, for example), and the correlation between the traits considered (uncorrelated information is preferred since the performance improvement is more pronounced in this case). In a cell phone that is equipped with a camera, it might be easier to combine the face and voice traits of a user, while in an ATM application it might be easier to combine the fingerprint and face traits of the user. In identification systems comprising of a large number of users (in the order of millions), an indexing mechanism may be facilitated using a multimodal approach [15]. Researchers are currently studying the performance gain that can be obtained using state-of-the-art commercial off-the-shelf (COTS) fingerprint and face systems, on a large population of individuals [19].

7. SUMMARY AND CONCLUSIONS

Multimodal biometric systems elegantly address several of the problems present in unimodal systems. By combining multiple sources of information, these systems improve matching performance, increase population coverage, deter spoofing, and facilitate indexing. Various fusion levels and scenarios are possible in multimodal systems. Fusion at the match score level is the most popular due to the ease in accessing and consolidating matching scores. Performance gain is pronounced when uncorrelated traits are used in a multimodal system. Incorporating user-specific parameters can further improve performance of these systems. With

the widespread deployment of biometric systems in several civilian and government applications, it is only a matter of time before multimodal biometric systems begin to impact the way in which identity is established in the 21st century (Figure 4).

REFERENCES

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, pp. 4–20, Jan 2004.

[2] M. Golfarelli, D. Maio, and D. Maltoni, "On the error-reject tradeoff in biometric verification systems," *IEEE Trans. on Patt. Anal. and Mach. Intell.*, vol. 19, pp. 786–796, July 1997.

[3] A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters*, vol. 24, pp. 2115–2125, Sep 2003.

[4] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Is independence good for combining classifiers?," in *Proc. of Int'l Conf. on Pattern Recognition (ICPR)*, vol. 2, (Barcelona, Spain), pp. 168–171, 2000.

[5] Y. Zuev and S. Ivanon, "The voting as a way to increase the decision reliability," in *Foundations of Information/Decision Fusion with Applications to Engineering Problems*, (Washington D.C., USA), pp. 206–210, August 1996.

[6] S. Prabhakar and A. K. Jain, "Decision-level fusion in fingerprint verification," *Pattern Recognition*, vol. 35, no. 4, pp. 861–874, 2002.

[7] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Face recognition using 2D and 3D facial data," in *Proc. of Workshop on Multimodal User Authentication*, (Santa Barbara, CA), pp. 25–32, Dec 2003.

[8] A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain, "Personal verification using palmprint and hand geometry biometric," in *Proc. of 4th Int'l Conf. on Audio and Video-based Biometric Person Authentication (AVBPA)*, (Guildford, UK), pp. 668–678, Jun 2003.

[9] A. K. Jain, S. Prabhakar, and S. Chen, "Combining multiple matchers for a high security fingerprint verification system," *Pattern Recognition Letters*, vol. 20, pp. 1371–1379, 1999.

[10] G. L. Marcialis and F. Roli, "Experimental results on fusion of multiple fingerprint matchers," in *Proc. of 4th Int'l Conf. on Audio and Video-based Biometric Person Authentication (AVBPA)*, (Guildford, UK), pp. 814–820, Jun 2003.

[11] A. Ross, A. K. Jain, and J. Reisman, "A hybrid fingerprint matcher," *Pattern Recognition*, vol. 36, pp. 1661–1673, Jul 2003.

[12] X. Lu, Y. Wang, and A. K. Jain, "Combining classifiers for face recognition," in *Proc. IEEE Int'l Conf. on Multimedia and Expo (ICME)*, vol. 3, (Baltimore, MD), pp. 13–16, Jul 2003.

[13] R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Transactions on PAMI*, vol. 12, pp. 955–966, Oct 1995.

[14] E. Bigun, J. Bigun, B. Duc, and S. Fischer, "Expert conciliation for multimodal person authentication systems using Bayesian Statistics," in *First International Conference on AVBPA*, (Crans-Montana, Switzerland), pp. 291–300, March 1997.

[15] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Transactions on PAMI*, vol. 20, pp. 1295–1307, Dec 1998.

[16] R. W. Frischholz and U. Dieckmann, "Bioid: A multimodal biometric identification system," *IEEE Computer*, vol. 33, no. 2, pp. 64–68, 2000.

[17] R. P. W. Duin and D. M. J. Tax, "Experiments with classifier combining rules," in *Proc. of 1st Workshop on Multiple Classifier Systems*, vol. LNCS 1857, (Cagliari, Italy), pp. 16–29, Springer, Jun 2000.

[18] D. M. J. Tax, M. V. Breukelen, R. P. W. Duin, and J. Kittler, "Combining multiple classifiers by averaging or by multiplying?," *Pattern Recognition*, vol. 33, pp. 1475–1485, Sep 2000.

[19] M. Indovina, U. Uludag, R. Snelick, A. Mink, and A. K. Jain, "Multimodal biometric authentication methods: A COTS approach," in *Proc. of Workshop on Multimodal User Authentication*, (Santa Barbara, CA), pp. 99–106, Dec 2003.

[20] P. Verlinde and G. Cholet, "Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application," in *Proc. of 2nd Int'l Conf. on Audio- and Video-based Person Authentication*, (Washington D.C., USA), pp. 188–193, March 1999.

[21] U. Dieckmann, P. Plankensteiner, and T. Wagner, "Sesam: A biometric person identification system using sensor fusion," *Pattern Recognition Letters*, vol. 18, no. 9, pp. 827–833, 1997.

[22] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. on Neural Networks*, vol. 10, pp. 1065–1074, 1999.

[23] T. K. Ho, J. J. Hull, and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. on Patt. Anal. and Mach. Intel.*, vol. 16, pp. 66–75, January 1994.

[24] J. Kittler, M. Hatef, R. P. Duin, and J. G. Matas, "On combining classifiers," *IEEE Transactions on Pat. Anal. and Mach. Intel.*, vol. 20, pp. 226–239, Mar 1998.

[25] A. K. Jain and A. Ross, "Learning user-specific parameters in a multibiometric system," in *Proc. of the International Conference on Image Processing (ICIP)*, (Rochester, USA), pp. 57–60, Sep 2002.

[26] L. Lam and C. Y. Suen, "Optimal combination of pattern classifiers," *Pattern Recognition Letters*, vol. 16, no. 9, pp. 945–954, 1995.

[27] L. Xu, A. Krzyzak, and C. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 22, no. 3, pp. 418–435, 1992.

[28] J. Daugman, "Combining multiple biometrics," http://www.cl.cam.ac.uk/users/jgd1000/combine/.