The latest research indicates using a combination of biometric avenues for human identification is more effective, and far more challenging.

Multibiometric Systems

By Anil K. Jain and Arun Ross

Traditionally, passwords (knowledge-based security) and ID cards (token-based security) have been used to restrict access to secure systems. However, security can be easily breached in these systems when a password is divulged to an unauthorized user or a card is stolen by an impostor. Furthermore, simple passwords are easy to guess by an impostor and difficult passwords may be hard to recall by a legitimate user. The emergence of biometrics has addressed the problems that plague traditional verification methods. Biometrics refers to the automatic identification (or verification) of an individual (or a claimed identity) by using certain physiological or behavioral traits associated with the person (see Figure 1). By using biometrics

Illustration by Sandra Dionisi

it is possible to establish an identity based on "who you are," rather than by "what you possess" (for example, an ID card) or "what you remember" (for example, a password). Current biometric systems make use of fingerprints, hand geometry, iris, retina, face, facial thermograms, signature, gait, palm print and voiceprint to establish a person's identity [4].

While biometric systems have their limitations they have an edge over traditional security methods in that they cannot be easily stolen or shared. Besides bolstering security, biometric systems also enhance user convenience by alleviating the need to design and remember passwords. Moreover, biometrics is one of the few techniques that can be used for negative recognition where the system determines whether the person is who he or she denies to be.

Biometric systems can operate in one of two modes—the identification mode, in which the identity of an unknown user is determined, and the verification mode, in which a claimed identity is either accepted (a genuine user) or rejected (an impostor). Biometric systems are being deployed in various applications including computer logins, ATMs, grocery stores, airport kiosks, and driver's licenses. The successful installation of biometric systems in these apºplications does not imply that biometrics is a solved problem. In fact, there is significant room for improvement in biometrics as suggested by the error rates shown in the table on the next page.

WHILE biometric systems have their limitations they have an edge over traditional security methods in that they cannot be easily stolen or shared. Besides bolstering security, biometric systems also enhance user convenience by alleviating the need to design and remember passwords. Moreover, biometrics is one of the few techniques that can be used for negative recognition where the system determines whether the person is who he or she denies *to be.*

Biometric systems installed in real-world applications must contend with a variety of problems. Among them are:

Noise in sensed data. A fingerprint with a scar and a voice altered by a cold are examples of noisy inputs. Noisy data could also result from defective or improperly maintained sensors (for example, accumulation of dirt on a fingerprint sensor) and unfavorable ambient conditions (for example, poor illumination of a user's face in a face recognition system). Noisy biometric data may be incorrectly matched with templates in the database resulting in a user being incorrectly rejected.

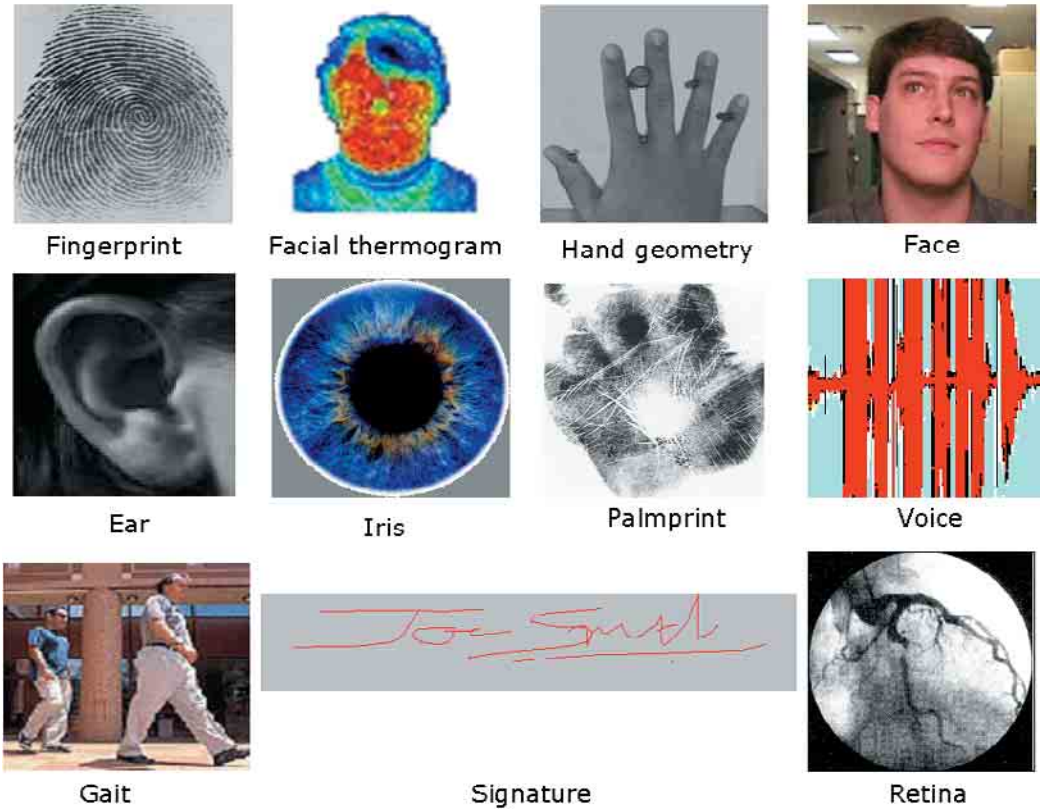

Figure 1. Examples of some of the biometric traits used for authenticating an individual. (Gait image taken from www.findbiometrics.com.)

Intra-class variations. The biometric data acquired from an individual during authentication may be very different from the data used to generate the template during enrollment, thereby affecting the matching process. This variation is typically caused by a user who is incorrectly interacting with the sensor, or when sensor characteristics are modified (for example, by changing sensors, that is, the sensor interoperability problem) during authentication.

Distinctiveness. While a biometric trait is expected to vary significantly across individuals, there may be large similarities in the feature sets used to represent these traits. Thus, every biometric trait has some theoretical upper bound in terms of its discrimination capability.

Non-universality. While every user is expected to possess the biometric trait being acquired, in reality it is possible for a subset of the users to not possess a particular biometric. A fingerprint biometric system, for example, may be unable to extract features from the fingerprints of certain individuals, due to the poor quality of the ridges (see Figure 2a). Thus, there is a failure to enroll (FTE) rate associated with using a single biometric trait. There is empirical evidence that about 4% of the population may have poor quality fingerprints that cannot be easily imaged by some of the existing sensors.

Spoof attacks. An impostor may attempt to spoof the biometric trait of a legitimately enrolled user in order to circumvent the system. This type of attack is especially relevant when behavioral traits such as signature and voice are used. However, physical traits like fingerprints are also susceptible to spoof attacks.

	Test	Test Parameter	False Reject Rate (FRR)	False Accept Rate (FAR)
Fingerprint	FVC 2002*	Users mostly in the age group 20-39	0.2%	0.2%
Face	FRVT 2002**	Enrollment and test images were collected in indoor environment and could be on different days	10%	1%
Voice	NIST 2000***	Text dependent	10-20%	2-5%

*Fingerprint Verification Competition; bias.csr.unibo.it/fvc2002
**Face Recognition Vendor Test; www.frvt.org/FRVT2002
***National Institute of Standards and Technology; www.nist.gov/speech/tests/spk/2000

State-of-the-art error rates associated with fingerprint, face, and voice biometric systems. The accuracy estimates of biometric systems depend on a number of test conditions.

Multibiometric Systems

Some of the limitations imposed by unimodal biometric systems (that is, biometric systems that rely on the evidence of a single biometric trait) can be overcome by using multiple biometric modalities [1, 2, 6]. Such systems, known as multibiometric systems, are expected to be more reliable due to the presence of multiple, fairly independent pieces of

evidence. These systems are also able to meet the stringent performance requirements imposed by various applications. Multibiometric systems address the problem of non-universality, since multiple traits can ensure sufficient population coverage. Furthermore, multibiometric systems provide anti-spoofing measures by making it difficult for an intruder to simultaneously spoof the multiple biometric traits of a legitimate user. By asking the user to present a random subset of biometric traits, the system ensures a live user is indeed present at the point of data acquisition. Thus, a challenge-response type of authentication can be facilitated using multibiometric systems.

A variety of factors should be considered when designing a multibiometric system. These include the choice and number of biometric traits; the level in the biometric system at which information provided by multiple traits should be integrated; the methodology adopted to integrate the information; and 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. In a cell phone 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. A commercial multibiometric system called BioID (www.bioid.com) integrates the face, voice, and lip movement of an individual.

The information presented by multiple traits may be consolidated at various levels.[1] At the feature extraction level, the feature sets of multiple modalities are integrated and a new feature set is generated; the new feature set is then used in the matching and decision-making modules of the biometric system. At the matching score level, the matching scores output by multiple matchers are integrated. At the decision level, the final decisions made by the individual systems are consolidated by employing techniques such as majority voting.

Although integration at the feature extraction level is expected to perform better than fusion at the other two levels, it is not always feasible for a number of reasons. First, most commercial systems do not provide access to information at this level. Second, the feature spaces of different biometric traits may not be compatible. For example, it is difficult to combine the minutiae feature set of a fingerprint image with the eigen-coefficients of a face image. Third, even if the feature sets were compatible, concatenation might result in a feature vector with a very large dimensionality leading to the "curse of dimensionality" problem. Fusion at the decision level is considered rigid due to the availability of limited information. In fact, the only type of information available at this level is an "Accept" or a "Reject" label in the verification mode, or the identity of the user in the identification mode.

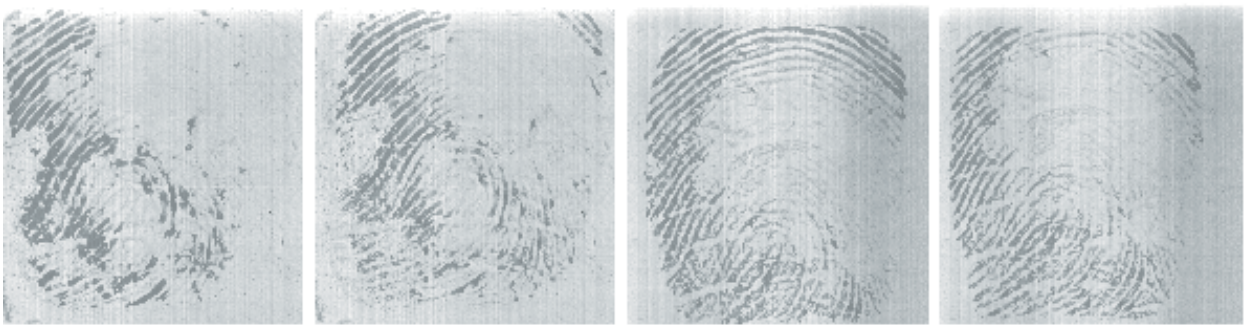

Figure 2a. The failure-to-enroll (FTE) problem as observed in fingerprints. The four impressions of a user's fingerprint shown here cannot be enrolled by most fingerprint systems, due to the poor image quality of the ridges. Consequently, alternate methods must be adopted in order to include this user in the system.

Due to the reasons stated here, fusion at the matching score level is usually preferred, as it is relatively easy to access and combine the scores presented by the different modalities. Note that fusion at this level is a practical compromise between fusion at the other two levels. In the context of verification, two distinct approaches 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) [9]. In the second

[1]Apart from the three levels mentioned here, fusion is also possible at the sensor level and rank level [3].

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. Our experiments suggest the combination approach performs better than the classification approach [7]; however, the approach of choice can vary depending on the database used for testing.

The multibiometric system we have designed uses the face, fingerprint, and hand geometry attributes of a person for recognition. Our database consists of 100 different users with each user providing five samples per biometric. Information from these three modalities was integrated at the matching score level using the combination approach [7]. Fingerprints were represented using minutiae features, and the output of the fingerprint matcher was a similarity score in the (0,100) range; face images were represented using eigen-coefficients, and the output of the face matcher was a distance score; hand-geometry images were represented by 14 feature values (corresponding to the lengths and widths of the fingers, as well as the width of the palm), and the output of the matcher was a distance score. Prior to combining the raw scores, a normalization scheme was employed to transform the face and hand geometry scores into similarity scores in the (0,100) range. If s_1, s_2, and s_3 represent the *normalized* scores pertaining to the fingerprint, face, and hand geometry modalities, respectively, then the final score, S_{fus}, was computed as, $S_{fus} = w_1s_1 + w_2s_2 + w_3s_3$. Here w_1, w_2 and w_3 are the weights associated with the three traits, and $w1+w_2+w_3=1$. In the first set of experiments, equal weights were assigned to all the three modalities. The improved matching performance as observed using the Receiver Operating Characteristic (ROC) curve, which plots the Genuine Accept Rate (GAR) against the False Accept Rate (FAR) at various matching thresholds, is shown in Figure 2b.

It is essential that different biometric traits be given different degrees of importance for different users. This is especially significant when biometric traits of some users cannot be reliably acquired. For example, users with persistently dry fingers may not be able to provide good quality fingerprints. Such users might experience higher false rejects when interacting with the fingerprint system. By reducing the weight of the fingerprint trait and increasing the weights associated with the other traits, the false reject error rate of these users can be reduced. The biometric system learns user-specific parameters by observing system performance over a period of time. This will appeal to that segment of the population averse to interacting with a system that constantly requests a user to provide multiple readings of the same biometric. The emphasis is on tuning the system variables automatically, yet appropriately, to attain performance gain. We compute user-specific weights by an exhaustive search technique in which various sets of weights are tried out on a training set of genuine and impostor scores, and selecting the weight set that results in the least error [5]. Let $w_{1,i}$, $w_{2,i}$ and $w_{3,i}$ be the weights associated with the i^{th} user in the database. The algorithm operates on the training set as follows:

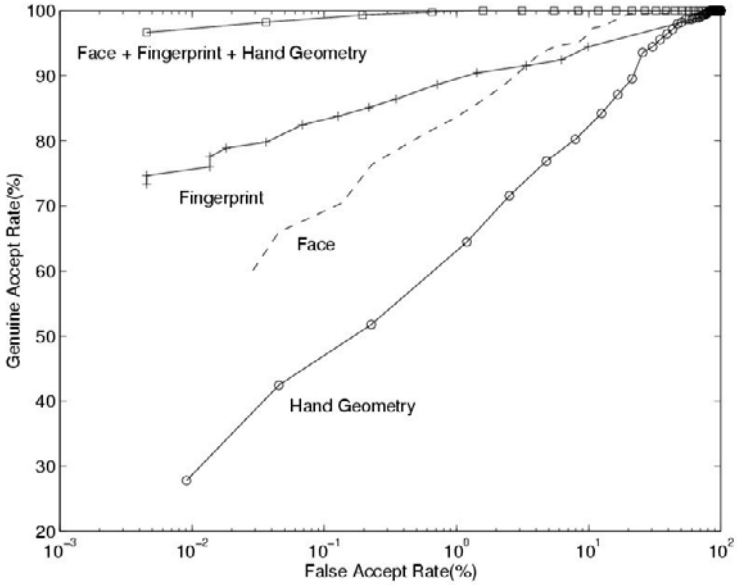

Figure 2b. The receiver operating characteristic (ROC) curve showing the performance gain when the simple sum rule is used to combine the matching scores of face, fingerprint, and hand geometry traits of 100 users (five samples per user).

- For the i^{th} user in the database, vary weights $w_{1,i}$, $w_{2,i}$ and $w_{3,i}$ over the range (0, 1), with the constraint $w_{1,i}+ w_{2,i}+w_{3,i} =1$. Compute $S_{fus} = w_{1,i}s_1+ w_{2,i}s_2+w_{3,i}s_3$. This computation is performed over all scores (that is, both genuine and impostor scores) associated with the i^{th} user.
- Choose that set of weights that minimizes the total error rate. The total error rate is the sum of the false accept and false reject rates pertaining to this user.

This technique was tested on a subset of 10 users who provided biometric data over a period of two months (approximately 30 samples per user per biometric). Figure 3 illustrates the case where reducing the face weight improves verification accuracy. Our exper-

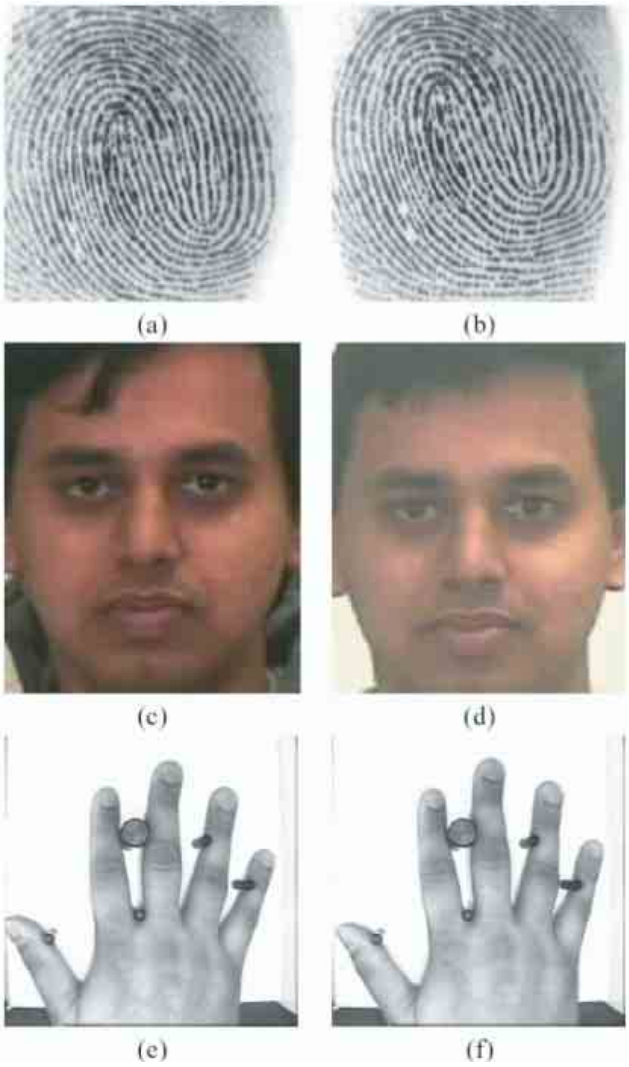

Figure 3. Eliminating false rejects by employing user-specific weights. (a), (c), and (e) are templates, and (b), (d), and (f) are the corresponding test samples (from the same user). At a matching threshold of 50, equal weighting of the three normalized matching scores (for fingerprint, hand geometry, and face) results in a false reject of this user, while user-specific weighting (weights of 0.6, 0.2, and 0.2 for fingerprint, hand and face, respectively) results in a correct acceptance of the user.

imental results indicate that employing user-specific weights further improves matching performance [5].

Conclusion

Multibiometric systems alleviate a few of the problems observed in unimodal biometric systems. Besides improving matching performance, they also address the problems of non-universality and spoofing. Multibiometric systems can integrate information at various levels, the most popular one being fusion at the matching score level where the scores output by the individual matchers are integrated. The simple sum rule results in improved matching performance, which can be further improved by employing user-specific biometric weights. User-specific weights aid in reducing the false reject rate, thereby enhancing user convenience.

It must be noted that deploying a multibiometric system introduces some overhead in terms of computational demands and costs. Therefore, it is important the cost versus performance trade-off is carefully studied before deploying these systems. Researchers from the National Institute of Science and Technology (NIST) used commercially available biometric products recently to acquire and test multibiometric data pertaining to 1,000 users [8]. This is an indication of the increased attention that multibiometric systems are receiving from the government (for various national identification programs currently under implementation such as the US-VISIT program) as well as from researchers (see Matt Turk's article in this section). C

References

1. Bigun, E.S., Bigun, J., Duc, B., and Fischer, S. Expert conciliation for multimodal person authentication systems using Bayesian statistics. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication.* (Crans-Montana, Switzerland, Mar. 1997), 291–300.
2. Brunelli, R., and Falavigna, D. Person identification using multiple cues. *IEEE Trans. on Pattern Analysis and Machine Intelligence 12,* 10 (Oct. 1995). IEEE, NY, 955–966.
3. Ho, T.K. Hull, J.J., and Srihari, S.N. Decision combination in multiple classifier systems. *IEEE Trans. on Pattern Analysis and Machine Intelligence 16,* 1 (1994), IEEE, NY, 66–75.
4. Jain, A.K., Bolle, R., and Pankanti, S. (Eds.). *Biometrics: Personal Identification in Networked Society.* Kluwer, Dordrecht, The Netherlands, 1999.
5. Jain, A.K., and Ross, A. Learning user-specific parameters in a multibiometric system. In *Proceedings of the IEEE International Conference on Image Processing* (Rochester, NY, Sept. 22–25, 2002), 57–60.
6. Kittler, J., Hatef, M., Duin, R., and Matas, J. On combining classifiers. *IEEE Trans. on Pattern Analysis and Machine Intelligence 20,* 3 (Mar. 1998). IEEE, NY, 226–239.
7. Ross, A., and Jain, A.K. Information fusion in biometrics. *Pattern Recognition Letters 24,* 13 (Sept. 2003), 2115–2125.
8. Snelick, R., Indovina, M., Yen, J., and Mink, A. Multimodal biometrics: Issues in design and testing. In *Proceedings of International Conference on Multimodal Interfaces* (Vancouver, B.C., Nov. 5–7, 2003).
9. Verlinde, P., and Cholet, G. Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application. In *Proceedings of the International Conference on Audio and Video-Based Biometric Person Authentication.* (Washington, D.C., Mar. 1999), 188–193.

Anil K. Jain (jain@cse.msu.edu) is a University Distinguished Professor in the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI.
Arun Ross (ross@csee.wvu.edu) is an assistant professor in the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV.

This work was funded in part by the Center for Identification Technology Research (CITeR), West Virginia University, under the NSF-IUCRC program.