Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

Feature Level Fusion Using Hand and Face Biometrics

Arun Ross[a] and Rohin Govindarajan[b]

[a] West Virginia University, Morgantown, WV 26506 USA
[b] Motorola Inc., Anaheim, CA 92806 USA

ABSTRACT

Multibiometric systems utilize the evidence presented by multiple biometric sources (e.g., face and fingerprint, multiple fingers of a user, multiple matchers, etc.) in order to determine or verify the identity of an individual. Information from multiple sources can be consolidated in several distinct levels, including the feature extraction level, match score level and decision level. While fusion at the match score and decision levels have been extensively studied in the literature, fusion at the feature level is a relatively understudied problem. In this paper we discuss fusion at the feature level in 3 different scenarios: (i) fusion of PCA and LDA coefficients of face; (ii) fusion of LDA coefficients corresponding to the R,G,B channels of a face image; (iii) fusion of face and hand modalities. Preliminary results are encouraging and help in highlighting the pros and cons of performing fusion at this level. The primary motivation of this work is to demonstrate the viability of such a fusion and to underscore the importance of pursuing further research in this direction.

1. INTRODUCTION

The pronounced need for establishing identity in a reliable manner has spurred active research in the field of biometrics.[1] It is now apparent that a single biometric is not sufficient to meet the variety of requirements - including matching performance - imposed by several large-scale authentication systems. Multibiometric systems seek to alleviate some of the drawbacks encountered by unibiometric systems by consolidating the evidence presented by multiple biometric traits/sources.[2] These systems can significantly improve the recognition performance of a biometric system besides improving population coverage, deterring spoof attacks, increasing the degrees-of-freedom, and reducing the failure-to-enroll rate.[3] Although the storage requirements, processing time and the computational demands of a multibiometric system are much higher (than a unibiometric system), the above mentioned advantages present a compelling case for deploying multibiometric systems in large-scale authentication systems.

Evidence in a multibiometric system can be integrated in several different levels as described below:

1. Sensor level: The raw data acquired from multiple sensors can be processed and integrated to generate new data from which features can be extracted. For example, in the case of face biometrics, both 2D texture information and 3D depth (range) information (obtained using two different sensors) may be fused to generate a 3D texture image of the face which could then be subjected to feature extraction and matching.

2. Feature level: The feature sets extracted from multiple data sources can be fused to create a new feature set to represent the individual. The geometric features of the hand, for example, may be augmented with the eigen-coefficients of the face in order to construct a new high-dimension feature vector. A feature selection/transformation procedure may be adopted to elicit a minimal feature set from the high-dimensional feature vector.

3. Match score level: In this case, multiple classifiers output a set of match scores which are fused to generate a single scalar score. As an example, the match scores generated by the face and hand modalities of a user may be combined via the simple sum rule in order to obtain a new match score which is then used to make the final decision.

Further author information: (Send correspondence to A. Ross)
A. Ross: ross@csee.wvu.edu
R. Govindarajan: rohin@motorola.com

Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

Author & Year	Biometric Modalities	Level of fusion
Kumar and Zhang 2003[4]	Face, Palm	Match score
Kumar et al. 2003[5]	Palm, Hand	Match score, Feature
Wang et al. 2003[6]	Face, Iris	Match score
Chang et al. 2003[7]	Face, Ear	Feature
Shakhnarovich and Darrell 2002[8]	Face, Gait	Match score
Ross and Jain 2001[2]	Face, Hand, Finger	Match score
Frischholz and Dieckmann 2000[9]	Face, Voice, Lip	Match score
Ben-Yacoub 1999[10]	Face, Voice	Match score
Hong and Jain 1998[11]	Face, Finger	Match score
Bigun et al. 1997[12]	Face, Voice	Match score
Kittler et al. 1997[13]	Face, Voice	Match score
Brunelli and Falavigna 1995[14]	Face, Voice	Match score

Table 1. A few multibiometric systems discussed in recent literature[4].[15]

4. Rank level: This type of fusion is relevant in identification systems where each classifier associates a rank with every enrolled identity (a higher rank indicating a good match). Thus, fusion entails consolidating the multiple ranks associated with an identity and determining a new rank that would aid in establishing the final decision. Techniques such as the Borda count may be used to make the final decision.

5. Decision level: When each matcher outputs its own class label (i.e., accept or reject in a verification system, or the identity of a user in an identification system), a single class label can be obtained by employing techniques like majority voting, behavior knowledge space, etc.

Fusion at the match score, rank and decision levels have been extensively studied in the literature. Fusion at the feature level, however, is a relatively understudied problem (see Table 1). Fusion at this level involves the integration of feature sets corresponding to multiple information sources. Since the feature set contains richer information about the raw biometric data than the match score or the final decision, integration at this level is expected to provide better authentication results. However, fusion at this level is difficult to achieve in practice because of the following reasons: (i) the feature sets of multiple modalities may be incompatible (e.g., minutiae set of fingerprints and eigen-coefficients of face); (ii) the relationship between the feature spaces of different biometric systems may not be known; (iii) concatenating two feature vectors may result in a feature vector with very large dimensionality leading to the 'curse of dimensionality' problem; and (iv) a significantly more complex matcher might be required in order to operate on the concatenated feature set. In this paper we discuss fusion at the feature level in 3 different contexts: (i) fusion of PCA and LDA coefficients of face; (ii) fusion of LDA coefficients corresponding to the R,G,B channels of a face image; (iii) fusion of face and hand modalities. We also introduce a distance measure known as the Thresholded Absolute Distance (TAD) to be used along with the Euclidean distance metric.

2. FEATURE LEVEL FUSION

In this work, feature level fusion is accomplished by a simple concatenation of the feature sets obtained from multiple information sources. Let $\boldsymbol{X} = \{x_1, x_2, \ldots x_m\}$ and $\boldsymbol{Y} = \{y_1, y_2, \ldots y_n\}$ denote feature vectors ($\boldsymbol{X} \in R^m$ and $\boldsymbol{Y} \in R^n$) representing the information extracted via two different sources. The objective is to combine these two feature sets in order to yield a new feature vector, $\boldsymbol{Z}$, that would better represent the individual. The vector $\boldsymbol{Z}$ is generated by first augmenting vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, and then performing feature selection on the resultant feature vector. The fusion of feature level data from any two biometric sources in this paper would follow a similar procedure. The different stages in this algorithm are described below.

2.1. Feature Normalization:

The individual feature values of vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ (i.e., the x_i's and y_i's) may exhibit significant variations both in their range and distribution. The goal of feature normalization is to modify the location (mean) and scale

Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

(variance) of the features values in order to ensure that the contribution of each component to the final match score is comparable.[16] Adopting an appropriate normalization scheme also helps address the problem of outliers in feature values. The simple min-max and the median normalization[17] techniques were tested in this work. Let x and x' denote a feature value before and after normalization, respectively. The min-max technique computes x' as, $x' = \frac{x - \min(F_x)}{\max(F_x) - \min(F_x)}$, where F_x is the function which generates x. The min-max technique is effective when the minimum and the maximum values of the component feature values are known beforehand. In cases where such information is not available, an estimate of these parameters has to be obtained from the available sample training data. The estimate may be affected by the presence of outliers in the training data and this makes min-max normalization sensitive to outliers. The median normalization scheme, on the other hand, is relatively robust to the presence of noise in the training data. In this case, x' is computed as, $x' = \frac{x - median(F_x)}{median(|(x - median(F_x))|)}$. The denominator is known as the Median Absolute deviation (MAD) and is an estimate of the scale parameter of the feature value. Although, this normalization scheme is relatively insensitive to outliers, it has a low efficiency compared to the mean and standard deviation estimators, i.e., when the score distribution is not Gaussian, the median and MAD are poor estimates of the location and scale parameters, respectively. In our experiments we used the median normalization scheme due to its robustness to outliers. Normalizing the feature values via this technique results in modified feature vectors $\boldsymbol{X'} = \{x'_1, x'_2, \ldots x'_m\}$ and $\boldsymbol{Y'} = \{y'_1, y'_2, \ldots y'_n\}$.

2.2. Feature Selection

Augmenting the two feature vectors, $\boldsymbol{X'}$ and $\boldsymbol{Y'}$, results in a new feature vector, $\boldsymbol{Z'} = \{x'_1, x'_2, \ldots x'_m, y'_1, y'_2, \ldots y'_n\}$, $\boldsymbol{Z'} \in R^{m+n}$. The 'curse-of-dimensionality'[18] dictates that the augmented vector need not necessarily result in an improved matching performance. Further, some of the feature values may be 'noisy' compared to the others. The feature selection process entails choosing a minimal feature set of size k, $k < (m+n)$, that improves classification performance on a training set of feature vectors. The sequential forward floating selection[19] technique is employed to perform feature selection on the feature values of $\boldsymbol{Z'}$. This results in a new feature vector $\boldsymbol{Z} = \{z_1, z_2, \ldots z_k\}$. The criterion function to perform feature selection is defined to be the *average* of the Genuine Accept Rate (GAR) at four different False Accept Rate (FAR) values (0.05%, 0.1%, 1%, 10%) in the ROC (Receiver Operating Characteristics) curve pertaining to the training data. The reason for choosing this criterion is explained below.

2.3. Criteria for Feature Selection

The feature selection algorithm relies on an appropriately formulated objective function to elicit the optimal subset of features from the complete feature set. In the case of a biometric system it is difficult to identify a single parameter that would characterize the matching performance across a range of FAR/FRR values. Note that the single-valued parameters commonly used in the biometric literature - the Equal Error Rate (EER) and the d-prime measure - do not summarize the matching performance across all matching thresholds. An alternative would be to define the objective function at a fixed GAR/FAR value. However, this would also result in maximizing the performance at only those specific points. In order to optimize the performance gain across a wide range of thresholds, we define the objective function to be the average of GAR corresponding to 4 different FAR's (0.05%, 0.1%, 1%, 10%).

2.4. Match Score Generation

Consider feature vectors $\{\boldsymbol{X_i}, \boldsymbol{Y_i}\}$ and $\{\boldsymbol{X_j}, \boldsymbol{Y_j}\}$ obtained at two different time instances i and j. The corresponding fused feature vectors may be denoted as $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$, respectively. Let s_X and s_Y be the *normalized* match (Euclidean distance) scores generated by comparing $\boldsymbol{X_i}$ with $\boldsymbol{X_j}$ and $\boldsymbol{Y_i}$ with $\boldsymbol{Y_j}$, respectively, and let $s_{match} = (s_X + s_Y)/2$ be the fused match score obtained using the simple sum rule.

Figure 1 shows the distribution of s_{match} - $P(S_{match})$ - under the genuine and impostor hypothesis. The critical region is defined as a range of scores, $[t - \epsilon, t + \epsilon]$, where the probability distributions of the genuine and impostor scores have substantial overlap. By utilizing the fused vectors $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$, the ambiguity presented by the critical region may be resolved.

In order to accomplish this we observe that in the case of genuine pairs, a high match score is typically the effect of a *few* feature values constituting the fused vector, while a similar score for an impostor pair is typically

Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

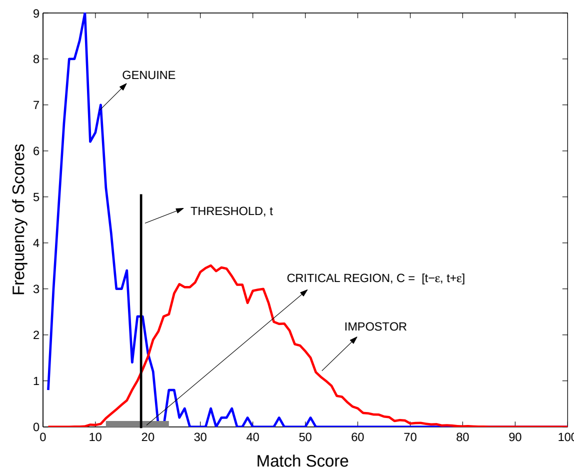


Figure 1. The genuine and impostor match score distributions indicating the presence of a critical region.

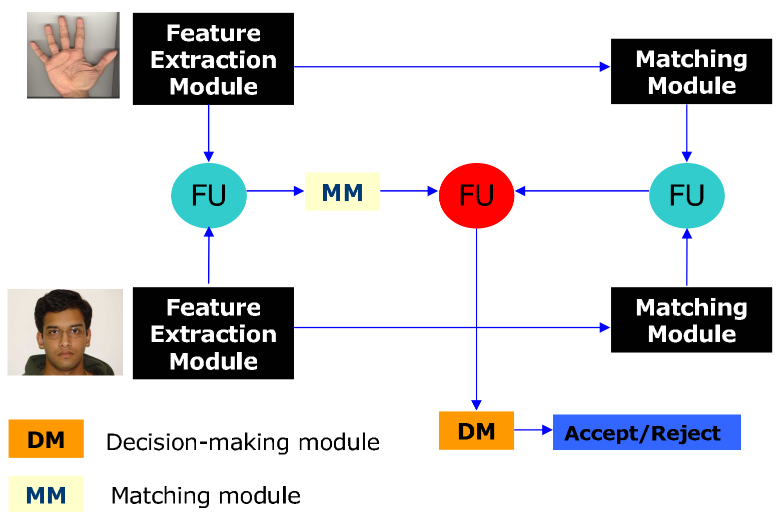


Figure 2. The flow of information when data from the feature level and match level are combined.

the cumulative effect of *all* feature values. This observation is fairly intuitive since the feature selection process eliminates features that are redundant as well as those that are correlated with other features. We, therefore, employ two different distance measures to compare $\boldsymbol{Z_i}$ and $\boldsymbol{Z_j}$:

$$s_{euc} = \sum_{r=1}^{k} (z_{i,r} - z_{j,r})^2 \qquad \text{(Euclidean distance)} \tag{1}$$

$$s_{tad} = \sum_{r=1}^{k} I(|z_{i,r} - z_{j,r}|, t) \qquad \text{(Thresholded absolute distance - TAD).} \tag{2}$$

Here, $I(y, t) = 1$, if $y > t$ (and 0 otherwise), and t is a pre-specified threshold. The TAD measure determines the *number* of normalized feature values that differ by a magnitude greater than t. The s_{euc} and s_{tad} values are consolidated into one feature level score, s_{feat}, via the simple sum rule.

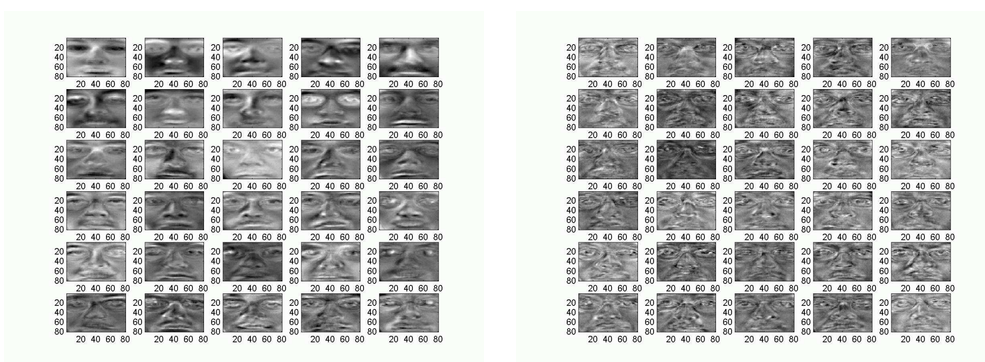
Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

We now have information both at the match score level (s_{match}) as well as the feature level (s_{feat}). Both these values are combined using the simple sum rule to obtain the final score s_{tot} (Figure 2). In the following experiments, s_{tot} is referred to as the score obtained by combining information at the match score level and the feature level.

3. EXPERIMENTAL RESULTS

A set of 500 face images and hand images were acquired from 100 users (5 biometric samples per user per biometric) at West Virginia University (WVU). The hand data was obtained using a commercial-off-the-shelf (COTS) hand geometry system installed at WVU. The feature set is a 9-byte value comprising of different geometric measurements of the hand. The face data was acquired using the Sony EVI-D30 PTZ color camera. Each face image was decomposed into its component R, G, B channels. Further, the grayscale rendition of the color image - I - was also computed. The Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were performed on these component images (i.e., R, G, B, I) in order to extract representational features. Thus, a component image was represented using PCA coefficients as well as LDA coefficients. Details of these methods may be obtained from the paper by Belhumeur et al.[20]

The proposed technique was tested on three different scenarios: (i) fusion of PCA and LDA coefficients corresponding to I (intra-modal); (ii) fusion of LDA coefficients corresponding to the R,G,B channels (intra-modal); (iii) fusion of face and hand modalities (inter-modal).



(a) (b)

Figure 3. The first 25 basis images extracted using the (a) PCA and (b) LDA techniques.

3.1. Fusion of PCA and LDA Coefficients

The basis images corresponding to PCA (eigenfaces) and LDA (fisherfaces) are shown in Figure 3. Each grayscale face image was represented using 25 eigen-coefficients and 27 fisher-coefficients. The feature fused vector had 28 features. The results of fusion are summarized in Figure 4. It is observed that the performance of the LDA-based matcher is much higher than that of the PCA-based matcher. The difference in performance is more pronounced at lower FAR values. For example, at a FAR = 0.01% the GAR using the PCA technique is $\sim$ 50% and the LDA technique is $\sim$ 80%. Therefore, this scenario represents a case where a weak classifier is combined with a strong classifier. The application of match level fusion in this situation is observed to degrade matching performance. However, combining the feature level and match score level information neither degrades nor improves the matching performance. The significance of the proposed scheme is, therefore, not borne out in this scenario, although it is seen to perform better than match level fusion. It must be mentioned that using alternate fusion schemes (other than the simple sum rule) might result in different performance curves.

Appeared in Proc. of SPIE Conference on Biometric Technology for Human Identification II, Vol. 5779, pp. 196-204, (Orlando, USA), March 2005

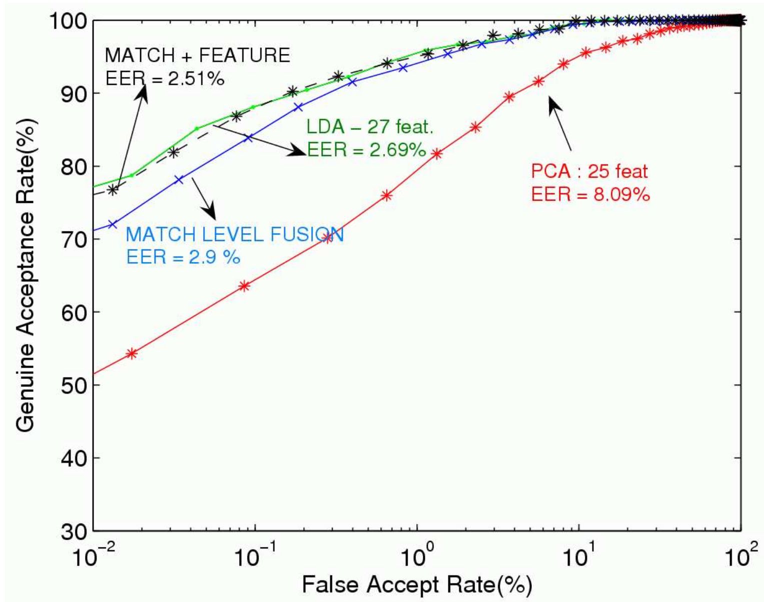


Figure 4. The results of fusion when combining the PCA and LDA coefficients of grayscale face images.

3.2. Fusion of R, G, B Channels

Experiments were also conducted in order to assess the efficacy of the proposed technique in an intra-modal fusion scenario involving the R,G and B color channels of the face images. In this case, three separate feature sets - denoted as LDA-R (18 features), LDA-G (32 features) and LDA-B (40 features) - were generated for a face image by performing LDA on each color channel independently. These three feature sets were then subjected to fusion at both the feature and match score levels. The feature fused vector consisted of 43 features. Figure 5 summarizes the results using this data. It is observed that the proposed scheme outperforms match score level fusion by a substantial margin thereby underscoring the significance of the proposed technique.

3.3. Fusion of Hand and Face Biometrics

In this case, the 9-byte hand feature set and the LDA-coefficients (27 features) of the grayscale face image were combined. Figure 6 presents the ROC curves for this scenario. The matching performance of the proposed scheme is seen to result in a marginally inferior performance compared to fusion at the match score level.

This experiment was also conducted using the Michigan State University (MSU) dataset consisting of hand and face information pertaining to 50 users with each user providing 5 samples of each biometric.[2] The hand feature set was extracted using the technique described by Jain et al.[21] In this case, the performance of the proposed fusion scheme was observed to be superior to that of match score level fusion (Figure 7). Currently, we are analyzing the reasons for such differences in performance across datasets. It is clear, however, that the performance of constituent matchers and the relationship between the feature sets generated by individual modalities have a large role to play in determining the efficacy of the proposed technique. A more formal framework may be necessary to better understand this phenomenon.

4. SUMMARY AND FUTURE WORK

A feature level fusion scheme to improve multimodal matching performance has been proposed. In our formulation, information at the feature level and match level are consolidated. The technique has been tested on inter-modal and intra-modal fusion scenarios consisting of both strong and weak biometric classifiers. In certain scenarios (viz., combining R, G, B channels) the performance gain has been substantial, thereby indicating the importance of pursuing research in this direction. However, it is difficult to predict the best fusion strategy given

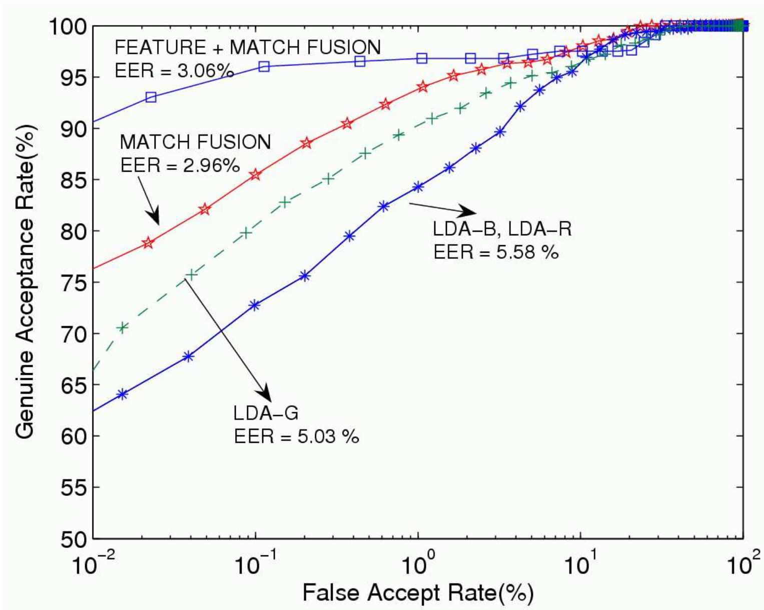


Figure 5. The results of fusion when combining the LDA coefficients corresponding to the R, G, B channels of a face image.

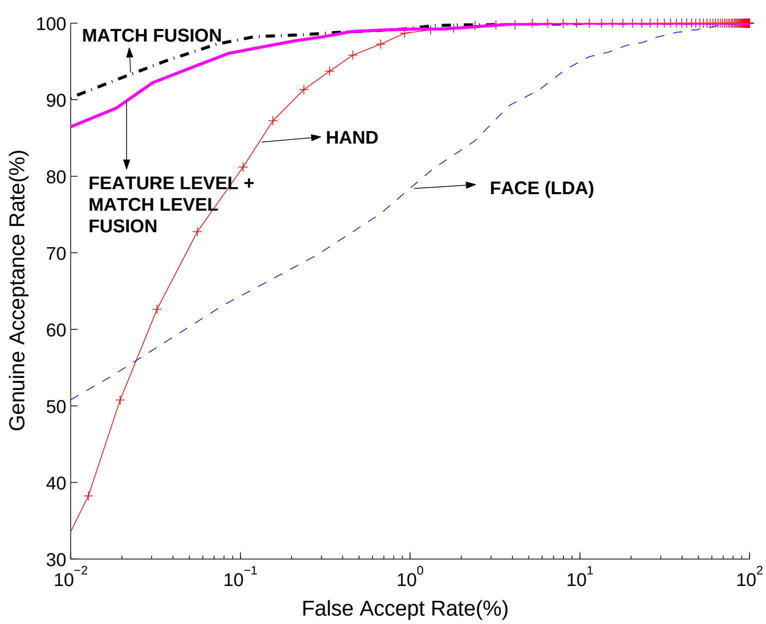


Figure 6. The results of fusion when combining the hand and face biometrics of a user (**WVU** dataset).

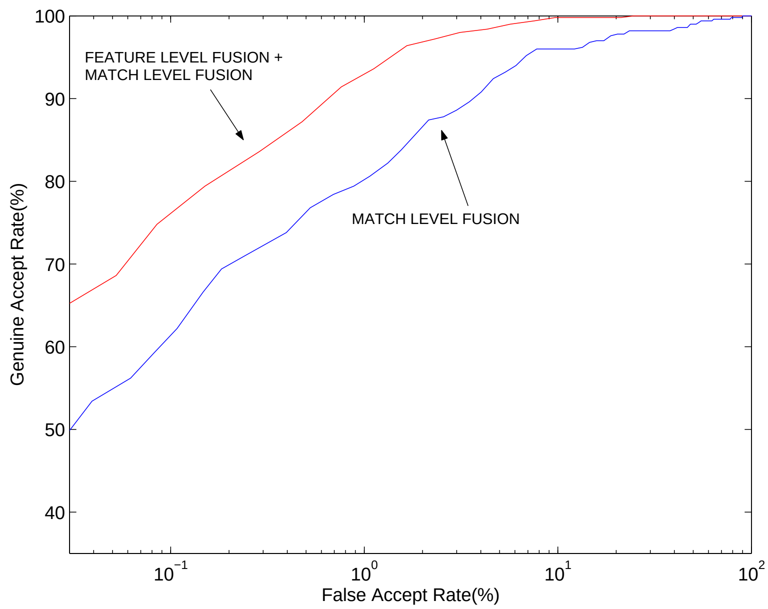


Figure 7. The results of fusion when combining the hand and face biometrics of a user (**MSU** dataset).

a scenario. Also, the algorithm as presented in this paper, does not allow incompatible feature sets (such as minutiae points of fingerprints and eigen-coefficients of face) to be combined.

The feature selection scheme ensures that redundant/correlated feature values are detected and removed before invoking the matcher. This is probably one of the key benefits of performing fusion at the feature level. Therefore, it is important that biometric vendors grant access to feature level information to permit development of effective fusion strategies. We are currently developing biometric indexing schemes using the ideas presented in this paper. Future work will include studying the effect of noisy data on the performance of the technique and the adoption of other biometric traits in this work (viz., fingerprint and iris).

REFERENCES

1. E. Rood and A. K. Jain, "Biometric research agenda: Report of the NSF workshop," in *Workshop for a Biometric Research Agenda*, (Morgantown, WV), July 2003.
2. A. Ross and A. K. Jain, "Information fusion in biometrics," *Pattern Recognition Letters* **24**, pp. 2115–2125, Sep 2003.
3. A. K. Jain and A. Ross, "Multibiometric systems," *Communications of the ACM* **47**, pp. 34–40, Jan 2004.
4. A. Kumar and D. Zhang, "Integrating palmprint with face for user authentication," in *Workshop on Multi Modal User Authentication (MMUA)*, pp. 107–112, 2003.
5. A. Kumar, D. C. M. Wong, H. Shen, and A. K. Jain, "Personal verification using palmprint and hand geometry biometric," in *Proc. of Int'l Conf. on Audio- and Video-based Person Authentication*, pp. 668–675, 2003.
6. Y. Wang, T. Tan, and A. K. Jain, "Combining face and iris biometrics for identity verification," in *Proc. of Int'l Conf on Audio- and Video-based Person Authentication*, pp. 805–813, June 2003.
7. K. Chang, K. Bowyer, V. Barnabas, and S. Sarkar, "Comparison and combination of ear and face images in appearance based biometrics," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**, pp. 1160–1165, 2003.
8. G. Shakhnarovich and T. Darrell, "On probabilistic combination of face and gait cues for identification," in *Proc. of Intl Conf. on Automatic Face and Gesture Recognition*, pp. 169–174, 2002.

9. R. Frischholz and U. Dieckmann, "Bioid: A multimodal biometric identification system," *IEEE Computer* **33**(2), pp. 64–68, 2000.
10. S. Ben-Yacoub, "Multi-modal data fusion for person authentication using svm," *Proc. of Int'l Conf on Audio- and Video-based Person Authentication* , pp. 25–30, 1999.
11. L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20**(12), p. 12951307, 1998.
12. E. Bigun, J. Bigun, B. Duc, and S. Fischer, "Expert conciliation for multi modal person authentication systems by bayesian statistics," in *Proc. of Int'l Conf on Audio- and Video-based Person Authentication*, pp. 311–318, 1997.
13. J. Kittler, G. Matas, K. Jonsson, and M. Sanchez, "Combining evidence in personal identity verification systems," *Pattern Recognition Letters* **18**(9), pp. 845–852, 1997.
14. R. Brunelli and D. Falavigna, "Person identification using multiple cues," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17**(10), pp. 955–966, 1995.
15. K. Chang, K. Bowyer, and P. Flynn, "Face recognition using 2d and 3d faces," in *Workshop on Multi Modal User Authentication (MMUA)*, pp. 25–32, 2003.
16. A. K. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognition* , 2005.
17. M. Indovina, U. Uludag, R. Snelick, A. Mink, and A. K. Jain, "Multimodal biometric authentication methods: A cots approach," in *Workshop on Multimodal User Authentication (MMUA)*, pp. 99–106, 2003.
18. G. Trunk, "A problem of dimensionality: A simple example," *IEEE Trans. Pattern Analysis and Machine Intelligence* **1**(3), pp. 306–307, 1979.
19. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters* **15**, pp. 1119–1124, November 1994.
20. P. N. Belhumeur, J. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence* **19**(7), pp. 711–720, 1997.
21. A. K. Jain, A. Ross, and S. Pankanti, "A prototype hand geometry-based verification system," in *Second International Conference on Audio and Video-based Biometric Person Authentication (AVBPA)*, pp. 166–171, (Washington, D.C., USA), March 1999.