AN EFFICIENT FUSION STRATEGY FOR MULTIMODAL BIOMETRIC SYSTEM

Nitin Agrawal, Hunny Mehrotra, Phalguni Gupta
Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, India

C. Jinshong Hwang
Department of Computer Science, Texas State University, San Marcos, Texas

Keywords: Face, Iris, Fusion, Support Vector Machines.

Abstract: This paper proposes an efficient multi-step fusion strategy for multimodal biometric system. Fusion is done at two stages i.e., algorithm level and modality level. At algorithm level the important steps involved are normalization, data elimination and assignment of static and dynamic weights. Further, the individual recognizers are combined using sum of scores technique. Finally the integrated scores from individual traits are passed to decision module. Fusion at decision level is done using Support Vector Machines (SVM). The SVM is trained by the set of matching scores and it classifies the data into two known classes i.e., genuine and imposters. The system is tested on database collected for 200 individuals and is showing a considerable increase in accuracy (overall accuracy 98.42%) compared to individual traits.

1 INTRODUCTION

Biometrics refers to the use of physiological or behavioral characteristics for recognition. These characteristics are unique to each individual and remain unaltered for a period of time. In recent years, biometrics authentication has seen considerable improvements in reliability and accuracy, with some of the traits offering good performance. However, even the best biometric traits till date are facing numerous problems; some of them are inherent to the technology itself. In particular, biometric authentication systems generally suffer from enrollment problems due to non-universal biometric traits, susceptibility to biometric spoofing or insufficient accuracy caused by noisy data acquisition in certain environments. One way to overcome these problems is the use of multi-biometrics. This approach also enables a user who does not possess a particular biometric identifier to still enroll and authenticate using other traits, thus eliminating the enrollment problems and making it universal (Gupta et al., 2006).

Among several available biometric traits, face and iris is gaining lots of attention due to ease of operation. Apart from improving the verification performance, the fusion of iris and face has several other advantages (Yunhong et al., 2003). Recognition using face is natural and easily accepted by the end users. Face recognition systems are less expensive as compared to other modalities available. On the other hand, iris is one of the most reliable and secure biometric trait. It has fast authentication even when searching in database with millions of templates. The recognition is contact free and unnoticed grabbing of images is not possible. However the accuracy of face recognition is affected by illumination, pose and facial expression (Zhao et al., 2000). The appearance of faces is directly affected by a person's facial expression and emotions. Whereas iris recognition system needs a well trained cooperative user for functionality. Further, iris images must meet stringent quality criteria, so the images of poor quality (e.g., iris with large pupil, or off center images) are rejected at the time of acquisition. Consequently, several attempts may be necessary to acquire the iris image, which not only delays the enrollment and verification, but also annoys the user. The combination of face and iris allows for simultaneous acquisition of face and iris images. Thus, in this particular case, no additional inconvenience is introduced. Finally, the use of the face recognizer in addition to the iris classifier, may allow people with imperfect iris images to enroll, reducing

Agrawal N., Mehrotra H., Gupta P. and Jinshong Hwang C. (2007).
AN EFFICIENT FUSION STRATEGY FOR MULTIMODAL BIOMETRIC SYSTEM.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications - IU/MTSV*, pages 178-183

the enrollment failure rate (Yunhong et al., 2003).

(Ross and Jain, 2003) have presented an overview of Multimodal Biometrics. The multimodal biometric system is used to overcome the limitations of a unimodal biometrics. These systems are basically used in applications where sufficient information from the data is not available. Thus in such a case the data from other biometric trait can be used to enroll the person. The fusion at classifier level is done in order to overcome the limitations of individual recognizers and increase the overall performance of each trait. There exist several fusion strategies at classifier level. A hybrid fingerprint matcher using minutiae point and reference point algorithm is proposed in (Ross et al., 2003). Another design scheme for classifier combination is discussed in (Prabhakar and Jain, 2000). The scheme stresses the importance of classifier selection in classifier combination. The combination of three recognizers for face recognition was proposed at matching score level architecture and the overall accuracy of the combined recognizer increased by 6.8% from the average accuracy of all three recognizers. Large amount of work has been done for fusion at trait level. An approach for score level fusion is proposed in (Dass et al., 2005). Experimental results are presented on face, fingerprint and hand geometry using product rule and coupla method. It was observed that fusion rules show better performance than individual recognizers. Common theoretical framework for combining classifiers using sum rule, median rule, max and min rule are analyzed in (Kittler et al., 1998) under the most restrictive assumptions and have observed that sum rule outperforms other classifiers combination schemes.

For the proposed implementation fusion is performed at two levels i.e., classifier level and modality level. At classifier level the individual recognizers for face and iris are combined at matching score level using weighted sum of score technique. Prior to combination the data is compared against some offset values and those values lying above a particular threshold (or below another threshold) are not taken into consideration for fusion as these persons can be clearly declared as authentic or imposter. Further after getting the combined matching score from the two traits, fusion is carried out at modality level using Support Vector Machines so that final decision can be made.

The paper is divided into four sections. The proposed fusion strategy is discussed in next section. In this section the individual recognizers are discussed briefly and the fusion strategy at classifier as well as trait level is explained in detail. The results are plotted to measure the accuracy and reliability of the system and it is found that the combined system using SVM gives an accuracy of 98.42%. The experimental results are given in Section 3 and conclusions are given in the last section.

2 FUSION STRATEGY

The individual recognizers for face and iris are combined at matching score level using weighted sum of scores technique and the combined scores of each trait are passed to decision module. At decision level, Support Vector Machines (SVM) is used to arrive at a final decision. The main idea to make use of these support vectors is to find an optimal separating hyperplane between data points of different classes in a high dimensional space (Burges, 1998). This approach is used for classification of matching scores into genuine and imposters. But prior to classification the matching scores generated from the individual algorithms are combined. The steps involved in generation of matching scores by the individual recognizers and detailed description of fusion strategy are given in this section.

2.1 Face Recognition

Face is one of the widely used biometric trait for recognition. For implementation purpose a digital image of face is captured and passed to the detection module. The face portion is extracted from the image using Gradient Vector Flow approach (Vatsa et al., 2003). The detected face image is further passed to feature extraction module which generates a set of feature values using a combination of Haar Wavelet (Hassanien and Ali, 2003) and Kernel Direct Discriminant Analysis (KDDA) (Juwei et al., 2003) as shown in Figure 1. The same sequences of steps are followed for extraction of feature values from the verification image. Database and query images are matched using Hamming Distance approach for Haar Wavelet and Bounding Box technique for KDDA. The output from the recognition module is distance value (F_{Haar}) in case of Haar Wavelet and matching score (F_{KDDA}) in case of KDDA.

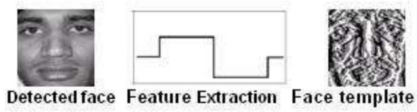

Figure 1: Generation of face template.

2.2 Iris Recognition

The iris recognition system is divided into four modules namely image acquisition, iris localization, feature extraction and matching. The images are acquired using 3CCD camera. The localization modules (Tisse,) delineates iris from the rest of the image. After localization the iris portion is transformed into a rectangular block known as strip (Daugman, 1993). The strip is further passed to the feature extraction module where the feature vectors are formed using Haar Wavelet (Hassanien and Ali, 2003) and Circular Mellin operators (Ravichandran and Trivedi, 1995). Database and query images are matched using Hamming Distance method. The recognition module returns two distance values I_{Haar} for Haar Wavelet and I_{Mellin} for Circular Mellin operators.

2.3 Fusion

The matching scores/distance values from Haar and KDDA in case of face recognition and Haar and Mellin in case of iris recognition are combined at matching score level using sum of scores technique (Ross and Jain, 2003). Further the combined values are plotted on SVM space to generate final decision. The important steps involved in fusion are normalization, data elimination, fusion at classifier or algorithm level and fusion at trait level.

2.3.1 Normalization

In this technique initially the scores are scaled to a common range (from 0 to 1). Several techniques are available for score normalization like Min-Max technique, Decimal Scaling, Median and Median Absolute Deviation (MAD), Double Sigmoid function and tanh estimators (Jain et al., 2005). Among the approaches named above min-max technique is used as it is simplest and suited for cases where the minimum and maximum bounds by a matcher are known. The scores are normalized using

$$F'_{Haar} = \frac{F_{Haar} - min(F_{Haar})}{max(F_{Haar}) - min(F_{Haar})} \tag{1}$$

where F_{Haar} is the matching score generated using Haar Wavelet for face recognition while F'_{Haar} is the score obtained after normalization. Similarly F'_{KDDA}, I'_{Haar} and I'_{Mellin} are obtained. Further the normalized score is subtracted from one if it is a dissimilarity score.

$$F''_{Haar} = 1 - F'_{Haar} \tag{2}$$

Thus I''_{Haar}, I''_{Mellin}, F''_{Haar} and F''_{KDDA} become the similarity scores. The matching scores are further rescaled so that threshold value becomes same for each algorithm.

$$NF_{Haar} = \begin{cases} c \times \frac{F'_{Haar}}{thresh} & F'_{Haar} < thresh \\ (1-c) \times \frac{F'_{Haar} - thresh}{1 - thresh} & otherwise \end{cases} \tag{3}$$

where thresh is the value of threshold for a particular recognizer and c is the value of threshold. Through experiments on IITK database, it has been found that the value of c is 0.5. Using the quantization scheme given above, values of NF_{KDDA}, NI_{Haar} and NI_{Mellin} can be obtained.

2.3.2 Data Elimination

This step is relevant from data reduction point of view. The matching scores generated from the four recognizers (two each for face and iris) are compared against the upper and lower offset values. For each algorithm there are two offset values. The lower offset value is the value slightly less than the threshold value below which FRR is zero for the particular recognizer whereas upper offset is the value slightly greater than the threshold value having no false acceptance. If the value of matching score from any recognizer is greater than the upper offset value of any recognizer then the candidate is declared as genuine at this level and further steps are not required for fusion of matching scores at trait level. Similarly for imposters if the value of matching score is less than the lower offset value then the candidate is truly declared as imposter. Here the matching scores which do not lie between upper and lower offset values are not used further for training the SVM classifier but if the matching score is used for testing then at this stage a person can be clearly declared as genuine or an imposter. Thus the scores used for training the SVM module should lie between the upper and lower offset values given by

$$ENF_{Haar} = OS_{Lower} < NF_{Haar} < OS_{Upper} \tag{4}$$

where ENF_{Haar} is the matching score obtained which is greater than lower offset value (OS_{Lower}) and less than upper offset value (OS_{Upper}). Similarly ENF_{KDDA}, ENI_{Haar} and ENI_{Mellin} are obtained. Data elimination stage is useful for two reasons. Firstly it improves the accuracy of individual recognizers as in some cases one particular recognizer gives very good accuracy while other recognizers fail. Secondly data elimination reduces the overall time complexity of the system. Thus only those candidates are considered for further processing which lie between the upper and lower offset values.

2.3.3 Fusion at Classifier or Algorithm Level

Individual recognizers are assigned weights based on performance. The value of weights can be assigned using static and dynamic approaches. In case of static approach the value of weight is assigned empirically on the basis of experimental results. But for dynamic assignment of weights combination of functions are used. Among available list of combinations, linear and exponential functions outperform the others. In this combination linear weightage is given to the recognizers if the value of matching scores is less than the threshold value but the value of weights is exponential after the matching score crosses a threshold value. After dynamic assignment of weights on matching score generated from the data elimination stage (ENF_{Haar}) the value becomes DF_{Haar}. Further static weights are assigned to all the four recognizers based on performance. As Haar performs better as compared to KDDA in case of face recognition so higher weightage of 0.8 is given to matching score generated by Haar Wavelet and lower weight value of 0.2 is assigned to KDDA. In case of iris recognition Haar and Mellin are assigned weight values of 0.7 and 0.3 respectively, as shown in (5). After assigning the value of weights the two recognizers for face as well as iris is combined using weighted sum of score technique.

$$\begin{aligned} Face &= \alpha \times DF_{Haar} + \beta \times DF_{KDDA} \\ Iris &= \gamma \times DI_{Haar} + \delta \times DI_{Mellin} \end{aligned} \tag{5}$$

where α, β, γ, δ are the value of static weight assigned and Face and Iris are the matching score generated after fusion while DF_{Haar}, DF_{KDDA}, DI_{Haar} and DI_{Mellin} are the matching scores from individual recognizers after assignment of dynamic weights.

2.3.4 Fusion at Trait Level

The fusion of individual recognizers generates combined matching scores for face and iris. The combined scores have to be integrated further to generate the final decision regarding acceptance or rejection. Multimodal Biometric System is treated as a pattern classification problem. This approach followed by (Verlinde et al., 2000) has compared various pattern classification techniques like Logistic Regression, Maximum a Posteriori, k-nearest neighbor classifiers, multilayer perceptrons, Binary decision Tress, Maximum Likelihood, Quadratic Classifiers and Linear Classifiers. The approach of Support Vector Machines (SVM) is compared with all above mentioned approaches and it is observed that SVM is showing maximum accuracy (Gutschoven and Verlinde, 2000). The matching scores generated from face and iris is passed to the SVM based classifier. The classifier finds the hyperplane that separates the genuine users from imposters and maximizing the distance of either class from hyperplane. The plane separating the data has to be obtained using some kernel function. There exist several kernel functions like linear, polynomial, radial basis function (RBF), and sigmoid kernels. From the available set of kernels polynomial function is used to segregate the data into two or more classes. The kernel non-linearly maps samples into higher dimensional space unlike linear kernels. The polynomial kernel is given by

$$K(x_i, x_j) = (\gamma Face^T Iris + r)^d, \gamma > 0 \tag{6}$$

where γ, r and d are the kernel parameters and *Face*, *Iris* are the values obtained after fusion at classifier level.

Thus, the combined matching scores of Face and Iris is plotted on the SVM space where x axis represents matching scores of face and y axis represents matching scores for iris. The kernel function separates the data into two known classes i.e., genuine and imposters. The process of feeding data to the SVM module is done to train the classifier with already existing matching scores. Next, if the test data has to be classified then the matching scores from the four individual recognizers is initially combined at classifier level and further the combined scores are passed to the polynomial function for classification. Thus the candidate's identity may be declared as authentic or forged based on the class to which it belongs to. The steps for fusion are summarized below **Step 1**: The individual recognizers generate matching scores by finding the similarity/dissimilarity between the feature sets

Step 2: The matching scores are normalized to a common range

Step 3: The normalized scores are further compared against lower and upper offset values to remove the data which can be clearly classified at this level

Step 4: After data elimination the scores are assigned dynamic and static weights

Step 5: These weighted scores are combined using sum of score technique

Step 6: the scores for face and iris are fused at decision level using Support Vector Machines

3 EXPERIMENTAL RESULTS

The results are obtained on database collected under the laboratory environment at Indian Institute of Technology Kanpur. It comprises of images of 200 persons with four images per person for face (200 $\times$ 4) and three images per person for iris (200 $\times$ 3). The face image is acquired using digital camera at a distance of about 30 cm. The database comprises of Asian faces of frontal view. The iris image is acquired from a 3CCD camera and the subject's eye is placed at a distance of about 9 cm from the camera lens. The source of light is placed at a distance of 12 cm (approx) from the user eye. The sample images of iris and face from IIT Kanpur database is shown in Figure 2.

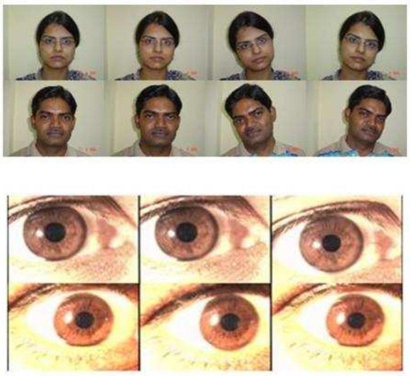

Figure 2: Sample images from IIT Kanpur database.

The results are computed at various levels of implementation to measure the robustness of the system. At first level it is observed that for face recognition, Haar Wavelet generates a dissimilarity score in range of 0.1 to 0.55 whereas KDDA generates a similarity score in range of 0 to 1. In case of iris recognition Haar Wavelet generates a dissimilarity score from 0 to 0.5 and Circular Mellin gives a dissimilarity score from 0 to 0.55. Thus the values cannot be integrated at this level, so the scores are normalized to a common range of 0 to 1 and dissimilarity scores are converted into similarity scores. The performance of individual recognizers is measured after normalization and the accuracy curves are shown in Figure 3. The individual accuracies are further enhanced after the data elimination stage. The candidate which lie below the lower offset values and above the upper offset values is declared as genuine or imposter. Thus the accuracy curve at this stage is shown in Figure 4. From the results it is evident that individual recognizers even after data elimination are not able to give good accuracy thus the results are further enhanced by combining the classifiers. The classifiers are combined at matching score level using weighted sum of score technique. The graph after fusion at classifier level is given in Figure 5.

The matching scores obtained after fusion for face and iris are passed to SVM module and results are plotted on the SVM hyperplane. The matching scores are classified with a polynomial kernel and results are shown in Figure 6 where black area represents the genuine region and green area represents the imposters. The candidates wrongly accepted (false acceptance) by the system are encircled whereas the candidates wrongly rejected (false rejection) are bounded by squares. The accuracy values obtained at various stages is shown in Table 1. At first level the accuracy values are obtained after normalization. The accuracy value of individual recognizer increases after data elimination. Further the individual recognizers are combined for face and iris. The matching scores for face and iris are plotted on SVM space and overall accuracy of the system is found to be 98.42% with FAR of 0.79% and FRR of 2.38%.

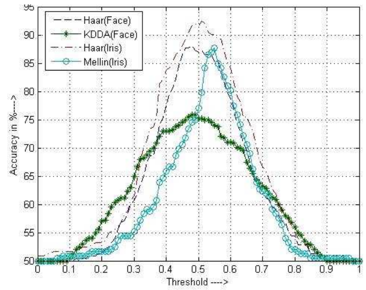

Figure 3: Accuracy graph of individual recognizers prior to fusion.

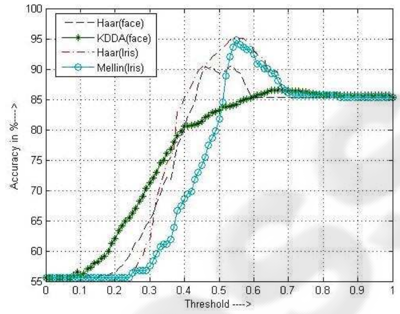

Figure 4: Accuracy graph of individual recognizers after data elimination.

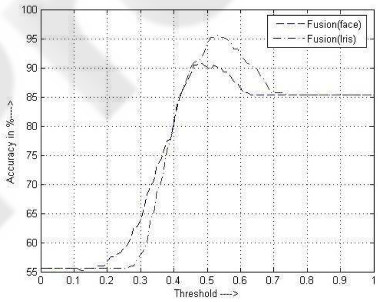

Figure 5: Accuracy graph after fusion at classifier level.

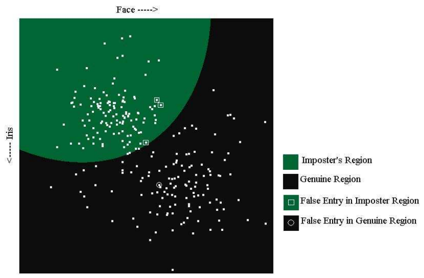

Figure 6: SVM Hyperplane.

4 CONCLUSION

In this paper the fusion of matching scores is performed at two levels at classifier level and at trait level. A single fusion strategy may not be suitable for some cases. Hence different fusion strategies are applied at different stages to get good results. The data elimination stage provides only that data which is used for deciding the classification hyperplane. Thus combination of techniques has been used and results are found to be very encouraging.

ACKNOWLEDGEMENTS

The study is supported by the Ministry of Communication and Information Technology, Government of India.

REFERENCES

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. In *Data Mining and Knowledge Discovery*.

Dass, S. C., Nandakumar, K., and Jain, A. K. (2005). A principal approach to score level fusion in multimodal biometrics system. In *Proceedings of ABVPA*.

Daugman, J. G. (1993). High confidence visual recognition of persons by a test of statistical independence. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Gupta, P., Rattani, A., Mehrotra, H., and Kaushik, A. K. (2006). Multimodal biometrics system for efficient human recognition. In *SPIE Defense and Security Symposium*.

Gutschoven, B. and Verlinde, P. (2000). Multimodal identity verification using support vector machines. In *Proceedings of the 3rd International Conference on Information Fusion*.

Hassanien, A. E. and Ali, J. M. (2003). An iris recognition system to enhance e-security environment based on wavelet theory. In *Advanced Modelling and Optimization Journal*.

Jain, A. K., Nandakumar, K., and Ross, A. (2005). Score normalization in multimodal biometric systems. In *The Journal of Pattern Recognition Society*.

Juwei, L., Plataniotis, K. N., and Venetsanopoulos, A. N. (2003). Face recognition using kernel direct discriminant analysis algorithms. In *IEEE Transactions on Neural Networks*.

Kittler, J., Hatef, M., Duin, R. P. W., and Mates, J. (1998). On combining classifiers. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Prabhakar, S. and Jain, A. (2000). Decision level fusion in biometric verification. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ravichandran, G. and Trivedi, M. (1995). Circular-mellin features for texture segmentation. In *IEEE Transaction on Image Processing*.

Ross, A. and Jain, A. K. (2003). Information fusion in biometrics. In *Pattern Recognition Journal*.

Ross, A., Jain, A. K., and Riesman, J. A. (2003). Hybrid fingerprint matcher. In *Pattern Recognition Journal*.

Tisse. Person identification technique using human iris recognition. In *Journal of System Research*.

Vatsa, M., Singh, R., and Gupta, P. (2003). Face detection using gradient vector flow. In *International Conference on Machine Learning and Cybernetics*.

Verlinde, P., Chollet, G., and Archeroy, M. (2000). Multimodal identity verification using expert fusion. In *Information Fusion*. Elsevier.

Yunhong, W., Tan, T., and Jain, A. K. (2003). Combining face and iris biometrics for identity verification. In *Proceedings of Fourth International Conference on AVBPA*.

Zhao, W., Chellappa, R., Rosefeld, A., and Phillips, P. (2000). Face recognition: A literature survey. In *Technical report, Computer Vision Lab, University of Maryland*.

Table 1: Accuracy values at various levels.

Stage	Recognizer	Accuracy	FAR	FRR
Prior to Fusion at classifier level	Face(Haar)	88.09	7.93	15.87
	Face(KDDA)	75.79	30.95	17.46
	Iris (Haar)	92.46	1.58	13.49
	Iris (Mellin)	87.69	2.38	22.22
After data elimination	Face(Haar)	90.47	6.34	12.69
	Face(KDDA)	86.51	13.49	13.49
	Iris (Haar)	95.05	0.85	7.01
	Iris (Mellin)	94.84	2.38	7.93
Fusion at Classifier level	Face (Haar + KDDA)	90.87	5.56	12.69
	Iris (Haar + Mellin)	95.62	0.79	5.97
Fusion at trait level (SVM)	Face + Iris	98.42	0.79	2.38