A Feature Information Based Approach for Enhancing Score-Level Fusion in Multi-Sample Biometric Systems

Sandeep Puthanveetil Satheesan*, Sergey Tulyakov and Venu Govindaraju
Center for Unified Biometrics and Sensors
University at Buffalo
Amherst, NY, USA 14260
{sputhanv,tulyakov,govind}@buffalo.edu

***Abstract*—Matching score fusion is a commonly used technique for improving the performance of biometric systems. In this paper we investigate the methods for fusing the scores obtained from matching individual video frames to a stored face template. Traditional fusion rules like sum and product does not account for the diversity of information contained in consecutive frames. Instead, we propose to use a quantitative measure of the shared information content between adjacent frame pairs to capture this information and enhance the score fusion performance. We conduct our experiments in a database of 132 person videos. The results show that application of information content to score level fusion can increase the performance of a fusion algorithm and hence make it more robust to errors. The developed matching score fusion method can be applied to other systems involving the multiple biometric samples or scans.**

***Keywords*—Multi-biometrics; score-level fusion; feature information content**

I. Introduction

Biometric systems are widely used in recognizing the identity of an individual. They employ the uniqueness in patterns of physical or behavioral characteristics of people like face, fingerprint, iris, voice, handwriting, and gait for identification.

Based on the kind of application, biometric systems are classified as verification and identification systems. In verification systems, the subject claiming to be an individual in the database submits a biometric sample to verify this claim. It is then matched against an enrolled template of the claimed-to-be individual and the result of match or non-match is obtained. In identification systems, an individual's submitted biometric is matched against all the enrolled templates to find if that person's biometric data is present in the database or not. These systems are mainly used in identifying criminals or in forensics.

Another classification of biometric systems is based on the number of modalities used in recognition [1]. While uni-modal systems employ only one modality, multi-modal systems use two or more modalities for improving the accuracy of recognition. These belong to a higher level of classification called multi-biometric systems [2] which explores the usage of different modalities (multi-modal), multiple samples of the same modality (multi-sample), multiple matching algorithms (multi-algorithm), multiple sensors for capturing the biometric (multi-sensor), or any combination of the above (hybrid) for information fusion. Since information from various sources is combined in a multi-biometric system it is less prone to fake biometric submissions [3] and hence more reliable.

Fusion of scores can be done at various levels which has got different implications in the overall computational complexity and storage. Feature level fusion tries to obtain different instances of the same feature and combine them to develop a better feature vector. Here the main aim is to reduce errors in feature extraction. Score level fusion is done after matching scores are generated. This can be applied to any biometric system and hence is useful in combining scores from different matchers. Score level fusion can be done either using simple predetermined rules or using machine learning algorithms like SVM [4]. Though the application of machine learning can be more accurate, its computational complexity can be a design issue in certain scenarios. Score level fusion with good predetermined rules can provide comparable performance at less computational expense. Another approach is to use decision level fusion in which decisions of different algorithms are fused using voting techniques. Obviously, feature level fusion reports higher performance due to accessibility to raw data but, where data storage and computation is a constraint, score level fusion is a better approach.

Diversity of biometric samples is an important factor that determines the performance of a system. It is generally agreed that the more diverse the set of biometric samples are, the more information that is used during fusion. This results in an increased performance of the matcher. Diversity of biometric samples is inversely related to the amount of redundant information they share. We measure the redundant information content in samples using a similarity metric.

In this work we refer to a verification multi-sample multi-biometric system in which information from multiple face video frames of the same person are combined using score level fusion. Since score level fusion alone cannot provide exact information content, we obtain it from the features. Once the amount of information content is calculated from frames, the features need not be stored and hence this does not create

*The author is currently with the Image and Spatial Data Analysis Division at National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL, USA 61801

any additional storage constraint. Calculation of information content creates additional computational complexity when compared to using simple fusion rules but nevertheless it is lesser than using machine learning methods. We present two novel score level fusion rules based on information content of biometric samples extracted from its features. Even though our discussion is based on a simple face recognition system, we believe that it is generic enough to be applied to other kinds of multi-sample biometric systems that use score level fusion.

The rest of the paper is organized as follows: In Section II, we present the details of previous work done in this domain concentrating on the use of information content of biometric samples and other related works. In Section III, we provide an overview of our approach wherein the similarity metric used is explained and the feature used in face recognition is described. In Section IV, we give a detailed description of the experiments carried out and the results obtained. In Section V, we present our conclusions based on the results obtained.

II. Previous Work

We define the information content in a video sequence by the diversity of biometric measurements obtained in different frames. For example, if a person does not move and all frames contain essentially similar biometric information, then the diversity of such a sequence will be close to zero. On the other hand, if the video contains highly variable biometric images, then the information content will be high.

It seems that application of information content has not been deeply studied in multi-sample biometrics. For example, in order to merge the matching results from different frames of the video, [5] apply simple sum rule. A more probabilistic approach is the use of conditional entropy [6]. In both cases, the individual frames are considered independently, and the similarity between consecutive frames is not measured.

Use of diversity measures have been proposed in the classifier ensemble research [7]. The fusion of classifiers in the ensembles might give different weights for the sets of similar or diverse classifiers, but typically such measures are used only to select a diverse classifier set. It is not clear if that research could be applicable to multi-sample biometric fusion.

Use of matching score vector correlation for improving the performance of biometric matchers was proposed in [8]. They used correlation measures to study diversity and similarity between matching scores obtained from multiple samples of the same person and applied it in score fusion. Authors consider the situation in which raw data is not available from biometric matchers. This approach has limitations in that it considers only the matching scores for determining information content. Our approach uses feature level details which should contain more information when compared to matching scores vector.

III. Overview Of Method

We propose that the performance of a biometric matcher improves when we take into account the amount of information content in the features used. If a sample adds redundant information to the pool, there is no gain in the overall information content. This accumulates the amount of correlated errors and hence reduces the performance. By giving more weight to the contribution of diverse frames in fusion, we intend to reduce this error.

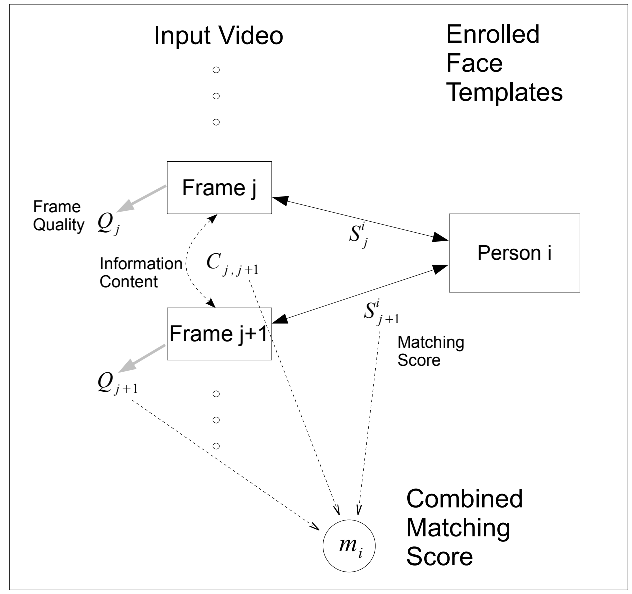

Fig. 1. Block diagram of the proposed system showing the metrics used in generating the fused matching score

A. Similarity Metric

Similarity of two biometric samples determines how much redundant information they share. In fusion, when two or more samples are fused, the information gain is an important factor that determines the matching algorithm performance. If the fused samples are having redundant information, the matcher's performance cannot be expected to be of superior quality.

We are using Pearson's correlation coefficient (1) on feature vectors to determine the amount of redundant information in two samples.

$$corr(X,Y) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{1}$$

where, X and Y are two feature vectors and $\bar{x}$ and $\bar{y}$ are their means respectively.

B. Feature Description

We extract Histogram of Oriented Gradients [9] from five different land mark points namely eyes, eyebrows, and mouth. After localizing the facial landmark points, HOG features were extracted from four overlapping windows centered near the facial landmark point each contributing to 10 orientations. During this step, Gaussian windowing and trilinear interpolation were used to increase the robustness of the descriptor. L2-norm was used to normalize the HOG feature vector. Each landmark point is now represented using a feature vector of length 40 and hence the combined feature length used for representing a face is 4x10x5 = 200.

C. Proposed Method

The proposed method of fusion is based on the following hypothesis:

- From a time dependent sequence of face images, the one with similar information content as a previously used image should contribute less in deciding the final matching score since it's adding less information to the pool.

Based on this we developed the following new rules for biometric fusion:

1) Sim-Rule-1: Equations (2) and (3) below defines this rule where, m_i is the fused matching score, S_j^i and S_{j+1}^i are respectively the matching scores obtained by comparing j^{th} and $(j+1)^{th}$ frame of the person to be verified with the enrolled template of the i^{th} person in the database, $C_{j,j+1}$ is Pearson's correlation value between j^{th} and $(j+1)^{th}$ frame from the video of the person to be verified. This measures the amount of shared information between these two frames. A value closer to 1 for $C_{j,j+1}$ indicates high amount of redundant information between these two frames and a value closer to 0 indicates that they contain more unique information. Here $n \geq 2$ denotes the number of frames being fused and the word *Sim* in the name stands for the use of similarity measure.

$$m_\sigma^i = \sum_{j=0}^{n-1} \begin{cases} \frac{S_j^i + S_{j+1}^i (1-|C_{j,j+1}|)}{2-|C_{j,j+1}|}, & \text{if } S_j^i > S_{j+1}^i \\ \frac{S_{j+1}^i + S_j^i (1-|C_{j,j+1}|)}{2-|C_{j,j+1}|}, & \text{otherwise} \end{cases} \tag{2}$$

$$m_i = \frac{m_\sigma^i}{n-1} \tag{3}$$

Let us analyze this rule. Initially let us consider the base case in which n = 2. This is a linear equation and the boundary conditions are well defined.

a) Case 1: If 0^{th} and 1^{st} frames are completely dissimilar ($C_{0,1} = 0$) i.e., the information content in both these frames is unique, then (3) becomes:

$$m_i = \frac{S_0^i + S_1^i}{2} \tag{4}$$

b) Case 2: If 0^{th} and 1^{st} frames are completely similar ($C_{0,1} = 1$), i.e., the information content in both these frames is same, then (3) becomes:

$$m_i = \begin{cases} S_0^i, & \text{if } S_0^i > S_1^i \\ S_1^i, & \text{otherwise} \end{cases} \tag{5}$$

Intuitively, when the two frames are completely similar, we need to consider only one of them for fusion since the remaining frame does not provide any additional information. When two frames are completely dissimilar the best we can do is to take the average of the scores. For values of $n > 2$, If all the frames are dissimilar, i.e. $C_{j,j+1} = 0, 0 \leq j \leq n-1$, (3) reduces to weighted sum rule. If all the frames are completely similar, i.e. $C_{j,j+1} = 1, 0 \leq j \leq n-1$, (3) reduces to sum rule with the exception that among each pair of scores considered, the greater one gets selected for fusion. In all other cases, it can be observed that scores from diverse frames receive more weight than the ones from similar frames.

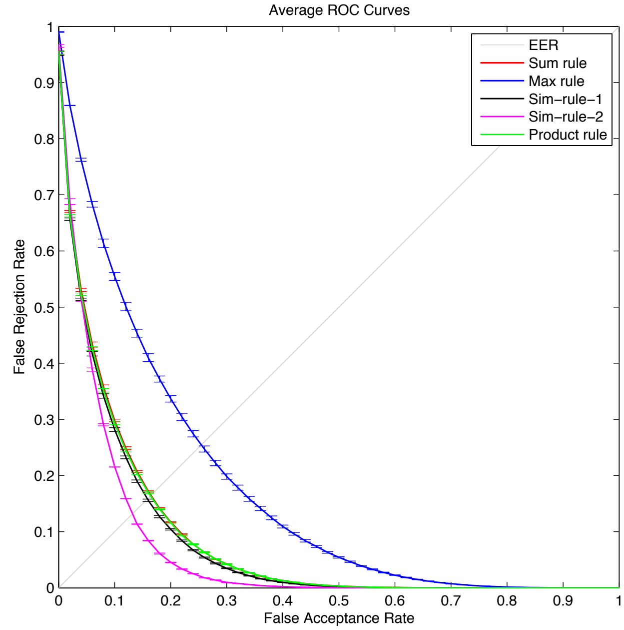

Fig. 2. Mean ROC curves obtained from 100 iterations of the experiment

2) Sim-Rule-2: In this method, the strategy is to boost the matching score based on the similarity of j^{th} and $(j+1)^{th}$ frame. Here, notations mean the same as in (2) and (3).

$$m_\sigma^i = \sum_{j=0}^{n-1} \begin{cases} S_j^i + (1-S_j^i)(1-|C_{j,j+1}|)S_{j+1}^i, \\ \text{if } S_j^i > S_{j+1}^i \\ \\ S_{j+1}^i + (1-S_{j+1}^i)(1-|C_{j,j+1}|)S_j^i, \\ \text{otherwise} \end{cases} \tag{6}$$

$$m_i = \frac{m_\sigma^i}{n-1} \tag{7}$$

Similar to the previous rule, let us consider the two cases for (7).

a) Case 1: If 0^{th} and 1^{st} frames are completely dissimilar ($C_{0,1} = 0$) i.e., the information content in both these frames are unique, then (7) reduces to (8). Here we boost the best score by a factor which is the contribution of the second best score to the fusion, where the $(1 - S_j^i)$ term sets the upper limit so that the resultant score is not greater than the maximum possible score.

$$m_i = \begin{cases} S_0^i + (1-S_0^i)S_1^i, & \text{if } S_0^i > S_1^i \\ S_1^i + (1-S_1^i)S_0^i, & \text{otherwise} \end{cases} \tag{8}$$

b) Case 2: If 0^{th} and 1^{st} frames are completely similar ($C_{0,1} = 1$), i.e., the information content in both these frames are same, then (7) reduces to (9). Here we discard the frame with the least score since it is not providing any additional information to the pool. The score difference between these two identical frames could be because of additional factors such as noise.

$$m_i = \begin{cases} S_0^i, & \text{if } S_0^i > S_1^i \\ S_1^i, & \text{otherwise} \end{cases} \tag{9}$$

Here again, when the two frames are completely similar, we consider only the best one among them for fusion. In this rule, when the two frames are completely dissimilar, we boost the best score among them by a factor which is the contribution of the remaining score. For values of $n > 2$, if all the frames are dissimilar, i.e. $C_{j,j+1} = 0, 0 \le j \le n-1$, (7) reduces to a weighted sum rule which from each pair of scores, the greater score is boosted by the contribution of the remaining score. If all the frames are similar, i.e. $C_{j,j+1} = 1, 0 \le j \le n-1$, (7) reduces to the sum rule with the exception that among each pair of scores considered, the greater one gets selected for fusion. In all other cases, it can be observed that scores from diverse frames receive more weight than the ones from similar frames. A detailed block diagram of the proposed system is in given in Fig. 1.

IV. Experiments

The experiments were conducted on a database of 132 person videos [10] of resolution 720x480 and frame rate 30 fps. From each video, at least 600 frames were extracted in such a way that each pair of frames is separated by 10 frames in the originally recorded video. This interleaving is done because in such high frame rates nothing much would have changed between two adjacent frames. The videos were taken from close proximity wherein the subjects are talking and there are considerable variations in facial expressions and head orientation. The subjects were videotaped in uniform lighting conditions which is a safe assumption for biometric systems.

In genuine matching, one frame of a subject was chosen as the enrolled template and a different randomly selected frame of the same subject was chosen as the test template. These two samples were ensured to be at least 10 frames apart to ensure enough variation in facial expressions and head orientation. In impostor matching, one frame of a subject was chosen as the enrolled template and a randomly selected frame of a different subject which was in turn selected at random was chosen as the test template. Inverse of normalized Euclidean distance between feature vectors was used to calculate the matching score. Using this method, for each subject 100 sets each of genuine matching scores and impostor matching scores were generated. Each of these sets contained 10 scores which were fused using different biometric fusion techniques.

We used sum rule, product rule, max rule, sim-rule-1, and sim-rule-2 for fusion and average ROC curves where plotted to compare their performance.

We ran the experiment 100 times, each time with a different subset of the data. The results are shown in TABLE I which clearly indicates that the newly proposed rules outperform product, sum and max rules. Among the newly proposed rules, sim-rule-2 gives the best results. In this rule, along with the similarity metric, we are using a boosting factor to include the contribution of both scores in a matching pair. This could explain the slight improvement in performance over sim-rule-1 which does not consider that. The mean ROC curves are plotted in Fig. 2.

TABLE I. Equal Error Rate For Different Fusion Rules

ERR (%)	Sim-Rule-1	Sim-Rule-2	Sum	Prod.	Max.
Mean	0.157	**0.130**	0.163	0.162	0.253
Std. Dev.	0.014	0.019	0.015	0.015	0.018
Best	0.124	**0.091**	0.128	0.127	0.211
Worst	0.193	**0.179**	0.198	0.198	0.296

V. Conclusion

In this paper we proposed a method to enhance the score level fusion rules with the measure of information content in biometric samples. The proposed approach is applicable to scenarios where data storage and computational complexity is a constraint. Experiments conducted on videos containing faces show that the two developed fusion rules gave a better performance over fusion rules which does not consider measure of information content. The method of boosting a score based on the amount of diverse information content in comparison with an adjacent frame has produced a positive impact on the fusion performance. The proposed technique could possibly be applied to other biometric modalities, if appropriate measures of information content in sample sets are defined. Current method uses predetermined fusion rules, which could produce suboptimal matching results. The future research could consider the trainable fusion methods incorporating original individual matching scores and measure of information content.

References

[1] B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan, "Studies of biometric fusion-executive summary," 2006.

[2] A. Ross, K. Nandakumar, and A. K. Jain, "Introduction to multibiometrics," in *Handbook of Biometrics*. Springer, 2008, pp. 271–292.

[3] C. Roberts, "Biometric attack vectors and defences," *Computers & Security*, vol. 26, no. 1, pp. 14–25, 2007.

[4] R. Singh, M. Vatsa, and A. Noore, "Intelligent biometric information fusion using support vector machine," in *Soft Computing in Image Processing*. Springer, 2007, pp. 325–349.

[5] M.-C. Cheung, M.-W. Mak, and S.-Y. Kung, "Multi-sample data-dependent fusion of sorted score sequences for biometric verification," in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, vol. 5. IEEE, 2004, pp. V–681.

[6] S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic recognition of human faces from video," *Computer Vision and Image Understanding*, vol. 91, no. 1, pp. 214–245, 2003.

[7] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine learning*, vol. 51, no. 2, pp. 181–207, 2003.

[8] X. Cheng, S. Tulyakov, and V. Govindaraju, "Utilization of matching score vector similarity measures in biometric systems," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. IEEE, 2012, pp. 111–116.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[10] N. Bhaskaran, I. Nwogu, M. G. Frank, and V. Govindaraju, "Lie to me: Deceit detection via online behavioral learning," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 24–29.