Fusion of Gaussian Mixture Densities for Face and Ear Biometrics Using Support Vector Machines

Dakshina Ranjan Kisku[1,*], Phalguni Gupta[2], Jamuna Kanta Sing[3], and Mita Nasipuri[4]

[1] Department of Computer Science and Engineering,
Asansol Engineering College
Asansol – 713305, India
[2] Department of Computer Science and Engineering,
Indian Institute of Technology Kanpur
Kanpur – 208016, India
[3, 4] Department of Computer Science and Engineering,
Jadavpur University
Kolkata – 700032, India
drkisku@ieee.org, pg@cse.iitk.ac.in,
{jksingh,mnasipuri}@cse.jdvu.ac.in

Abstract. This paper presents a multimodal biometric system for face and ear biometrics which convolves face and ear images with Gabor wavelet filters for extracting enhanced Gabor features from the corresponding images which are characterized by spatial frequency, spatial locality and orientation. Gaussian Mixture Model (GMM) is applied to the Gabor responses for measurements and Expectation Maximization algorithm is used to estimate density parameters in GMM. It produces two sets of feature sets which are fused using Support Vector Machines. Experiments on two different databases reveal its usefulness towards robust multimodal fusion.

Keywords: Multimodal biometrics, Face, Ear, Gabor wavelet filter, Gaussian Mixture Model, Support Vector Machines.

1 Introduction

Unimodal biometric systems may not be able to meet the desired performance requirements due to lack of viable characteristics. Advances in biometrics security have increased the possibility of using identification system based on multiple biometrics identifiers to combat efficiently with counter spoofing of unauthorized users. Multimodal biometric system integrates multiple sources of information obtained from different biometric cues. It takes advantage by collecting the relevant constraints together from individual biometric matchers by validating its pros and cons independently. It can overcome some of the limitations in single biometrics by fusing individual sources of information together. Experimental result reflects that the identity established by such an integrated biometric system is more reliable than that due to

* Corresponding author.

T.-h. Kim et al. (Eds.): FGIT 2010, LNCS 6485, pp. 344–351, 2010.

the single biometrics. There exist some multimodal biometrics with various levels of fusion [1], [2], namely, sensor level, feature level, matching score level, decision level and rank level. They have been found advantages over monomodal biometrics. In [3] a novel fusion approach of face and voice has been proposed where hyperbolic tangent is used for normalization and weighted geometric average is used for fusion. In [4] a multimodal biometrics fusion of face and voice with several fusion techniques has been discussed. A set of statistical learning and neural network based fusion strategies has been proposed in [5] for face and speech. A set of three score level fusion strategies for face, fingerprint and hand geometry has been presented in [6]. In [2] a fusion approach has been proposed at feature level showing significant improvements in experimental results. However, they are lacking in some respects such as robust feature extraction techniques and fusion strategies. Further, features are not well-characterized and fusion techniques are not working properly for the change in probabilities of data distributions.

This paper has proposed a fusion strategy of face and ear biometrics using Support Vector Machines (SVM). The technique uses Gabor wavelet filters [7] for convolution with the face and ear images. Gabor wavelet filters extract facial features and ear features as wavelet coefficients from the spatially enhanced face and ear images respectively where each feature point is characterized by spatial frequency, spatial location and orientation. These characterizations are viable or robust to the variations that occur due to facial expressions, pose changes and non-uniform illuminations. GMM [8] is applied to the Gabor face and Gabor ear responses for further characterization to create measurement vectors of discrete random variables. These two vectors of discrete variables are fused using SVM. Fusion of density parameters using SVM [8], [9] depends on the decision function in feature spaces. We validate the technique using two databases, each containing face and ear images. These databases are IITK multimodal database [15] and a database consisting of BANCA face dataset [13] and Technical University of Madrid (TUM) ear dataset [14]. Experimental result exhibits better accuracy obtained from the fusion approach.

The paper is organized as follows. Section 2 presents face and ear image localization and Gabor wavelets extraction. Next section discusses density estimation using GMM and Expectation-Maximization (EM) algorithm. Fusion of mixture densities using SVM is presented in Section 4. Experimental results are discussed in Section 5. Concluding remarks are made in Section 6.

2 Subject Localization and Gabor Wavelets

To locate the facial region for feature extraction, three landmarks positions on both the eyes and mouth are automatically localized by applying the technique proposed in [10]. A rectangular region is formed around the landmarks positions for Gabor characterization. This rectangular region is then cropped from the original face image which is constituted by facial part itself and background. For localization of ear region, triangular fossa and antitragus [11] are detected manually on ear image. Ear localization technique [12] has been used. Using these landmarks positions, ear region is cropped from ear image. After geometric normalization, image enhancements are performed on face and ear images. Histogram equalization is done for photometric normalization of face and ear images having uniform intensity distribution.

In the proposed approach, the evidences are obtained from the GMM [8] estimated scores which are computed from spatially enhanced Gabor face and Gabor ear responses. Two-dimensional Gabor filter [7] refers to a linear filter whose impulse response function is defined as the multiplication of harmonic function and Gaussian function. The Gaussian function is modulated by a sinusoid function. The Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function. Gabor function [7] is a non-orthogonal wavelet and can be specified by the frequency of the sinusoid and the standard deviations in x and y directions.

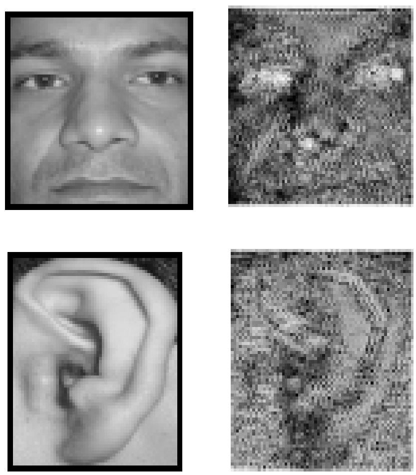

Fig. 1. Gabor Responses of Face and Ear Images

For the computation, 180 dpi gray scale images with the size of 200 × 220 pixels are used. For Gabor face and Gabor ear representations, face and ear images are convolved with the Gabor wavelets [7] for capturing substantial amount of variations among face and ear images in the spatial locations in spatially enhanced form. Gabor wavelets with five frequencies and eight orientations are used for generation of 40 spatial frequencies. Convolution generates 40 spatial frequencies in the neighbourhood regions of the current spatial pixel point. For the face and ear images of size 200 × 220 pixels, 1760000 spatial frequencies are generated. Infact, the huge dimension of Gabor responses could cause the performance degradation and slow down the matching process. In order to validate the multimodal fusion system GMM [8] further characterizes these higher dimensional feature sets of Gabor responses and density parameter estimation is performed by Expectation-Maximization (EM) algorithm [8]. For illustration, a face and an ear image from IITK multimodal database and their corresponding Gabor face and Gabor ear responses are shown in Fig. 1.

3 Density Estimation

In order to obtain better accuracy and performance, Gaussian mixture models (GMM) [8] representation has been used for the feature refinement in the proposed fusion for face and ear biometrics. The feature vectors extracted from Gabor face and Gabor ear responses can be further characterized by Gaussian distribution. Quantitive measurements for face and ear are defined by two parameters: mean and standard deviation or variability among features. Suppose, the measurement vectors are the discrete random variable x_{face} for face and the variable x_{ear} for ear. For the general case, where the feature vectors obtained from face and ear are multidimensional, the probability density function of the normal distributions is Gaussian functions [8]:

$$p(x_{face/ear}, \mu_{face/ear}, \Sigma) = \frac{1}{\sqrt{(2\pi)^{L_{face/ear}} \left|\Sigma_{face/ear}\right|}} \exp^{\frac{(x_{face/ear} - \mu_{face/ear})^T}{2\Sigma(x_{face/ear} - \mu_{face/ear})}} \quad (1)$$

where μ is the mean, Σ is the covariance matrix and L is the dimension of feature vector. Covariance matrix is the generalization to higher dimensions of the concept of the variance of a random variable. If the random variable measurements are not characterized by simple Gaussian distribution, it can be defined with multiple Gaussian components, called Gaussian Mixture Models (GMM) [8]:

$$p(x_{face}) = \sum_{m=1}^{M} \pi^{m} p(x_{face}, \mu_{face}^{(m)}, \Sigma_{face}^{(m)}) \quad (2)$$

and

$$p(x_{ear}) = \sum_{m=1}^{M} \pi^{m} p(x_{ear}, \mu_{ear}^{(m)}, \Sigma_{ear}^{(m)}) \quad (3)$$

where M is the number of Gaussian mixtures and $\pi^{(m))}$ is the weight of each of the mixture. The model of each user is the final values of $\pi^{(m)}$, $\mu^{(m)}$, $\Sigma^{(m)}$ and M, which increase the database size.

In order to estimate the density parameters of GMMs, the Expectation Maximization algorithm (EM) is adopted [8]. Each EM iteration consists of two steps – Estimation (E) and Maximization (M). The M-step maximizes a likelihood function which is refined in each iteration by the E-step.

The GMM parameters can be divided into two categories: one contains the individual mixture densities by incorporating the prior probabilities, whereas the other one contains the kernel parameter defining the form of mixture density.

4 SVM Fusion of Mixture Densities

The principle of SVM [8], [9] relies on a linear separation in a high dimensional feature space where data are mapped to consider the eventual non-linearity of the problem. To get a good level of generalization capability, the margin between the separator hyperplane and the data is maximized. A SVM classifier is trained with matching score vectors m_i, each of dimensions M. The decision surface for pattern classification is as:

$$f(\boldsymbol{m}) = \sum_{i=1}^{M} \alpha_i y_i K(\boldsymbol{m}, \boldsymbol{m}_i) + b \tag{4}$$

where α_i is the Lagrange multiplier associated with pattern $\boldsymbol{m}_i$ and $K(\cdot,\cdot)$ is a kernel function that implicitly maps the matching vectors into a suitable feature space. If $\boldsymbol{m}_k$ is linearly dependent on the other support vectors in feature space, i.e.

$$K(\boldsymbol{m}, \boldsymbol{m}_k) = \sum_{\substack{i=1 \\ i \neq k}}^{M} c_i K(\boldsymbol{m}, \boldsymbol{m}_i) \tag{5}$$

where c_i are scalar constants, then the decision surface (1) can be written as

$$f(\boldsymbol{m}) = \sum_{\substack{i=1 \\ i \neq k}}^{M} \alpha_i y_i K(\boldsymbol{m}, \boldsymbol{m}_i) + b \tag{7}$$

From Equation (7), one can get decision function is

$$D(f(\boldsymbol{m})) = sign\left\{ \sum_{\substack{i=1 \\ i \neq k}}^{M} \alpha_i y_i K(\boldsymbol{m}, \boldsymbol{m}_i) + b^* \right\} \tag{8}$$

Equation (8) is solved for α_i and b^* in its dual form with a standard QP solver which together with decision function (4), avoids manipulating directly the elements of f and starting the design of SVM for classification from the kernel function.

In [10], the fusion strategy relies on the computation of the decision function D. The combined score $FS_T \in \boldsymbol{M}$ of the multimodal pattern $\boldsymbol{m}_T \in \boldsymbol{M}^R$ can be calculated as:

$$FS_T = \sum_{\substack{i=1 \\ i \neq k}}^{M} \alpha_i y_i K(\boldsymbol{m}, \boldsymbol{m}_i) + b^* \tag{9}$$

Parameters can be adjusted to get various operating points. These operating points and the combined scores of the entire database are used to find the recognition rate of the proposed fusion approach.

5 Experimental Results

The proposed multimodal system has been tested on two databases viz. IITK multimodal database [15] and a database which contains face images from BANCA database [13] and ear images from TUM dataset [14]. In IITK database, there are 1200 images having 2 face and 2 ear images per individual. The face images are taken under control environment with change of ±20 degree in head pose. We have used frontal view faces only with uniform lighting and illumination condition and the near consistent facial expressions. These face images have acquired in different sessions.

The ear images are captured with high-resolution camera under control environment with uniform illumination and invariant pose. The face and ear biometrics are statistically and physiologically different from each other and independent for an individual. However, both of these physiological patterns are widely accepted and challenging in multimodal biometrics. One face and one ear image for each client are labeled as target and the remaining face and ear image are labeled as probe. For determining GMM estimated scores produced from Gabor responses, we use the entire database of face and ear images. Gaussian scores are generated from each of the two biometric modalities. Using GMMs, the density scores are produced from Gabor responses.

Beside the available IIT Kanpur database [15], we have used another database which has been built with the help of BANCA face database [13] consisting of 20×52 face images obtained from 52 subjects, each having 20 face images and of TUM ear database [14] consisting of 102 ear images taken from 17 subjects. The face images are presented with changes in pose, illumination and facial expression. On the other hand, each ear image is taken with a grayscale CCD camera and has a resolution of 384×288 pixels and 256 grayscales. Six ear instances of the left profile from each subject are taken under uniform, diffuse lighting conditions and slight changes in head position.

Six face and six ear images are considered for the creation of chimeric database for the experiment. Face images of BANCA database are taken randomly. For uniform experimental setup, face and ear images are normalized by histogram equalization. Uniform resolution and scaling are applied to all the face and ear images. A total of 6×17 images is collected separately for each face and ear modality.

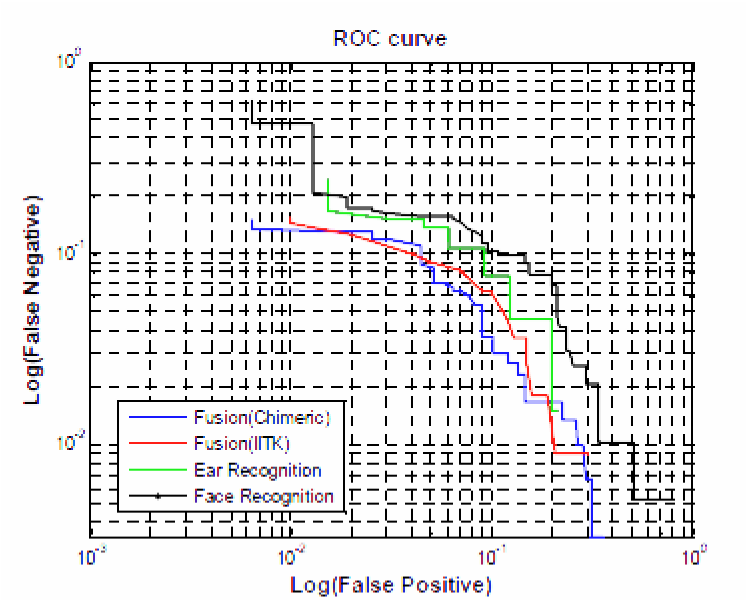

Fig. 2. Receiver Operating Characteristics Curves

Experimental results reveal that the fusion approach is better than the individual matching of face and ear biometrics. It achieves 96.49% recognition rate for IITK dataset. When other dataset is used, recognition rate is found to be 97.99%. The performance of individual face and ear matchers determined on IITK dataset only since it has viable effect to computation than that of other database. Face matcher achieves 91.96% as recognition rate. On the other hand, ear matcher achieves 93.35% recognition rate. The robust performances are exhibited when SVM based fusion uses Gabor wavelets and GMMs. ROC curves for the proposed fusion approach as well as for the individual face and ear biometrics are shown in Fig. 2. Table 1 shows Equal Error Rates (EER) and Recognition Rates (RR) for different methods which are determined on two multimodal databases.

Table 1. Equal Error Rates (EER) and Recognition Rates (RR) Determined on Two Multimodal Databases are shown

Method	Database	RR (%)	EER (%)
Multimodal Approach – I	IITK (Face + Ear Datasets)	96.49	3.51
Multimodal Approach - II	Chimeric (BANCA Face Dataset + TUM Ear Dataset)	97.99	2.01
Face Recognition	IITK Face Database	91.96	8.04
Ear Recognition	IITK Ear Database	93.35	6.65

6 Conclusion

This paper has proposed a multimodal biometrics system for face and ear biometrics. The system has been tested in two multimodal databases. Gabor filters are used to extract enhanced face and ear features which are viable and robust to different variations. E-estimator and M-estimator in GMM are used to estimate the density parameters representing the high dimensional Gabor face and Gabor ear responses. Feature sets obtained from the individual estimators are fused by SVM. Experimental results reveal its efficiency with respect to its performance for large database.

References

1. Jain, A.K., Ross, A.K.: Multibiometric Systems. Communications of the ACM 47(1), 34–40 (2004)
2. Rattani, A., Kisku, D.R., Bicego, M., Tistarelli, M.: Robust Feature-Level Multibiometric Classification. In: Proceedings of the Biometric Consortium Conference- A special issue in Biometrics, pp. 1–6 (2006)
3. Brunelli, R., Falavigna, D.: Person Identification using Multiple Cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(10), 955–966 (1995)
4. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(3), 226–239 (1998)
5. Ben-Yacoub, S., Abdeljaoued, Y., Mayoraz, E.: Fusion of Face and Speech Data for Person Identity Verification. IEEE Transactions on Neural Networks 10(5), 1065–1075 (1999)

6. Ross, A., Jain, A.K.: Information Fusion in Biometrics. Pattern Recognition Letters 24(13), 2115–2125 (2003)
7. Lee, T.S.: Image Representation using 2D Gabor Wavelets. IEEE Transaction on Pattern Analysis and Machine Intelligence 18, 959–971 (1996)
8. Bredin, H., Dehak, N., Chollet, G.: GMM-based SVM for Face Recognition. In: IEEE International Conference on Pattern Recognition, pp. 1111–1114 (2006)
9. Gutschoven, B., Verlinde, P.: Multi-modal Identity Verification using Support Vector Machines (SVM). In: Proceedings of the 3rd International Conference on Information Fusion (2000)
10. Smeraldi, F., Capdevielle, N., Bigün, J.: Facial Features Detection by Saccadic Exploration of the Gabor Decomposition and Support Vector Machines. In: 11th Scandinavian Conference on Image Analysis, pp. 39–44 (1999)
11. Iannarelli, A.: Ear Identification. In: Forensic Identification series, Paramont Publishing Company, Fremont (1989)
12. Chang, K., Bowyer, K.W., Sarkar, S.: Comparison and Combination of Ear and Face Images in Appearance- based Biometrics. IEEE Transaction on Pattern Analysis and Machine Intelligence 25(9), 1160–1165 (2003)
13. Kisku, D.R., Rattani, A., Grosso, E., Tistarelli, M.: Face Identification by SIFT-based Complete Graph Topology. In: 5th IEEE International Workshop on Automatic Identification Advanced Technologies, pp. 63–68 (2007)
14. Carreira-Perpiñán, M.A.: Compression Neural Networks for Feature Extraction: Application to Human Recognition from Ear Images. In: M.Sc. Thesis, Faculty of Informatics, Technical University of Madrid, Spain (1995)
15. Kisku, D.R., Mehrotra, H., Gupta, P., Sing, J.K.: Probabilistic Graph-based Feature Fusion and Score Fusion using SIFT Features for Face and Ear Biometrics. In: International Symposium on Optics and Photonics, vol. 7443, p. 744306 (2009)