◆◆ Scientific
◆◆ Research

Multimodal Belief Fusion for Face and Ear Biometrics

Dakshina Ranjan KISKU[1], Phalguni GUPTA[2], Hunny MEHROTRA[3], Jamuna Kanta SING[4]

[1]*Department of Computer Science and Engineering, Dr. B. C. Roy Engineering College, Durgapur, India*
[2]*Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, India*
[3]*Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela, India*
[4]*Department of Computer Science and Engineering, Jadavpur University, Kolkata, India*
Email: *drkisku@ieee.org, pg@cse.iitk.ac.in, hunny.mehrotra@nitrkl.ac.in, jksing@ieee.org*

Abstract: This paper proposes a multimodal biometric system through Gaussian Mixture Model (GMM) for face and ear biometrics with belief fusion of the estimated scores characterized by Gabor responses and the proposed fusion is accomplished by Dempster-Shafer (DS) decision theory. Face and ear images are convolved with Gabor wavelet filters to extracts spatially enhanced Gabor facial features and Gabor ear features. Further, GMM is applied to the high-dimensional Gabor face and Gabor ear responses separately for quantitive measurements. Expectation Maximization (EM) algorithm is used to estimate density parameters in GMM. This produces two sets of feature vectors which are then fused using Dempster-Shafer theory. Experiments are conducted on two multimodal databases, namely, IIT Kanpur database and virtual database. Former contains face and ear images of 400 individuals while later consist of both images of 17 subjects taken from BANCA face database and TUM ear database. It is found that use of Gabor wavelet filters along with GMM and DS theory can provide robust and efficient multimodal fusion strategy.

Keywords: multimodal biometrics, gabor wavelet filter, gaussian mixture model, belief theory, face, ear

1. Introduction

Recent advancements of biometrics security artifacts for identity verification and access control have increased the possibility of using identification system based on multiple biometrics identifiers [1–3]. A multimodal biometric system [1,3] integrates multiple source of information obtained from different biometric cues. It takes advantage of the positive constraints and capabilities from individual biometric matchers by validating its pros and cons independently. There exist multimodal biometrics system with various levels of fusion, namely, sensor level, feature level, matching score level, decision level and rank level. Advantages of multimodal systems over the monomodal systems have been discussed in [1,3].

In this paper, a fusion approach of face [4] and ear [5] biometrics using Dempster-Shafer decision theory [7] is proposed. It is known that face biometric [4] is most widely used and is one of the challenging biometric traits, whereas ear biometric [5] is an emerging authentication technique and shows significant improvements in recognition accuracy. Fusion of face and ear biometrics has not been studied in details except the work presented in [6]. Due to incompatible characteristics and physiological patterns of face and ear images, it is difficult to fuse these biometrics based on some direct orientations. In-

stead, some form of transformations is required for fusion. Unlike face, ear does not change in shape over the time due to change in expressions or age.

The proposed technique uses Gabor wavelet filters [8] for extracting facial features and ear features from the spatially enhanced face and ear images respectively. Each extracted 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. Prior to feature extraction, some preprocessing operations are done on the raw captured face and ear images. In the next step, Gaussian Mixture Model [9] is applied to the Gabor face and Gabor ear responses for further characterization to create measurement vectors of discrete random variables. In the proposed method, these two vectors of discrete variables are fused together using Dempster-Shafer statistical decision theory [7] and finally, a decision of acceptance or rejection is made. Dempster-Shafer decision theory based fusion works on changed accumulative evidences which are obtained from face and ear biometrics. The proposed technique is validated and examined using Indian Institute of Technology Kanpur (IITK) multimodal database of face and ear images and using a virtual or chimeric database of face and ear images collected from

BANCA face database and TUM ear database. Experimental results exhibit that the proposed fusion approach yields better accuracy compared to existing methods.

This paper is organized as follows. Section 2 discusses the preprocessing steps involved to detect face and ear images and to perform some image enhancement algorithms for better recognition. The method of extraction of wavelet coefficients from the detected face and ear images has been discussed in Section 3. A method to estimate the score density from the Gabor responses which are obtained from the face and the ear images through Gabor wavelets has been discussed in the next section. This estimate has been obtained with the help of Gaussian Mixture Model (GMM) and Expectation Maximization (EM) algorithm [9]. Section 5 proposes a method of combining the face matching score and the ear matching score which makes use of Dempster-Shafer decision theory. The proposed method has been tested on 400 subjects of IITK database and on 17 subjects of a virtual database. Experimental results have been analyzed in Section 6. Conclusions are given in the last section.

2. Face and Ear Image Localization

This section discusses the methods used to detect the facial and ear regions needed for the study and to enhance the detected images. To locate the facial region for feature extraction and recognition, three landmark positions (as shown in Figure 1) on both the eyes and mouth are selected and marked automatically by applying the technique proposed in [10]. Later, a rectangular region is formed around the landmark positions for further Gabor characterization. This rectangular region is then cropped from the original face image.

For localization of ear region, Triangular Fossa [11] and Antitragus [11] are detected manually on ear image, as shown in Figure 1. Ear localization technique proposed in [6] has been used in this paper. Using these landmark positions, ear region is cropped from ear image. After geometric normalization, image enhancement operations are performed on face and ear images. Histogram equalization is done for photometric normalization of face and ear images having uniform intensity distribution.
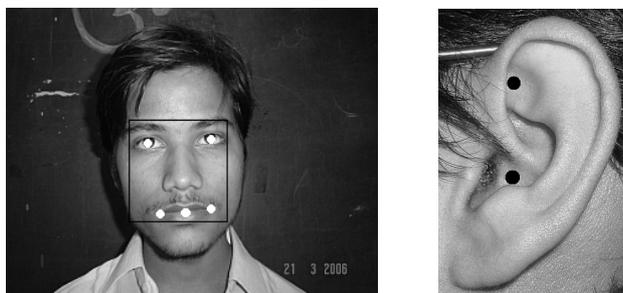

Figure 1. Landmark positions of face and ear images

3. Gabor Wavelet Coefficients

In the proposed approach the evidences are obtained from the Gaussian Mixture Model (GMM) estimated scores which are computed from spatially enhanced Gabor face and Gabor ear responses. Two-dimensional Gabor filter [8] refers a linear filter whose impulse response function is the multiplication of harmonic function and Gaussian function. The Gaussian function is modulated by a sinusoid function. The convolution theorem states that 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 [8] is a non-orthogonal wavelet and it can be specified by the frequency of the sinusoid and the standard deviations in both x and y directions.

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 [8] 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, Gaussian Mixture Model (GMM) further characterizes these higher dimensional feature sets of Gabor responses and density parameter estimation is performed by Expected Maximization (EM) algorithm. For illustration, some face and ear images from IITK multimodal database and their corresponding Gabor face and Gabor ear responses are shown in Figure 2(a) and 2(b) respectively.

4. Score Density Estimation

Gaussian Mixture Model (GMM) [9] is used to produce convex combination of probability distribution and in the subsequent stage, Expectation Maximization (EM) algorithm [9] is used to estimate the density scores. In this section, GMM is described for parameter estimation and score generation.

GMM is a statistical pattern recognition technique. The feature vectors extracted from Gabor face and Gabor ear responses can be further characterized and described by Gaussian distribution. Each 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 modality and variable x_{ear} for ear
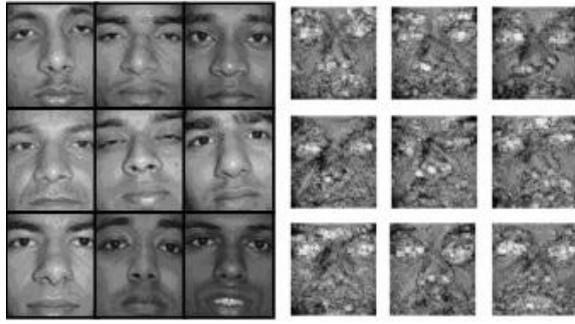
(a) Face images and their gabor responses
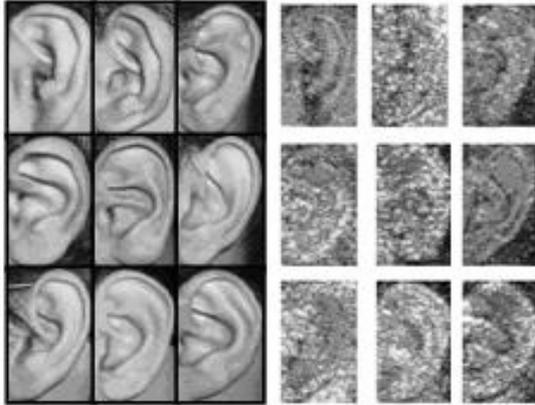
(b) Ear images and their gabor responses

Figure 2. Face and ear images and their gabor responses

modality. GMM is of the form of a convex combination of Gaussian distributions [9]):

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

and

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

where M is the number of Gaussian mixtures and $\pi^{(m)}$ is the weight of each of the mixture. In order to estimate the density parameters of GMM, EM has been used. Each of the EM iterations consists of two steps – Estimation (E) and Maximization (M). The M-step maximizes a likelihood function that is refined in each iteration by the E-step [9].

5. Fusing Scores by Dempster-Shafer Theory

The fusion approach uses Dempster-Shafer (DS) decision theory [7] to combine the score density estimation obtained by applying GMM to Gabor face and ear responses for improving the overall verification results. DS

decision theory is considered as a generalization of Bayesian theory in subjective probability and it is based on the theory of belief functions and plausible reasoning. DS decision theory can be used to combine evidences obtained from different sources of system to compute the probability of an event. Generally, DS decision theory is based on two different ideas such as the idea of obtaining degrees of belief for one question from subjective probabilities for a related query and Dempster's rule for fusing such degrees of belief while they depend on independent items of information or evidence [7].

DS theory combines three function ingredients: the basic probability assignment function (*bpa*), the belief function (*bf*) and the plausibility function (*pf*). Let t^{Face} and t^{Ear} be two transformed feature sets obtained from the clustering process for the Gabor face and Gabor ear responses, respectively. Further, $m(t^{Face})$ and $m(t^{Ear})$ are the *bpa* functions for the Belief measures $Bel(t^{Face})$ and $Bel(t^{Ear})$ for the individual traits respectively. Then the belief probability assignments *(bpa)* $m(t^{Face})$ and $m(t^{Ear})$ can be combined together to obtain a Belief committed to a feature set $C \, \epsilon \, \Theta$ according to the following combination rule [13] or orthogonal sum rule

$$m(C) = \frac{\displaystyle\sum_{\Gamma^{Face} \cap \Gamma^{Ear} = C} m(\Gamma^{Face}) m(\Gamma^{Ear})}{1 - \displaystyle\sum_{\Gamma^{Face} \cap \Gamma^{Ear} = \varnothing} m(\Gamma^{Face}) m(\Gamma^{Ear})}, \quad C \neq \varnothing. \quad (3)$$

The denominator in Equation (3) is normalizing factor, which denotes the amounts of conflicts between the belief probability assignments $m(r^{Face})$ and $m(r^{Ear})$. Due to two different modalities used for feature extraction, there is an enough possibility to conflict the belief probability assignments and this conflicting state is being captured by the two bpa functions. The final decision of user acceptance and rejection can be established by applying threshold to *m(C)*.

6. Experimental Results

The proposed multimodal biometrics system is tested on the two multimodal databases, namely, IIT Kanpur multimodal database of face and ear images and virtual multimodal database consisting of face images taken from BANCA face database and ear images taken from TUM ear database.

6.1. Test Performed on IIT Kanpur Database

The results are obtained on multimodal database collected at IIT Kanpur. Database of face and ear consists of 400 individuals' with 2 face and 2 ear images per person. The face images are taken in controlled environment with maximum tilt of head by 20 degree from the origin. However, for evaluation purpose frontal view faces are used with uniform lighting, and minor change in facial

expression. These face images are acquired in two different sessions. The ear images are captured with high-resolution camera in controlled environment with uniform illumination and invariant pose. The face and ear biometrics are statistically different from each other for an individual. One face and one ear image for each client are labeled as target and the remaining face and ear images are labeled as probe.

Table 1 illustrates that the proposed fusion approach of face and ear biometrics using DS decision theory increases recognition rates over the individual matching. The results obtained from IITK multimodal database indicates that the proposed fusion approach with feature space representation using Gabor wavelet filter and GMM outperforms the individual face and ear biometrics recognition while DS decision theory is applied as fusion rule. This fusion approach achieves 95.53% recognition rate with 4.47% EER. It has been also seen that FAR is significantly reduced to 3.4% while it is compared with the individual matching performances for face and ear biometrics. Receiver Operating Characteristic (ROC) curve is plotted in Figure 3 for minute illustrations about the computed errors and recognition rates. The proposed fusion approach is also compared with the technique discussed in [6] and it is found to be a robust fusion technique for user recognition and authentication while the combination of Gabor wavelet filter, GMM and DS decision theory is used.
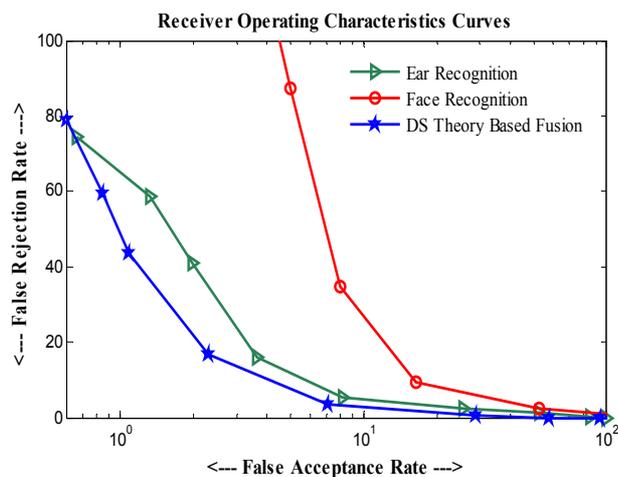

Figure 3. ROC curves for IIT Kanpur database

Table 1. Error rates of different methods

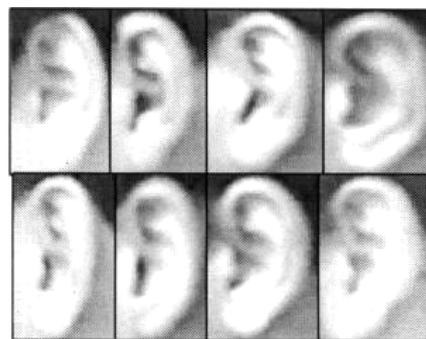| Methods	FRR (%)	FAR (%)	EER (%)	Recognition Rate (RR) (%)
Face recognition	8.26	7.82	8.04	91.96
Ear recognition	7.60	5.70	6.65	93.35
DS based fusion	5.55	3.40	4.47	95.53

Figure 4. Sample ear images from TUM database

6.2. Test Performed on Virtual Database

To verify the usefulness of the proposed technique evaluated with more than one multimodal database, a virtual multimodal database of face and ear images taken from BANCA database [4] and TUM database [12], respectively, is created. In reality, there does not exist any multimodal database of face and ear images of same subjects, except the IIT Kanpur database, for experiment. BANCA face database [4] consists of 20×52 face images obtained from 52 subjects, each having 20 face images. The face images are presented with changes in pose, in illumination and in facial expression. The BANCA face database considered as one of the complex databases for experiment. On the other hand, ear images of Technical University of Madrid ear database [12] are taken with a grayscale CCD camera. Each ear image 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. There are 102 ear images in the database of TUM which are taken from 17 different individuals.

Six face and six ear images are considered for the creation of virtual 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. Apply uniform resolution and scaling to all the face and ear images. A total of 6×17 images is collected separately for each face and ear modality. Some sample ear and face images of TUM ear database and BANCA database are shown in Figure 4 and Figure 5 respectively.

Three images per person are used for enrollment in the face and ear verification systems. For each individual, 3 pairs of face-ear are used for training the fusion classifier. For evaluation purpose, remaining 3 pairs of face-ear are used for verification and match score generation. The performance of individual matchers as well as the DS theory based fusion technique is presented through Equal Error Rate (EER), False Reject Rate (FRR), False Accept Rate (FAR) and recognition rate.

Table 2 illustrates the performance of the DS based fusion classifier as well as the individual modality of face

Figure 5. Sample face images from BANCA database

Table 2. Error rates determined from virtual multimodal database are shown

Methods	FRR (%)	FAR (%)	EER (%)	Recognition Rate (RR) (%)
Face recognition	6.18	5.44	5.81	94.19
Ear recognition	5.88	4.04	4.96	95.04
DS based fusion	4.7	2.98	3.84	96.16

and ear biometrics in terms of EER, FRR, FAR, recognition rates. The results obtained from the chimeric or virtual database indicate that individual face and ear biometrics perform well while Gabor wavelets and GMM are used for feature characterization and score density estimation respectively. Further, when DS decision theory is used to fuse the individual scores obtained from face and ear biometrics, it performs better than the individual matchers. However, due to the less number of subjects in virtual database in comparison to that in the IIT Kanpur multimodal database, results obtained from virtual database show better performance over the IIT Kanpur database.

DS decision theory based fusion approach achieves the recognition rate of 96.16% from the virtual database with 3.84% EER. It has also been seen that the FAR is significantly reduced to 2.98% while it is compared with that of the FAR obtained from IIT Kanpur database and that of the FAR determined from individual matchers. Receiver Operating Characteristics (ROC) curves are shown in Figure 6 for the individual face and ear biomet-

rics along with the DS based fusion scheme.

Multimodal fusion of face and ear biometrics are rarely available except the work presented in [6], which has used appearance based techniques to fuse these two biometric traits. Test performed on IIT Kanpur database consists of 400 subjects with the proposed fusion approach exhibits robust performance while it is compared with the performance based on virtual database having very less number of subjects. When IIT Kanpur database is used for evaluation, a pair of face-ear is used for training the fusion classifier and another pair of face-ear is used for verification. Therefore, two images per person are used for face and ear modalities in IIT Kanpur database against 6 images per person are taken in virtual database. The recognition rate obtained from virtual multimodal database shows the robustness and efficacy for the proposed fusion method while small numbers of subjects are used with more instances for a single subject. Test performed on virtual database also exhibits the invariant performance with various facial expressions, pose changes, illumination changes in BANCA face database. This could not be analyzed for IIT Kanpur database because face images in that database are almost uniform and not much variation in facial expressions and pose changes.

In [6], the authors have proposed a multimodal fusion of face and ear biometrics using principal component analysis and this principal component analysis is used for extracting eigen-face and eigen-ear. Further, these eigen-face and eigen-ear are fused for recognition. This fusion classifier has been achieved 90.09% recognition rate with 197 subjects of face and ear images. In contrast, the proposed fusion classifier has achieved 95.53% and 96.16% recognition rates for IIT Kanpur and virtual databases respectively. The use of Gabor wavelets for feature characterization and GMM for score density estimation with Dempster-Shafer decision theory based fusion technique has provided higher recognition rates with the substantial variations in the databases.
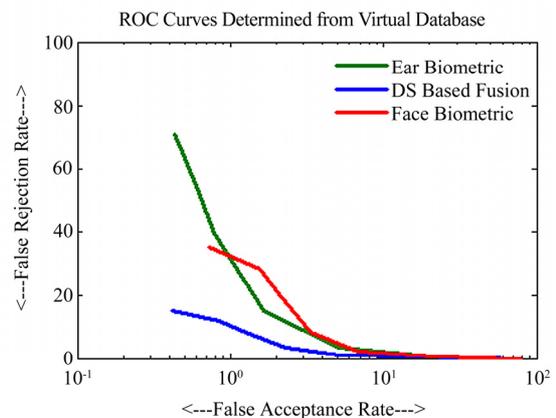

Figure 6. ROC curves for different matchers

7. Conclusions

The proposed fusion strategy combines information that has been extracted through Gabor wavelet filters and Gaussian Mixture Model estimator. Gabor wavelet filters have been used for extraction of spatially enhanced face and ear features which are viable and robust to different variations. Using E-estimator and M-estimator in GMM, reduced feature sets have been extracted from high dimensional Gabor face and Gabor ear responses through parameter estimation. These reduced feature sets are fused together by Dempster-Shafer decision theory. It has been found that the technique exhibits increase in accuracy and significant improvement over the existing methodologies.

REFERENCES

[1] A. K. Jain and A. K. Ross, "Multibiometric systems," Communications of the ACM, Vol. 47, No.1, pp. 34–40, 2004.

[2] A. K. Jain, A. K. Ross, and S. Prabhakar, "An introduction to biometrics recognition," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, No. 1, pp. 4–20, 2004.

[3] A. Rattani, D. R. Kisku, M. Bicego, and M. Tistarelli, "Robust feature-level multibiometric classification," Proceedings of the Biometric Consortium Conference – A special issue in Biometrics, pp. 1–6, 2006.

[4] D. R. Kisku, A. Rattani, E. Grosso, and M. Tistarelli, "Face identification by SIFT-based complete graph topology," Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies, pp. 63–68, 2007.

[5] B. Arbab-Zavar, M. S. Nixon, and D. J. Hurley, "On model-based analysis of ear biometrics," First IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–5, 2007.

[6] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor "Comparison and combination of ear and face images in appearance-based biometrics," Transaction on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp. 1160–1165, 2003.

[7] N. Wilson, Algorithms for Dempster-Shafer theory, Oxford Brookes University.

[8] T. S. Lee, "Image representation using 2D Gabor wavelets," IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 18, pp. 959–971, 1996.

[9] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for Gaussian Mixtures," Neural Computation, Vol. 8, No. 1, pp. 129–151, 1996.

[10] F. Smeraldi, N. Capdevielle, and J. Bigün, "Facial features detection by saccadic exploration of the gabor decomposition and support vector machines," In the 11th Scandinavian Conference on Image Analysis, pp. 39–44, 1999.

[11] A. Iannarelli, Ear Identification, Forensic Identification series, Fremont, Paramont Publishing Company, California, 1989.

[12] M. A. Carreira-Perpiñán, "Compression neural networks for feature extraction: Application to human recognition from ear images," MSc thesis, Faculty of Informatics, Technical University of Madrid, Spain, 1995.

[13] D. R. Kisku, M. Tistarelli, J. K. Sing, and P. Gupta, "Face recognition by fusion of local and global matching scores using DS theory: An evaluation with uni-classifier and multi-classifier paradigm," In the Proceedings of IEEE Computer Soceity Conference on Computer Vision and Pattern Recognition Workshop, pp. 60–65, 2009.