Pyramid Histogram of Oriented Gradients based Human Ear Identification

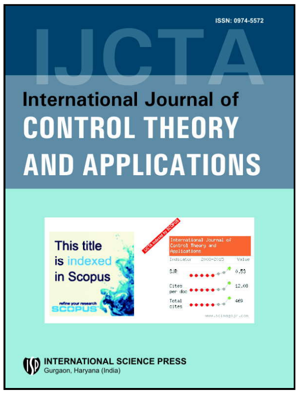
Pyramid Histogram of Oriented Gradients based Human Ear Identification

Article *in* International Journal of Control Theory and Applications · April 2017

3 authors, including:

Partha Pratim Sarangi
Seemanta Engineering College, Jharpokharia

B. S. P. Mishra
KIIT University



International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 15 • 2017

Pyramid Histogram of Oriented Gradients based Human Ear Identification

Partha Pratim Sarangi[1], B.S.P. Mishra[1] and Sachidanada Dehuri[2]

[1] *School of Computer Engineering KIIT University, Bhubaneswar , Emails: ppsarangi@gmail.com, mishra.bsp@gmail.com*
[2] *Department of ICT FM University, Balasore, Email: satchi.lapa@gmail.com*

Abstract: Nowadays, ear biometrics is regarded as an effective method for human identification using 2D side face images which has inspired the attention of researchers. In this paper, we explore a new local feature extraction technique pyramid histogram of oriented gradients (PHOG) to make simple, fast and high performance ear identification system. We also explore here two distance measures: the histogram intersection and the Chi-squared statistics for calculating similarity between images. However, the global feature descriptor of the ear image is significantly large. To reduce the dimension of the PHOG descriptor, linear discriminant analysis (LDA) has been used to remove noise and avoid overfitting, while the identification performance only slightly decreases. The comparative study is carried out in a ear identification system using large number of images from public ear databases, namely IIT Delhi ear database and University of Notre Dame ear database (Collections E). The experimental results demonstrate the effectiveness of our method which increases its performance over the existing techniques.

Keywords: Biometrics, Ear identification, Pyramid histograms of oriented gradients, Linear discriminant analysis.

1. INTRODUCTION

In modern society, the importance of the automated identification of individual is growing for highly reliable law-enforcement and commercial security applications. Biometrics deals with personal identification based on their unique physiological and behavioural characteristics. The face, fingerprint and iris have appeared as the three most successful physiological biometric modalities. Similarly, voice, signature, and gait have emerged as the three most popular behavioural biometric traits. In some applications like tracking and surveillance systems face biometric is generally considered than figureprint and iris. On the other hand, face biometric suffers from change in shape and expression over time and other challenging issues like varying illumination, poor contrast and non-cooperation by users causing limited recognition performances. In recent years, the problems of face biometric attracted researcher's attention to an emerging biometric modality the ear structure for human recognition. The ear structure is a passive biometrics which can be easily captured in a similar way like face and does not require much user cooperation. First, in [20] Iannarelli studied the uniqueness of human ear to identify criminals in forensic applications. The forensic science literature reported that major changes of the ear shape

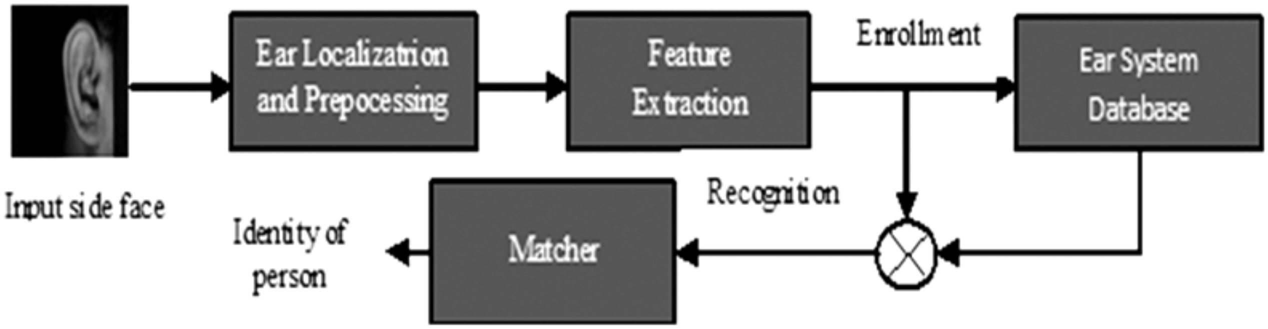


Figure 1: Principle of an ear identification system

occurs from four months to the age of eight after which the shape of the ear is observed to be stable up to the age of seventy then again it is stretched out. Hence, the recognition performance is comparable with face since its stable shape for long time. The performance of ear biometrics can be further improved when it is combined with face or other biometric cues in multimodal biometric systems. First research on computer based ear biometrics was performed by Burge and Burger [1]. In [23] authors provide extensive literature for recent survey of ear biometrics.

Basically, an ear identification system based on 2D side face images comprises three major segments as follows: ear localization, feature extraction and ear recognition shown in Figure 1. Key to the success of ear recognition is ear feature representation and recognition technique. In literature, there exist a large number of feature representation techniques starting from Principal Component Analysis [2, 3], Kernel Fisher Discriminant Analysis [14], 2D Gabor Filter [16], Independent Component Analysis [5], locally linear embedding (LLE) [7], Local Similar Binary Patterns [10], SIFT local features [11]. In addition, feature extraction using Force field energy functional in [4] and geometric features in [8, 9] have been proposed for ear recognition. In [6] authors have given importance to classification and proposed a multi-matcher based ear recognition system. Even though several feature extractors for ear identification have been employed in literature, still the performance of these existing techniques need to improve in presence of change in illumination, partial occlusion and degradation of input profile face images. The objective of this paper is to minimize the influence of these factors and design an efficient ear recognition system. To deal with the above issues, authors explore a simple and computationally fast local feature extraction technique pyramidal histogram of orientation gradients (PHOG) to represent ear image. Further, for matching localized ear images we employed two histogram based dissimilarity measures like histogram intersection and the Chi-squared statistics to match histograms of sub-regions of two PHOG descriptors. The Chi-squared statistics measure achieves better performance than other common matching measures. Additionally, we apply a linear dimensionality reduction (LDA) to remove noise, make the recognition more efficient by extracting discriminant features and which have large inter-class variance and small intra-class variance. While in the experiments, it has been observed that performance of proposed ear identification system with reduced feature vectors slightly decreases with fast processing. After transformation the reduced PHOG feature vectors contain negative values and the used histogram dissimilarity measures are not good for comparison hence, we employed distance based similarity measures like Euclidean distance, City block distance and normalized cosine distance. The proposed algorithm validates using IIT Delhi ear database and University of Notre Dame ear database (Collections E). The experimental results exhibit improved recognition rate in our method than other known state-of-the-art techniques.

The rest of the paper is organized as follows. Section II describes PHOG in details as well as proposed algorithm. We present experimental results and discussion in section III. Section IV finalizes the paper with concluding remarks.

2. PROPOSED ALGORITHM

In this section, we describe a robust local feature extraction technique PHOG to represent ear shapes. As PHOG features are histogram of edge directions of blocks and finally, all feature vectors are concatenated to represent the shape of the ear. As the dimension of PHOG descriptor is large, it is indeed necessary to select discriminant features by removing redundant features. For efficient classification purpose, LDA has used in our experiments.

2.1. Preprocessing

One of the important steps of the proposed algorithm is preprocessing in which ear localization or detection from side face images of human subjects affects the performance of ear identification. Further, localized ear images require image preprocessing operations to enhance quality of the images which results in extracting better features. Our ear localization technique discussed in [18] has been used to segment or detect the ear portion from the side face images. In this scheme, non-ear edges were pruned and from this result using Hausdorff distance based template matching technique rectangle is drawn around the ear portion of the side face images. Then, detected ear portion is cropped to get ear image which is used for further processing to normalize the effect of illumination and is enhanced the contrast of the localized ear images for improving the performance of ear identification. In this work, Gaussian low pass filter has been used to neutralize the variation of illumination and noise which significantly improved the performance. Other popular technique to avoid illumination variation can be used for further improvement of the ear identification. Adaptive histogram equalization (AHE) is used here to improve the contrast of ear images and to normalize the effect of shadow in the ear images. During enhancement of ear images with AHE, the images are divided into number of non-overlapping regions of almost equal sizes. Then, histogram equalization performs on each region by using an estimate of the CDF. Subsequently, the neighboring enhanced regions are combined using bilinear interpolation for mapping a pixel which in turn eliminates the artificially induced boundaries. It is important to mention that the proper parameter setting of AHE strongly influences the results.

2.2. Pyramid histogram of orientation gradients (PHOG)

In this paper, we have applied PHOG descriptor to represent the ear biometric for human recognition. The pyramid of histogram of orientation gradients (PHOG) features are used to represent local shape descriptor. PHOG was proposed by Bosch et al. [15] and has been efficiently used in object classification. In pyramid HOG, an image is decomposed into sequence of increasingly fine sub-regions termed as cells at several pyramid levels. In the sequel, level l of the pyramid has 2^l number of cells and at each level the cells are of different grid resolutions. Then PHOG descriptor is generated by combining histogram of edge orientations of each cell to achieve more local discriminative representation of objects than global measure. The technique for extracting PHOG descriptor is mainly motivated by the histograms of oriented gradients (HOG) of Dalal and Tiggs [12]. In [13] introduced spatial pyramid to represent each image efficiently as sub-dividing it into a sequence of increasingly finer spatial grids by repeatedly doubling the number of divisions in each sub-regions at different levels. In [12] used HOG descriptor by dividing the image into number of blocks then each block is arranged into number of cells by combining neighborhood pixels. Consequently, for every block a histogram of edge orientations is computed. The histogram bins are evenly distributed over $0 - 180^0$ or $0 - 360^0$, depending on the gradient magnitude is positive or negative. Finally, the edge orientations are quantized and the histograms of each block are normalized to compensate for illumination variation. The results of normalized histograms of all blocks are concatenated to describe shape descriptor.

In [15] described PHOG descriptor to represent an image by its local shape and spatial information of the shape. In this case, the image is partitioned into a sequence of sub-regions of different resolutions by repeatedly doubling the division of area of interest at each level of the pyramid. The sub-regions of different levels, called cells and for each cell a HOG vector are computed. The combination of these HOG feature vector represents the final PHOG descriptor. Here local shape is described by the counts occurrences of edge orientations in cells at each level. Hence, PHOG represents both edge direction and location. Furthermore, PHOG descriptor is more efficient to slight change in shape of the object due to clutter. In building the pyramid the image at level l is divided into 2^l cells along both directions of 2D axis. Consequently, level 0 is represented by a K-vector corresponding to the K bins of the histogram, level 1 by a 4K-vector and so on. Hence, the PHOG descriptor of the entire image is a vector of size $K * \sum_{l \in L} 4^l$. For example, for levels up to L = 3 and K = 9 bins the PHOG descriptor dimension will be 765. For any application, it is necessary to limit the number of levels of the pyramid to L = 3 to prevent over fitting. The PHOG descriptor is normalized to unity compensate for illumination.

2.3. Dimension Reduction

In PHOG, by increasing number of levels and bin size the dimensionality of the descriptor is enlarged. However, dimensionality reduction is indeed important for selection of discriminant features. Different machine learning algorithms such as principal component analysis (PCA), independent component analysis (ICA), linear discriminant analysis (LDA) etc. can be used for dimensionality reduction. In this paper, to improve the performance of the classifier, LDA has been used to reduce the size of the PHOG descriptor. In LDA, discriminant directions are calculated and the feature vector is projected into the lower dimensional space while protecting the discriminative information. LDA ensures that the hyper-planes minimize the with-in class scatter while maximizing the between-class scatter.

2.4. Methodology

The proposed ear recognition system consists of four major steps: Preprocessing, Feature Extraction, Dimensionality reduction and Classification. The block diagram of the proposed system is shown in Figure 2. This section describes the proposed system in the following subsections.

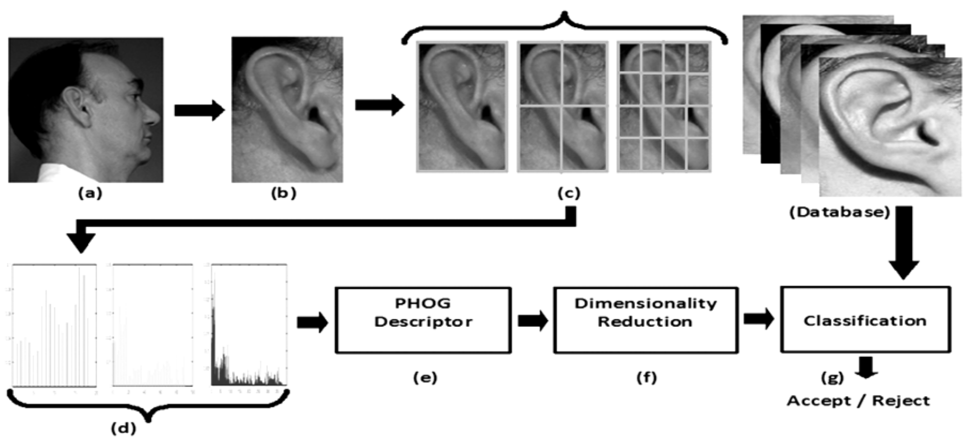


Figure 2: Block diagram of proposed ear recognition system

3. EXPERIMENTS AND RESULT

Two groups of experiments for our ear identification system are carried out using PHOG and reduced PHOG features to validate the results. We have simulated the proposed algorithm using IITD and UND-E databases. The used databases, experimental details and the results are discussed in following subsections.

3.1. Databases

Experimental results were evaluated using two ear databases: IIT Delhi ear database [22] and University of Notre Dame database (Collections E) [21]. Table 1 provides short description of three databases.

3.1.1. IIT Delhi ear database

The IITD ear image database [22] includes two set of ear images, database 1 contains 125 subjects and database 2 have 221 subjects, which is combination of first 125 subjects from first set and remaining ear images from second set. All the subjects are in the age group 14-58 years. The datasets provides automatically cropped and normalized ear images of size 180 x 50.

3.1.2. University of Notre Dame (Collections E) database

University of Notre Dame database, Collection E (UND-E) [21] consists of 464 side face images collected from 114 subjects, 3 to 9 samples per subject. The images are collected on different days with different conditions of pose and illumination. A huge intra-class variation is observed in these images due to pose variation and different illumination conditions. The ear images are automatically cropped using ear localization technique described in [18].

Table 1
Description of two databases used in the experiments

Name	*Classes*	*Total Samples*	*Samples per class(min/max)*	*Explanation*
IITD dataset I (Version 1.0)	125	493	3/6	Cropped grayscale ear images of size 50 x 180 pixels
IITD dataset II (Version 1.0)	221	793	3/6	Cropped grayscale ear images of size 50 x 180 pixels
UND Dataset (Collection E)	118	464	3/6	Images are affected by illumination and pose variations

3.2. Parameters Tuning

In our human ear identification experiments, the results are summarized in Table 2, 3 and 4. The first step of our experiments is preprocessing which includes Gaussian low pass filter to deal with the variation of illumination and noise. For all experiments the window size of low pass filter is [7 × 7]. Then local image enhancement technique adaptive histogram equalization (AHE) has been used to enhance the contrast of the ear image. The ear image is divided into number of tiles and dimensions of the tiles are varied from 8 x 8 to 20 x 20. In the experiments, we were realised the optimal tile size of 8 x 8 and 16 x 16 for IITD and UND-E databases respectively. Similarly, the PHOG features with 12 and 14 orientation bins in the range of [0, 360] for IITD and UND-E databases respectively and level of pyramid is 4 for both the databases. Similarly, the PHOG with reduced features using LDA revealed best results in 8 orientation bins [0, 360] degree with level 4.

3.3. Distance measures

We investigated and implemented here two distance measures for histogram [19]: Chi-squared statistics (χ^2), histogram intersection, as shown in equation (1) and (2) respectively. These two metrics are used for similarity measure between two PHOG descriptors which contain histograms for each block of different levels to compare the performance of our method using other popular distance measures.

$$\chi^2(S,M) = \sum_{j=1}^{B}\left(\sum_{i=1}^{P}\frac{\left(S_{i,j}-M_{i,j}\right)^2}{S_{i,j}+M_{i,j}}\right) \tag{1}$$

$$D(S,M) = \sum_{j=1}^{B}\left(\sum_{i=1}^{P}\min\left(S_{i,j},M_{i,j}\right)\right) \tag{2}$$

where $S_{i,j}$ and $M_{i,j}$ represent index of the bin i from the sub-region j. In the work, equation (1) and (2) are used to compare two images a sample *(S)* and a model *(M)* which compute dissimilarity between two feature vectors. The lower the value of the χ^2, then more closer a sample *(S)* to be model *(M)*. Similarly, lower the distance value of the histogram intersection measure between two feature vectors is more dissimilar from a sample *(S)* to be model *(M)*. In the experiments, the best results are obtained with Chi-square statistics.

3.4. Ear identification system

The performance of ear identification is tested using the process described by Kumar and Wu [16]. In IITD database, most of the subjects have number of samples less than 4. Hence, we considered three samples for each subject and out of them one sample was used for testing while the other two were considered for training. Similarly, in UND-E database we considered six samples for each subject due to more variation of illumination in the images. As all subjects do not contain six samples, we added few more samples of some subjects by (-5^0 and 5^0) rotating samples of the existing images. Here six test sets were considered, out of them one sample was considered randomly as test sample and remaining five samples were considered as training samples to improve performance. The average rank-one and rank-five recognition rates were reported from all the experiments in the Table 3 and 4 for ear identification with PHOG feature vectors and reduced PHOG feature vectors using LDA. The average recognition rate is plotted in figure 3, as function of the rank in a sorted list of scores. The rank-1 refers to first minimum distance (best result) and rank-5 refers to fifth minimum distance of the model (probe) image with the sample images of the subjects present in the database.

3.5. Results and Discussion

After preprocessing, we evaluated the effect of extracting PHOG descriptors at multiple scales with [0, 360] degree orientation bins and the use of different distance measures for identification of subjects. To remove redundant features, remove noise and avoid overfitting in recognition, we used LDA for dimensionality reduction of PHOG descriptor. The PHOG features were extracted at different levels over the whole normalized ear image. Then PHOG features are then processed by LDA. Recognition was done using nearest neighbor classifier with Euclidean distance, City block distance, normalized cosine, histogram intersection and Chi-square statistic measure. In case of reduced PHOG features after transformation values are negative hence, we could not use histogram intersection and Chi-square statistic measure. By using distance metrics we calculate a distance matrix containing normalized distance between each pair of images. Then by using the distance matrix and the given gallery image and probe image lists, our algorithm calculates the classification accuracy in terms of rank. A rank curve is a cumulative matching score for the probe images with the gallery images. The best performance result of the proposed method with optimized parameters of AHE and PHOG algorithms for two databases of IITD and UND-E revealed in Table 2. The Table 3 and 4 show the results with the average of correct recognition rate in percentage for two databases as values of recognition accuracy in terms of Rank-1 and Rank-5. It can be seen on the Table 3, the average recognition accuracy of PHOG features with χ^2 similarity measure outperforms other similarity measures. Similarly, City block distance measure recognition rate is very closer to histogram intersection measure. From the experimental results, it is clearly observed that PHOG features have high illumination invariant capability and easily manage small rotation of (-5^0 and 5^0) in the input ear image to recognize genuine subject. Experimental results in Table 4 indicates the recognition rates in percentage for reduced PHOG feature vectors using LDA, here it is clearly observed that results using reduced PHOG features are slightly more than PHOG feature and similarly, ear identification system with reduced PHOG features takes less processing time than that of the algorithm using PHOG features. In Table 4 average recognition rate using Euclidean distance measure outperforms other two distance measures. The cumulative match characteristic (CMC) curves of the identification experiments using PHOG features represent average recognition rates verses sorted ranks in two datasets are shown in figure 3. It is observed for best cases that proposed method using PHOG descriptor classifies 118 cases correctly and 7 cases incorrectly in IID (125) database. Similarly, it classifies 210 cases correctly and 11 cases incorrectly in IID (221) database. In UND-E database, it classifies 114 cases correctly and 4 cases incorrectly.

Table 2
Evaluation of optimal parameters: tile size in AHE, number of bins and level of pyramid of HOG for best performance of identification rate in percentage.

Method/Database		*Tile Size / Number of bins / Level*	*Maximum Rank-1 Recognition rate (%)*
PHOG	IITD (125)	[8 X 8], 12, 4	94.97
	IITD (221)	[8 X 8], 12, 4	95.07
	UND-E	[16 X 16], 16, 4	96.66

Table- 3
Average recognition rates (%) in the IITD (125), (221) and UND-E databases were obtained using PHOG features.

Database/Measures		*Euclidean*	*City block*	*Cosine*	*Histogram Intersection*	*Chi-squared statistics*
IITD (125)	Rank-1	90.13	90.60	90.13	89.33	93.60
	Rank-5	92.53	94.13	92.53	92.00	95.33
IITD (221)	Rank-1	93.36	93.97	93.36	92.31	93.97
	Rank-5	96.53	96.83	96.53	95.93	96.86
UND-E	Rank-1	96.05	96.33	96.05	95.48	96.35
	Rank-5	99.44	99.72	99.44	99.44	99.44

Table 4
Average recognition rates (%) in the IITD and UND-E databases were obtained using PHOG feature vectors with LDA for feature extraction.

Database/Measures		*Euclidean*	*City block*	*Cosine*
IITD (125)	Rank-1	90.57	89.47	90.33
	Rank-5	96.27	95.24	95.20
IITD (221)	Rank-1	93.86	90.89	91.25
	Rank-5	96.83	95.02	96.08
UND-E	Rank-1	96.76	96.07	95.92
	Rank-5	99.44	98.55	99.15

4. CONCLUSION

In this paper, we explore the use of PHOG descriptor to represent localized ear images for human identification. We also used here two dissimilarity measures for histograms: the histogram intersection and the Chi-squared statistics for calculating matching distance between gallery and probe images. It has been observed that Chi-squared statistics measure obtained better results as compared to other distance measures in the experiments. As global PHOG descriptor for local sub-regions is indeed large that needs dimensionality reduction using LDA to identify discriminative features to avoid overfitting and to decrease computational time and storage. Performance of reduced feature vectors slightly increases with less processing time. A series of experiments have been carried out over two standard datasets namely IIT Delhi ear database and University of Notre Dame ear database (Collection E) to show the effectiveness of the proposed system. Experimental results reveal that the proposed technique achieves significantly improved performance over the existing techniques.

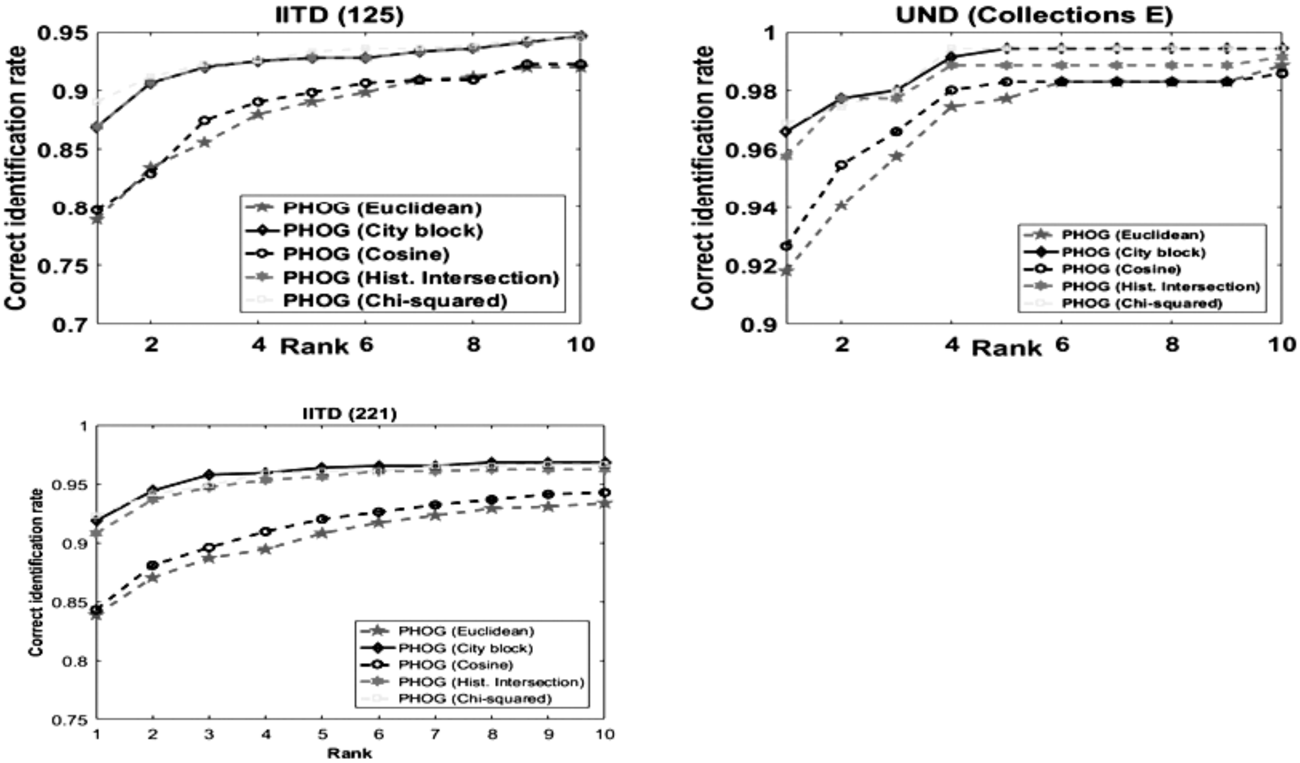


Figure 3: Classification performance in two ear databases

In summary, we conclude that our ear identification system can accomplish good performance in terms of speed and accuracy. For further improvement, we will focus on feature level fusion to add robustness to our scheme.

REFERENCES

[1] M. Burge, W. Burger, " Ear biometrics in computer vision," In: Proc. of Int'l Conference on Pattern Recognition (ICPR' 00), vol. 02, pp. 822-826, 2000.

[2] B. Victor, K.W. Bowyer, and S. Sarkar, "An Evaluation of Face and Ear Biometrics", Proc. Int'l Conf. Pattern Recognition, pp. 429-432, Aug. 2002.

[3] K. Chang, K. W. Bowyer, S. Sarkar, B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics." IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1160-1165, 2003.

[4] D. Hurley, M. Nixon, J. Carter, "Force field energy functionals for image feature extraction," In Proc. of Int'l Conf.Image and Vision Computing, vol. 20(5-6), pp. 311-317, 2002.

[5] Zhang, H, Mu, Z, Qu, W, Liu, L, Zhang, C, "A novel approach for ear recognition based on ICA and RBF network," In: Proc. of 4th Int'l Conference on Machine Learning and Cybernetics (CMLC' 05), pp. 4511-4515, 2005.

[6] L. Nanni, A. Lumini, "A multi-matcher for ear authentication," Pattern Recognition Letters 28(16), 2219-2226, 2007.

[7] Xie, Z, Mu, Z, "Ear recognition using LLE and IDLLE algorithm," In: 19th Int. Conf. on Pattern Recognition (ICPR), pp. 1–4, 2008.

[8] M. Choras, " Ear biometrics based on geometrical feature extraction," Electronic letters on computer vision and image analysis 5(3), 84-95, 2005.

[9] D. Shailaja, P. Gupta, "A simple geometric approach for ear recognition," In: Proc. Int'l Conference on Information Technology (ICIT), pp. 164-167, 2006.

[10] Y. Guo, Z, Xu, "Ear recognition using a new local matching approach". Fifteenth IEEE Int. Conf. on Image Processing (ICIP), pp. 289–292, 2008.

[11] J. Bustard, M. Nixon, "Robust 2d ear registration and recognition based on sift point matching," In: Proc. of Int'l Conference on Biometrics: Theory, Applications and Systems (BTAS' 08), pp. 1-6, 2008.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, 2005, pp. 886 –893, 2007.

[13] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," In Proc. ICCV, 2005.

[14] Li Yuan and Mu Zhichun, "Ear Recognition Based on Gabor Features and KFDA," Hindawi Publishing Corporation

[15] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," In: Proceedings of the ACM International Conference on Image and Video Retrieval, 2007.

[16] A. Kumar and C. Wu, "Automated human identification using ear imaging," Pattern Recognition, vol. 45, no. 3, pp. 956–968, 2012.

[17] S. Prakash, and Phalguni Gupta, "An efficient ear recognition technique invariant to illumination and pose," Telecommunication Systems vol. 52 (3), pp. 1435-1448, 2013.

[18] P. P. Sarangi, M. Panda, B. S. P. Mishra, and S. Dehuri, "An Automated Ear Localization Technique Based on Modified Hausdorff Distance," In Proc. of Int'l Conference on Computer Vision and Image Processing (CVIP), IIT Roorkee, 2016.

[19] G. Snedecor, W. Cochran, "Statistical Methods," ed 6. Ames, Iowa, 1967.

[20] Iannarelli, A.,"Ear Identification," Paramont Publishing, 1989.

[21] University of Notre Dame Ear Database, Collection E. http://www.nd.edu/~cvrl/CVRL/DataSets.html

[22] IIT Delhi Ear Database Version 1, http://web.iitd.ac.in/~biometrics/Database_Ear.htm.

[23] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison and M. S. Nixon, A survey on ear biometrics. ACM computing surveys (CSUR), 45(2), 22, 2013.