

EAR RECOGNITION AND OCCLUSION

B. S. El-Desouky¹, M. El-Kady², M. Z. Rashad³, Mahmoud M. Eid⁴

¹ Mathematics Department, Faculty of Science, Mansoura University, Egypt
b_desouky@yahoo.com

² Mathematics Department, Faculty of Science, Helwan University, Egypt
mamdouh_elkady@cic-cairo.com

³ Computer Science Department, Faculty of Computer science, Mansoura University,
Egypt
magdi_z2011@yahoo.com

⁴ Mathematics Department, Faculty of Science, Mansoura University, Egypt
m_m_3ed@hotmail.com

ABSTRACT

Personal identification using 2D ear images still has many problems such as occlusion mostly caused by hair, earrings, and clothes. To avoid this problem, we propose to divide the ear image into non-overlapping equal divisions and identify persons through these non-occluded parts separately and then combine outputs of the classification of these parts in abstract, rank, and measurement level fusion. Experimental results show that the increasing of recognition rate through combining small parts of non-occluded divisions of ear image.

KEYWORDS

Ear recognition, LDA (Linear Discriminante Analysis), DCT (Discrete Cosine Transform), combining classifiers

1. INTRODUCTION

Ear recognition is one of the recent uni-modal biometric techniques. Ear recognition system considers better than many other famous biometric systems because ear images are relatively small so it need low computation time. Also, ear shape didn't affected by expression, mode, or health. And the rate of change of the ear size is small. But, ear recognition systems, still face many problems such as occlusion, illumination, and pose.

Our proposed algorithm aims to overcome the occlusion problem based on detection of persons based on small part of the ear image. In this algorithm, the ear image is divided into non-overlapping equal parts called divisions, blocks, or patches (figure 1). In each division we extract the features and then match each division separately. Then combine outputs of the classification stage of these small parts at abstract, rank, and measurements level.

Chang et al. [1] compared ear and face recognition rates using a principal component analysis (PCA) technique on faces and ear images. In a multimodal experiment for combining ear and face images, the recognition rate was 90.9%. Zhang et al. [2] combined the left and right ears to increase the recognition rate. They achieved recognition rate of 93.3% by using one ear image (left or right) and 95.1% by combining left and right ear images.

Many other researches identify persons based on occluded ear images. Arbab et al. try to solve the occlusion problem through using Hough transform and the recognition rate was suitable for example the recognition rate was (96%, 93%, 90%, and 83%) when the occlusion was (10%, 20%, 30%, and 40%) [3]. Maria et al. proposed HERO (Human Ear Recognition against Occlusions) technique and achieved good recognition rate when they used complete and occluded ear images. The achieved recognition rate ranged from 90% to 98% when using complete ear images and 84% to 97% when using occluded ear images [4]. Yuan et al. improved Non-Negative Factorization with Sparseness Constraints (INMFSC) by imposing an additional constraint on the objective function of NMFSC. NMF (Non-Negative Matrix Factorization) is a part-based image representation method. It factorizes the image database into two matrix factors whose entries are all non-negative and produces a part-based representation of images because it allows only additive, not subtractive, combinations of basis components. The recognition rate ranged from 91% to 93% when USTB ear database (24 subjects 4 image per subject) used and when the ear image occluded the results fluctuated from 48% to 92% [5]. Bustard et al. extracted the features from ear images using SIFT (Scale Invariant Feature Transform) based on 2D distance algorithm and compare the results with PCA. The recognition rate of the original ear images was 100%. And in case of 20% occlusion from above the results was 92% and when the occlusion was 20% from side the results was 66% [6, 7].

In this study, first we exclude the occluded parts then extract the features from the remaining divisions and match each division separately. After that, we make the combination at different classifier levels such as, abstract (decision), rank, and score (measurement) levels.

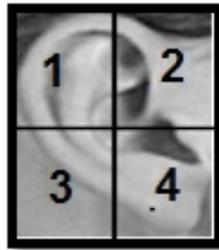


Figure 1. 4x4 blocks on 2D ear image

2. METHODOLOGY

In this section, we present the proposed method: the first part presents the feature extraction methods, while, the second part introduces the combination techniques used in our research.

2.1 Feature Extraction Methods

Feature extraction methods used to represent images in another space. In the following sections, we will explain the two feature extraction methods used in this paper.

2.1.1 LDA (Linear Discriminant Analysis)

Has been successfully used as a classification technique for a number of problems, including speech recognition; face recognition, and multimedia information retrieval. While PCA takes the complete training data as one entity, LDA's goal is to find an efficient way to represent the vector space by using class information (class is defined as a collection of data belonging to a particular entity, for example a collection of images belonging to a person.) [9, 10]. The ear images in the

training set are divided into the corresponding classes. LDA then compute a set of vectors W as follows:

$$W = \max \left| \frac{W^T S_b W}{W^T S_w W} \right| = \max \left| \frac{S_b}{S_w} \right|, \quad (1)$$

where S_b is the between-class scatter matrix and S_w is the within-class scatter matrix, defined by:

$$S_i = \frac{1}{N_i} \sum_{k=1}^M (x_k - m_i)(x_k - m_i)^T, \quad (2)$$

$$S_w = \sum_{i=1}^M \frac{1}{M} S_i, \quad (3)$$

$$S_b = \sum_{i=1}^M \frac{1}{M} (m_i - m)(m_i - m)^T, \quad (4)$$

where N_i is the number of training samples in class i , M is the number of distinct classes, m_i is the mean vector of the samples belonging to class i . S_w represents the scatter of features around the mean of each ear class and S_b represents the scatter of features around the overall mean for all ear classes [8, 9]. Solve for the generalized eigenvectors (V) and eigenvalue (λ) of the within class and between class scatter matrices.

$$S_b V = S_w V \lambda. \quad (5)$$

Then, sort the eigenvectors according to eigenvalue and project all the original (i.e. not centered) images onto the Fisher basis vectors

$$Y = W V. \quad (6)$$

2.1.2 DCT (Discrete Cosine Transform)

Is a well-known signal analysis tool used in compression standards due to its compact representation power. A detected and enhanced ear image is divided into 8x8 blocks. Then, apply DCT equation (equation 7) on each block. DCT will push the lower frequencies toward the upper left corner of each block. Then, extract the information from each block using zig-zag scan. To fuse the local information, the extracted features from 8x8 pixels blocks can be combined at the feature level or at the decision level [10].

$$B_{pq} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N}, \quad (7)$$

Where

$$0 \leq p \leq M-1, 0 \leq q \leq N-1, \quad \begin{cases} p=0 \\ \sqrt{2/M} \\ \sqrt{2/M} \end{cases} \quad \begin{cases} p=0 \\ 1 \leq p \leq M-1 \end{cases}, \quad \begin{cases} q=0 \\ \sqrt{2/N} \\ \sqrt{2/N} \end{cases} \quad \begin{cases} q=0 \\ 1 \leq q \leq N-1 \end{cases}.$$

In this paper, the simple nearest neighbour classifier is applied using Euclidean, Cosine, and City block metrics.

2.2 Combining Classifiers

Combining classifiers technique used to increase the recognition rate and it achieves good results when the classifiers used in combination are diversified or independent. In many researches the combination achieved in many levels, outputs of different samples, outputs of one classifier using different parameters, or outputs of classifiers using different feature extraction methods. In this research, combining classifiers in abstract, rank, and score level is used.

2.2.1 Decision (Abstract) Level Fusion

Combining classifiers by combining the outputs of different classifiers for a test image. In our case, we combined outputs of nearest neighbor classifiers trained by different blocks that correspond to different regions on an ear image. And in this research we used majority voting. Many researches focus on producing a pool of classifiers and select the most diverse and accurate classifiers. The most diverse ensemble, Giacinto and Roli [11], uses the double fault measure (DF) [12] and the Q statistics [13] form a pair wise diversity matrix for a classifier pool and subsequently to select classifiers that are least related. Other abstract level fusion methods are clustering and selection [14], and thinning the ensemble [15].

2.2.2 Rank Level Fusion

Combine the classifiers through sort the output of each classifier (subset of possible matches) in decreasing order of confidence so each class has its own rank; the fusion can be done through counting the ranks of each class and the decision will be the class of the highest rank.

2.2.3 Score (Measurement) Level Fusion

Combine the vectors represents the measures between the test image and the training images (output of each classifier). Then, compute minimum, maximum, mean, median, sum, and product of the used classifiers and the decision is determined by the class that has the minimum value. Assume that, the problem of classifying an input pattern Z into one of m possible classes based on the evidence provided by R different classifiers. Let \vec{x}_1 be the feature vector (derived from the input pattern Z) presented to the i th classifier. Let the outputs of the individual classifiers be $P(\cdot | \vec{x}_1)$, i.e., the posterior probability of the pattern Z belonging to class \cdot given the feature vector \vec{x}_1 . Let $c \in \{1, 2, \dots, m\}$ be the class to which the input pattern Z is finally assigned. The following rules is used to determine the class c [9]: $c = \arg \max_j \max_i P(\cdot | \vec{x}_1)$, $c = \arg \max_j \min_i P(\cdot | \vec{x}_1)$, $c = \arg \max_j \text{med}_i P(\cdot | \vec{x}_1)$, $c = \arg \max_j \text{avg}_i P(\cdot | \vec{x}_1)$, and $c = \arg \max_j \prod_i P(\cdot | \vec{x}_1)$.

3. PROPOSED ALGORITHM

In our proposed algorithm, first, we identify persons using only the non-occluded parts using single classifiers. Then, we combine outputs of the classifiers of each part in abstract, rank, or measurement level as in figure 2. From the figure also we note the black block is occluded. In abstract level, we used the whole ensemble, most diverse, thinning ensemble, or clustering techniques to choose the suitable ensemble then apply majority voting technique on the selected ensemble. Also, in measurement level we used min, max, sum, median, mean (average), or product rules to compute the final decision. And in rank level we used borda count technique.

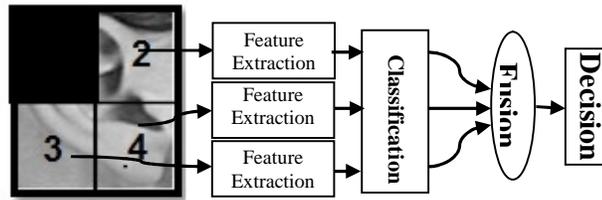


Figure 2. Proposed Algorithm

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed ear recognition system, we will make two experiments using ear images database. The ear images database consists of 102 grayscale images (6 images for each of 17 subjects) in PGM format. Six views of the left profile from each subject were taken under uniform, diffuse lighting. Slight changes in the head position were encouraged from image to image [11].

In our experiments we used three and four training images and we identify persons using non-occluded parts of the ear images. In the first experiment we used two non-occluded parts while in the second experiment we used three non-occluded parts. And, in all experiments we used LDA and DCT as feature extraction methods.

Table 1: Recognition rate (%) of ear images using three training images and two non-occluded parts

		Part 1&2		Part 1&3		Part 1&4		Part 2&3		Part 2&4		Part 3&4	
		LDA	DCT										
Single Classifier	Euclidean	94	90	96	94	96	90	96	90	92	92	96	92
	City Block	94	90	96	94	94	92	96	90	94	92	96	96
	Cosine	96	92	96	96	100	96	96	90	92	90	96	96
Abstract Level	Whole Ensemble	94	88	98	94	98	94	90	88	88	88	96	90
	Most Diverse	90	92	96	94	90	84	96	80	88	80	96	94
	Thinning Ensemble	96	94	96	98	96	94	96	94	90	88	96	94
	Clustering	96	88	98	96	96	94	88	94	88	88	90	88
Rank Level	Rank	90	86	98	98	96	94	98	96	94	90	96	92
Measurement Level	Min	88	88	98	98	92	88	92	96	90	88	92	84
	Max	90	90	98	94	96	92	92	90	92	90	96	94
	Mean	96	90	98	96	96	96	98	94	94	92	94	90
	Median	92	90	98	94	98	94	96	94	92	92	94	90
	Product	94	90	98	94	98	94	96	90	94	92	94	92
	Sum	92	90	98	94	98	94	96	94	92	92	94	90

Table 2: Recognition rate (%) of ear images using four training images and two non-occluded parts

		Part 1&2		Part 1&3		Part 1&4		Part 2&3		Part 2&4		Part 3&4	
		LDA	DCT										
Single Classifier	Euclidean	94	91	94	91	94	91	94	94	94	94	91	91
	City Block	94	91	94	91	91	91	94	91	94	94	91	97
	Cosine	94	94	94	94	97	94	94	94	94	94	91	97
Abstract Level	Whole Ensemble	94	91	94	94	97	94	91	88	91	88	94	91
	Most Diverse	94	88	91	94	94	91	91	94	91	91	94	91
	Thinning Ensemble	94	91	94	94	97	91	91	94	97	91	97	94
	Clustering	94	91	94	97	97	94	94	97	97	91	97	94
Rank Level	Rank	94	91	97	97	100	97	97	97	100	91	94	94
Measurement Level	Min	94	88	97	97	97	88	97	97	100	94	97	94
	Max	94	91	97	88	94	94	91	91	94	88	94	94
	Mean	94	91	97	97	100	100	97	97	100	97	97	91
	Median	94	91	97	91	100	94	97	94	100	94	94	94
	Product	94	91	97	91	100	97	97	94	100	94	100	94
	Sum	94	91	97	91	100	94	97	94	100	94	94	94

Table 3: Recognition rate (%) of ear image using three non-occluded parts

		Parts 1&2&3				Parts 1&2&4				Parts 1&3&4				Parts 2&3&4			
		LDA		DCT		LDA		DCT		LDA		DCT		LDA		DCT	
		T_4	T_3	T_4	T_3												
Single Classifier	Euclidean	94	96	91	92	94	94	94	94	94	96	91	94	94	94	94	92
	City Block	94	96	94	94	94	94	91	92	94	96	91	94	94	96	94	94
	Cosine	94	96	94	94	94	94	94	94	94	96	94	96	94	94	94	90
Abstract Level	Whole Ensemble	94	96	94	94	94	92	94	92	97	96	94	94	97	92	94	92
	Most Diverse	91	96	94	80	94	90	88	84	94	90	91	84	91	88	91	80
	Thinning Ensemble	94	96	94	98	97	96	91	94	97	96	94	98	97	96	94	94
	Clustering	94	98	97	94	97	96	94	88	97	98	94	98	94	90	94	88
Rank Level	Rank	97	98	97	96	97	94	97	94	97	98	97	96	97	96	97	94
Measurement Level	Min	97	96	97	96	100	92	94	88	100	98	94	92	100	94	100	92
	Max	97	96	88	92	94	94	91	92	97	98	94	96	91	92	91	90
	Mean	97	98	97	96	100	96	97	94	100	98	100	96	97	98	100	94
	Median	97	96	91	92	94	94	94	94	100	98	100	98	97	98	94	94
	Product	94	92	94	94	97	96	94	92	97	96	94	94	97	96	91	92
	Sum	97	96	91	92	94	94	94	94	100	98	100	98	97	98	94	94

Table 4. Recognition Rate of the state-of-the-art methods for occluded Ear recognition systems

Arbab et al. [3]	10% Occlusion	96%
	20% Occlusion	93%
	30% Occlusion	90%
	40% Occlusion	83%
	50% Occlusion	66%
Yuan et al. [5]	10% Occlusion	92%
	20% Occlusion	85%
	30% Occlusion	67%
	40% Occlusion	58%
Bustard et al. [6,7]	20% Occlusion From above	92%
	30% Occlusion From above	74%
	10% Occlusion From the side	92%
	20% Occlusion From the side	66%

From table 1 and table 2 we note that, combining classifiers in measurement level achieved recognition rate better than abstract or rank level and the mean rule achieved the best result in the measurement level. Also, we note that, the rank level achieved results better than abstract level and thinning ensemble and clustering techniques achieved the best results in abstract level.

For example in table1, table 2, and table 3 we note that, the mean rule in measurement level achieved recognition rate approximately equal to single classifiers in about 37.5% of all cases and it outperforms the single classifiers in about 62.5% of all cases. Also, in rank level the recognition rate is approximately equal to single classifiers in about 35% while better than single classifiers in about 57.5 % of all results. Finally, in abstract level thinning ensemble and clustering techniques achieved results approximately near to the single classifiers' results in about 62.5% and 40% respectively of all cases while they achieved results better than single classifiers in about 25% and 32.5% of all cases respectively.

Recognition Rates of the state-of-the-art methods for occluded ear images are shown in Table 4. In this table, [3], identified persons based on occluded ear images and the occlusion was from top and it ranged from 10% to 50% of the size of the ear, [6, 7] also identified persons based on occluded ear images and their occlusion was from top (20% and 30%) and side (10% and 20%), and [5], used occluded ear images in their experiments and their occlusion rates fluctuated from 10% to 40%. Results show that the proposed fusion algorithm increases the recognition rate in comparison to the state-of-the-art methods.

5. CONCLUSION

In this study, we tried to overcome the occlusion problem in ear recognition systems. In our proposed algorithm we divide the ear images into four equal blocks (parts) and identify persons based on these non-occluded small parts. Then, combine outputs of the classifiers in abstract, rank, and measurement level to increase the recognition rate and decrease the error and avoid the effect of the occlusion.

From table 1, table 2, and table 3 we note that, identifying persons based on combining non-occluded parts achieved results better than single classifiers. Also, we note that measurement level achieved results better than abstract or ranks level and the mean rule achieved the best

results in measurement level. Finally, our algorithm overcomes the occlusion problems in ear recognition system. Future work includes, dividing ear images into a non-uniform parts and using many other classifiers and combination techniques.

6. REFERENCES

- [1] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and Combination of Ear and Face Images in Appearance-based Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(9): 1160-1165, September 2003.
- [2] D. J. Hurley, M. S. Nixon, and J. N. Carter, "Force Field Feature Extraction for Ear Biometrics," *Computer Vision and Image Understanding*, 98: 491-512, June 2005.
- [3] B. Arbab-Zavar, and M. S. Nixon, (2007) On Shape-Mediated Enrolment in Ear Biometrics. In: *International Symposium on Visual Computing (ISVC)*, 26-28 November 2007, Nevada. Vol. 2
- [4] M. De Marsico, M. Nappi, R. Daniel, "HERO: Human Ear Recognition against Occlusions", *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on, pp. 178 – 183, 2010.
- [5] L. Yuan, Z. Mu, Y. Zhang, K. Liu, "Ear Recognition using Improved Non-Negative Matrix Factorization," *icpr*, vol. 4, pp.501-504, 18th International Conference on Pattern Recognition (ICPR'06), Vol. 4, 2006.
- [6] D. Bustard, M. S. Nixon, "Robust 2D Ear Registration and Recognition Based on SIFT Point Matching", *Biometrics: Theory, Applications and Systems (BTAS 2008) 2nd IEEE International Conference on*, pp. 1 – 6, 2008.
- [7] D. Bustard "Toward Unconstrained Ear Recognition From Two-Dimensional Images", *Journal IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans - Special issue on recent advances in biometrics*, Volume 40 Issue 3, May 2010
- [8] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *Proc. of the Fourth European Conference on Computer Vision Cambridge-UK*, 1: 45-58, April 1996.
- [9] A. Martinez, A. Kak, "PCA versus LDA", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(2): 228-233, February 2001.
- [10] E. H. Kemal , R. Stiefelhagen, "Local Appearance Based Face Recognition Using Discrete Cosine Transform", *13th European Signal Processing Conf. (EUSIPCO 2005)*, Antalya, Turkey, 2005.
- [11] G. Giacinto and F. Roli. Design of effective neural network ensembles for image classification processes. *Image Vision and Computing Journal*, 19(9-10):699-707, 2001.
- [12] D. Ruta and B. Gabrys. Analysis of the correlation between majority voting error and the diversity measures in multiple classifier systems. *Proc. of 4th International Symposium on Soft Computing*, 2001.
- [13] G. U. Yule. On the association of attributes in statistics. *Philosophical Tran. of the Royal Society of London, A*, 194:257-319, 1900.
- [14] G. Giacinto F. Roli. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 22:25-33, 2001.
- [15] R. E. Banfield, L. O. Hall, K. W. Bowyer, W. P. Kegelmeyer. A new ensemble diversity measure applied to thinning ensembles. *Proc. of 4th International Workshop on Multiple Classifier Systems, LNCS 2709*, 306-316, 2003.
- [16] M. Á. Carreira-Perpiñán, "Compression Neural Networks for Feature Extraction: Application to Human Recognition from Ear Images" (in Spanish), M.S. thesis, Faculty of Informatics, Technical University of Madrid, Spain, 1995.