Ear Recognition Using Block-Based Principal Component Analysis and Decision Fusion

Alaa Tharwat^{1(✉)}, Abdelhameed Ibrahim², Aboul Ella Hassanien³,
and Gerald Schaefer⁴

¹ Electrical Engineering Department, Suez Canal University, Ismailia, Egypt
engalaatharwat@hotmail.com

² Computer Engineering and Systems Department,
Mansoura University, Mansoura, Egypt

³ Faculty of Computers and Information, Cairo University, Giza, Egypt

⁴ Department of Computer Science, Loughborough University, Loughborough, UK
gerald.schaefer@ieee.org

Abstract. In this paper, we propose a fast and accurate ear recognition system based on principal component analysis (PCA) and fusion at classification and feature levels. Conventional PCA suffers from time and space complexity when dealing with high-dimensional data sets. Our proposed algorithm divides a large image into smaller blocks, and then applies PCA on each block separately, followed by classification using a minimum distance classifier. While the recognition rates on small blocks are lower than that on the whole ear image, combining the outputs of the classifiers is shown to increase the recognition rate. Experimental results confirm that our proposed algorithm is fast and achieves recognition performance superior to that yielded when using whole ear images.

Keywords: Ear recognition · Principal component analysis (PCA) · Feature fusion · Classifier fusion

1 Introduction

Ear recognition systems are a relatively recent biometric technique, and are challenging to implement in practice due to difficulties controlling occlusions, pose, illumination etc. Ears have played a significant role in forensic science, especially in the United States, where an ear classification system based on manual measurements has been used for more than 40 years [5]. Using a collection of over 10,000 ears, they were found to be distinguishable based on only 12 measurements.

Chang *et al.* [3] compared ear recognition with face recognition using a standard PCA technique on face and ear images, and reported accuracies of 71.6 % and 70.5 % for ear and face recognition, respectively. They also presented results with varying lighting which resulted in lower recognition accuracies of 64.9 % and 68.5 % for face and ear images, respectively. Combining ear and face images lead

to a significant improvement and an accuracy of 90.9 %. Kumar and Zhang [7] employed different feature extraction methods and different classification algorithms, namely feed-forward artificial neural networks and three classifiers based on a nearest neighbour rule. The experiments they performed yielded recognition rates ranging from 76.5 % to 94.1 %. Alaa *et al.* [13] used feature combination to improve the performance of an ear recognition system, and achieved recognition rates between 85.9 % and 96.1 %.

Principal component analysis (PCA) is widely used for dimensionality reduction, feature extraction, compression, visualisation, and other tasks. PCA finds the c principal orthonormal vectors which describe an eigenspace. In many applications, c is much smaller than the original dimensionality of the data, while the computation of PCA can be implemented using eigenvalue decomposition (EVD) of the covariance matrix of the data matrix [16]. However, PCA requires relatively high computational complexity and memory requirements, especially for large datasets [1, 11].

To address this, Golub and van Loan [4] used Jacobi's method which diagonalises a symmetric matrix and requires about $O(d^3 + d^2n)$ computations, where n represents the number of feature vectors or samples and d^d represents the dimensionality of the vectors. Roweis [9] proposed an expectation maximisation (EM) algorithm for PCA, which is shown to be computationally more effective compared to the EVD method for PCA. However, calculating PCA based on EM is still expensive, while the EM algorithm may not converge to the global maximum but only a local one, and is dependent on the initialisation. The power method can also be used to find leading eigenvectors, and is less expensive, but can compute only one most leading eigenvector [10]. Also, Skarbek [12] and Liu *et al.* [8] proposed eigenspace merging where it is not necessary to store the covariance matrix of previous training samples.

In this paper, we propose a fast and accurate ear recognition system through fusion at classification level and feature level and PCA on subimages. Our algorithm aims to decrease the dimensionality and hence decrease the complexity of a PCA-based algorithm by dividing the ear image into small blocks. PCA is then applied on the image blocks separately and classification performed using a minimum distance classifier. The outputs of these classifiers at abstract, rank, and score level are combined, while we also investigate combining the block features at the feature level. Experimental results confirm that our proposed algorithm is fast and achieves recognition performance superior to that yielded when using whole ear images.

The rest of the paper is organised as follows. In Sect. 2, we summarise some of the background on principal component analysis and current fusion methods. Section 3 then details our proposed algorithm. Section 4 gives experimental results, while Sect. 5 concludes the paper.

2 Background

2.1 Principal Component Analysis

Principal component analysis (PCA) is a popular linear subspace method that finds a linear transformation which reduces the d -dimensional feature vectors to h -dimensional feature vectors with $h < d$. It is possible to reconstruct the h -dimensional feature vectors from the d -dimensional reduced representation with some finite error known as reconstruction error. Of the resulting h basis vectors, the first one is in the direction of the maximum variance of the given data, while the remaining basis vectors are mutually orthogonal and maximise the remaining variance. Each basis vector represents a principal axis. These principal axes can be obtained by the dominant/leading eigenvectors (i.e. those with the largest associated eigenvalues) of the measured covariance matrix of the original data matrix. In PCA, the original feature space is characterised by these basis vectors and the number of basis vectors is usually much smaller than the dimensionality d of the feature space [12].

For an ear image $\Gamma(M \times N)$, where M and N are the width and height of the image, it is first transformed into a vector of length $M \times N$. The feature matrix of K training ear images is then given by $\Gamma = [\Gamma_1, \Gamma_2, \dots, \Gamma_K]$ and the average of the training set is calculated as $\psi = \frac{1}{K} \sum \Gamma_i$. The average is subtracted, i.e. $\phi_i = \Gamma_i - \psi$, and the data matrix created as $A = [\phi_1, \phi_2, \dots, \phi_K]((M \times N) \times K)$.

The covariance matrix of A is then calculated as

$$C = AA^T. \quad (1)$$

Next, the eigenvalues (λ_k) and eigenvectors (V_k) of C are computed and the eigenvectors sorted according to the corresponding eigenvalues. Dimensionality reduction is then achieved by retaining only the top h eigenvectors to yield a projection matrix P . For an ear image T (of the same size as the training images), it is first mean-normalised by $\phi_T = T - \psi$, and then transformed into the “eigen-ear” components, i.e. projected into ear space, by

$$\omega = P^T \phi_T. \quad (2)$$

Considering the computational complexity of PCA [15], for an ear image set of K images of dimensions $M \times N$, calculating the mean image is of $O(KMN)$ and subtracting it from the data matrix also of $O(KMN)$, while the complexity of calculating the covariance matrix is $O(K(MN)^2)$. Identifying the eigenvalues and eigenvectors of the covariance matrix then requires $O(K(MN)^3)$, whereas sorting the eigenvectors according to their eigenvalues can be done in $O(K(MN)\log_2(MN))$ (using a merge sort algorithm). Since only the first h eigenvectors are considered, computation of the reduced eigenspace is carried out in $O(hMN)$. Finally, projecting the images into this eigenspace requires $O(KLMN)$. Consequently, the overall computational complexity is

$$\begin{aligned} O_{\text{PCA}} &= O(KMN) + O(KMN) + O(K(MN)^2) + O(K(MN)^3) + \quad (3) \\ &\quad O(K(MN)\log_2(MN)) + O(hMN) + O(KhMN) \\ &= O(K(MN)^3). \end{aligned}$$

2.2 Fusion Methods

Combining different and independent resources can increase the accuracy of biometric (or other) systems. Misclassification of some samples by a method or classifier can be compensated by combining different resources which in turn can be performed at different levels. There are various approaches for such fusion methods, of which we summarise the most common in the following.

Multi-instance systems use various sensors to capture samples. In multi-sensorial systems, samples from the same instance are captured using two or more different sensors (e.g., both visible light and infra-red cameras) are combined in a sensor level fusion approach to increase the robustness of the biometric system [6].

Combination at feature level can lead to improved performance as more information is available (compared to fusion at classification level, which is discussed below). Fusion of features is usually implemented by concatenating two or more feature vectors, i.e. if $f_1 = \{x_1, \dots, x_n\}$ and $f_2 = \{y_1, \dots, y_m\}$ are two feature vectors of lengths n and m , respectively, then the fused feature vector $f = \{x_1, \dots, x_n, \dots, y_1, \dots, y_m\}$ is obtained by concatenation of f_1 and f_2 .

Fusion at classification level, or classifier fusion, can improve recognition performance compared to simple individual classifiers. In general, we can distinguish three levels of fusion here, namely:

- *Abstract Level Fusion:* Abstract or decision level fusion can be seen as making a decision by combining the outputs of different classifiers for a test sample. It is the simplest fusion method and majority voting is the most commonly employed method here.
- *Rank Level Fusion:* Here, the outputs of each classifier (a subset of possible matches) are sorted in decreasing order of confidence so that each class has its own rank. Fusion can be performed by summing up the ranks of each class and the decision is given by the class of the highest rank.
- *Score (Measurement) Level Fusion:* Fusion rules on the vectors are derived to represent the distance between the test image and the training images. Thus, the output of each classifier is represented by scores or measurements. Fusion at this level combines the vectors of scores, and the decision is given by the class that has the minimum value. Assume that we want to classify an input pattern Z into one of m possible classes based on the evidence provided by R different classifiers. Let x_i be the score derived for Z from the i -th classifier, and let the outputs of the individual classifiers be $P(\omega_j|x_i)$, i.e., the posterior probability of pattern Z belonging to class ω_j given the scores x_i . If $c = \{1, 2, \dots, m\}$ is the class to which Z is finally assigned, then this can be

done by the following rules [14]:

$$\begin{aligned}
 c &= \arg \max_j \max_i P(\omega_j | x_i) \\
 c &= \arg \max_j \min_i P(\omega_j | x_i) \\
 c &= \arg \max_j \text{med}_i P(\omega_j | x_i) \\
 c &= \arg \max_j \text{avg}_i P(\omega_j | x_i) \\
 c &= \arg \max_j \prod_i P(\omega_j | x_i)
 \end{aligned} \tag{4}$$

3 Proposed Algorithm

In this paper, we propose a fast and accurate ear recognition system based on principal component analysis (PCA) and fusion at classification and feature levels. Figure 1 gives an overview of our proposed algorithm.

The first step in our approach is to divide each ear image into non-overlapping blocks. In each block, PCA features are extracted and each subimage is matched separately using a minimum distance classifier.

In the first model, the features of the blocks are combined at feature level. Here, the PCA algorithm is applied on each subimage and then the features of all blocks combined to form a feature vector. Classification is performed using a minimum distance classifier.

In the first model, the outputs of classifiers are combined at abstract, rank, or score levels. For decision level fusion, majority voting method is used to combine results from all blocks. For rank level fusion, the Borda count method is employed to combine the ranks obtained by the individual classifiers. Finally, for score level fusion, minimum, maximum, product, mean, and median rules are used to combine the scores.

Considering the computational complexity of our approach, the first step in our algorithm is to divide the ear images, which are of size $M \times N$, into Q

Fig. 1. Overview of our proposed ear recognition algorithm

equal-sized non-overlapping blocks. Consequently, the size of each image block will be $(M \times N)/Q$, and the computational complexity to perform PCA on an image block is thus of $O(K(MN/Q)^3)$. For all Q sub images, the computational complexity is thus of

$$O_{\text{blockPCA}} = O(QK(MN/Q)^3) = O(K(MN)^3/Q^2), \quad (5)$$

and hence significantly smaller, by a factor of Q^2 , compared to the $O(K(MN)^3)$ of performing PCA on the full images. However, the recognition rate of each block will be lower compared to the recognition rate of the whole ear image. Thus, in our approach, the fusion techniques discussed above are applied to increase the recognition performance.

4 Experimental Results

In our experiments, we use a dataset of 102 ear images, 6 images for each of 17 subjects [2]. In particular, six views of the left profile from each subject were taken under uniform diffuse lighting conditions, while there are slight changes in the head position from image to image.

For our experiments, a minimum distance classifier is used based on three different metrics, namely Euclidean, city-block, and cosine metrics. In our first experiment, we applied classical PCA based on the whole ear image using 2 and 4 of the images for training, respectively. The results of this are summarised in Tables 1 and 2 in terms of recognition performance and computation time, respectively.

In the second experiment, the proposed PCA algorithm is applied. Here, the ear images are divided into into 4, 9, and 16 blocks, respectively to reduce the

Table 1. Recognition rates [%] using different minimum distance classifiers and PCA on whole ear images

Metric	2 Training images	4 Training images
Euclidean	89.7059	94.1176
City-block	88.2353	94.1176
Cosine	89.7059	94.1176

Table 2. Comparison of required CPU times of applying PCA on whole ear images and on image blocks

Blocks	2 Training images	4 Training images
1 (whole image)	607	615
$2 \times 2 = 4$	36	39
$3 \times 3 = 9$	9	10
$4 \times 4 = 16$	2	3

Fig. 2. Recognition rates [%] of the proposed feature fusion model at feature level

Fig. 3. Recognition rates [%] of the proposed classifier fusion model at abstract level

computational complexity of the proposed model. The resulting computation times are given in Table 2. As we can see from there, the time required for PCA calculation on blocks is very small compared to that of performing PCA on the whole images, while computation time decreased with decreasing block sizes.

After dividing the ear image into blocks, the features of all blocks are combined into one feature vector to perform feature level fusion, while matching is performed using minimum distance classifiers. The obtained results are given in Fig. 2. As shown there, feature fusion of image blocks can lead to better recognition performance compared to utilising the whole ear images. Moreover, the accuracies of different block sizes are approximately the same.

The final experiment is conducted to investigate the proposed classifier fusion models. Here, the features that result from each block are matched separately, while we combine the results of the classifiers at abstract level, rank level, and score level, respectively. The corresponding recognition results are shown in Figs. 3, 4, and 5, respectively.

From Figs. 3, 4, and 5, we can see that our proposed classifier fusion model achieves accuracies that exceed those obtained based on the whole ear images. Moreover, the accuracy of our model is in general inversely proportional to the number of blocks employed. When the number of blocks is increased, some of the resulting blocks will contain only background or small parts of the ear, which

Fig. 4. Recognition rates [%] of the proposed classifier fusion model at rank level

Fig. 5. Recognition rates [%] of the proposed classifier fusion model at score level

decreases the accuracy for these parts and hence affects the overall accuracy. This problem can in particular be observed for abstract level fusion, while for rank and score level fusion, using many ranks or scores may compensate the problem. Score level fusion leads to recognition performance that is significantly better compared to abstract and rank level fusion. Not surprisingly, abstract level fusion yields the lowest accuracy, as it is based purely on the decisions without further information.

Our proposed method also performs much better than some reported in the literature including that by Kumar and Zhang [7], which uses PCA to extract features from ear images and gives a recognition rate of 71.6 % on the same database, and [13], which combines PCA with linear discriminant analysis (LDA) and discrete cosine transform (DCT), respectively and yields a classification performance of 63.8 % on the database.

5 Conclusions

In this paper, we have presented an algorithm to identify persons using 2D ear images based on principle component analysis. Crucially, the computational complexity of PCA is addressed by partitioning the images into small blocks and

performing PCA on the subimages separately. We then combine the blocks at feature and classification level, respectively, with the latter leading to the best results and significantly improved performance compared to performing PCA-based recognition based in the whole ear images. In addition to this increased classification accuracy, our approach also significantly reduces the computation time required, hence giving a fast and accurate ear recognition algorithm as demonstrated by a series of experimental results.

References

1. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? *J. ACM* **58**(3), 11 (2011)
2. Carreira-Perpinan, M.: Compression neural networks for feature extraction: Application to human recognition from ear images. MS thesis, Faculty of Informatics, Technical University of Madrid, Spain (1995)
3. Chang, K., Bowyer, K.W., Sarkar, S., Victor, B.: Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(9), 1160–1165 (2003)
4. Golub, G.H., van Loan, C.F.: *Matrix Computations*, vol. 3. JHU Press, Baltimore (2012)
5. Iannarelli, A.V.: *Ear Identification*. Paramont Publishing Company, Fremont (1989)
6. Jain, A., Nandakumar, K., Ross, A.: Score normalization in multimodal biometric systems. *J. Pattern Recogn.* **38**(12), 2270–2285 (2005)
7. Kumar, A., Zhang, D.: Ear authentication using log-Gabor wavelets. In: *Defense and Security Symposium*. pp. 65390A–65390A (2007)
8. Liu, L., Wang, Y., Wang, Q., Tan, T.: Fast principal component analysis using eigenspace merging. In: *IEEE International Conference on Image Processing (ICIP)*, vol. 6, pp. VI-457 (2007)
9. Roweis, S.: EM algorithms for PCA and SPCA. In: *Advances in Neural Information Processing Systems*, vol. 10, pp. 626–632. MIT Press, Cambridge (1998)
10. Schilling, H.A., Harris, S.L.: *Applied Numerical Methods for Engineers Using MATLAB*. Brooks/Cole Publishing Co, Pacific Grove (1999)
11. Shlens, J.: A tutorial on principal component analysis. *arXiv preprint [arXiv:1404.1100](https://arxiv.org/abs/1404.1100)* (2014)
12. Skarbek, W.: Merging subspace models for face recognition. In: Petkov, N., Westenberg, M.A. (eds.) *CAIP 2003*. LNCS, vol. 2756, pp. 606–613. Springer, Heidelberg (2003)
13. Tharwat, A., Hashad, A., Salama, G.: Human ear recognition based on parallel combination of feature extraction methods. *Mediterr. J. Comput. Netw.* **6**(4), 133–137 (2010)
14. Tharwat, A., Ibrahim, A.F., Ali, H.A.: Multimodal biometric authentication algorithm using ear and finger knuckle images. In: *7th International Conference on Computer Engineering & Systems*, pp. 176–179 (2012)
15. Toygar, Ö., Acan, A.: Boosting face recognition speed with a novel divide-and-conquer approach. In: Aykanat, C., Dayar, T., Körpeoğlu, I. (eds.) *ISCIS 2004*. LNCS, vol. 3280, pp. 430–439. Springer, Heidelberg (2004)
16. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)