

EXPRESSION-INDEPENDENT FACE RECOGNITION USING BIOLOGICALLY INSPIRED FEATURES

REZA EBRAHIMPOUR, AHMAD JAHANI, ALI AMIRI AND MASOOM NAZARI

*Brain & Intelligent Systems Research Lab, Department of Electrical and Computer Engineering,
Shahid Rajaee Teacher Training University, P.O.Box:16785-163, Tehran, Iran.
rebrahimpour@ipm.ir, ahd_jhi@yahoo.com, evinar@gmail.com, innocent1364@gmail.com
Brain and Intelligent Systems Research Laboratory
Tehran, Iran
<http://www.bislab.ir>*

Abstract

This paper presents an effective two-dimensional Expression-Independent face recognition method, based on features inspired by the human's visual ventral stream. A feature set is extracted by means of a feed-forward model, which contains illumination and view invariant C2 features from all images in the dataset. Then, these C2 feature vectors which derived from a cortex-like mechanism passed to a standard Nearest Neighbor classifier. We evaluated the proposed approach on JAFEE database. The results show that this model is an efficient and high accurate face recognition algorithm that is robust to facial expressions. Experiments indicate that the proposed approach maintains high recognition rate and outperforms the other alternative methods such as PCA and 2DPCA. The improvement in performance than PCA and 2DPCA based methods is about 5% and 4.5% respectively.

Keywords: HMAX Model; Principal Component Analysis; Expression-Independent Face Recognition.

1. Introduction

Nowadays recognition of faces is one of the active research fields for computer and machine vision researchers. From 1980 up to now, several researchers have evaluated the task of face recognition. The previously face recognition methods used the geometry of key points (like the eyes, nose and mouth) and their geometric relationships (angles, length, ratios, etc). Kanade attempted to implement an automatic face recognition system 30 years ago. After 1975 many researchers have investigated face recognition [1]. In 1991, Turk and Pentland introduced the novel idea of applying principal component analysis (PCA) for face recognition task [2]. Later, algorithms inspired by PCA were proposed by Valentin et al. [3], Ebrahimpour et al. [4] and Samal et al. [5] surveyed the Neural-Network-Based, Mixture of Experts based and the Feature-Based techniques, respectively, Pantic and Rothkrantz [6] surveyed an automatic facial expression analysis, Yang et al. reviewed face detection techniques [7], and Zhang et al. [8], reviewed many of the latest techniques. As we know, face appearance varies with changes of illumination, pose, facial expression, and so on. Therefore, appearance-based face recognition task become a challenging problem with these variations. This paper attempts to propose an Expression-Independent face recognition method which is invariant to variations of facial expression. Bronstien et al. [9] proposed a 3D face recognition which was invariant to the facial expression by introducing the isometric-invariant representation of the facial surface. Li et al. used AAM to recognize individuals with varying face expression. Liu et al. [10] measured two types of the asymmetric facial information, density difference and edge orientation. They showed that this information could obtain individual differences which are stable to the changes of facial expressions. Elad and Kimmel [11] proposed an efficient isometric transformation framework of a non-rigid object on a manifold. This framework overcame the disadvantage of taking a rigid transformation of an existing isometric transformation using multidimensional scaling (MDS). Wang and Ahuja [12] decomposed facial expression features using a higher-order singular value decomposition (HOSVD) technique on the expression subspace to perform face recognition and facial expression recognition at the same time in the subspace. In this paper, we use HMAX model to implement an Expression-Independent face recognition method, using biologically inspired features and Nearest Neighbor classifier. Also, we utilize two dimensional Principal Component Analysis (2DPCA) to reduction the feature vectors dimension. Experimental results show that the rate of recognition using this method is significantly higher than traditional methods such as PCA based methods (e. g. PCA, 2DPCA and so on).

The rest of this paper is organized as follows: Sections 1 and 2 review used feature extraction techniques. Section 3 describes the construction of the proposed Expression-Independent face recognition method. Section 4 shows our experimental results. Finally, in section 5 we conclude and discuss about the future works.

1. Principal Component Analysis (PCA)

Commonly, PCA referred to use eigenfaces. This technique pioneered by Kirby and Sirivich in 1988. An image can be represented as a vector of pixel values (for example suppose a 256×256 pixel grey-scale image; it can be viewed as a vector containing 65536 values). An image vector can be used not only in its original space, but also in many other subspaces in which the image vector is reduced using various mathematical/statistical maps such

as PCA, Linear Discriminant Analysis (LDA) and so on. PCA transforms image vectors into different subspace (also called feature space) and acts as a feature extraction stage, which described in the following. Consider a face image vector containing N pixels, and we have M images for training our system. After vectorizing images, we store all of them into an image matrix, and then we compute the mean matrix as following:

$$X = \begin{bmatrix} X_1^1 & X_2^1 & \dots & X_N^1 \\ X_1^2 & X_2^2 & \dots & X_N^2 \\ \cdot & & & \\ \cdot & & & \\ X_1^M & X_2^M & \dots & X_N^M \end{bmatrix} \tag{1}$$

$$\bar{M} = \frac{1}{M} \begin{bmatrix} X_1^1 + X_2^1 + \dots + X_N^1 \\ X_1^2 + X_2^2 + \dots + X_N^2 \\ \cdot \\ \cdot \\ X_1^M + X_2^M + \dots + X_N^M \end{bmatrix} \tag{2}$$

Then we subtract all training samples from \bar{M} :

$$Y = \begin{bmatrix} X^1 - \bar{M} \\ X^2 - \bar{M} \\ \cdot \\ \cdot \\ X^M - \bar{M} \end{bmatrix} \tag{3}$$

Then we compute the covariance matrix between training image vectors, as following:

$$A = YY^T \tag{4}$$

To find eigenvalues and eigenvectors we have to solve the following equation:

$$AV = \Lambda V \tag{5}$$

Where Λ is vector of eigenvalues of the covariance matrix. Then, we rearrange the eigenvectors according to their corresponding eigenvalues in descending order.

$$V = [V_1, \dots, V_{N-1}] \tag{6}$$

Figure 1 shows a typical eigenvector obtained using PCA.

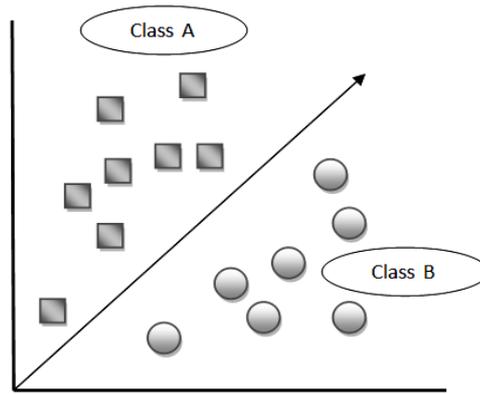


Fig. 1. Effective PCA data separation

2. Two-Dimensional Principal Component Analysis (2DPCA)

In the traditional PCA, two-dimensional face images are transformed into one-dimensional vectors before computing the covariance matrix. It is worth noting that the evaluation of covariance matrix become more difficult due to the big size of training samples. Moreover, computing the eigenvectors of a large covariance matrix is a very time-consuming task. Recently, a new technique called Two-Dimensional Principal Component Analysis (2DPCA) was proposed by Yang in 2004 [1] for face identification. It’s a Principal idea to calculate the covariance matrix, based on two-dimensional original training image matrices. As a result, less time is required to determine the corresponding eigenvectors. Moreover, 2DPCA leads to higher recognition rates than its counterpart (traditional PCA) [23].

2.1. Major Headings Basic Steps of 2DPCA Algorithm [1]

Suppose that there are M face images in the training set. Each image can be denoted by X^i ($m \times n$ matrix, ($i=1, 2...M$)). Let $V \in R^{n \times d}$ be a matrix with orthonormal columns, projecting X into V yields an m by d matrix $Y = X V$.

In 2DPCA, the total scatter of the projected samples is used to determine a good projection matrix V. The following criterion is adopted:

$$J(V) = trace \left\{ E \left[(Y - EY)(Y - EY)^T \right] \right\} \tag{7}$$

If we replace Y with XV, yields:

$$J(V) = trace \left\{ E \left[(XV - E(XV))(XV - E(XV))^T \right] \right\} \tag{8}$$

As we know: $trace(AB) = trace(BA)$
Then we have:

$$J(V) = trace \left\{ V^T E \left[(X - E(X)) (X - E(X)) V \right] \right\} \tag{9}$$

We can define the image covariance matrix as follows:

$$COV = \left[(X - E(X)) (X - E(X))^T \right] \tag{10}$$

Consider M training face images, and then COV is computed by:

$$COV = \frac{1}{M} \sum_{k=1}^M [(X_k - \bar{X})^T (X_k - \bar{X})] \tag{11}$$

Where \bar{X} is the mean of all images.

It has been proved that the optimal value for the projection matrix $V_{optimum}$ is composed by the orthonormal eigenvectors V_1, V_2, \dots, V_d of COV corresponding to the d largest eigenvalues. The size of COV matrix is only n by n . Hence, computing its eigenvectors is very efficient. Also, as well as PCA, the value of d can be controlled by setting a threshold as following:

$$\frac{\sum_{i=1}^d \lambda_i}{\sum_{i=1}^n \lambda_i} \geq \theta \tag{12}$$

Where $\lambda_1, \lambda_2, \dots, \lambda_n$ are the n biggest eigenvalues of COV matrix and θ is a pre-set threshold. It is worth noting that 2DPCA returns a 2D matrix for each image matrix.

Authors are encouraged to have their contribution checked for grammar. Abbreviations are allowed but should be spelt out in full when first used. Integers ten and below are to be spelt out. Italicize foreign language phrases (e.g. Latin, French).

The text is to be typeset in 10 pt roman, single spaced with baselineskip of 13 pt. Text area is 5 inches in width and the height is 8 inches (including running head). Final pagination and insertion of running titles will be done by the publisher. Upon acceptance, authors are required to submit their data source file including postscript files for figures.

3. Proposed Method

In this paper, a new method for Expression-Independent face recognition based on biologically motivated feature extraction is proposed. First, feature vectors of train and test are extracted from face images, using HMAX model (this model is proposed by Poggio and Serre (1996) in Masachossete Inistitute of Technology (MIT) [13]. This model is illustrated in Figure 2.

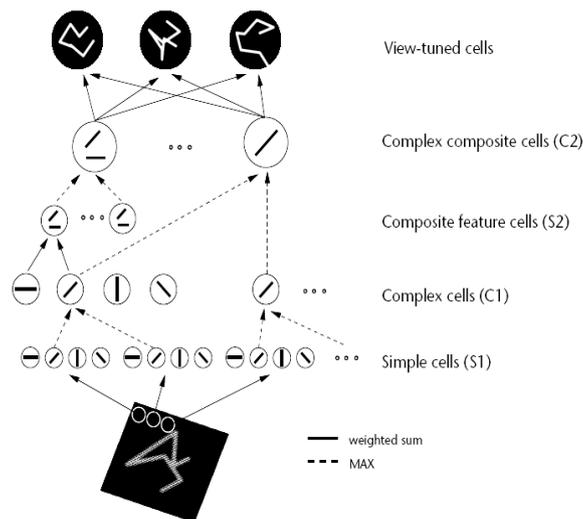


Fig 2. The schematic illustration of HMAX model [14]

The HMAX feature extraction model follows the feed-forward mechanism of object recognition in human and primate's cortex. Biological findings show that the visual process in human and primates is hierarchical. The first aim of this hierarchy is building invariance to position and scale, then to viewpoint and other transformations.

Along the hierarchy, response of receptive fields of the neurons is measured and used for constructing this model. This model consists of four layers of computational units where simple S cells alternate complex C cells. The inputs of the S units are combined with a bell-shaped function to increase selectivity. The C units pool their inputs through a maximum (MAX) operation, and it cause to increasing of invariance. These layers are derived from the experiments in monkey IT cortex (See fig 2).

Finally, derived feature vectors from HMAX, passed to a Nearest Neighbor (NN) classifier. Biologically motivated feature extraction, describes a feature extraction system that derives from a feed-forward model of visual cortex, which gives illumination and view invariant C2 features from all images in the dataset [13]. The sketch of the proposed method is shown in Figure 3.

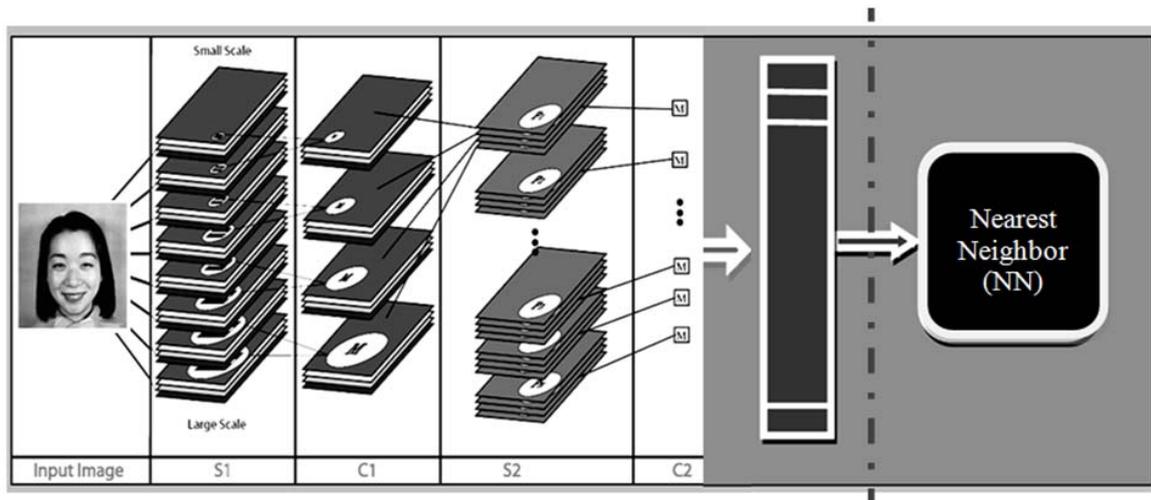


Fig 3. The sketch of the proposed of model

3.1. Feature Extraction

Feature extraction means extracting powerful features in different cases of recognition. The way of feature extraction is very important for implementing an effective and accurate face recognition method. In this paper, we use biologically motivated feature extraction, which is inspired by ventral stream of Human's Visual System (HVS). Recent surveys on ventral stream of visual cortex of human and primates led to different theories about dividing the visual cortex into different levels in which the information is encoded in that levels. Visual processing in ventral stream of cortex is hierarchical and feed-forward process. In the first step, the receptive fields of Simple neurons (S units) combine their inputs with a bell-shaped tuning function to increase the selectivity, and the Complex neurons (C units) pull their inputs with maximum operation, and this is resulted in increasing of invariance to scale and other transformations.

Riesenhuber and Poggio proposed a model that is motivated from a quantitative theory of ventral stream of visual cortex [15]. This system has high performance in complex scenes. In these scenes object recognition must be robust respect to pose, scale, position, rotation and Image condition (Lighting, camera characteristics and resolution). The simplest version of the model consists of four layers that try to summarize a core of well-accepted facts about the ventral stream in the visual cortex.

First, simple S units combine their inputs with a bell-shaped function to increase selectivity, and complex C units pool their inputs through a maximum operation, to increase the invariance. Simple neurons divide in two part, S1 units and S2 units.

3.1.1. S1 Units

S1 units take the form of Gabor functions, which have been shown to provide a good model of cortical simple cell receptive fields [16]. Gabor functions are described by the following equation:

$$F(x, y) = \exp\left(-\frac{(x_0^2 + \gamma^2 y_0^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right) \quad (13)$$

$$x_0 = x \cos \theta + y \sin \theta \text{ and } y_0 = -x \sin \theta + y \cos \theta \quad (14)$$

All filter parameters, *i.e.* the aspect ratio, $\gamma = 0.3$, the orientation θ , the effective width σ , the wavelength λ as well as the filter sizes s were adjusted so that the tuning properties of the corresponding S1 units match to the bulk of V1 parafoveal simple cells [17-19]. Setting the parameters is done using biological experiments on monkey to obtain the tuning properties.

In this experiment, S1 units are 3×3 to 17×17 pixels in steps of 2 pixels. To keep the number of units tractable we considered 4 orientations ($0^\circ, 45^\circ, 90^\circ$ and 135°) thus leading to 32 different S1 receptive field types total (8 scales \times 4 orientations).

3.1.2. C1 Units

These units are more complicated and have larger receptive fields than S1 units. These units have response to bars or edges anywhere into their receptive field. C1 units pool the outputs of S1 units using a MAX operation, which is the response r of a complex unit corresponds to the response of the strongest of its m afferents (x_1, x_2, \dots, x_m) from previous S1 layer such that:

$$r = \max_{j=1 \dots m} x_j \quad (15)$$

3.1.3. S2 Units

In this layer, units pool the outputs of the C1 layer from spatial neighborhood for each orientations, the behavior of S2 units is as Radial Basis Function (RBF) units [13].

The response of each S2 unit depends on Euclidean distance between a new input and prototypes that before stored such that:

$$r = \exp(-\beta \|X - P_i\|^2) \quad (16)$$

Where β is the sharpness value of tuning, P_i is the sorted prototype, and X is the new input image.

3.1.4. C2 Units

After extracting S2 units using HMAX model, C2 units are extracted as shift and scale-invariant units by taking a global maximum (See Eq.1) over all scales and positions for S2 units. HMAX model in S2 layer measures the rate of matching between a stored prototype and an input image to find the best matching and discards the rest. The result is N C2 feature vector that N corresponds to the number of prototypes extracted during learning stage which is described in the following subsection.

3.2. The learning stage

In this stage, N prototypes are selected by a simple sampling. These prototypes are selected randomly at various sizes and positions from a target set of positive images. These prototypes are extracted from C1 layer outputs in four orientations (*i.e.* $n \times n \times 4$ elements). In this experiment, patches of 4 different sizes are extracted ($n=3, 5, 7, 9, 11, 13, 15, 17$).

3.3. The Classification stage

Images are propagated through the architecture described in figure 2. The C2 Standard Model Feature (SMFs) vectors are passed to a Nearest Neighbor (NN) classifier to classify the face images notwithstanding their facial expressions.

4. Experimental Results

To prove our claims, the proposed HMAX based method is implemented on JAFEE database. JAFEE contains 213 images of 7 facial expressions (6 basic facial expressions and one neutral) posed by 10 Japanese female models. Each image has been rated on 6 emotion adjectives by 60 Japanese subjects. As shown in Figure 4, JAFEE images were taken against a homogeneous background with extreme expression variation, and the image size is of 256×256 . For this database, we downsize the images to 100×100 sub-images. To determine the

optimal number of training and testing samples, we set the training and testing ratio as 3:17, 5:15 and 10:10 respectively, and then we evaluate the recognition rate of each arrangement.



Fig 4. Some samples of JAFEE database.

Our method is based on extracting features from face images using Standard Model Features (SMFs) and classifying the test images using Nearest Neighbor classifier. The best rate of recognition, which was the average of 20 times run, for 10 training samples and 10 testing samples was **98.00%**. As a comparison, we also give the results of implementation other methods such as PCA and 2DPCA with Nearest Neighbor classifier. Results and comparisons are shown in table I. Clearly, the Biologically-motivated Expression-Independent face recognition technique achieved the highest recognition rate in these experiments. The difference in performance of this method than PCA and 2DPCA based approaches is about 5% and 4.5% respectively. Therefore, this rate of recognition is significantly higher than other mentioned methods.

Table I. The recognition rates of the proposed method, PCA and 2DPCA methods with different number of training samples.

Number of Samples	Train=3 Test=17	Train=5 Test=15	Train=10 Test=10
PCA	75.00%	79.30%	93.00%
2DPCA	76.90%	82.00%	93.50%
Proposed Method	83.50%	90.70%	98.00%

5. Conclusion and Future Works

We proposed a simple but effective method to make applicable face recognition task in situations where we like to recognize face images regardless their facial expressions. Experiments show that the proposed method is not only feasible, but also better in recognition performance than traditional methods. We compared our method with some of the well known face recognition methods, to demonstrate the superior performance, and experimental results supported our claim that it is better than other common methods of expression-independent face recognition. By applying our method, the Expression-Independent face recognition rate was **98.00%**. Therefore, the performance increased significantly than the mentioned methods.

In future work, we plan to explore more complicated methods for Expression-Independent face recognition, using different combination of feature extractors and classifiers to gain to the best performance this task.

References

- [1] Lee, H.S.; Kim, D. (2008). Expression-invariant face recognition by facial expression transformations. *Pattern Recognition*, **29**, pp. 1797–1805.
- [2] Turk, M.; Pentland, A. (1991). Face recognition using eigenfaces, *Proc. CVPR*, **11**, pp. 586-591.

- [3] Valentin, D.; Abdi, H. (1994). Connectionist models of face processing: a survey. *Pattern Recognition*, **27**, pp. 1209–1230.
- [4] Ebrahimpour, R.; Taheri Makhsoos, N.; Hajiany, A. (2009). Face Recognition by Combining Neural Network based on Mixture of Experts. *Journal of Technology and Education*, **5(2)**, pp. 129-136.
- [5] Samal, A.; Iyengar, P.A. (1992). Automatic recognition and analysis of human faces and facial expressions: a survey. *Pattern Recognition*, **25**, pp. 65–77.
- [6] Pantic, M.; Rothkrantz, L. J. M. (2000). Automatic analysis of facial expressions: the state of the art. *IEEE Transactions. Pattern Anal. Mach. Intell*, **22**, pp. 1424–1445.
- [7] Yang, M. H.; Kriegman, D.; Ahuja, N. (2002). Detecting faces in images: a survey. *IEEE Trans. Pattern Anal. Mach. Intell*, **24**, , pp. 34–58.
- [8] Zhang, D.; Yang, J. (2004). Two-Dimensional PCA; A New Approach to Appearance-Based Face Representation and Recognition. *IEEE Trans on PAMI*, **26**, pp. 131-137.
- [9] Bronstein, A.; Bronstein, M.; Kimmel, R. (2003). Expression-invariant 3D face recognition. In: *Proc. Audio- and Video-Based Biometric Person Authentication*, **2**, pp. 62–70.
- [10] Liu, Y.; Schmidt, K.; Cohn, J.; Mitra, S. (2003). Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding*, **91(1)**, pp. 138–159.
- [11] Elad, A.; Kimmel, R. (2001). On bending invariant signatures for surfaces. *IEEE Trans. Pattern Anal. Machine Intell*, **25(10)**, pp. 1285–1295.
- [12] Wang, H.; Ahuja, N. (2003). Facial expression decomposition. In: *Proc. IEEE Internat. Conf. on Computer Vision*, pp. 958–964.
- [13] Serre, T.; Wolf, L.; Bileschi, S.; Riesenhuber, M.; Poggio, T. (2007). Robust Object Recognition with Cortex-like Mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, **29(3)**, pp. 411-426.
- [14] Riesenhuber, M.; Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neurosis*, **2**, pp. 1019–1025.
- [15] Serre, T.; Kouh, M.; Cadieu, C.; Knoblich, U.; Kreiman, G.; Poggio, T. (2005). A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex. *AI Memo 2005-036/CBCL Memo 259*. Massachusetts Inst. Of Technology, Cambridge.
- [16] Turk, M. A.; Pentland, A.P.; Cognit., J. (1991). Eigen faces for recognition. *Neuroscience*, **3(1)**, pp. 71–86.
- [17] DeValois, R.; Albrecht, D.; Thorell, L. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vis. Res*, **22**, pp. 545–559.
- [18] DeValois, R.; Yund, E.; Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vis. Res*, **22**, pp. 531–544.
- [19] Schiller, P. H.; Finlay, B. L.; Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex III. Spatial frequency. *Neurophysiol*, **39(6)**, pp. 1334–1351.