2009 Ninth International Conference on Intelligent Systems Design and Applications

View-Independent Face Recognition with Biological Features based on Mixture of Experts

Alireza hajiany
Electrical and Computer Engineering
Shahid Rajaee University
Tehran, Iran
alireza.hajiany@gmail.com

Nina Taheri Makhsoos
Electrical and Computer Engineering
Shahid Rajaee University
Tehran, Iran
ninataheri82@ieee.org

Reza Ebrahimpour
Electrical and Computer Engineering
Shahid Rajaee University
Tehran, Iran
ebrahimpour@ipm.ir

Abstract— **The proposed view-independent face recognition model based on Mixture of Expert, ME, uses feature extraction, C1 Standard Model Feature, C1 SMF, motivated from biology on the CMU PIE dataset. The strength of the proposed model is using fewer training data as well as attaining high recognition rate since C1 Standard Model Feature and the combining method based on ME were jointly used.**

Keywords-C1 Standard Model Feature; Mixture of Expert; view-independent face recognition; CMU PIE dataset;

I. INTRODUCTION

Computer face recognition has received tremendous amounts of attention during last decades. A challenging task is to achieve face recognition under the constraint that the face has only been previously observed from different angles. Various models in view-independent face recognition can be categorized into three classes of Multiview, 3D Model and View-Invariant methods. Earlier methods focused on constructing invariant features [1] or synthesizing a prototypical view (frontal view) after a 3D model is extracted from the input image. A recent survey of approaches to 3D face recognition is provided in [2]. Such methods work well for small rotation angles, but they fail when the angle is large, say 60°, causing some important features to be invisible.

Most proposed methods are based on using a number of multiview samples. It seems that, in these methods, the most direct way of recognition is by simply storing a sufficient number of different views associated with each face, and then comparing the unknown image with all these views. Some models of associative memories propose that the huge memory capacity of the brain may be used for such a direct approach to recognition [3,4]. Although useful, especially for the recognition of highly familiar faces, this direct approach by itself is insufficient for recognition in general. The main reason is the problem of generalization, which is, recognizing a face under a novel viewing direction. An example of this multiview approach is the work of Beymer [5], which models faces with templates from 15 views, sampling different poses from the viewing sphere. The recognizer consists of two main stages, a geometrical alignment stage where the input is registered with the model views and a correlation stage for matching. The main limitations of these methods are the need for many different views per person in the database, dependence on lighting variations or facial expressions and the high computational cost, due to iterative searching involved.

In this paper, we propose a neural computational model for view-independent face recognition which is based on ME architecture and feature extraction method is inspired from cortex. The used feature extraction model is a fraction of hierarchical model suggested by Poggio at el [6] which has been extremely successful in object recognition [7]. This hierarchical model comply the standard model of object recognition in primate cortex encompassing several stages so that each stage simulates one part of visual cortex. The standard model is comprised of several computational layers of simple and complex cell units creating a growth in complexity as the layers progress from V1 to inferior temporal cortex [6]. The outputs of the first layer of complex cell unit are employed as feature extraction method.

The remainder of this paper is organized as follows. Section 2 and Section 3 briefly described C1 SMF and ME. It is followed by the description of our proposed model by details in Section 4. Section 5 presents experimental results and comparisons with previously published approach to the same problem. Section 6, finally draws conclusion and summarizes.

II. S1 AND C1 STANDARD FEATURE MODELS

The recent studies in object recognition for primate visual cortex have proved that feedforward path way of object recognition is performed in preliminary processing in the ventral stream [6,8]. Some theories are arisen from aforementioned studies which encompass common facts based on empirical evidence such as hierarchical process, growth in receptive fields of the neurons and feedforward processing [7].

The standard model is a model covering all aforesaid facts for object recognition in visual cortex. This model was introduced by Poggio et. al [6]. The simplest form of this model is proposed in [9] for object recognition. This computational model has hierarchical structure and consists of simple cells S and complex cells C which is alternately used in four layers. It is furthermore invariance to size, scale, position and etc in each layer. Both functions used at S and C layers have biological evidence.

978-0-7695-3872-3/09 $26.00 © 2009 IEEE
DOI 10.1109/ISDA.2009.57

TABLE I. TYPE SUMMARY OF THE S1 AND C1 SMFS PARAMETERS

C_1 layer			S_1 layer		
Scale band $\mathcal{S}$	Spatial pooling grid ($N_S \times N_S$)	Overlap $\Delta_{\mathcal{S}}$	filter size s	Gabor σ	Gabor λ
Band 1	8×8	4	7×7 9×9	2.8 3.6	3.5 4.6
Band 2	10×10	5	11×11 13×13	4.5 5.4	5.6 6.8
Band 3	12×12	6	15×15 17×17	6.3 7.3	7.9 9.1
Band 4	14×14	7	19×19 21×21	8.2 9.2	10.3 11.5
Band 5	16×16	8	23×23 25×25	10.2 11.3	12.7 14.1
Band 6	18×18	9	27×27 29×29	12.3 13.4	15.4 16.8
Band 7	20×20	10	31×31 33×33	14.6 15.8	18.2 19.7
Band 8	22×22	11	35×35 37×37	17.0 18.2	21.2 22.8

At S1 layer, a battery of Gabor filters taking from Hubel and Wiesel classical model is applied to images leading to extract bars, gratings and edges of the images [10]. This algorithm is taken from V1 in primary visual cortex. C1 layer behaviour is similar to Max operation bringing about invariance to scale and size [9]. Gabor functions are employed in S1 units [11]. The used equation is as follows:

$$F(x, y) = \exp\left(-\frac{(x_0^2 + \gamma y_0^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} x_0\right), \qquad (1)$$

$$\begin{aligned} x_0 &= x \cos\theta + y \sin\theta \\ y_0 &= -x \sin\theta + y \cos\theta \end{aligned} \qquad (2)$$

where the orientation θ, aspect ratio $\gamma = 0.3$, effective width σ, wavelength λ and the filter sizes s were adjusted so that the tuning properties of the corresponding S1 units match the bulk of V1 parafoveal simple cells based on data from two groups [12].The outputs of S1 units are inserted into C1 units which have greater receptive fields. In these units, MAX operator is used as equation 3 [12]:

$$r = \max x_j, \quad j = 1, \ldots, m \qquad (3)$$

where r is the greatest response of previous units and m is the number of S units covering for each C1 unit. This operator can be applied for local spatial neighboring with overlapping among its neighbors Δ_S as well as between two adjacent scales. Each pair of two adjacent scales brings about a band. Thus, 8 bands would be created. More details of scales, neighbors and bands are given in Table 1. The response of these units is strictly invariant to scale and position.

III. MIXTURE OF EXPERTS

From a computational point of view, according to the principle of divide and conquer, a complex computational task is solved by dividing it into a number of computationally simple tasks and then combining the solutions to those tasks. In supervised learning, computational simplicity is achieved by distributing the learning task among a number of experts, which in turn divides the input space into a set of subspaces. The combination of experts is said to constitute a combination of classifiers.

Mixture of experts is one the most famous methods in the category of dynamic structures of combining classifiers, in which the input signal is directly involved in actuating the mechanism that integrates the outputs of the individual experts into an overall output [13]. The experts are technically performing supervised learning in that their individual outputs are combined to model the desired response. There is, however, a sense in which the experts are also performing self–organized learning; that is they self–organize to find a good partitioning of the input space so that each expert does well at modeling its own subspace, and as a whole group they model the input space well. The learning algorithm of the mixture structure is described in [14].

For improve the performance of the expert networks devise a modified version of ME in which each expert is an MLP, instead of linear networks [15,16]. In order to match the MLP networks, the learning algorithm is corrected by using an estimation of the posterior probability of the generation of the desired output by each expert. Using this new learning method, the MLP expert networks' weights are updated on the basis of those estimations and this procedure is repeated for the training data set. The learning procedure is described in [17], and is briefly described in the following paragraphs.

Each expert is a one-hidden-layer MLP, that computes an output vector O_i as a function of the input stimuli vector x and a set of parameters such as weights of hidden and output layer and a sigmoid function as the activation function. It is assumed that each expert specializes in a different area of the face space. The gating assigns a weight g_i to each of the experts' outputs, O_i. The gating network determines the g_i as a function of the input vector x and a set of parameters such as weights of the hidden layer, the output layer and a sigmoid function as the activation function. The g_i can be interpreted as estimates of the prior probability that expert *i* can generate the desired output y. The gating network is composed of two layers: the first layer is an MLP network, and the second layer is a softmax nonlinear operator as the gating network's output. The gating network computes O_g, which is the output of the MLP layer of the gating network, then applies softmax function to get:

$$g_i = \frac{\exp(O_{gi})}{\sum_{j=1}^{N} \exp(O_{gi})} \quad i = 1, 2, \ldots, 5 \qquad (4)$$

$$O_T = \sum_i O_i g_i \quad i = 1, 2, \ldots, 5 \qquad (5)$$

The "normalized" exponential transformation of Eq. (4) may be viewed as a multi-input generalization of the logistic function. It preserves the rank order of its input values, and is a differentiable generalization of the "winner-takes-all" operation of picking the maximum value, so referred to as softmax.

The weights of MLPs are learned using the back-propagation, BP, algorithm, in order to maximize the log likelihood of the training data given the parameters. Assuming that the probability density associated with each expert is Gaussian with identity covariance matrix; MLPs obtain the following online learning rules:

$$\Delta w_{y} = \mu_e h_i (y - O_i)(O_i(1-O_i))Oh_i^T \quad (6)$$

$$\Delta w_{h} = \mu_e h_i w_y^T (y - O_i)(O_i(1-O_i)) \; Oh_i(1-Oh_i) \quad (7)$$

$$\Delta w_{yg} = \mu_g (h - g)(O_g(1-O_g))Oh_g^T \quad (8)$$

$$\Delta w_{hg} = \mu_g w_{yg}^T (h - g)(O_g(1-O_g)) \; Oh_g(1-Oh_g)x_i \quad (9)$$

where μ_e and μ_g are learning rates for the experts and the gating network, respectively, Oh_i is the output of expert network's hidden layer, and hi is an estimate of the posterior probability that expert i can generate the desired output y:

$$h_i = \frac{g_i \exp(-\frac{1}{2}(y - O_i)^T (y - O_i))}{\sum_j g_j (-\frac{1}{2}(y - O_i)^T (y - O_i))} \quad (10)$$

This can be thought of as a softmax function computed on the inverse of the sum squared error of each expert's output, smoothed by the gating network's current estimate of the prior probability that the input pattern was drawn from expert i's area of specialization. As the network's learning process progresses, the expert networks "compete" for each input pattern, while the gating network rewards the winner of each competition with stronger error feedback signals. Thus, over time, the gate partitions the face space in response to the expert's performance.

IV. PROPOSED MODEL

A model for view-independent face recognition is proposed in this section. In this model, the standard model feature, inspired from visual cortex, is employed [7]. Result evaluations in [8] proved that in tasks containing limited clutter scene, instead of using C1 and C2 SMFs, using only C1 is sufficient and even lead to obtain better result. Since images in the same view of the face have restricted clutter, using C1 SMF is better suited for face recognition. Since the acquired information, resulted in processing our visual environment, in cortex is tremendous, and as a result many processes are required to recognise a typical object, this huge obtained data cannot be directly implemented and inserted into artificial systems. Therefore, PCA is employed to reduce data dimension [18].

According to the structure of the proposed model shown in Fig. 1, the ME is employed. The ME networks are not biased to prefer one class of faces to another; in that, the network itself partitions the face space into subspaces and decides which subspace should be learned by which expert [16].

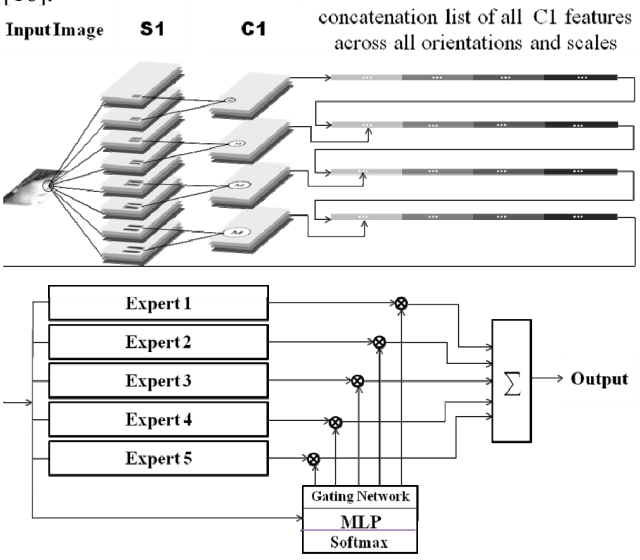

Figure 1. Structure of proposed model.

V. EXPERIMENTAL RESULTS AND COMPARISONS WITH PREVIOUSLY PUBLISHED APPROACH

The dataset used for the structure of the proposed model includes 180 images of 20 people taken from PIE dataset [19]. Fig. 2 shows the example of training and testing PIE dataset. All people have 9 images on the views of angels ($\pm 90^\circ, \pm 67.5^\circ, \pm 45^\circ, \pm 22.5^\circ, 0^\circ$). Views of angles ($\pm 90^\circ, \pm 45^\circ, 0^\circ$) are used for training purpose and ($\pm 67.5^\circ, \pm 22.5^\circ$) for test. The size of each image employed in this dataset is 640×486 which the face region is cropped and then the image size is changed to 48×48.

To extract features, all filters listed in Table 1 are first applied on the images. In that, 64 different filters in 16

variant sizes, which each size contains four orientations, are applied on each image. Consequently, two types of maximizing are performed on the images obtained from previous step. The first one is to take a max over the outputs of two sizes of adjacent filters on the basis of pixel-by-pixel comparison. As a result, each size pair brings about one band. The size of each band is given in Table 1. Therefore, 8 bands along with four orientations ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) are achieved. In the next step, the max operation is taken over each pixel with the neighborhood size of Δ_S pixels. The output obtained from the last step is sampled in steps of Δ_S pixels.

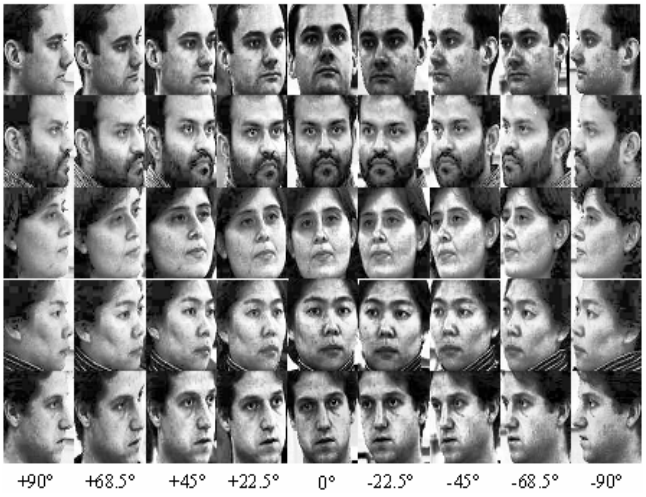

Figure 2. Examples of face images, taken from the CMUPIE dataset, used to train and test our proposed model.

C1 SMF is formed as linear and consecutive vectors to pass to classifiers. The dimension of the final vector is required to diminish since this vector has 1916 length as well as high computational load. Therefore, PCA is used to meet this requirement. 30 valuable components are selected to use for the experts. Eigenspaces in gating network are taken using C1 SMFs and amount of selected components are the same as before.

As the number of hidden neurons changes from 15 to 35 in steps of 5 neurons and the number of components alter from 10 to 40 in steps of 5 components, 30 components are chosen on the basis of taking average over 10 run times. The result is given in Table 2 with respect to number of hidden neurons and variant component for experts. For other structure's parameters, η_g and η_e are chosen 0.5 and 0.1 respectively.

Some experiments were carried out to compare the result with related works proposed for view-independent face recognition. In compared works, dataset used for test and training are similar to the dataset employed for the proposed model. Finally, we would like to compare performance of the proposed model with two of the most related works in the literature for view independent face recognition, [20] and [21]. These models were implemented and tested on our dataset under the same condition as our experiment with the intermediate unseen views. The results are tabulated in Table 3. It can be observed that the proposed model gives better performance than the view-based eigenspaces method of [20] and the fusion method of [21]. The comparison result is presented in Table 3. As it is seen in Table 3, the best recognition rate using overlap eigenspace with TDL is 91.58% [22], whereas the proposed model has attained the recognition rate of 92.5%.

TABLE II. RECOGNITION RATES OF DIFFERENT TOPOLOGIES OF THE PROPOSED MODEL EXPERTS. EACH RESULT IS THE AVERAGE OF TEN TIMES TESTING THE PROPOSED MODEL, EACH TIME TRAINED WITH DIFFERENT RANDOM INITIAL WEIGHTS

	Number of Hidden Layer				
	15	20	25	30	35
Number of Components 10	45	68.5	74	78.5	80.5
15	49.5	76	83.7	83.7	81.7
20	55.2	83.7	88.2	90	90.5
25	56.7	83.7	90	90.5	88.2
30	58	87	**92.5**	90.2	90
35	58.7	87.2	90	91.5	91.5
40	61.2	87	90	88.2	90.2

In all related works [17,22,23,24,25], to train the network, 15 artificial samples were generated by altering contrast, brightness as well as by making samples blur owing to shortage of training data samples in CMU PIE dataset. In contrast, in the present work, there is no need to create these artificial samples, and the network can be trained with a high-performance using fewer samples even one sample for each angle. Therefore, the advantage of using fewer samples even one sample as well as the high recognition rate benefiting over related works can be named as two capabilities of the proposed model.

TABLE III. THE COMPARISON BETWEEN THE PROPOSED METHOD AND THE MOST RELATED WORKS IN THE LITERATURE IMPLEMENTED AND TESTED ON OUR DATASET.

Method	Recognition Rate
Proposed Model	92.5%
Overlapping Eigenspaces with TDL [22]	91.58%
Fusion of pose-invariant face-identification experts [21]	85.63%
View-based Eigenspaces [20]	79.44%
Single-view Eigenspaces [24]	80.51%
Global Eigenspace [23]	77.14%
Global Eigenspace with TDL [25]	84.62%
Overlapping Eigenspace [22]	81.04%

VI. CONCLUSION

A model for view-independent face recognition, extraction based on ME, was presented. C1 SMF feature extraction is applied on CMU PIE dataset. Experiments were carried out with a training set on the views of angles ($0°, \pm 45°, \pm 90°$) and a test on the intermediate views of angles ($\pm 22.5°, \pm 67.5°$). Results have shown that embedding the feature extraction method inspired from biology is beneficial way for this task. The strength of the proposed model includes using one sample for each view so that there is no need to generate extra training data, making the convergence of the ME network faster and attaining the higher recognition rate in comparison with other related works for view-independent face recognition.

REFERENCES

[1] Wiskott, L., Fellous, J.M., and von der Malsburg, C., "Face Recognition by Elastic Bunch Graph Matching", IEEE Trans. Patt. Anal. Mach. Intell, 1997, vol.19, pp. 775–779.

[2] Bowyer, K., Chang, P., and Flynn, A., "Survey of Approaches to Three-Dimensional Face Recognition", In: Proc. of the IEEE International Conference on Pattern Recognition, 2004, pp. 358–361.

[3] Hopfield, J.J., "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", In Proceedings of the National Academy of Sciences, USA, 1982, pp. 2554-2558.

[4] Kohonen, T., Associative Memories: A System Theoretic Approach, Springer, Berlin, 1978.

[5] Beymer, D.J., "Face Recognition under Varying Pose", Technical Report 1461, MIT AI Lab, Massachusetts Institute of Technology, Cambridge, MA, 1993.

[6] Riesenhuber, M., Poggio, T, "Hierarchical models of object recognition in cortex", Nature Neuroscience, vol.2, no.11, 1999, pp.1019–1025.

[7] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, Tomaso Poggio, "Robust Object Recognition with Cortex-Like Mechanisms", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.29, no.3, 2007.

[8] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, T. Poggio, 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, 2005.

[9] T. Serre, Learning a Dictionary of Shape-Components in Visual Cortex: Comparison with Neurons, Humans, and Machines, PhD dissertation, Massachusetts Inst. of Technology, Cambridge, Apr. 2006.

[10] D.H. Hubel and T.N. Wiesel, "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", J. Physiology, vol.160, 1962, pp. 106-154.

[11] D. Gabor, "Theory of Communication", J. IEE, vol.93, 1946, pp. 429-459.

[12] T. Serre, M. Riesenhuber, "Realistic Modeling of Simple and Complex Cell Tuning in the HMAX Model, and Implications for Invariant Object Recognition in Cortex", CBCLPaper 239/AIMemo 2004-017, Massachusetts Inst. of Technology, Cambridge, 2004.

[13] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, USA, 1999.

[14] Jacobs, R., Jordan, M., Nowlan, S., and Hinton, G., "Adaptive Mixtures of Local Experts", Neural Comput3, 1991, pp. 79–87.

[15] R. Ebrahimpour, E. Kabir and M.R. Yousefi, "Face Detection Using Mixture of MLP Experts", Neural Processing Letters, 2007, pp. 69-82.

[16] N. Taheri M., A. Hajiany, R. Ebrahimpour G. Sepidnam, "A Modified Mixture of MLP Experts for face recognition", in Proceedings of Image Processing, Computer Vision, & Pattern Recognition, IPCV08, Las Vegas, Nevada, USA, 2008, pp. 740-745.

[17] R. Ebrahimpour, E. kabir, M. Yousefi, "Teacher-directed learning in view-independent face recognition with mixture of experts using single-view eigenspaces", Franklin Institute, 345, 2008, pp. 87–101.

[18] Sirovich L, Kirby M, "Low-dimensional procedure for characterization of human faces", J Opt Soc Am, vol.4, 1987, pp. 519-524.

[19] Sim T, Baker S, Bsat M, "The CMU Pose, Illumination, and Expression Database", IEEE Trans Pattern Anal Mach Intell, vol.25, no.12, 2003, pp. 1615-1618.

[20] Moghaddam B, Pentland A, "Probabilistic Visual Learning for Object Representation", IEEE Trans on Pattern Analysis and Machine Intelligence, vol.19, no.7, 1997, pp. 696-710.

[21] Kim T, Kittler J, "Combining Classifier for Face Identification at Unknown Views with a Single Model Image", IEEE Trans Circuits and Systems for Video Technology, vol.16, no.9, pp.1096-1106.

[22] R. Ebrahimpour, E. kabir, M. Yousefi, "Teacher-directed learning in view-independent face recognition with mixture of experts using overlapping eigenspaces", Computer Vision and Image Understanding, vol.111, 2008, pp. 195-206.

[23] R. Ebrahimpour, E. kabir, M. Yousefi, "View-independent Face Recognition with Mixture of Experts", Neuro computing, vol.71, 2008, pp. 1103-1107.

[24] R. Ebrahimpour, E. Kabir, M.R. Yousefi, "View-based eigenspaces with mixture of experts for view-independent face recognition", Lecture Notes Comput. Sci, vol.4472, 2007, pp. 131-140.

[25] R. Ebrahimpour, E. Kabir, M.R. Yousefi, "Teacher-directed learning with mixture of experts for view-independent face recognition", Lect. Notes Comput. Sci., vol.4362, 2007, pp. 601–611.