ACSIJ Advances in Computer Science: an International Journal, Vol. 5, Issue 1, No.19 , January 2016
ISSN : 2322-5157
www.ACSIJ.org

ACSIJ
WWW.ACSIJ.ORG

C_3 Effective features inspired from Ventral and dorsal stream of visual cortex for view independent face recognition

Somayeh Saraf Esmaili [1], Keivan Maghooli [2] and Ali Motie Nasrabadi [3]

[1] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
s.saraf@srbiau.ac.ir

[2] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.
k_maghooli@srbiau.ac.ir

[3] Department of Biomedical Engineering, Faculty of Engineering, Shahed University, Tehran, Iran.
nasrabadi@shahed.ac.ir

Abstract

This paper presents a model for view independent recognition using features biologically inspired from dorsal and ventral stream of visual cortex. The presented model is based on the C_3 features inspired from Ventral stream and Itti's visual attention model inspired from the dorsal stream of visual cortex. The C_3 features, which are based on the higher layer of the HMAX of the ventral stream of visual cortex, are modified to extract important features from faces in various viewpoints. By Itti's visual attention model, visual attention points are detected from faces in various views of faces. Effective features are extracted from these visual attention points and the view independent C_3 effective features (C_3EFs) are created from faces. These C_3EFs are used to distinct multi-classes of different subjects in various views of faces. The presented model is tested using FERET face datasets with faces in various views of faces, and compared with C_2 features (C_2SMFs) and C_3 features of standard HMAX model (C_3SMFs). The results illustrated that our presented view independent face recognition model has high accuracy and speed in comparison with standard model features, and can recognize faces in various views by 97% accuracy.

Keywords: *view independent face recognition, HMAX model, Itti's visual attention model, C_3 effective features, ventral stream, dorsal stream.*

1. Introduction

In recent years, providing computational algorithms for face recognition in these different situations especially in various views of faces have been the fundamental issues [1-3]. Computational view independent face recognition is one challenging work that human visual system can do it with high-speed performance, easily. Thus based on need, recently the study of brain mechanisms of the human visual system have been more considered.

The visual system is organized in two functionally specialized processing pathways in the visual cortex. One pathway (from the primary visual cortex to the parietal cortex for controlling eye movements and visual attention) is named dorsal stream, and other pathway (from the primary visual cortex towards the inferior temporal lobe including V_1, V_2, V_4, Posterior Infer temporal (PIT), Anterior Infer temporal (AIT) and Fusiform Face Area (FFA)) is named ventral stream, which processes detail of objects and faces in different conditions [4-6]. The dorsal and ventral streams are not completely independent and there are interactions. For example, area V_4 is interconnected with some visual attention areas in dorsal stream [7-9].

Partial analogies of the ventral stream cortex have been used in many computing models in canonical computer vision. Among these models, HMAX is one powerful computational model that models the object recognition mechanism of human ventral visual stream in visual cortex [10-13], that first proposed by Poggio et al., is based on experimental results in neurobiology. The extracted bio-inspired C_2 features of this model can recognize and classify objects based on the mechanism of human ventral visual stream in visual cortex. Then Serre et al. established HMAX model and considered learning ability of the model for real-world object recognition in [13]. The features extracted by this model are C_2 standard model features (SMFs). The C_2SMFs of the ventral stream have been used for multi-class face recognition [14-16]. In recent years, different development models of the ventral stream HMAX model have been presented to enhance the efficiency of the model and in all these models, some feature extraction methods are considered [17-20]. Leibo et al. in [20], extended HMAX model and added new S_3 and C_3 layers. They found that the performance of the model on view independent within-category identification tasks on different objects was increased and the C_3 features of the extended HMAX model performed significantly

better than C_2 and it was independent to viewpoints. Also, there is some visual attention model which based on visual attention regions in the dorsal stream of the visual cortex and these models were applied in many applications such as target detection, object recognition, object segmentation and robotic localization [21-24]. These visual attention models use low-level visual features such as colour, intensity and orientation to form saliency maps and find focus of attention locations. The basic computational model of visual attention was proposed by Itti in 1998 which are the basic model for the most of the new models and used bottom-up visual cortex features in the where path [22].

In this paper, a model based on new C_3EFs inspired from Ventral and dorsal stream of visual cortex for the goal of view independent face recognition are presented. For experimental analysis, the FERET faces datasets in various views of faces are utilized and view independent face recognition task by the SVM classifiers on the new C_3EFs of them are done and they are compared with the results of other extracted features.

The rest of the paper is organized as follows. Section 2 illustrates the C_2SMFs and C_3SMFs features of ventral stream model and, also the visual attention model of dorsal stream for detecting important facial regions. Our proposed view independent face recognition model is presented in section 3. Section 4 is presented experimental evaluation including dataset description, results and detailed discussions. In the final section, we sum up with a conclusion.

2. Material and methods

2.1 The C_2SMFs features of ventral stream model

The C_2SMFs features of HMAX model are inspired by ventral stream of visual cortex, and it created by four layers (S_1, C_1, S_2, C_2) [12, 13]. S_1 features resemble the simple cells found in the V_1 area of the primate visual cortex and consists of Gabor filters. C_1 features imitate the complex cells in V_1&V_2 area of cortex and have the same number of feature types (orientations) as S_1. These features pools nearby S_1 features (of the same orientation) to reach the position and scale invariance over larger local regions, and as a result can also subsample S_1 to reduce the number of features. The values of C_1 features are the value of the maximum S_1 features (of that orientation) that comes within a max filter [13]. S_1 features imitate the visual area V_4 and posterior infer temporal (PIT) cortex. They contain RBF-like units which tuned to object-parts and compute a function of the distance between the input C_1 patches and the stored prototypes. In human visual system, these patches correspond to learning patterns of previously seen visual images and store in the synaptic weights of the neural cells. The S_2 features learn from the training set of K patches ($P = 1,..., K$) with various $n \times n$ sizes ($n \times n$ = 4×4, 8×8, 12×12 and 16×16) and all four orientations at random positions (Thus a patch P of size $n \times n$ contains $n \times n \times 4$ elements). Then S_2 features, acting as Gaussian RBF-units, compute the similarity scores (i.e., Euclidean distance) between an input pattern X and the stored prototype P : $f(X) = \exp(-\|X - P\|^2 / 2\sigma^2)$, with σ chosen proportional to patch size. The C_2 features imitate the inferotemporal cortex (IT) and perform a max operation over the whole visual field and provide the intermediate encoding of the stimulus. Thus, for each face image, the C_2 features vector is computed and used for face recognition. This vector has robustness properties. The lengths of C_2 features vector are equal to the number of random patches extracted from the images and have the property of shift and scale independent.

2.2 The C_3SMFs features of ventral stream model

In the implementation of the S_3 and C_3 layers from developed HMAX model which was proposed by Leibo et al. in [20], the response of a C_2 cell (associating templates w at each position t) was given by Eq. (1):

$$S_3 = \exp\left(-\frac{1}{2\sigma}\sum_{j=1}^{n}(w_{t,j} - x_j)^2\right) \qquad (1)$$

Then, the S_3 features corresponding to all layers were extracted and the maximum of these values were utilized as (Eq. (2)). The C_3 features imitate the view independent properties in the FFA of the IT [20]. These original features are named C_3SMFs in this paper.

$$C_3 = \max(S_3^{\,i}) \qquad (2)$$

In this paper, we used the C_2SMFs and C_3SMFs features vector for view independent face recognition and present a model which can be extracted effectively C_2 and C_3 features from face image in different viewpoints.

2.3 Visual attention model in dorsal stream of visual cortex

In face images, some facial regions are more attentive and helpful regions to face recognition, such as eyes, nose and mouth that have been demonstrated by the results of psychophysical studies in paper [25]. Human can find distinctive information from face images in a short time and with high accuracy. These detected features have so much local information to recognize the similarity of face images and can track and match to the similar face images. So, visual attention models inspired from human visual

systems in visual attention regions of dorsal stream can detect salient points from face images. Visual attention Itti's model biologically models the visual attention regions in the posterior cortex and specifies the locations of salient points from a colour image simulating saccadic eye movements of human vision [22]. We utilized this model in our proposed model to find automatically important regions of the face by adjusting its parameters.

3. Proposed view independent face recognition model

In the ventral stream HMAX model, the S_2 features learned from the randomly extracted patches. So, maybe some of the extracted special features from cropped face images such as extracted features from the forehead and cheek regions are not useful features. Since, these features caused CPU usage in the system and make the system very slow achieving the effective features vector. Then, it seems necessary to detect best features. Also, HMAX model only models the ventral stream and the connections between the visual attention regions in the posterior cortex to the ventral stream are not considered. In the proposed model, in order to model the human-like face recognition system, we extract the C_2 and C_3 features from visual attention points and achieve the C_2EFs and C_3EFs for view independent face recognition.

Generally, by combining the hierarchical ventral stream features with visual attention model, the feature extraction model for face recognition system is proposed as follows (Fig. 1 illustrates different visual cortex layers in two dorsal and ventral streams, and also the whole structure of our proposed feature extraction model for view independent face recognition inspired from them):

1) For each face image, the colour features, intensity features and orientation features are extracted (inspired from the regions in the primary visual cortex, which showed by red dashed line ▬ in Fig. 1) as represented in Itti's Visual attention model [22].

$$R = r - \frac{g+b}{2}$$
$$G = r - \frac{r+b}{2}$$
$$B = b - \frac{r+g}{2}$$
$$Y = \frac{g+b}{2} - b - \frac{|r-g|}{2}$$
$$I = \frac{R+G+B}{3} \quad (3)$$

2) colour saliency map, intensity saliency map and orientation saliency map are founded as Eq. (4)-(9). $M_f(s)$ represents of F feature map in s scale. F included I intensity features, C colour features. A function of $N(0)$ is used for created normalization map where the symbol Θ represents interpolation of the coarser image to the finer scale and point by point subtraction. In this system, $c = \{2,3\}$ and $s = c + d$, where $d = \{2,3\}$.

$$F_{I,c,s} = N(|M_I(c) \Theta M_I(s)|)$$
$$F_{C,c,s} = N(|M_C(c) \Theta M_C(s)|)$$
$$= N(|M_{RG}(c) \Theta M_{RG}(c)| + |M_{BY}(s) \Theta M_{BY}(s)|) \quad (4)$$

I and C are the saliency maps of intensity and colour, respectively (in Eq. (5)).

$$I_{c,s} = \sum_{c=2}^{3} \sum_{s=c+2}^{3} N(|M_I(c) \Theta M_I(s)|)$$
$$C_{c,s} = \sum_{c=2}^{3} \sum_{s=c+2}^{3} N(|M_{RG}(c) \Theta M_{BY}(s)|) \quad (5)$$

Also, the Gabor filters which generated in S_1 are used as orientation features where $O(c,s,\theta)$ denotes the orientation saliency map of θ by c and s operation of scales (Eq. (6)).

$$O(c,s,\theta) = |O(c,\theta) \Theta O(s,\theta)|$$
$$= \sum_{\theta=\{0^\circ,45^\circ,90^\circ,135^\circ\}} \sum_{c=2}^{3} \sum_{s=c+2}^{3} N(O(c,s,\theta)) \quad (6)$$

Then the features are combined by Eq. (7) to create salient points (SP) [22, 24].

$$SP = (C + I + O)/3 \quad (7)$$

Then a winner take all network (WTN) is used to detect N salient points and N attention points (inspired from the regions in the dorsal stream of the visual cortex, which showed by blue dotted line ▬ ▪ ▬ in Fig. 1).

3) Create S_1 and C_1 features from each face image (inspired from the regions in the primary visual cortex, which showed by red dashed line ▬ in Fig. 1).

4) Extract N patches p_i $(i = 1,..., N)$ in four orientations and $n \times n$ best patch sizes from C_1 features of each face image by using detected attention points as the central pixel of them to create effective S_2 features. During recognition, from each test face image, N patches are created as X_i $(i = 1,..., N)$ patches and the distance between the patches p_i and X_i are calculated according to the Eq. (8).

$$V_k = \exp\left(\frac{-\left\| X_i - P_i \right\|^2}{2\sigma^2}\right) \quad k = 1,2,..., N \quad (8)$$

Which σ is proportional to the patch size and N dimensional vectors create for each face image. The set of V_k ($k = 1,2,\ldots,N$) forms S_2EFs (inspired from the V4 and

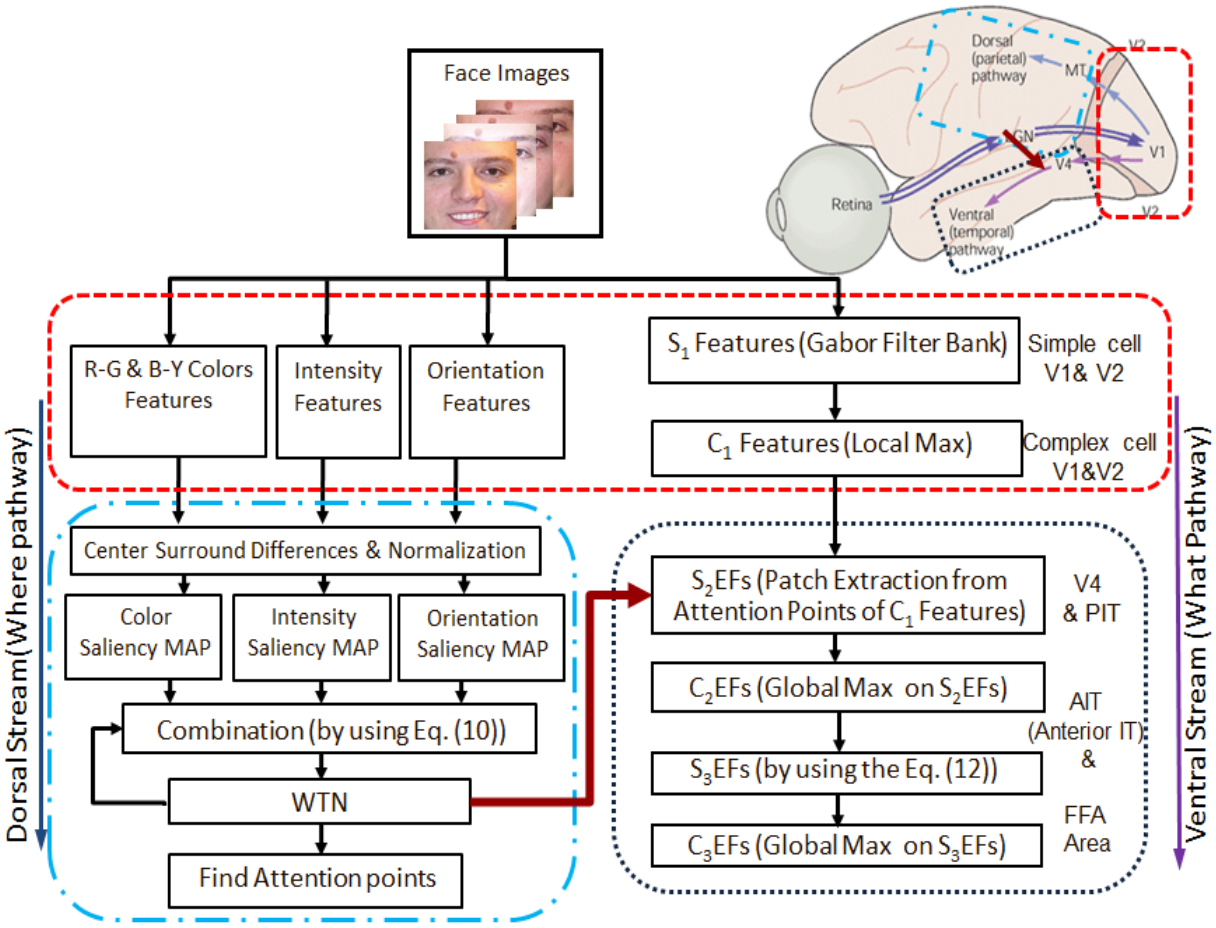

Fig 1 The structure of proposed view independent face recognition model.

PIT area in the ventral stream of the visual cortex, which showed by dark blue dotted line ▪▪▪, also the interaction between the V_4 area in ventral stream by the dorsal stream of the visual cortex has been showed by red arrows → in Fig. 1).

5) Obtain C_2EFs by general maximum on S_2EFs to create N -dimensional C_2EFs vector from distinctive regions of faces (inspired from AIT area in ventral stream visual cortex which showed by dark blue dotted dot ▪▪▪ in Fig. 1).

6) Create S_3EFs Features on C_2EFs (the response of a C_2EFs cell) by using the Eq. (9) to extract S_3EFs corresponding to all layers (inspired from the FFA area in the ventral stream visual cortex which showed by dark blue dotted dot ▪▪▪ in Fig. 1). During recognition, from each test face image, C_2EFs of them are created as x_i $(i = 1,..., N)$ responses and the distance between these responses and associating templates w at each position t are calculated according to the Eq. (9).

$$S_3 EFs = \exp(-\frac{1}{2\sigma}\sum_{j=1}^{n}(w_{t,j} - x_j)^2) \qquad (9)$$

7) Obtain C_3 EFs by using the maximum on the S_3EFs as following (Eq. (10)) (inspired from the view independent regions in the FFA area of the ventral stream visual cortex which showed by dark blue dotted dot - - - in Fig. 1).

$$C_3 EFs = \max(S_3^{\,i} EFs) \qquad (10)$$

8) Do step 1 to 6 for all images to extract C_3EFs vector from distinctive regions of faces.

9) Feed C_3EFs and C_2EFs vectors to SVM and classify face images with the goal of view independent face recognition.

4. Experimental Analysis

4.1 Image Dataset

To demonstrate the feasibility of our proposed face recognition system, experiments on the subset of the colour FERET database are organized [26]. The subset contains ten classes (unique subjects) of face images, with variations in pose, expressions and scales. It consist 200 face images (10 classes, each with 20 images). The proposed model is tested using a 10-fold cross validation strategy. So that, for each fold 18 images of each individual (180 face images) are selected as training samples and the rest 2 images of each individual as test images (20 face images). In the pre-processing method, all images are cropped manually to remove complex background and then they are resized to the dimensions of 140×140. Fig. 2 shows a series of 10 different classes of people from FERET database which used in this work. All experiments are performed entirely in a 32 Matlab2012 experimental environment (characteristic of a computer system is Intel core 2 duo processor (2.66 GHz) and 4 GB RAM). Fig. 3 shows twenty samples of images from one subject in varied view points and expression.

Fig. 2 Images from ten classes of cropped FERET dataset.

Fig. 3 Images of one subject in varied viewpoints and expressions.

4.2 Results and discussion

In proposing a view independent face recognition model, to achieve the effective features vector, at first the saliency maps of colour, intensity and orientations are extracted and the feature map's weights of them are adjusted to create the saliency maps from face images and select the attended locations of salient points as important regions of them. Fig. 4, shows the original images of faces with seventy attention points and saliency maps on them. Fig. 4(a-c) and also Fig. 4(d-f) shows the original face images of two individuals in three situations of viewpoints. In the saliency toolbox, local max is selected for normalization. As shown in Fig. 4, by visual attention model, important and important regions of faces such as nose, eyes, lip and mole are selected as attention points however the faces are in varied viewpoints.

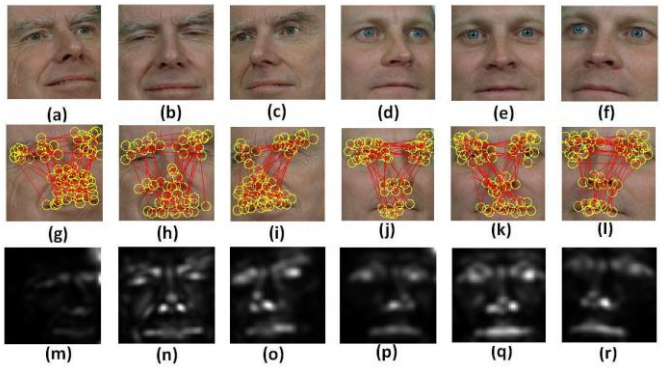

Fig. 4 (a) To (f) are original images. (g) To (l) are face images with selected sixty attention points. (m) To (k) are saliency maps of face images.

After finding attention points from face images, it is necessary to convert colour images to grey scale for extracting C_2 and C_3 features. The biological S_1 features of face images are created by convolving with 64 Gabor filters. So, there are 64 S_1 features for each face image. Fig. 5 shows these S_1 features for one original grey-scale face image after convolving with 64 Gabor filter. Also, Fig. 6 shows the C_1 features in band 1 and in four orientations ($\theta = 0^\circ$, 45°, 90° and 135°). For each orientation in band 1, there are two 7 × 7 and 9 × 9 filters. At first, Maximum responses to them are calculated for each pixel with 8 × 8 grid sizes. Then, the C_1 features are created by maximum on corresponding pixels from two images in each orientation.

The C_2EFs and C_3EFs of the images are extracted using the proposed view independent face recognition model that presented in last section. Then, SVM classifiers with RBF kernel are trained using these C_2 features vectors and the class labels, implemented using LIBSVM [27]. In this approach, an SVM is constructed for each class by

discriminating that class against the remaining 9 classes. The number of SVMs used in this approach is 10. In the testing phase, the features are extracted using proposed method and the classification is done on test data using the test SVM classifiers.

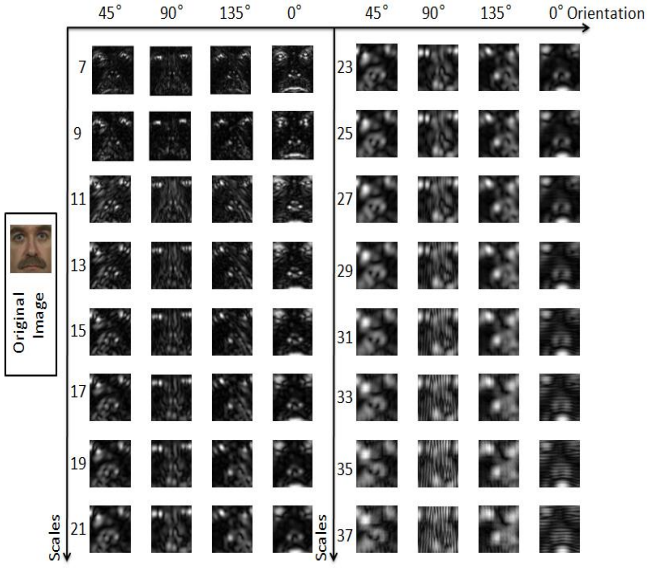

Fig. 5 S_1 features created by 64 Gabor filters.

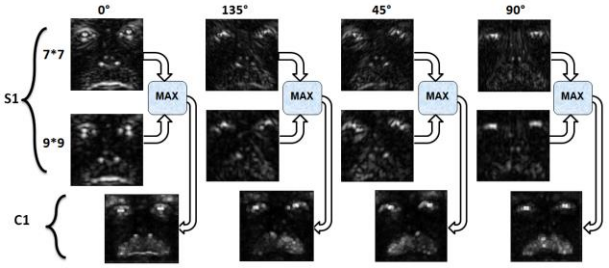

Fig. 6 C_1 features created in band 1 and in four orientations.

In order to find proper sizes of patches for extracting features, the representation of dissimilarity matrices (RDMs) are created. For obtaining this goal, we extract proper prototype patches with different patch sizes (4×4, 8×8, 12×12 and 16×16) and attain the best biological features. These features create the RDMs which have view independent identity specific between features of eight viewpoints of ten subjects. Scheme to extract hierarchical features from 10 subjects in 8 viewpoints and create RDMs is shown in Fig. 7, and Fig. 8 shows a comparison of the created RDMs in different patch sizes of C_3EFs in equal extracted features (N=40). The best RDMs can be created with the extraction prototype patches in 12×12 patch size from each of the cropped face images (Fig. 8(c)). It shows created RDMs on C_3EFs features in 12×12 patch size have high correlation in 15 diagonals parallel to the main diagonal that represent the view-independent specific of C_3EFs in 12×12 patch size in comparison with others patch sizes. As in other patch sizes (Figure 8(a), Figure 8(b) and Figure 8(d)), some pixels of these diagonals are not observable on the same background. In Figure 8 each pixel in vertical and horizontal of RDMs represents one subject in one viewpoint.

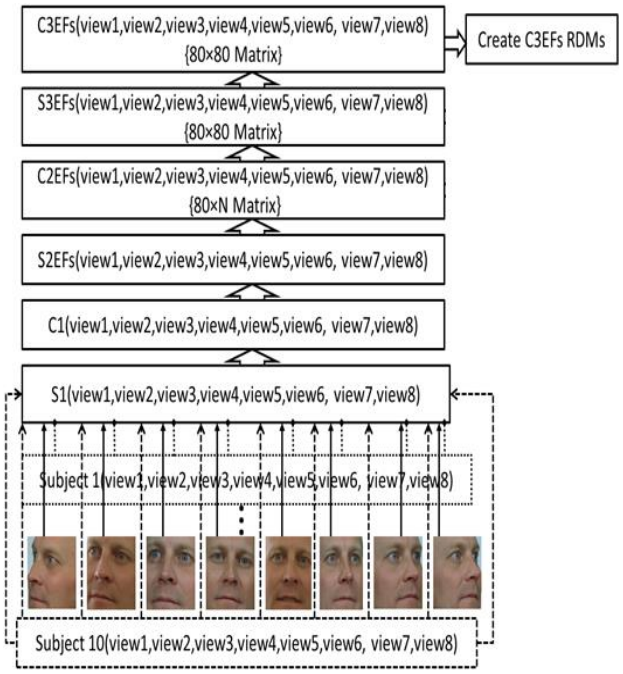

Fig. 7 Scheme to extract hierarchical features and create RDMs.

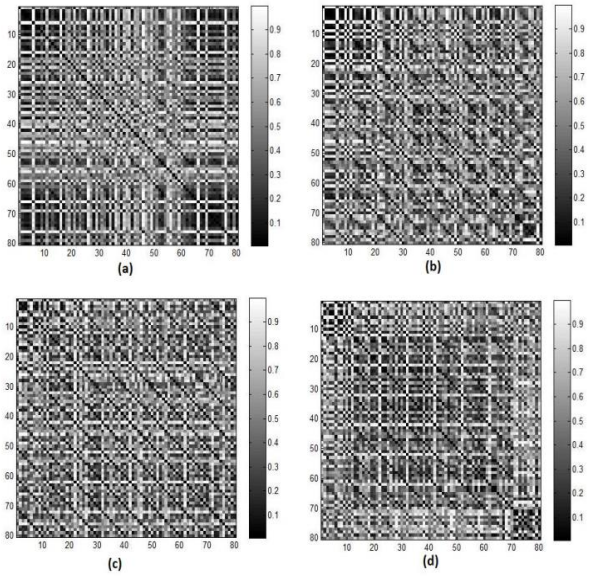

Fig. 8 RDMs on C_3EFs of 80 face images (8 viewpoints of 10 subjects) in (a) 4×4 patch size (b) 8×8 patch size (c) 12×12 patch size (d) 16×16 patch size.

Also, in order to demonstrate feasibility of the proposed view independent face recognition model, we compared the performance of the C_2EFs vector and C_3EFs vector (with selected attention points) of the proposed model in best patch size with the performances of the C_2SMFs and C_3SMFs (without attention points). So, *N* patches of C_2SMFs, C_3SMFs, C_2EFs and C_3EFs in best patch size from the same face of test and train images are extracted and fed into SVM classifiers to compare the recognition rate of them by ten-fold cross validation. Fig. 9 shows the accuracy rate of face recognition using SVM classifiers on the C_2SMFs, C_2EFs and the C_3EFs of proposed model in various numbers of extracting features. Given the results in Table 1 and Fig. 9, the recognition rates of C_3EFs and C_2EFs are better than C_2SMFs and C_2SMFs. Also, the results in table 1 show that 80 number (*N*=80) of the C_3EFs are enough to get good performance (around 97%) against; 300 number of the C_3SMFs are required to get this good face recognition rate. So, face recognition by C_2SMFs and C_3SMFs needs more extracted features and more extracting time.

The extracting time is the average computing time of each face image for extracting features that obtained by tic-toc function in Matlab2012 software. For example, the extracting time of C_3EFs is total computing time for selecting attention points and extracting C_3 features from them. Given by the results in the table 1, the extracting time to enhance the good recognition accuracy near 97% by using C_3SMFs (300 number of features) are more than others. So, by using C_2EFs and C_3EFs, the appropriate view independent face recognition rate in lesser time is achieved.

From these results, the C_2EFs of the proposed model showed a marginal improvement over the C_2SMFs on FERET face database which mainly deals with variances in viewpoints. The advantages of C_2 features intolerance to variations and the advantages of visual attention model to select the attention points are complementary to each other and mutually enhancing the C_2EFs to view independent face recognition.

Table 1: Accuracy rates of face recognition using SVM classifiers on extracted different features.

Features	**Recognition accuracy (mean ± standard deviation)%**	**Extracting time (Sec)**
C_2EFs (N=80)	94 ± 2.108	5.863
C_3EFs (N=80)	97± 2.852	6.052
C_2SMFs (N=80)	89±3.944	4.592
C_3SMFs (N=80)	91±3.162	4.923
C_2SMFs (N=300)	93.5 ± 2.838	9.624
C_3SMFs (N=300)	97 ± 3.162	10.357

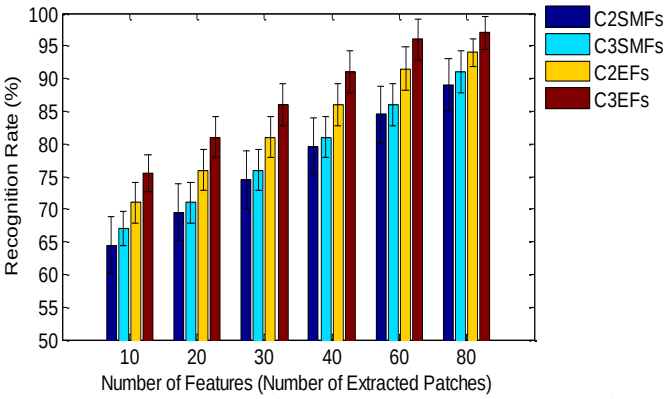

Fig. 9 The comparisons plot of computed recognition accuracy rate of C_3EFs, C_2EFs, C_2SMFs and C_2SMFs in different number of extracted features.

5. Conclusions

In this paper, we described a biologically-motivated framework for view independent face recognition, which the proposed C_3 EFs was inspired from the ventral and dorsal stream of cortex. In fact, we proposed a model to extract new view independent features, using visual attention model and ventral stream model for the goal of view independent face recognition. By visual attention model, we specified the set of attention points from salient points on a gallery of face images in varied viewpoints and by ventral stream features, we extracted proper view independent features from the set of attention points and then ran SVM classifiers on the vectors of features obtained from the input images.

Through the experimental results, we proved that the C_3EFs by using attention points in comparison to the same number of C_2SMFs improved the face classification accuracy under varying facial expressions and viewpoints. Likewise, in the proposed model, face recognition needs lesser time since we do not only use feature selection method on C_2 features, but also upgrade recognition rate with the appropriate number of attention points which selected from important regions of the face.

In the future work, we will propose the model that can detect and recognize multi-classes of faces with complicated and varied backgrounds. Also in this study, we used only SVM classifier and did not examine other classifiers. Then, in future we research directions for classification changes, the use of artificial neural network and fuzzy classification to enhance the efficiency measure.

Acknowledgments

Support of Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran is gratefully acknowledged. This study was extracted from Ph.D. thesis that was done in Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran.

References

[1] Z. Fan, and B. Lu, "Multi-View Face Recognition with Min-Max Modular Support Vector Machines", in Proceedings of the 1st International Conference on Advances in Natural Computation, Changsha, China, 2007, pp. 396-399.

[2] T.K. Kim, and J. Kittler, "Design and Fusion of Pose-Invariant Face-Identification Experts", IEEE Trans. Circuits and Systems for Video Technology, Vol. 16, No.9, 2006, pp. 1096–1106.

[3] K. R Singh., M. A. Zaveri, and M. M. Raghuwanshi M. M., "Illumination and Pose Invariant Face Recognition: A technical Review", IJCISIM, Vol. 2, 2010, pp. 029–038.

[4] R.H. Wurtz, and E.R. Kandel, Central visual pathways. In: E.R .Kandel, J.H. Schwartz, T.M. Jessell, editors. Principles of neural science. 4th ed. New York: McGraw-Hill, p. 523–547, 2000.

[5] E. Kobatake, and K. Tanaka, "Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex", Journal of Neurophysiology, Vol.71, No.3, 1994, pp. 856–867.

[6] Ungerleider L. G., J. V. Haxby, "‘‘What’’ and ‘‘where’’ in the human brain", Current Opinion in Neurobiology, Vol.4, No.2, pp. 157-165, 1994.

[7] L.G. Ungerleider, and L. Pessoa, "What and where pathways", Scholarpedia, Vol. 3, No.11, 2008, pp. 5342.

[8] J.M. Brown, "Visual streams and shifting attention", Progress in Brain Research, Vol. 176, 2009, pp. 47-63.

[9] R. Farivar, "Dorsal–ventral integration in object recognition", Brain Research Reviews, Vol.61, No.2, , 2009, pp. 144–153.

[10] M. Riesenhuber, and T. Poggio, "Hierarchical Models of Object Recognition in Cortex", nature neuroscience, Vol. 2, No.11, 1999, pp. 1019–1025.

[11] M. Riesenhuber, and T. Poggio, "Models of object recognition," nature neuroscience, Vol. 3, No. Suppl, 2000, pp. 1199–1204.

[12] T. Serre, and M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, and T. Poggio, A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex, Technical Report, MIT, Massa- chusetts, USA., 2005.

[13] T. Serre, and L. Wolf, S. Bileschi, S. Bileschi, M. Riesenhuber, "Robust object recognition with cortex-like mechanisms", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.29, No. 3, 2007, pp. 411–426.

[14] J. Lai, and W.X. Wang, "Face recognition using cortex mechanism and svm", in 1st international conference intelligent robotics and applications, Wuhan, 2008, pp. 625–632.

[15] E. Meyers, and L. Wolf, "Using Biologically Inspired Features for Face Processing", International Journal of Computer Vision, Vol. 76, 2008 ,pp. 93–104.

[16] P. Pramod Kumar, and P. Vadakkepat, "Hand posture and face recognition using a fuzzy-rough approach", International Journal of Humanoid Robotics, Vol. 7, No. 3, 2010, pp. 331–356.

[17] S. Dura-Bernal, T. Wennekers, and S. L. Denham, "Top-Down Feedback in an HMAX-Like Cortical Model of Object Perception Based on Hierarchical Bayesian Networks and Belief Propagation", PLoS ONE, Vol. 7, No. 11, 2012.

[18] J. Mutch, and D.G. Lowe, "Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields", International Journal of Computer Vision, Vol. 80, No. 1, 2008, pp. 45-57.

[19] C. Thériault, N. Thome, M. Cord, "Extended Coding and Pooling in the HMAX Model", IEEE Transactions on Image Processing, Vol. 22, No. 2, 2013, pp. 764 – 777.

[20] J. Z. Leibo, J. Mutch, and T. Poggio, "Why the Brain Separates Face Recognition from Object Recognition", in In Advances in Neural Information Processing Systems (NIPS), Granada, Spain, 2011.

[21] J. Han, and K.N. Ngan, "Unsupervised extraction of visual attention objects in color images", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, No. 1, 2006, pp. 141-145.

[22] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual-attention for rapid scene analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, No. 11, 1998, pp. 1254–1259.

[23] L. Itti, and C. Koch, "Computational modeling of visual attention", Nature Reviews Neuroscience, Vol. 2, No. 3, 2001, pp. 194-203.

[24] D. Walther, and C. Koch, "Modeling attention to salient proto-objects", Neural Networks, Vol. 19, No. 9, 2006, pp. 1395–1407.

[25] F. Kano, and M. Tomonaga, "Face scanning in chimpanzees and humans: Continuity and discontinuity", Animal Behaviour, Vol. 79, 2010, pp. 227.

[26] P.J. Phillips, H. Wechsler, J. Huang, P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms", Image and Vision Computing, Vol. 16, No. 5, 1998, pp. 295-306.

[27] C.C. Chang, C.J. Lin, LIBSVM: A library for support vector machines, Available at http://www.csie.ntu.edu.tw/cjlin/libsvm.

Somayeh Saraf Esmaili, was born in 1984. She received her B.S, M.S. and Ph.D degree in Bioelectric Engineering from Science and Research Branch of Islamic Azad University, Tehran, Iran in 2006, 2009 and 2015 respectively. Since 2011, she has been a lecturer in the Garmsar Branch of Islamic Azad University, Iran. Her current main research includes Bio-inspired Computing and Applications in face recognition.

Keivan Maghooli, is the Assistant Professor at Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. He was a Ph.D. Scholar at mentioned Department. He received M.S. degree in Biomedical Engineering from Tarbiat Modares University, Tehran, Iran. He has more than 10 years of experience, including teaching graduate and undergraduate classes. He also leads and teaches modules at B.Sc., M.Sc. and Ph.D. levels in Biomedical Engineering. His current research interests are pattern recognition, biometric and data mining.

Ali Motie Nasrabadi, received a B.S. degree in Electronic Engineering in 1994 and his M.S. and Ph.D. degrees in Biomedical Engineering in 1999 and 2004, respectively, from Amirkabir University of Technology, Tehran, Iran. Since 2005, he has been Associate Professor in the Biomedical Engineering Department at Shahed University, in Tehran, Iran. His current research interests are in the fields of biomedical signal processing, nonlinear time series analysis and evolutionary algorithms. Particular applications include: EEG signal processing in mental task activities, hypnosis, BCI, epileptic seizure prediction and visual attention models.