Optik 126 (2015) 2818–2824

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Combining two visual cortex models for robust face recognition

Somayeh Saraf Esmaili [a], Keivan Maghooli [a,*], Ali Motie Nasrabadi [b]

[a] Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
[b] Department of Biomedical Engineering, Faculty of Engineering, Shahed University, Tehran, Iran

ARTICLE INFO

Article history:
Received 18 May 2014
Accepted 2 July 2015

Keywords:
Efficient C_2 features
Visual cortex
Visual attention points
Feature extraction
Face recognition

ABSTRACT

This paper describes a new robust face recognition system based on combining C_2 features of HMAX model and visual attention model inspired from the behavior of neurons in ventral and dorsal stream of visual cortex. By visual attention model, visual attention points are selected which are discriminative areas of face images. C_2 features are extracted from these visual attention points to create efficient C_2 features and have robustness properties. These efficient C_2 features are used to distinct multi-classes of faces. For experimental analysis, we use cropped Caltech faces database and do face recognition task by support vector machine (SVM) classifier on efficient C_2 features of them. The results illustrate our proposed face recognition system using efficient C_2 features are effective and can recognize faces in various illumination and expression conditions with high accuracy and speed in comparison with original C_2 features.

1. Introduction

Human visual system can recognize faces, despite changes in illumination, size, position, expressions and viewpoint. Computational invariant face recognition is one challenging work that human visual system can do it with high-speed performance, easily. Thus, providing a model of invariant face recognition is recently one of the fundamental issues. Based on need, recently the study of brain mechanisms of human visual system and visual attention has been more considered.

The visual system divided into two main processing streams; the dorsal stream (from the primary visual cortex to the parietal cortex) for controlling eye movements and visual attention, and the ventral stream (from the primary visual cortex toward the inferotemporal lobe including V_1, V_2, V_4, Posterior Inferotemporal (PIT) and Anterior Inferotemporal (AIT)) which processes detail of objects and faces [1–3]. Both the dorsal and ventral streams are not completely independent and in higher areas such as the Prefrontal Cortex (PFC) and V_4 are interacting through interaction affect connections [4–7].

Partial analogies of the ventral stream cortex have been used in many computing models in canonical computer vision. However, most models only imitate the ventral stream and not consider the connections between visual attentions and ventral stream areas. Then, improvements will need to integrate visual attention model and hierarchical ventral stream model. Among these models, HMAX is one powerful computational model that models the object recognition mechanism of human ventral visual stream in visual cortex [8–11]. In [11], they proposed HMAX model with learning capability and increased the performance of the model. The features extracted by this model are original C_2 features. The C_2 features of ventral stream HMAX model were used for face recognition [12,13] and hand written recognition [14,15]. In recent years, different developed models of ventral stream HMAX model were presented to enhance the efficiency of the model. In all these models, the appropriate feature selection methods have been proposed [12–18]. The sparse localized feature model (SLF) is proposed by [17], which is the extend version of the HMAX model to extract sparse features. This model in compare to HMAX model has better performance and more accordance with the visual cortex. Also, there are some computational models of visual attention which account for bottom-up visual attention based on visual attention areas in the posterior cortex and these models were applied in many applications such as target detection, object recognition, object segmentation and robotic localization [19–23]. These visual attention models use low-level visual features such as color, 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 [19]. The hybrid C_2

* Corresponding author at: Department of Biomedical Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran. Tel.: +98 2144474321; fax: +98 2144474319.

E-mail addresses: s.saraf@srbiau.ac.ir (S.S. Esmaili), k_maghooli@srbiau.ac.ir (K. Maghooli), nasrabadi@shahed.ac.ir (A.M. Nasrabadi).

http://dx.doi.org/10.1016/j.ijleo.2015.07.010
0030-4026/

features of HMAX model and spectral residual approach is proposed in [24] to increase performance of object recognition HMAX model.

In this paper, we propose a new combination method of visual attention algorithm and C_2 features to access the attended locations of salient points of face images by visual attention model and extract efficient C_2 features from these regions for robust face recognition. Also, this proposed method considers the connections between ventral and dorsal stream which has previously been demonstrated biologically.

The rest of the paper is organized as follows. Section 2 illustrates the original C_2 features of HMAX model, the visual attention model for selecting distinctive facial areas and our proposed combination face recognition algorithm. The experimental evaluation including database description, results and detailed discussions is presented in Section 3. Section 4 provides the final section offers conclusions.

2. Face recognition algorithm

2.1. The original C2 features of HMAX model

In the HMAX model inspired by ventral stream [10], S_1 features, resemble the simple cells found in the V_1 area of the primate brain consists of Gabor filters, described by

$$G(x, y) = \exp\left(-\frac{\hat{x}^2 + \gamma^2\hat{y}^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}\hat{x} + \varphi\right) \quad (1)$$

where

$$\hat{x} = x\cos\theta + y\sin\theta, \hat{y} = -x\sin\theta + y\cos\theta \quad (2)$$

and (x, y) refer to the pixel in a 2D coordinate system, and the parameters affecting the filter outputs are: θ (orientation), γ (aspect ratio), σ (effective width) and λ (wavelength). Then, S_1 features imitate the simple cells in V_1&V_2 area of cortex [25] and generate by applying Gabor filters in four orientation {from 0° to 167.5°, in steps of 22.5°} and 16 scales (S) {$7\times7, \ldots, 35\times35, 37\times37$} (divided into 8 bands). In order to determine σ and λ, we use the following equation according to [10]:

$$\sigma = 0.0036S^2 + 0.35S + 0.18 \quad \text{and} \quad \lambda = \frac{\sigma}{0.8} \quad (3)$$

where S is scale and varies from 7 to 37 by steps 2 and parameter γ and φ is set to 0.5. The parameters used for Gabor filters are tabulated in Table 1.

C_1 features imitate the complex cells in V_1&V_2 area of cortex [25] 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 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 simply (of that orientation) that falls within a max filter [8]. S_1 features imitate the visual area V_4 and posterior infer temporal (PIT) cortex.

Table 1
Parameters of Gabor filters used in paper.

Band	S_1 layer				C_1 layer
	Filter size (S)	γ	λ	σ	Grid size
1	7, 9	0.5	3.5, 4.6	2.8, 3.6	8
2	11, 13	0.5	5.6, 6.8	4.5, 5.4	10
3	15, 17	0.5	7.9, 9.1	6.3, 7.3	12
4	19, 21	0.5	10.3, 11.5	8.2, 9.2	14
5	23, 25	0.5	12.7, 14.1	10.2, 11.3	16
6	27, 29	0.5	15.4, 16.8	12.3, 13.4	18
7	31, 33	0.5	18.2, 19.7	14.6, 15.8	20
8	35, 37	0.5	21.2, 22.8	17, 18.2	22

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 learn patterns of previously seen visual images and store in the synaptic weights of the neural cells. S_2 features learn from the training set of K patches ($P = 1, \ldots, K$) with various $n\times n$ sizes ($n\times n = 4\times4, 8\times8, 12\times12$ and 16×16) and all four orientations at random positions (Thus a patch P of size $n\times n$ contains $n\times n\times4$ 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 $P : f(X) = \exp\left(-||X-P||^2/2\sigma^2\right)$, with σ chosen proportional to patch size. 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 images a features vector of C_2 is computed and used for face recognition. These features have robustness properties. The lengths of C_2 features are equal to the number of random patch extracted from the images and have the property of shift and scale invariant.

In this paper, we use these C_2 features for robust face recognition and present new combination algorithm which can be extracted efficient C_2 features from face image.

2.2. Visual attention model for selecting distinctive facial areas

In face images, some facial areas are more attention and helpful areas to face recognition, such as eyes, nose and mouth that by the results of paper [26] has been demonstrated. Human can select distinctive information from face images in short time and with high accuracy. These selected features have so much local information to recognize similarity of face images and can tracked and matched to the similar face images. Then visual attention models inspired from human visual systems can select salient points from face images. Visual attention Itti's model biologically models the visual attention areas in the posterior cortex and specifies the locations of salient points from a color images simulating saccades of human vision [19,20]. In this paper, we use this model by adjusting its parameters to find automatically important areas of face.

In Itti's visual attention model, at first features of colors, intensity and orientation are extracted accordance with the neurons in the visual attention areas in retina and in dorsal stream cortex. First of all, the original red, green and blue (rgb) color features from the input image are converted to the normalized color features red, green, blue and yellow ($RGBY$) color components to create the color feature map using Eq. (4). The intensity is calculated to convert RGB to the intensity (I) [19].

$$\begin{aligned} R &= r - \frac{g+b}{2} \\ G &= r - \frac{r+b}{2} \\ B &= b - \frac{r+g}{2} \\ Y &= \frac{g+b}{2} - b - \frac{|r-g|}{2} \\ I &= \frac{R+G+B}{3} \end{aligned} \quad (4)$$

Then, center-surround differences of each of the components are calculated using the equations in (5) and (6). $M_F(s)$ represents of F feature map in s scale. F included I intensity features and C color features. A function of $N(0)$ is used for created normalization map where the symbol Θ represents interpolation of the coarser image