Neurocomputing 205 (2016) 349–359

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A biologically inspired model mimicking the memory and two distinct pathways of face perception

Xuanyang Xi [a], Peijie Yin [b], Hong Qiao [a,c,*], Yinlin Li [a], Wensen Feng [d]

[a] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[b] Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] CAS Center for Excellence in Brain Science and Intelligence Technology (CEBSIT), Shanghai 200031, China
[d] School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

ARTICLE INFO

Article history:
Received 22 December 2015
Received in revised form
16 March 2016
Accepted 1 April 2016
Communicated by Xianbin Cao
Available online 10 May 2016

Keywords:
Biologically inspired model (BIM)
Face perception
Memory
Component-based

ABSTRACT

In this paper, we propose a face perception model to mimic the biological mechanism of face perception and memory in human brain. We are mainly inspired by the fact that there are two functionally and neurologically distinct pathways after the early face perception and they both interact to process the changeable features of faces. Accordingly, our model consists of three perception parts, facial structure perception, facial expression perception and facial identity perception, which are all component-based. The structure perception has a feed-forward projection to the expression and identity perception, while the expression affects the identity perception with a modulation process. We embody the three parts referring to three bio-inspired computational models. For the facial structure perception, we utilize a cascaded-CNN (convolutional neural networks) approach to estimate the center locations of key facial components. For the facial expression perception, we propose a novel approach which exploits convolutional deep belief networks (CDBN) to spontaneously locate the places containing the most discriminative information and synchronously complete the feature learning and feature selection. For the facial identity perception, we propose an approach which adopts the hierarchical max-pooling (HMAX) model to encode notable characteristics of facial components and utilizes a new memory formation integrating the preliminary decision, expression modulation and final decision process. We evaluate our model through a series of experiments and the experimental results demonstrate its rationality and effectiveness.

1. Introduction

In the last several decades, a series of biologically inspired computational models for visual cognition are proposed based on the study of neurophysiology and psychology. They greatly promote not only the understanding of brain mechanism but also the development of computer vision. In 1959, Hubel and Wiesel [24] firstly suggest that there are two categories of cells in visual cortex, simple cells and complex cells. They study the receptive fields and functional architectures of the two cells, and propose a hierarchical model of the visual nervous system [25]. Fukushima [14] proposes a self-organizing neural network, called neocognitron, which has a modular structure similar to the hierarchical model proposed by Hubel and Wiesel. The modular structure is composed of two layers of cells connected in a cascade. The first layer consists of “S-cells” which are similar to the simple cells, and the second layer consists of “C-cells” which are similar to the complex cells. Neocognitron can reduce the problem of the shift in position and the distortion in shape. Derived from the function of prestriate visual cortex (areas V2 and V3) and inferior temporal (IT) cortex, Perrett and Oram [41] propose a model for the shape processing and object recognition. The model adopts a parallel processing framework based on a series of hierarchical steps, and completes the pooling operation through associative learning rules at each stage. Another famous visual computational model is the HMAX [45,48] which mimics the structure and function of the ventral stream in the primate visual cortex. It is a hierarchical feed-forward model composed of simple cells and complex cells alternately. Based on a MAX-like operation, it can extract position-invariant and scale-invariant features used for recognition. Inspired by the top-down recognition mechanism, Hinton et al. [17] propose a multilayer generative probability model, deep belief networks (DBN), in which layers encode statistical dependencies among the units in the next layer, and give a fast learning

* Corresponding author at: State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: hong.qiao@ia.ac.cn (H. Qiao).

http://dx.doi.org/10.1016/j.neucom.2016.04.032
0925-2312/

algorithm which maximizes the likelihood of training data one layer each time.

On one hand, researchers have done some meaningful works to extend the HMAX model. Motivated by the property of neurons in the area V4 of macaque monkeys, Cadieu et al. [6] propose a quantitative hierarchical model which increases more modular units at the top of the HMAX model. The modular unit is composed of a simple cell and a complex cell. The model can achieve visual shape selectivity and invariance properties of area V4. Qiao et al. [43] introduce the mechanism of active memory and association into the HMAX, which provides a top-down control function in the improved model. Meanwhile, the model exploits semantic and episodic features in the practical procedure. Later, they further improve the model and introduce the mechanism of preliminary cognition and active adjustment both occurring in the IT cortex [44]. In addition, the attention mechanism [56,4], memory processing mechanism [30], feedback mechanism [10], sparsification [23] and lateral inhibition mechanism [37] have also been considered to extend the HMAX model.

On the other hand, researchers have also done some meaningful works to extend the application of DBN. Liu et al. [31] propose a unified loopy framework, boosted deep belief network, which combines the boosting technique and DBN. It is used for facial expression recognition and gets a decent performance. For high-dimensional realistic images, Lee at al. [28] present a generative model, convolutional deep belief networks. It shrinks higher layers with probabilistic max-pooling and learns hierarchical representations in an unsupervised way. Huang et al. [22] propose a novel representation combining hand-crafted descriptors and additional features learned by the CDBN, and utilize them for face verification. Lee el al. [29] use the CDBN to learn unsupervised feature from the spectrograms of time-domain signals, and get a good performance for multiple audio classification tasks. In addition, DBN is also used for 3D object recognition [38], image classification [26], natural language call routing [47], learning a covariance kernel [19] and so on. These applications all obtain decent performances.

Recently inspired by the biological mechanism of face perception, we introduced the expression factor into a traditional face recognition method and got a successful improvement [60]. After this, we further investigated the biological mechanism of face perception and memory in human brain. Based on these, we propose a face perception model in this paper which can give a more plausible solution to the face perception procedure from biological perspective. Each aspect of the model is almost inspired by relevant neurophysiological and psychophysical research findings. As for the face perception in human brain, neurobiologists point out that there are two functionally and neurologically distinct pathways after the early perception which is responsible for encoding the structural information of faces. The two distinct pathways perceive relatively invariant aspects and changeable aspects of faces, and both interact to process the changeable features of faces such as the features used for identity recognition under different expressions. The invariant aspects include facial identity while the changeable aspects include facial expression. As for the memory in human brain, neuropsychologists suggest that the memory consists of three processing stages: encoding, storing and recall. The recall stage starts with a search and retrieval process and then continues with a decision or recognition process. Accordingly, our face perception model consists of three perception parts, facial structure perception, facial expression perception and facial identity perception. The structure perception has a feed-forward projection to the expression and identity perception. And the expression perception affects the identity perception with a modulation process.

Practically, we embody the three perception parts referring to three bio-inspired computational models presented recent years. For the facial structure perception, we utilize a cascaded-CNN approach to estimate the center locations of key facial components. Then guided by the location estimation results, a component-based division is carried out. For the facial expression perception, we propose a novel approach which utilizes the CDBN to complete the feature learning and feature selection synchronously. It spontaneously locates the places which contain the most discriminative information for expression perception and represents them effectively. For the facial identity perception, we propose an approach which adopts the HMAX to encode notable characteristics of facial components and utilizes a new memory formation integrating the preliminary decision, expression modulation and final decision process. As semantic patches from key and non-key components penetrate the entire model, the model is actually component-based. Through a series of comprehensive experiments, we study the property of our model and demonstrate its rationality and effectiveness.

In short, this paper has four major contributions summarized as below.

(1) We propose a face perception model which mimics the biological mechanism of face perception and memory in human brain. Our model is inspired by relevant neurophysiological and psychophysical research findings, and is implemented with a component-based form.
(2) We propose a novel approach which utilizes the CDBN to spontaneously locate the places containing the most discriminative information and synchronously complete feature learning and feature selection for the facial expression perception.
(3) We propose an approach which adopts the HMAX to encode notable characteristics of facial components and utilizes a new memory formation for the facial identity perception which integrates the preliminary decision, expression modulation and final decision process.
(4) Through a series of comprehensive experiments, we study the property of our model and demonstrate its rationality and effectiveness.

The remaining parts of this paper are organized as follows. Section 2 briefly reviews the biological evidence motivating the proposed model. Section 3 explains the implementation of our model in detail. Section 4 evaluates the model through a series of experiments and analyzes its characteristics. Section 5 makes a conclusion with respect to the paper.

2. Biological evidence

Our model is mainly inspired by the biological mechanism of face perception and memory in human brain. In this section, we briefly review the related biological evidence from four aspects. The detailed descriptions are stated below.

2.1. Two distinct visual pathways in the face perception

For face perception, Bruce and Young [5] firstly propose a multi-route model. They suggest that there are seven distinct types of information that people derive from faces, which are labeled pictorial, structural, visually derived semantic, identity specific semantic, name, expression and facial speech codes. Meanwhile, these different facial attributes are processed by distinct functional routes. Later, Haxby et al. [16] present a compatible model for the visual perception of faces, which comprises three bilateral regions corresponding to three different functions. The early perception encodes the structural information of faces which is taken as input for the latter perception.