Neurocomputing 74 (2011) 2008–2017

Contents lists available at ScienceDirect

Neurocomputing

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

Hebbian-based neural networks for bottom-up visual attention and its applications to ship detection in SAR images

Ying Yu [a,b], Bin Wang [b,c,*], Liming Zhang [b]

[a] School of Information Science and Engineering, Yunnan University, Kunming 650091, PR China
[b] Department of Electronic Engineering, Fudan University, Shanghai 200433, PR China
[c] The Key Lab of Wave Scattering and Remote Sensing Information (Ministry of Education), Fudan University, Shanghai 200433, PR China

ARTICLE INFO

Available online 19 February 2011

Keywords:
Visual attention
Saliency map
Principal component analysis
Discrete cosine transform
Hebbian neural networks

ABSTRACT

This paper proposes a bottom-up attention model based on pulsed Hebbian neural networks. The salience of the visual input can be generated through the networks using a simple normalization process, which can be calculated rapidly. Moreover, visual salience in this model can be represented as binary codes that mimic neuronal pulses in the human brain. Experimental results on psychophysical patterns and eye fixation prediction for natural images prove the effectiveness and efficiency of the model. In an arduous task of detecting ships in synthetic aperture radar (SAR) images, there are large amounts of data to be processed in real time. As a fast and effective technique for saliency detection, the proposed model is applied to ship detection in SAR images and its robustness against speckles is further proved.

1. Introduction

It is believed that there exist information bottlenecks along the visual pathway because the human brain has limited neural resources [25,30]. Therefore, instead of fully processing the massive sensory input in parallel, a serial visual mechanism known as attention selection has evolved [24,25,30]. Only the selected part of visual inputs is allowed to reach high-level cortical processing such as short-term memory, visual awareness, recognition and learning [6,8]. This is demonstrated by our blindness to unattended visual inputs even when they are salient [38].

Attention selection has been divided into two categories. One is primarily driven in a bottom-up manner and the other is controlled by top-down cues [26]. While top-down attention is largely task-dependent, bottom-up attention is scene-dependent, i.e., it only depends on the salience of an input scene [25,28,43]. Since bottom-up attention is independent of the particular task and operates very rapidly, it can make our eyes always gaze towards salient objects in a clustered scene [8,28,43].

Recent years have witnessed the fast development of interdisciplinary studies focused on computer vision in relation to robotics, cognitive science and neuroscience. Computational modeling of bottom-up visual attention has attracted intensive investigations in area of neural information processing and computer engineering. Itti et al. [27] proposed a biologically plausible model of bottom-up attention selection. The model implements and expands the generation mechanism of attention selection proposed by Koch and Ullman [28]. Inspired by feature-integration theory [43], Itti's model decomposes an input image into three channels: intensity, color and orientation. Simulating the lateral surround suppression among cortical cells, a center–surround operation produces a set of feature maps. These feature maps are then normalized and combined across scales to create conspicuity maps for each channel. The normalized conspicuity maps are further linearly combined to form the overall saliency map. Itti's model has been shown to be successful in detecting salient objects and predicting human fixations. It remains the most popular attention model and the yardstick to measure the performance of other models. However, Itti's model is ad-hoc designed because many parameters must be tuned by hand. Besides that, it demands high computational cost.

A number of recent papers (e.g., [4,5,14,15,20]) have attempted to address the question of what attracts human visual attention in an information theoretic way. Bruce and Tsotsos [4,5] proposed a bottom-up attention model based on the principle of information maximization sampled from a scene. This model uses Shannon's self-information to measure saliency. Harel et al. [20] proposed a graph-based visual saliency model. This model first forms activation maps on feature channels, and then normalizes them in a way that highlights conspicuity. Gao et al. [14,15] proposed a discriminant center–surround model that defines saliency using mutual information between center and surround. These models are able to achieve good performance and consistency

* Corresponding author at: Department of Electronic Engineering, Fudan University, Shanghai 200433, PR China. Tel./fax: +86-21-55664210.
E-mail addresses: yuying.mail@163.com (Y. Yu), wangbin@fudan.edu.cn (B. Wang), lmzhang@fudan.edu.cn (L. Zhang).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.06.026

with psychophysics but demand high computational cost. Different from these spatial domain models, Fourier transform (FT)-based approaches have been recently proposed [3,18,19,22]. The gist of these FT-based approaches is that, to some extent, the operation of flattening the amplitude spectrum is almost equivalent to the center–surround operation in the visual space. These approaches can detect salient areas with less computational cost.

It is well known that Hebbian-learning rule commonly exists among neurons in the human brain [21] and hence, Hebbian-based neural networks have been deeply investigated. However, the relationship between Hebbian-based neural networks and selective visual attention has seldom been investigated. In this paper, we only use simple feedforward Hebbian-based neural networks to produce visual salience. A lot of previous studies have suggested that Hebbian-learning in neural network can find principal components of sensory data and the principal components of natural scenes can capture the spatial correlation of the visual input [12]. In this paper, we describe the visual salience in terms of spatial correlation in the visual space and obtain the information about visual salience within a simple normalization process for the principal component analysis (PCA) coefficients. Our proposed computational model is called pulsed PCA transform. It is worth stating that the visual salience in our model can be represented as binary codes that mimic neuronal pulses in the human brain. Moreover, since discrete cosine transform (DCT) can be considered as a "completely developed" PCA transform [1,11,17], our PCA-based model can be extended to a DCT-based framework in engineering applications. Note that DCT is data-independent and has many fast algorithms for its calculation. Thus, our model can calculate a saliency map easily and rapidly.

Satellite-based synthetic aperture radar (SAR) is a powerful tool for ship detection because it can work in all weather conditions, day and night [7]. However, speckles and heterogeneous regions in SAR images pose great challenges on automatic detection of ships [7,10,47]. On the other hand, a fast surveillance system needs a fast and effective detection algorithm because it usually needs to process large amounts of SAR data in real time. Using our attention model to generate saliency maps of SAR images rapidly, the ship signatures can be easily detected. The analysis of the detection performance over real SAR images confirms the robustness of the proposed model.

The following sections are organized as follows. Section 2 gives an overview of the proposed architecture of bottom-up visual attention as well as its biological plausibility. Section 3 presents experimental results, where our model is compared with other state-of-the-art models. Section 4 introduces a visual attention-based ship detection approach for SAR images. Finally, discussion and conclusion are presented in Section 5.

2. Model architecture

Hebbian-learning rule commonly exists among neurons in the human brain. In this section, we propose a bottom-up attention model based on Hebbian-based neural networks. We will explain how such frameworks relate to visual salience.

2.1. What is visual saliency?

Li [29] has hypothesized that the primary visual cortex (V1) creates a bottom-up saliency map of visual space and the contextual influence is necessary for saliency computation. For example, a vertical bar is salient in a context of horizontal but not vertical bars. Each neuron in V1 is tuned to a visual feature such as orientation and color. The dominant contextual influence in V1 is iso-feature suppression, i.e., nearby neurons tuned to similar features are linked by intra-cortical inhibitory connections [30,32]. Besides Li's V1 hypothesis, a number of recent studies (e.g., [4,5,23,14,15,42,48]) have attempted to describe the visual salience in terms of self-information, surprise, interest, innovation, center–surround discrimination, etc. These studies provided a general idea that higher information entropy accounts for higher visual salience.

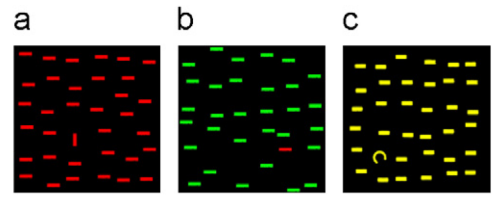

Fig. 1. Examples for where visual saliency happens. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

Our visual environment is highly structured and hence there exists much information redundancy in the visual input. It has been shown that human visual system can reduce redundancy of visual input and the dominant form of visual input redundancy arises from second-order rather than higher order input statistics [2,31,35,37]. In this paper, we describe the visual salience in terms of statistical correlation in the visual space. For example in Fig. 1(a), in order to make all bars overlapped, we can move a horizontal bar to the location of another horizontal bar, but we need to move and rotate the vertical bar to overlap other bars. This means that the vertical bar has low spatial correlation with its surroundings and hence it is salient. Similarly, a red bar among green bars as in Fig. 1(b) and a curve among bars as in Fig. 1(c) are salient objects. According to such an interpretation of visual salience, in the next subsection we attempt to capture highly correlated components in visual space and suppress them so as to highlight salient visual features.

2.2. P^2CA model

It has been noted that the principal components of natural images with stationary statistics reflect global features in the visual space, and that all the redundancy reflected in the second-order correlations between pixels is captured by the PCA coefficients of the image [12]. A lot of studies have shown that Hebbian-learning in neural networks can find the principal components of incoming sensory data [13,34,36,46]. Hence, we attempt to use simple feedforward connections to generate the visual salience of an input scene. According to the analysis in previous subsection, a location with high spatial correlation with its surroundings is suppressed and a salient location is then highlighted by whitening (normalizing) the output of the networks.

In engineering applications, the feedforward connections can be represented by a set of PCA projection vectors that are easily obtained using some efficient numerical methods such as eigenvalue decomposition or the QR algorithms [16]. After all PCA projection vectors are obtained, we reshape the n-pixel input image X into an n-dimensional vector x, and calculate its saliency information as

$$p = \mathrm{sign}(Cx), \tag{1}$$

$$f = \mathrm{abs}(C^{-1}p) \tag{2}$$

where C is an $n \times n$ PCA transformation matrix that comprises n principal components vectors. The notation "sign(.)" is the signum function, and "abs(.)" is the absolute value function.