Information Sciences 309 (2015) 1–10

Contents lists available at ScienceDirect

Information Sciences

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

Visual acuity inspired saliency detection by using sparse features

Yuming Fang [a], Weisi Lin [b], Zhijun Fang [a,*], Zhenzhong Chen [c], Chia-Wen Lin [d], Chenwei Deng [e]

[a] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang 330032, Jiangxi, China
[b] School of Computer Engineering, Nanyang Technological University, 639798 Singapore, Singapore
[c] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[d] Department of Electrical Engineering, National Tsing Hua University, Hsinchu 330013, Taiwan
[e] School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 3 June 2014
Received in revised form 25 February 2015
Accepted 4 March 2015
Available online 11 March 2015

Keywords:
Visual attention
Saliency detection
Human visual system
Sparse features
Human visual acuity

ABSTRACT

In this paper, we propose a new computational model of visual attention based on the relevant characteristics of the Human Visual System (HVS) and sparse features. The input image is first divided into small image patches. Then the sparse features of each patch are extracted based on the learned independent components. The human visual acuity is adopted in calculation of the center-surround differences between image patches for saliency extraction. We choose the neighboring patches for center-surround difference calculation based on the relevant characteristics of the HVS. Furthermore, the center-bias factor is adopted to enhance the saliency map. Experimental results show that the proposed saliency detection model achieves better performance than the relevant existing ones on a large public image database with ground truth.

1. Introduction

When looking at a visual scene, an observer cannot capture all visual information in the scene with detail at the same time. The visual information in a scene is usually much more than that the human central neural system can process. Selective attention in the Human Visual System (HVS) is an important mechanism to focus on some particular visual information while ignoring the other in visual scenes [39,43]. There are two types of visual attention mechanisms in the HVS: bottom-up and top-down. Bottom-up attention is a involuntary and task-independent perceptual processing for salient region selection [16,21,31,36,46], while top-down attention is a voluntary perceptual processing influenced by prior knowledge such as tasks to be performed, and feature distribution of targets [4,18,28].

In this paper, we focus on bottom-up visual attention modeling. Currently, bottom-up saliency detection is widely used in various visual processing applications such as retargeting, retrieval, and coding. Based on the Feature Integration Theory (FIT) [39], various computational models of visual attention have been proposed to predict human fixation regions [5,7,12,27,42]. Most of these existing models extract the saliency map of the input image by computing the center-surround differences between image patches. However, these models rarely consider other useful characteristics of the HVS. In this

* Corresponding author.
E-mail addresses: fa0001ng@e.ntu.edu.sg (Y. Fang), wslin@ntu.edu.sg (W. Lin), zjfang@gmail.com (Z. Fang), zzchen@whu.edu.cn (Z. Chen), cwlin@ee.nthu.edu.tw (C.-W. Lin), cwdeng@bit.edu.cn (C. Deng).

http://dx.doi.org/10.1016/j.ins.2015.03.004
0020-0255/

study, we propose a acuity-based saliency detection model for images. Sparse features are extracted to represent the image patches for center-surround difference calculation. Additionally, an important characteristic of the HVS, the human acuity, is adopted to weight the center-surround differences between image patches for saliency detection.

Existing studies show that the sparse representation of image statistics exists in the primate visual system. In particular, sparse coding is regarded as an efficient coding strategy in the primary visual cortex [8,35]. With such image representation, sparse features of images can be obtained. In this study, we extract sparse features of image patches to compute center-surround differences between image patches.

It has been demonstrated that the HVS is highly space-variant in processing visual information because the retina in the human eye has different densities of cone photoreceptor and ganglion cells [41]. The density of cone receptor and ganglion cells accounts for the visual acuity. The highest density of cone receptors is found in the fovea of retina. With greater retinal eccentricity from the fovea region, the density of the cone receptors becomes lower and thus the visual acuity decreases. We propose to incorporate the characteristics of the HVS in this regard (i.e. the human visual acuity) into the proposed saliency detection model. To be more specific, the human visual acuity is used to weight the center-surround feature differences between image patches for saliency calculation.

In essence, the proposed model first divides the input image into small image patches, whose sparse features are extracted based on the learned basis. The center-surround differences between image patches are then calculated by the extracted sparse features. By using the human visual acuity to weight the center-surround differences, we compute the saliency map of the input image. Besides, existing studies demonstrate that the center-bias exists during human fixation and we integrate this factor into the proposed saliency detection model. Different from existing studies which use the Euclidean distance or the Gaussian distribution of Euclidean distance to weight the center-surround differences [7,42], the proposed model uses a more reasonable weighting method by the human visual acuity for saliency extraction. Additionally, we choose the neighboring image patches for center-surround difference calculation based on the characteristics of the HVS. Experimental results on a large public database with ground truth show the better performance of the proposed saliency detection model over other existing ones.

The reminder of this paper is organized as follows. In Section 2, we introduce the related research work in the research literature. Section 3 describes the proposed model in detail. In Section 4, the experimental results are provided to demonstrate the performance of the proposed model. The final section concludes the paper.

2. Related work

Currently, many computational models of visual attention have been proposed for various multimedia processing applications during the past decades. One of the earliest saliency detection models proposed by Itti et al. is built based on the behavior and the neuronal architecture of the primates' early visual system [16]. That model computes saliency map by calculating the multi-scale center-surround differences from low-level features of intensity, color and orientation. Based on Itti's model, Harel et al. devised a Graph-based Visual Saliency (GBVS) model through utilizing a dissimilarity measure for saliency extraction [11]. Different from Itti's model, GBVS model uses the graph theory to compute the saliency map based on low-level features. Gao et al. calculated the saliency map through a defined center-surround discriminant [5]. The saliency value of each image pixel is computed by the power of a Gabor-like feature set [5]. Goferman et al. designed a context-aware saliency detection model by including more context information in the saliency map [7]. In that study, color information in Lab color space is used for feature extraction [7].

Some studies of visual attention modeling try to calculate the saliency map of the input image in transform domain. Hou et al. built a visual attention model by using the concept of Spectral Residual (SR), and claimed that the SR model can be implemented by log spectra representation of images [12]. Later, Guo et al. found that the phase spectrum is the main factor for the saliency detection in that model [12] and thus designed a phase-based saliency detection model based on Fourier Transform (FT) [10]. In that study, the saliency map is estimated by Inverse Fourier Transform (IFT) on a constant amplitude spectrum and the original phase spectrum [10]. The wavelet coefficients are adopted to build the saliency detection model in [14]. Both local and global contrast are calculated from the wavelet coefficients for saliency estimation [14]. In [34], Murray et al. proposed a saliency detection model by using Inverse Wavelet Transform on the multi-scale features [34]. In that study, the scale-weighting function has been optimized to replicate psychophysical data on color perception [34]. Wang et al. proposed a saliency detection model based on the information maximization principle [42]. In that study, they defined the *Site Entropy Rate* to measure the saliency for images [42].

Recently, there are some spatiotemporal saliency detection models proposed for video sequences. Itti et al. proposed to use a Bayesian model to detect surprising events which attract human attention. In that study, surprising events are calculated by the differences between posterior and prior beliefs for observers [15]. Ma et al. designed a saliency detection model by considering both the top-down and bottom-up mechanisms for video summarization [29]. Zhai et al. proposed a spatiotemporal saliency detection model by fusing spatial and temporal saliency linearly [45]. Le Meur et al. extended their previous saliency detection model from the spatial domain to the spatiotemporal domain [22]. The achromatic, chromatic and temporal features are used to estimate the saliency map of video frames. Li et al. proposed spatiotemporal saliency detection models based on multi-task learning techniques [24,25]. Seo et al. introduced the concept of self-resemblance to measure visual saliency for video frames [36].