J. Vis. Commun. Image R. 30 (2015) 201–207

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Efficient saliency analysis based on wavelet transform and entropy theory ☆

Xiaolong Ma [a,b], Xudong Xie [a,*], Kin-Man Lam [b], Yisheng Zhong [a]

[a] TNlist and Department of Automation, Tsinghua University, Beijing, China
[b] Centre for Signal Processing, Department of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong

ARTICLE INFO

Article history:
Received 16 July 2014
Accepted 15 April 2015
Available online 24 April 2015

Keywords:
Saliency detection
Human vision system
Preprocessing technique
Color information
Bottom-up approach
Fixation–prediction
Wavelet transform
Entropy theory

ABSTRACT

Saliency detection has extensive applications in daily life. In this paper, an efficient saliency-detection method based on wavelet transform and entropy theory is proposed. In the algorithm proposed in this paper, salient regions are viewed as uncommon regions in the background of an image. The uncommon regions can be caused by differences in color, orientation, texture, shape, or other factors. Considering the fact that wavelet coefficients can represent the local features of an image in different scales and orientations, the wavelet transform is therefore employed to identify the salient regions. Unlike those conventional wavelet-based methods, our proposed method need not perform the inverse wavelet transformation; this can reduce the computational requirements. In addition, because the different factors (i.e. color, orientation, texture, shape, etc.) stimulate different aspects of the human visual system, a saliency-map combination scheme based on the entropy theory is devised in this paper, which can evaluate the influence or significance of the different factors. Experimental results show that our method, based on wavelet transformation and entropy theory, can achieve excellent performance in terms of the area under the receiver operating characteristic curve (AUC) score, the linear correlation coefficient (CC), the normalized scan-path saliency (NSS) score, and visual performance, as compared to existing state-of-the-art methods.

1. Introduction

Saliency detection [1–4] is a useful preprocessing technique [5,6] for object detection, egocentric action recognition [7], image- and video-quality assessment [8,9], image segmentation [10], image interpolation [11], and image compression [12]. Saliency-detection methods can be classified into two categories: the top-down approach [13] and the bottom-up approach [14,15]. Detection conspicuity is likely to reflect bottom-up saliency, while identification conspicuity contains a top-down saliency component [16]. The top-down approach is related to task demands, and the bottom-up approach involves the analysis of the behavior of the human visual system (HVS) during free viewing [17].

Many bottom-up-based saliency-detection methods have been proposed during the last two decades. Inspired by the putative neural mechanism, and hypothesizing that some visual inputs are salient in certain background contexts, a center-surround mechanism was proposed to detect saliency across scales [18–20]. Moreover, local complexity [21] and local differences [22–24] to the surrounding areas are also employed to describe saliency. Among the local features, color information is the most important, and different color channels – including YCbCr [25], RGBYI [26], and Lab [27] – have been used to produce saliency maps.

In this paper, an efficient saliency-detection method, based on the bottom-up approach, is proposed. The key idea is that salient regions can be reflected by their difference from the surrounding areas, and wavelet coefficients can be used to represent the degree of difference. In other words, the larger the wavelet coefficients are in a region, the higher the saliency of that region is. The two-dimensional (2D) entropy theory [28] is employed to obtain a combined saliency map based on the different color channels (e.g. YCbCr, RGBYI, and Lab) being used.

Our proposed algorithm is evaluated based on human-eye fixations prediction. Therefore, we compare our algorithm with five state-of-the-art fixation–prediction methods, namely Soft Image Abstraction method (SIA) [4], Saliency Using Natural statistics method (SUN) [29], Attention based on Information Maximization (AIM) [30], Non-Parametric Low-Level method (NPLL) [31], Quaternion Spectral Saliency method (QSS) [32], and

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.
* Corresponding author. Fax: +86 010 62786911.
E-mail address: go0up@163.com (X. Xie).

http://dx.doi.org/10.1016/j.jvcir.2015.04.008
1047-3203/

Hypercomplex Fourier Transform (HFT) [28], in terms of the AUC score, CC score [33], and NSS score [33]. Experimental results show that our method can always achieve the best performance in terms of both the quantitative measurements and visual appearance.

This paper is organized as follows. In Section 2, our proposed method is presented, which includes the color channels used (Section 2.1), the wavelet transform, the wavelet-coefficient enhancement scheme (Section 2.2), and saliency-map combination based on the 2D entropy theory (Section 2.3). In Section 3, the experimental results are evaluated and compared with different state-of-the-art algorithms. Finally, a summary of our algorithm and possible future work are presented in Section 4.

2. The algorithm

A natural image can be classified into salient regions and non-salient regions. As salient regions are quite different from their surrounding areas in the spatial domain, methods based on the center-surround mechanism [34–38] have been proposed for salient-region detection. The center-surround difference of a salient region may be caused by various factors, such as color, orientation, texture, and shape, which lead to the uncommon regions in the image. Thus, an ideal saliency detector should have the ability to consider different features to determine saliency under various conditions. Due to the time–frequency analysis property of the wavelet transform [39–41], the wavelet coefficients are an effective means of representing salient regions. In fact, a salient region which is different from its surrounding areas in an image will stand out at a suitable scale and orientation in the wavelet domain, with a suitable color channel. Therefore, our algorithm considers different color channels so as to handle the effect of color variations. Meanwhile, wavelet coefficients at specific scales and orientations can effectively represent the difference between a salient region and its surrounding areas due to the texture and orientation information, to an extent. An example is illustrated in Fig. 1, where the original images are shown in Fig. 1(a), and the corresponding images decomposed in the wavelet domain (the horizontal coefficients of the Cr channel using Haar wavelet decomposition after a Gaussian filter smoothing) are shown in Fig. 1(b).

In Fig. 1(a), the salient regions in the six images are caused by the center-surround difference caused by color (the first two images), texture orientation (the 3rd image), shape (the 4th and 5th images), and density (the 6th image), respectively. From Fig. 1(b), it can be seen that if a proper color channel and wavelet coefficients at suitable scales and orientations are selected, the coefficients in the wavelet domain can be used preliminarily to detect saliency. Therefore, in our algorithm, wavelet transform is used for feature extraction and saliency analysis. Previous wavelet-based saliency-detection algorithms [31] consider the middle- or high-frequency components in the determination of saliency. However, the salient regions illustrated in Fig. 1 are actually of low-frequency contents. In our proposed model, we consider the mechanism of HVS, which is sensitive to color, orientation, texture, and shape. The wavelet transform can make these regions stand out from their backgrounds, which facilitate saliency detection. The wavelet transform is applied to the different color channels, with different scales and orientations, to obtain different wavelet coefficients. All the operations are performed in the wavelet domain only, and the inverse wavelet transformation is not required.

From the above discussion, we can see that different orientations, scales, and colors under different illumination conditions will affect the salient regions of an image (see Fig. 2). Therefore, we compute the wavelet coefficients at different scales and orientations for different color channels, so as to obtain the saliencies under various conditions. Because the different factors have different impacts on the HVS, the 2D entropy theory is employed to evaluate the degree of impact of each of the different factors in order to produce a combined saliency map. Then, this saliency map can be used for human-eye fixations prediction. Fig. 3 shows the procedures of our proposed method, which will be described in detail in the following sections.

2.1. Color-channels transformation

To detect the salient regions in an image, suitable color channels should first be selected. Luo et al. [42] have shown that salient regions may become more obvious and be detected more easily in some specific color channels. From Fig. 4, we can see that images with different conditions have different contrasts in the various color channels. In our method, an image in the RGB format is first converted to other color channels: YCbCr color channels [25], RGBYI color channels [26], and Lab color channels [27]. The YCbCr color channels and Lab color channels have been widely used for saliency detection, while the RGBYI color channels [26] were proposed recently. Thus, we will briefly introduce the RGBYI color channels. Denote r, g, and b as the red, green, and blue channels, respectively, of an input image. Then, the gray-scale image I can be computed using (1). As low luminance disturbs saliency detection [43], four broadly tuned color channels, namely R,

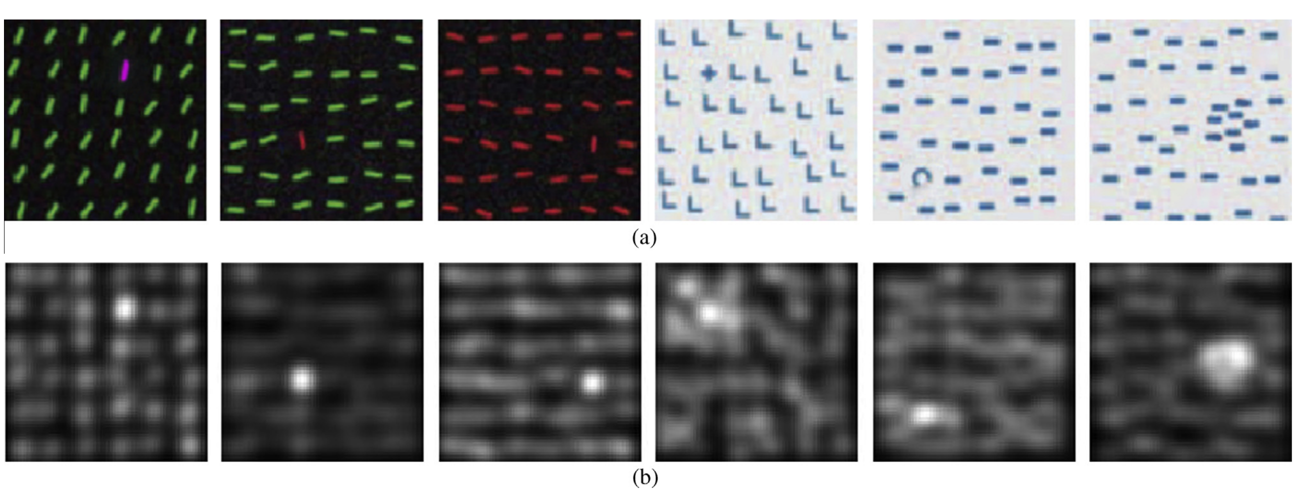

Fig. 1. (a) The original images and (b) the transformed images in the wavelet domain.