J. Vis. Commun. Image R. 30 (2015) 64–74

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

An effective vector model for global-contrast-based saliency detection ☆

Linfeng Xu *, Liaoyuan Zeng, Huiping Duan

School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

ARTICLE INFO

Article history:
Received 7 July 2014
Accepted 22 March 2015
Available online 30 March 2015

Keywords:
Visual attention
Saliency model
Global contrast
Vector model
Feature vector
Weighted mean vector
Chromatic double opponency
Similarity distribution

ABSTRACT

The saliency detection methods based on global contrast can generate full-resolution saliency map with uniformly highlighted regions and defined boundaries. For the images consisting of large salient objects, the use of unweighted sum of the color distances in the existing global-contrast-based methods may result in the detection of the background instead of the outstanding objects. In this paper, we propose a new global-contrast-based saliency detection method, called LRSW method, by deriving a new vector model which uses the weighted mean vector and contains the features of CIELAB color, chromatic double opponency, and similarity distribution. By using the vector model, the proposed method can significantly increase the detection precision and suppress the background in the saliency map, especially for large salient objects. The experimental results on the MSRA benchmark images show the effectiveness of the proposed method which outperforms the existing methods on visual saliency detection in terms of precision and recall.

1. Introduction

Humans can scan a complex visual environment quickly and focus their attention on the salient regions before cognition process is completed. This selective visual ability allows brain and visual system to break through the bottleneck of information-processing, because it is hypothesized that human visual system only concentrates on the most unusual parts of the massive sensory incoming information [1]. In the computer vision field, it is critical to simulate this ability to extract saliency maps because the maps are key to the applications in images and videos including perceptual video retargeting [2,3], perceptual object segmentation [4,5], adaptive coding [6], object recognition [7,8], and image retrieval [9].

Visual saliency is a perceptual state or quality that makes an item prominent from its neighborhoods. In the last decade, many visual attention models were proposed in the areas of neurobiology, psychophysics, cognition, and computer vision. The models can be implemented in the manner of bottom-up, top-down, or hybrid.

The bottom-up saliency detection can be performed based on center-surround contrast [10], common similarity [11–13], information theory [14,15], graph model [16], or learning methods [17,11], which is shown to be quite effective for predicting attention in free viewing of natural scenes. Thus, there has been much interest in bottom-up model study. The bottom-up models can be classified into local and global methods [18]. The local-contrast-based methods explore features from the neighborhoods according to the "center-surround" hypothesis [10,19,20,16,21,22,17]. The saliency maps generated by such methods usually highlight the object boundary instead of the entire object. To extract the local contrast features, the severe down-sampling lowers the map resolution. On the contrary, global-contrast-based methods integrate all the information features in the whole image [23–27,18]. Generally, these methods can generate full resolution saliency maps and evenly highlighted salient regions.

Following the bottom-up manner, we propose a novel saliency detection method, called LRSW method which can significantly improve the detection accuracy and suppress the non-salient background in the saliency map, especially for large salient region detection. The LRSW method is based on global contrast and proposed by developing a new vector model which is constructed by integrating the weighted mean vector and the feature vector of CIELAB color, chromatic double opponency, and similarity distribution. The vector model can enlarge the angle between the vector of the salient pixel and the mean vector, which facilitates the process of distinguishing the salient. The experimental results show that the LRSW method has significant improvements over the existing detection methods in terms of precision and recall on the MSRA dataset.

The remainder of this paper is organized as follows. The related work of saliency detection is introduced in Section 2. In Section 3,

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

* Corresponding author. Fax: +86 28 61830064.
E-mail addresses: lfxu@uestc.edu.cn (L. Xu), lyzeng@uestc.edu.cn (L. Zeng), huipingduan@uestc.edu.cn (H. Duan).

http://dx.doi.org/10.1016/j.jvcir.2015.03.011

the basic vector model is derived to perform the global-contrast-based saliency detection. In Section 4, we analyze the basic vector model. In Section 5, we develop the new vector model by modifying the construction of the basic model. The experimental results are shown in Section 6. Finally, conclusion is drawn in Section 7.

2. Related work

One of the most popular local-contrast-based computational model was proposed by Itti et al. [10]. In the model, three multi-resolution extracted local feature contrasts, i.e., luminance, chrominance, and orientation, were combined to produce a saliency map. Besides Itti's model, lots of local-contrast-based saliency detection methods were proposed due to the obvious biological supporting. Ma and Zhang [19] used local contrast analysis to generate the saliency map, and proposed a fuzzy growing method to extract salient objects. Extending Itti's model, Walther and Koch [20] inferred proto object regions which were used to achieve object recognition. Harel et al. [16] used Itti's method to create feature maps and proposed a graph-based approach to form activation maps, which were then normalized and combined to the final saliency map. Han et al. [21] constructed a Markov Random Field (MRF) model to segment salient objects, merging the seed values from Itti's saliency map and some low-level features. Gao and Vasconcelos [22] proposed a saliency detector which used the power of a Gabor-like feature set to discriminate the visual appearance between the center and surround regions in an image. Liu et al. [17] learned a Conditional Random Field (CRF) to combine a set of features, i.e., multi-scale contrast, center-surround histogram, and color spatial distribution, for saliency object detection.

Global-contrast-based methods were investigated frequently in saliency detection because of the advantages, such as full resolution, uniformly highlighted regions, and defined boundaries. The global contrast feature can be computed by the frequency or spatial analysis.

In the frequency analysis, Hou and Zhang [23] obtained saliency maps from the spectral residual of the log-spectrum. The data used by this method may not provide enough information for accurate saliency detection. Achanta et al. [24] proposed a frequency-tuned technology to estimate the center-surround contrast by using the color and luminance features. This work was extended in [25] to solve the object scale problem, which varied the bandwidth of the center surround filtering near image borders using the symmetric surrounds.

In the spatial analysis, Zhai and Shah [26] detected the spatial saliency of a pixel by computing the contrast to all the other pixels and using 1-D histogram of a specific color channel (e.g., red channel) or the luminance for efficiency. This method did not use the relationship between different color channels. Goferman et al. [27] proposed a context-aware saliency detection method which used local–global low-level features, multi-scale enhancement technique, and high-level factors. Extending the spatial saliency in [26], Cheng et al. [18] proposed a histogram-based method and a color quantization technique to make the processing more efficient. In Cheng's work, in order to highlight the entire objects uniformly, the spatial information is further incorporated to produce region-based contrast maps. Recently, Perazzi et al. [28] abstract images into perceptually uniform regions and measure the variance of the spatial distribution for these regions.

Much work discussed above needs efficient processing approaches to reduce the computational complexity. In addition, without the relationship between the saliency value and the feature, inappropriate features may be selected to impair the performance as analyzed in Section 4.2. Furthermore, in the existing global-contrast-based methods, the use of unweighted sum of the color distances can easily result in detecting large salient objects as the background.

3. The basic vector model for global-contrast-based saliency detection

In this section, via representing the color distance of pixels in the form of dot product of feature vectors, we derive the basic vector model for global-contrast-based saliency detection.

3.1. The vector model derived from Euclidean distance

As discussed in [26,18], for global-contrast-based saliency detection methods, the saliency value of a pixel p in an image I is defined as the pixel's color contrast to all the other pixels in the image, which can be formulated as

$$V_g(p) = \sum_{\forall q \in I} \mathcal{D}(I_p, I_q) \tag{1}$$

where I_p denotes the color value of pixel p in different color spaces, such as RGB, CIELUV, and CIELAB, and $\mathcal{D}(I_p, I_q)$ represents the difference between the color values of pixels p and q. In order to reduce the computational complexity, different methods are adopted, such as histogram-based speed up [26,18] and color channel quantization [18].

Representing $\mathcal{D}(I_p, I_q)$ as the squared Euclidean distance between the two color values, we rewrite (1) as

$$\begin{aligned} V_g(p) &= \sum_{\forall q \in I} \|\boldsymbol{x}_p - \boldsymbol{x}_q\|^2 = \sum_{\forall q \in I} (\boldsymbol{x}_p - \boldsymbol{x}_q)^T (\boldsymbol{x}_p - \boldsymbol{x}_q) \\ &= N\boldsymbol{x}_p^T\boldsymbol{x}_p - 2N\boldsymbol{x}_p^T\boldsymbol{\mu}_x + \sum_{\forall q \in I} \boldsymbol{x}_q^T\boldsymbol{x}_q \end{aligned} \tag{2}$$

where $\boldsymbol{x}_p = (x_1, x_2, x_3)^T$ is the color (feature) vector of pixel p, N denotes the pixel number of the image I, and $\boldsymbol{\mu}_x = \frac{1}{N}\sum_{\forall q \in I} \boldsymbol{x}_q$ represents the mean vector of color vectors of all the pixels in the image. Since the third term in (2) is a constant for a given image, it has no effect on the relative saliency values of pixels. So in the following, we use the value of $\boldsymbol{x}_p^T\boldsymbol{x}_p - 2\boldsymbol{x}_p^T\boldsymbol{\mu}_x$ to compute the saliency value of each pixel. The basic vector model is expressed as

$$S_g(p) = \boldsymbol{x}_p^T\boldsymbol{x}_p - 2\boldsymbol{x}_p^T\boldsymbol{\mu}_x = |\boldsymbol{x}_p|(|\boldsymbol{x}_p| - 2|\boldsymbol{\mu}_x|\cos\theta) \tag{3}$$

where θ is the angle between vectors $\boldsymbol{x}_p$ and $2\boldsymbol{\mu}_x$, and $2|\boldsymbol{\mu}_x|\cos\theta$ is the scalar projection of $2\boldsymbol{\mu}_x$ onto $\boldsymbol{x}_p$. The saliency value of a pixel is only related to the vector $\boldsymbol{x}_p$ and the mean vector $\boldsymbol{\mu}_x$. The concise relationship simplifies the analysis of this vector model. The difference in (3) may generate negative values. The values of all the pixels in the saliency map are normalized to the range $[0, 1]$ when the saliency values are computed (the minimal value is normalized to zero).

3.2. "Frequency-tuned" method: an approximate vector model with Gaussian filtering

In [24], a frequency-tuned algorithm (FT) was proposed to extract full resolution saliency maps with low computational complexity. The saliency map was obtained by computing the Euclidean distance between the mean vector of all the pixels in an image and a Gaussian blurred vector of each pixel, which can be formulated as

$$V_{FT}(p) = \|\boldsymbol{\mu}_x - \mathcal{G}\{\boldsymbol{x}_p\}\| \tag{4}$$

where $\mathcal{G}\{\boldsymbol{x}_p\}$ denotes the Gaussian filtered version of pixel p. Since $\boldsymbol{\mu}_x$ is a constant for an image, the saliency value of each pixel can be