Color conspicuity map based on wavelet low-pass pyramid for popping out contours of salient objects

Zhiqiang Li, Tao Fang, Hong Huo, and Julian Zhu
Shanghai Jiaotong University, Institute of Image Processing and Pattern Recognition, Shanghai 200240, China
E-mail: lizhiqiangmy@yahoo.com.cn

Abstract. In Itti's model, which was one of the representative saliency models proposed in 1998, a Gaussian pyramid is used to analyze color information in scene images, and to generate a color conspicuity map. In this conspicuity map, some important objects can be located by salient areas, but their contours cannot be described clearly and perfectly. In this work, a wavelet low-pass pyramid is used to generate a color conspicuity map, and the contours of important objects pop out perfectly from salient areas. Experimental results validate the superiority of the proposed method. © *2010 Society of Photo-Optical Instrumentation Engineers.*
[DOI: 10.1117/1.3425655]

1 Introduction

In recent years, many saliency models have been proposed to simulate human visual attention. These models have gathered much attention for the successful and rapid search of important objects in scene images[1–7] According to different salient features from image segmentation[5–7] or direct pixels of an image,[1–4] most previous saliency models can be divided into two categories: region- and pixel-based approaches. As for region-based saliency, a representative saliency model was presented by Aziz and Mertsching.[5] In this model, the original image is segmented into fragments by using an image segmentation algorithm first. Based on the segmented image, a color contrast map is generated in terms of the color theory of human vision, a symmetry map is constructed using a novel scanning-based method, and a size contrast map is generated using a new algorithm proposed by Aziz. Eccentricity and orientation maps are computed using the moments of segmented regions, respectively. As for pixel-based saliency,[1–4] a classic and representative saliency model is from Itti, Koch, and Niebur.[1] In this model, each input image is processed in parallel through three feature channels including intensity, color, and orientation. The outputs of these channels are ultimately combined to form a saliency map, which indicates the locations of important objects.

In Ref. 5, Aziz and Mertsching compared their results with the results of the model in Ref. 1 and concluded that the salient regions in the color conspicuity map from the model by Itti, Koch, and Niebur could not reflect the contour of important objects. In our experiment, as is seen

later, a similar phenomenon can also be found. To solve this problem, a new version of generating color conspicuity maps based on wavelet low-pass pyramids, instead of the Gaussian pyramid in the Ref. 1 model, is proposed in this work. In addition, several heuristic techniques are used to improve the results. In the color conspicuity map generated by our method, the contours of important objects pop out perfectly from salient areas.

2 Algorithm for Generating Color Conspicuity Map

Let R, G, and B be the red, green, and blue components of the input image, respectively. Let $r(x,y)$, $g(x,y)$, and $b(x,y)$ denote the values at location (x,y) in R, G, and B channels, respectively. An intensity image I is produced by

$$i(x,y) = [r(x,y) + g(x,y) + b(x,y)]/3, \tag{1}$$

where $i(x,y)$ is the intensity value at location (x,y) in I. When the intensity value of a pixel in a scene image is very small, the color information of the pixel is hardly perceived. Thus, when $i(x,y)$ is smaller than $1/10$ of the maximum over the whole intensity image I, the values of $r(x,y)$, $g(x,y)$, and $b(x,y)$ are set to be zero. In terms of R, G, and B, four new color component images are constructed and denoted by RN, GN, BN, and YN, respectively. Let $rn(x,y)$, $gn(x,y)$, $bn(x,y)$, and $yn(x,y)$ denote the values at location (x,y) in RN, GN, BN, and YN, respectively. RN, GN, BN, and YN are defined as follows:

$$rn(x,y) = r(x,y) - [g(x,y) + b(x,y)]/2, \tag{2}$$

$$gn(x,y) = g(x,y) - [r(x,y) + b(x,y)]/2, \tag{3}$$

$$bn(x,y) = b(x,y) - [r(x,y) + g(x,y)]/2, \tag{4}$$

$$yn(x,y) = g(x,y) + r(x,y) - |r(x,y) - g(x,y)| - b(x,y). \tag{5}$$

When the values of $rn(x,y)$, $gn(x,y)$, $bn(x,y)$, and $yn(x,y)$ are negative, these values are set to zero. In terms of RN, GN, BN, and YN, four image pyramids $RN(k)$, $GN(k)$, $BN(k)$, and $YN(k)$ are constructed respectively, where k is the number of levels in the pyramid. Unlike the model of Itti, Koch, and Niebur, in this work, a wavelet low-pass pyramid is used to generate an image pyramid instead of the Gaussian pyramid.

Before generating a color conspicuity map, the color feature map needs to be constructed. Let $RN(c)$, $GN(c)$, $BN(c)$, and $YN(c)$ denote the images on level c, respectively. Let $RN(s)$, $GN(s)$, $BN(s)$, and $YN(s)$ denote the images on level s, respectively. $c \in \{2,3,4\}$ and $s = c + p$, $p \in \{3,4\}$. To generate color feature maps, these images are resized to a finer size, respectively. In the Ref. 1 model, the resizing size is level 4 of the pyramid. In this work, the finer size is level c in the size of the pyramid image. When the size of the original image is too small and the pyramid images from original image are resized to the size of level 4, the size of the resized pyramid images may be smaller than 1×1. Let $RN(s)^*$, $GN(s)^*$, $BN(s)^*$, and $YN(s)^*$ be the resized images, respectively. Let $rn(s)^*(x,y)$, $gn(s)^*(x,y)$,
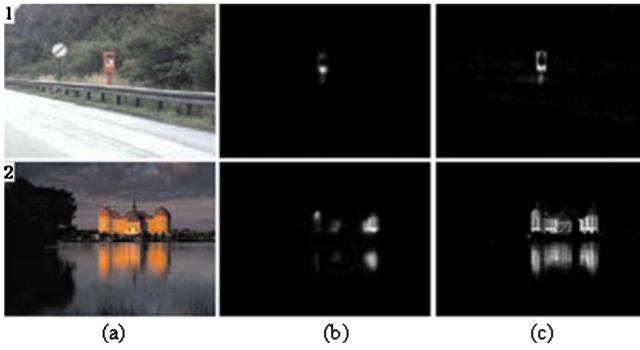
Fig. 1 (a) Input images. Row 1: the image from Ref. 1. Row 2: the image from Ref. 8. (b) The color conspicuity maps generated by Gaussian pyramid. (c) The color conspicuity maps generated by wavelet low-pass pyramid. (Color online only.)

$bn(s)^*(x,y)$, and $yn(s)^*(x,y)$ denote the values at the location (x,y) in $RN(s)^*$, $GN(s)^*$, $BN(s)^*$, and $YN(s)^*$, respectively. Let $rn(c)(x,y)$, $gn(c)(x,y)$, $bn(c)(x,y)$, and $yn(c) \times (x,y)$ denote the values at location (x,y) in $RN(c)$, $GN(c)$, $BN(c)$, and $YN(c)$, respectively. 12 color feature maps denoted by $RG(c,s)$ and $BY(c,s)$ can be generated by

$$rg(c,s)(x,y) = \left| [rn(c)(x,y) - gn(c)(x,y)] - [rn(s)^*(x,y) - gn(s)^*(x,y)] \right|, \quad (6)$$

and

$$by(c,s)(x,y) = \left| [bn(c)(x,y) - yn(c)(x,y)] - [bn(s)^*(x,y) - yn(s)^*(x,y)] \right|, \quad (7)$$

where $rg(c,s)(x,y)$ and $by(c,s)(x,y)$ are the values at location (x,y) in $RG(c,s)$ and $BY(c,s)$, respectively.

In the model by Itti, Koch, and Niebur, a color conspicuity map is constructed by integrating all color feature maps that are resized to the size of level 4. In this work, the color feature maps are resized to the size of the original image and then are used to generate color conspicuity maps by additional operations. Thereafter, a smart skill is adopted as follows: we square each element of the color conspicuity map to generate a new map. Because of this square operation, the range of values in the color conspicuity map is stretched, which results in few redundant salient areas in the color conspicuity map.

Wavelet low-pass pyramid. The level 0, the base of the pyramid, is the original image. The i'th-level image is obtained from the $(i-1)$'th-level image by: 1. translating the $(i-1)$'th-level image by wavelet transform, and 2. extracting the low-pass part of the resultant image from 1. as the i'th-level image.

Gaussian pyramid. The level 0, the base of the pyramid, is the original image. The i'th-level image is obtained from the $(i-1)$'th-level image by: 1. using a Gaussian filter kernel to convolute the $(i-1)$'th-level image, and 2. downsampling the image from 1.
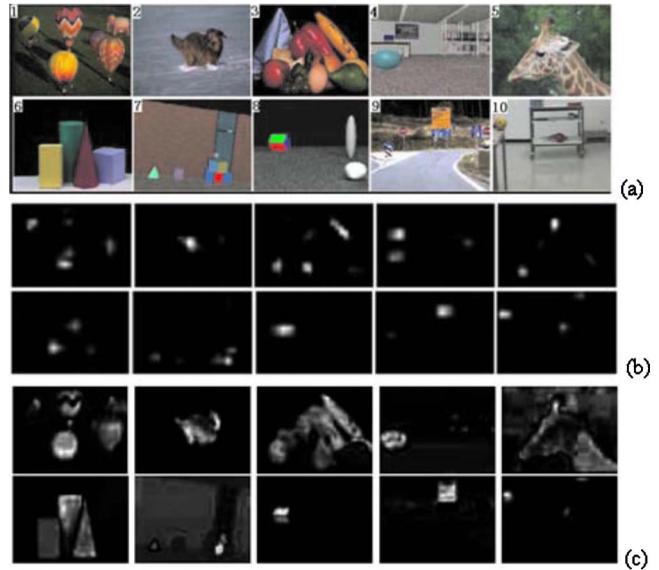

Fig. 2 (a) Input images from Ref. 5. (b) The color conspicuity maps generated by Gaussian pyramid. (c) The color conspicuity maps generated by wavelet low-pass pyramid. (Color online only.)

3 Experimental Results

In our experiment, original images from Refs. 1, 5, and 8 are used as input images to generate color conspicuity maps. These images are shown in Figs. 1(a) and 2(a). Because the color conspicuity maps based on the Harr wavelet function are similar to the ones based on other wavelet functions, only the color conspicuity maps based on the Harr wavelet function are generated. Figures 1(b) and 2(b) represent the color conspicuity maps from the Ref. 1 model based on the Gaussian pyramid. It is easy to see that all salient regions in these color conspicuity maps point to important objects, such as the red telephone, golden building, balloon, dog, etc. However, the contours of these important objects cannot be represented by the salient regions. These salient regions have many fragments that cover part of the object. Figures 1(c) and 2(c) show the color conspicuity maps based on a wavelet low-pass pyramid. Comparing the color conspicuity maps with the ones in Figs. 1(b) and 2(b), the contours represented by salient regions based on the wavelet low-pass pyramid are better than ones based on the Gaussian pyramid. These objects, such as the postbox,
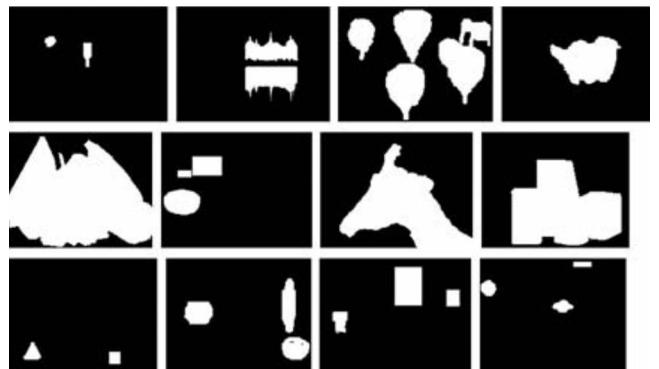

Fig. 3 Labeled maps corresponding to Figs. 1(a) and 2(a).

Table 1 Comparison data corresponding to criterion.

Image	S area	Our model H units	Our model F units	Ref. 1 model H units	Ref. 1 model F units
Fig. 1(a).1	709	562	91	442	72
Fig. 1(a).2	56,639	56,092	110	23,184	105
Fig. 2(a).1	7488	7011	12	3216	10
Fig. 2(a).2	2986	2704	15	712	8
Fig. 2(a).3	14,874	10,896	35	1032	23
Fig. 2(a).4	1859	862	13	348	124
Fig. 2(a).5	7646	7608	8215	331	16
Fig. 2(a).6	8992	6802	14	316	12
Fig. 2(a).7	426	367	125	339	134
Fig. 2(a).8	2234	214	12	314	19
Fig. 2(a).9	1973	384	26	279	21
Fig. 2(a).10	611	127	6	213	15

building, animal, etc., can be distinctly perceived from the color conspicuity maps in Figs. 1(c) and 2(c).

For the quantitative comparison of our results with the ones from Ref. 1, a comparison criterion is set. All of the labeled maps corresponding to images in Figs. 1(a) and 2(a) are manually annotated in the following way: 20 individuals are invited to view an image. The time for viewing the image is one second. After that, the individuals are asked what they see first. If one object in the image is noticed first by most of the individuals, this object is viewed as the salient one and labeled to generate a labeled map, as shown in Fig. 3. In the labeled map, the white areas represent the salient areas. Then, the salient fragments in color conspicuity maps from our model and the model from Itti, Koch, and Niebur are analyzed. If the salient units (pixels) in the conspicuity map are also salient in the labeled map, they are defined as hit units, else as false units. In the same image, if the number of hit units is nearly the same in these two models, then the smaller the number of false units, the better the model. If the number of false units is nearly the same, then the bigger number of hit units corresponds to the better model. The comparative results are listed in Table 1. The term "S area" means the number of the labeled salient units (the white pixels in the labeled map), "H units" is the number of hit units, and "F units" is the number of false units. As a whole, the results indicate that our model excels the model in Ref. 1.

4 Conclusion

The salient regions in the color conspicuity map from the model from Itti, Koch, and Niebur based on the Gaussian pyramid can only cover part of each object. In this work, the proposed method based on a wavelet low-pass pyramid generates the color conspicuity map, and the salient regions in this map can describe the contours of the objects in an image perfectly. Therefore, our method has the ability of improving the quality of scene analysis and object search.

References

1. L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998).
2. X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," *IEEE Intl. Conf. Computer Vision Patt. Recog.*, pp. 1–8 (2007).
3. J. K. Tsotsos and S. M. Culhane, "Modeling visual attention via selective tuning," *Artif. Intell.* **78**(1), 507–545 (1995).
4. B. A. Olshausen, C. H. Anderson, and D. C. V. Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information," *J. Neurosci.* **13**, 4700–4719 (1993).
5. M. Z. Aziz and B. Mertsching, "Fast and robust generation of feature maps for region-based visual attention," *IEEE Trans. Image Process.* **17**(5), 633–644 (2008).
6. C. Carson, S. Belongie, H. Greespan, and J. Malik, "Blobworld: image segmentation using expectation maximization and its application to image querying," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(8), 1026–1038 (2002).
7. J. Z. Wang, J. Li, and G. Wiederhold, "Simplicity: semantics-sensitive integrated matching for picture libraries," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9), 947–963 (2001).
8. See http://bcmi.sjtu.edu.cn/~houxiaodi/.