TELKOMNIKA, Vol.11, No.11, November 2013, pp. 6956~6963
e-ISSN: 2087-278X

A New Detection Model for Saliency Map

Zhongshan Chen[1,2]
[1]School of Information Engineering, Yancheng Institute of Technology, Yancheng, China
[2]Display Center, School of Electronic Science & Engineering, Southeast University, Nanjing, China
e-mail: ycddzjb@163.com

Abstract

Visual attention is a mechanism which filters out redundant visual information and focus on the most relevant parts when observing an image, many bottom-up computational models of visual attention have been devised to get the saliency map for an image. In this paper, a new visual attention model is proposed. Based on the experimental results obtained in this study, as compared with existing bottom-up visual model, the proposed model has better visual detection performance and low computational complexity.

Keywords: *visual attention mechanism (VAM), saliency map, image feature, Itti's model, bottom-up*

1. Introduction

Visual attention is a mechanism which can effectively eliminate the interference of redundant visual information and focus on the interested region of an image when observing an image, which significantly reduce the complexity of the information processing and improve the processing speed. In order to simulate which elements of a visual scene are likely to attract the attention of human observers, many domestic and foreign researchers put forward many visual attention models were proposed. There are two different approaches in visual attention mechanism: bottom-up and top-down process [1, 7, 11, 16]. The bottom-up process, which is data-driven and task-independent, can be considered as a function of primitive selective attention in HSV since human selectively attends such a salient area according to various lower visual features in input scene. The top-down process, on the other hand, is a directed task-driven process of focusing attention on one or more objects which are relevant to the observer's goal, which is related to the recognition processing influenced by the prior knowledge such as the feature distribution of the target, the context of the visual scene, and so on [12, 13, 15]. In most cases, the top-down process is thought to be based on the context provided by bottom-up process.

This study focused on the bottom-up approach. During the past several decades, many computational models of visual attention have been proposed [2-6], [8, 10], [14-15]. In the 1980s, Treisman developed the well-known Feature-Integration Theory (FIT) [16]. Itti et al. devised a visual attention model based on the behavior and the neuronal architecture of the primates' early visual system [10]. Harel et al. proposed a graph-based visual saliency (GBVS) model by using a better dissimilarity measure for saliency based on Itti's model [6]. Hou et al. devised a saliency detection model based on a concept defined as spectral residual (SR) [8]. Guo et al. found that Hou's model was caused by phase spectrum and they designed a phase-based saliency detection model [5], which model achieves the final saliency map by inverse Fourier transform (IFT) on a constant amplitude spectrum and the original phase spectrum of the image. Bruce et al. described visual attention based on the principle of maximizing information [2]. Liu et al. used the technology of machine learning to achieve the saliency map for images [14]. Gao et al. calculated the center-surround discriminant for saliency detection [3]. The saliency value for a location is obtained by the power of a Gabor-like feature set to discriminate the center-surround visual appearance [3]. Gopalakrishnan et al. built a saliency detection model based on the color and orientation distributions in images [4]. Recently, a saliency detection model by Valenti et al. is advanced through calculating the center-surround differences of edges, color, and shape for images [15]. The rest of this paper is organized as follows. In Section II, we elucidate the three steps model encompassing extraction of early

Received June 3, 2013; Revised August 1, 2013; Accepted August 14, 2013

visual features, Construction of the conspicuity maps and establishment of saliency map. In Section III, shows the experiment results by comparing the proposed model with other existing methods. The final section concludes the paper by summarizing our findings.

2. The Proposed Visual Attention Model

Itti's model is believed as a classic by many scholars. According to HVS attention mechanism, Itti's model combined the visual feature extraction, center-surround difference and normalization, linear combination to measure region saliency. Itti's model is strongly influenced by the feature, the feature maps are obtained through calculating the multi-scale center-surround differences for color, orientation,and intensity channels [10]. The final saliency map is achieved by the mean value of these feature maps, as shown in Figure1.

However, the saliency maps obtained by Itti's models are not accurate enough. Figure 2 shows the result of Itti's model [20]. The result shows that the saliency maps contain some error points which are out of saliency object.

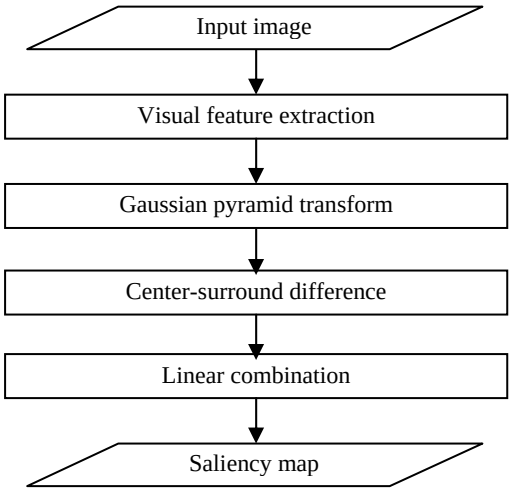


Figure 1. Itti's Model

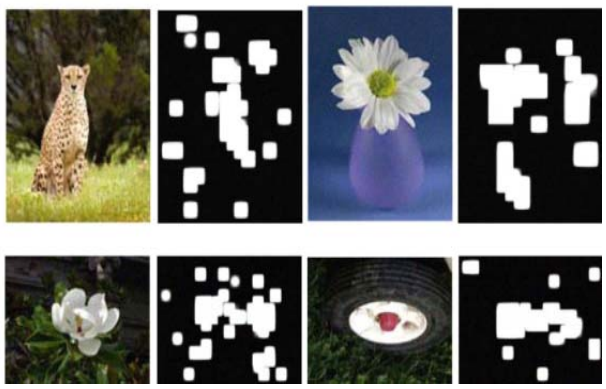


Figure 2. Saliency Map Outcomes (Itti's Model)

To solve this problem, an improved model based on Itti's model is proposed to obtain the saliency map. The proposed saliency map extraction overall work flow is depicted in Figure 3.

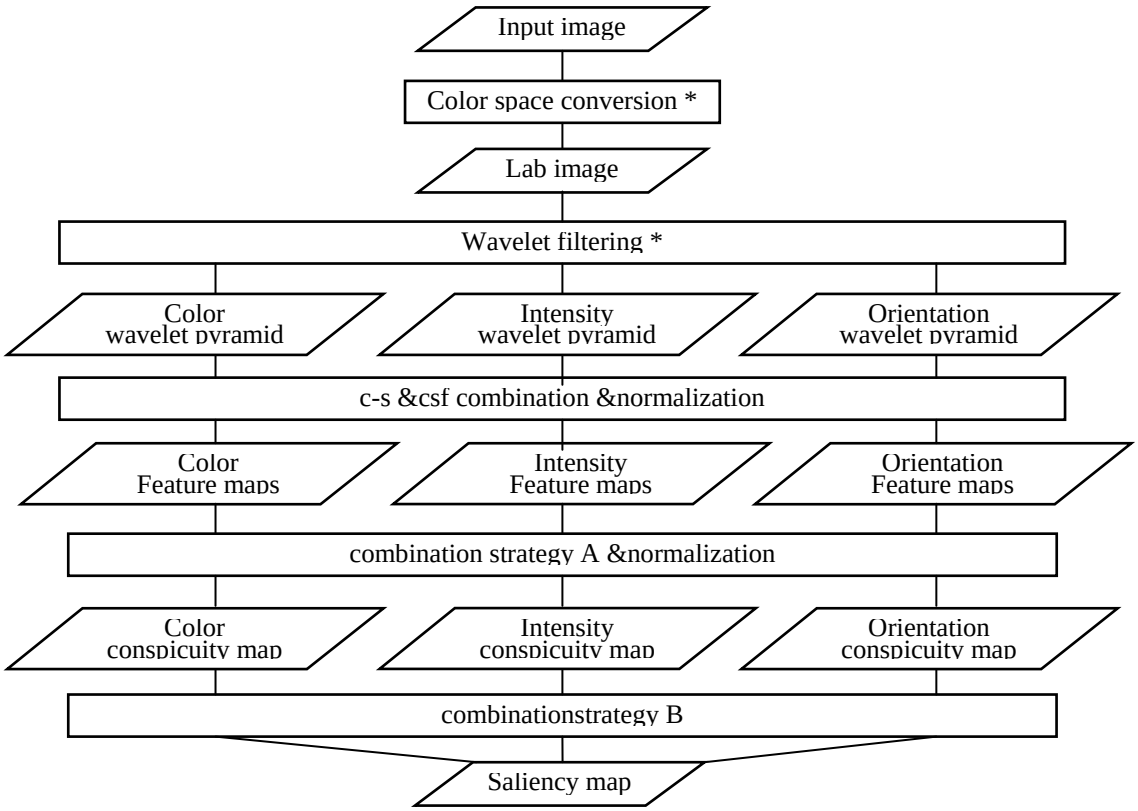


Figure 3. An Improved Model based on Itti's Model

The steps are described as follows:

Step 1: Color space conversion

The RGB model, depending on the equipment, is mainly used to color display, which is difficult for image processing and analysis. This study selects the CIELab model. The CIELab model isn't relevant to the equipment, so it is more accord with HVS. The CIELab model is conversed from RGB model as the Equation (1):

$$L = 0.2126 * R + 0.7152 * G + 0.0722 * B$$
$$a = 1.4749 * (0.2213 * R - 0.3390 * G + 0.1177 * B) + 128$$
$$b = 0.6245 * (0.1949 * R + 0.6057 * G - 0.8006 * B) + 128 \quad (1)$$

Step 2: Visual Feature Extraction

R, G and B are red, green and blue channels of the input image respectively, then:

The intensity feature I is equal to L. Two color channels are created: a is for red/green and b is for blue/yellow. It has been proved that these two color pairs can cover the entire visible light [9].Orientation features are obtained from I by a set of Gabor filters as Equation (2):

$$G(\lambda,\theta,\varphi,\sigma,\gamma,x,y)=\exp\left(\frac{x'^2+\gamma^2*y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\varphi\right)\right) \quad (2)$$

Where, x'= x*cos θ+y*sin θy'= - x*sin θ+y*cos θ

The parameters in this study are as follows: aspect ratioγ=1, standard deviation$\sigma=\pi$, phaseφ=0, 8 orientations $\theta\in\{0,$**Error!**,**Error!**,**Error!**,**Error!**,**Error!**,**Error!**,**Error!**$\}$,which is the preferred orientation.

There are 40 Gabor filters with 5 scales & 8 directions, each line is for the same scale, each column is for the same direction in this study.

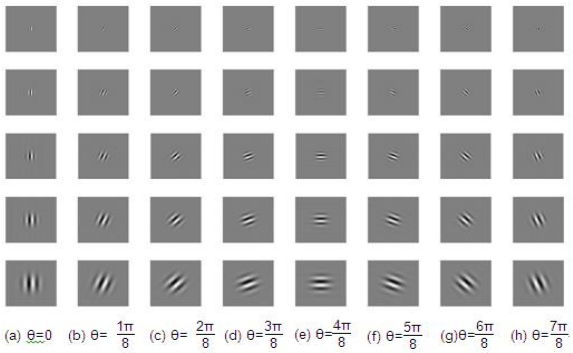


Figure 4. Images of 40 Gabor Filters (5 scales & 8 orientations)

Step 3: multi-scale pyramid

Gaussian multi-scale pyramid isn't accorded with multi-channel decomposition properties of HVS. As wavelet transform utilizes the different sensitivity of HVS in the response frequency band and spatial orientation selection, it decomposes the original image into independent frequency bands and different spatial orientation. And wavelet transform is more in accordance with HVS. Therefore, we replaced the Gaussian pyramid in Itti's model with wavelet-based multi-scale transform. The scale factor of wavelet decomposition is smaller, which shows that the signal frequency is higher and the details of the image are given. The scale is bigger, the signal frequency higher and the rough of the image.

Step 4: Center-Surround Difference

Attention is firstly attracted by the most significant part in an image according to HVS attention mechanism. In this approach, Itti's visual attention model use difference of Gaussian (DOG) with Equation (3) to calculate the most salient point for the purpose of the measurement of the interest of a region.

$$DOG(x,y)=\frac{1}{2\pi\sigma_c}\exp\left(-\frac{x^2+y^2}{2\sigma_c^2}\right)-\frac{1}{2\pi\sigma_s}\exp\left(-\frac{x^2+y^2}{2\sigma_s^2}\right) \quad (3)$$

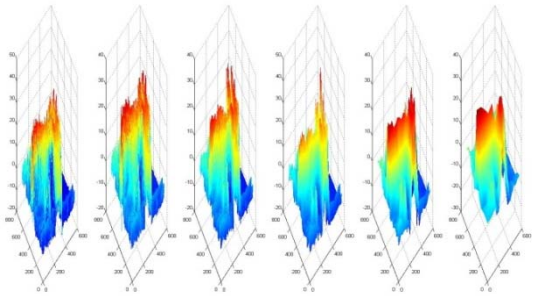
The DOG is simply the difference between two Gaussian distributions with different σ values, where the center c of the model corresponds with the excitatory center Gaussian, and the surround s to the inhibitive surround Gaussian.

Step 5: combination strategies

Now the problem we confronted is how to combine these feature maps into 4 conspicuity maps (L, a, b, O) which show the saliency of each feature and combine 4 conspicuity maps into a saliency map. Combination strategy is very critical. Itti proposed linear combination strategy, which obtain the saliency by calculating the average of 3 conspicuity maps simply.

A: combination between the same features

According to the similarity between the feature maps of the same visual feature as shown in Figure 5.



(a) scale=1 (b) scale=2 (c) scale=3 (d) scale=4 (e) scale=5 (f) scale=6

Figure 5. The Similarity between the Multi–scale Graphs of the Same Visual Feature

The feature weights Wi are calculated by the following method:

(1) Normalize all feature maps to the range [0 1], in order to eliminate across-scale amplitude differences.

(2) Convert each map into the corresponding frequency spectrogram.

(3) For each pixel P, calculate its visual spatial frequency f by Equation (4):

$$f = \sqrt{f_x^2 + f_y^2} \quad (4)$$

Here, $f_x = \frac{u'}{\alpha}$ $f_y = \frac{v'}{\alpha} u' = \left|u - \frac{m}{2}\right|$ $v' = \left|v - \frac{n}{2}\right|$

Where,f_x is horizontal spatial frequency,f_y is vertical spatial frequency, α is angle of view, whichdepends on the experiment, m & n is the size of the image, (u v) is the value of the point Q in the spectrum graph corresponding to the point P in the map.

(4) For each pixel P, calculate its the contrast sensitivity function A(f) as Equation (5), proposed by Manos and Sakrison, as show Figure 6:

$$A(f) = 2.6*(0.0192+0.114*f)*e^{-(0.114*f)^{1.1}} \quad (5)$$

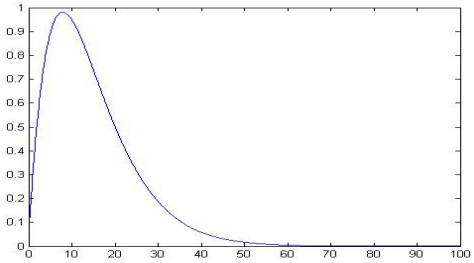


Figure 6. Contrast Sensitivity Function

(5) The weight Wi of each feature map is the averagevalue of all A(f) as Equation (6):

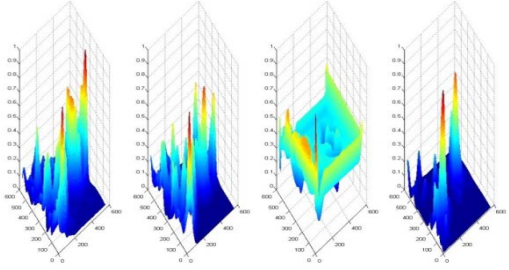
$$W_k = \frac{1}{m*n} \sum_{j=1}^{n} \sum_{i=1}^{m} A(f)_{(i,j)} \quad (6)$$

Where m and n are the size of a map.

(6) The conspicuity map C is given as Equation (7), which combines all feature maps F of a same visual feature.

$$C = \frac{\sum_{i=1}^{m*n} W_i * N(F_i)}{\sum_{i=1}^{m*n} W_i} \quad (7)$$

Where m & n are respectively the centre & surrounding factor N(.) is a normalization operator.



(a)L (b)RG (c)BY (d)O

Figure 7. The Dissimilarity between the Conspicuity Graphs of the Different Visual Feature

B: combination between the different features

According to the dissimilarity between the conspicuity maps of the diffrent feature, as shown in Figure 7.

(1) Normalize each conspicuity map to the range [0 1], in order to eliminate amplitude differences due to dissimilar feature extraction mechanisms;

(2) For each conspicuity map, find its global maximum M and the average$\overline{m}$ of all the local maxima except for M;

(3) Globally multiple the conspicuity map as Equation (8):

$$w_i = (Mi - \overline{mi})2 \qquad i \in \{L, a, b, O\} \tag{8}$$

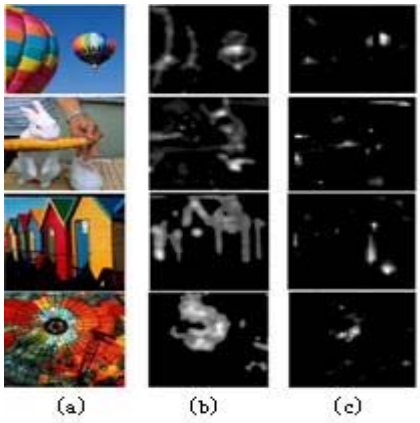
The total saliency S is given as Equation (9), which is a combination of intensity conspicuity map L, two color conspicuity maps a and b, an orientation conspicuity map O.

$$S = \frac{\sum_{i\in\{L,a,b,O\}} \omega_i * C_i}{\sum_{i\in\{L,a,b,O\}} \omega_i} \tag{9}$$

3. Experimental Results

The test images, 120 natural color images with various contents, are selected from the Corel Photo Library. The size of the images is 256*256. The simulation is performed by Matlab2010 on Intel(R) 2.4GHz-Microsoft Windows XP platform.

From the experimental results, as the illustrated example ("balloon", "rabbit", "house" and "dome") shown in Figure 8, it is seen that our proposed model produces the better visual effect than Nitti's model.



(a) original images, (b)saliency map by the proposed method, (c)saliency maps obtained by Itti's model

Figure 8. Performance Comparison:

In order to verify the credibility of the algorithm, we track and record the real eye movements by eye movement tracking experiments. Experimental settings are shown in Figure 9. We use iView X infrared eye movement tracking system (analysis software: iView v1.7.39) fromSensoMotoric Instruments company (Germany), the display is Dell 19 inch monitor (M993). There are 18 participants, participants are normal vision, no color blindness or color weakness. The testers need to ensure that the head fixed during the test, the eyes and picture center is located at the same horizontal line, the observation distance is 4 times the monitor height, about 1.2 meters. 100 natural color images with various contents have been selected. The image is displayed on the center of the screen, the rest part of the screen is filled with blank background, and they are seen in random order by the testers for 10 seconds each in a viewing period.

Figure 9. iView X Infrared Eye Movement Tracking System in the Experiment

Figure 10 shows the eye tracking results from 4 illustrated mages ("balloon", "rabbit", "house" and "dome"), the concern degree of highlight areas is higher. Through comparing the eye tracking results with the concern degree by our method, as shown in Figure 10, the saliency by our method indeed is very lose to ROIs in the image.

(a) original images, (b) the concern degree by our method, (c) theeye tracking results of the experiments

Figure 10. Performance Comparison

4. Conclusion

In this study, a new visual attention model is proposed. Based on the experimental results, as compared with Itti's model, the proposed model has better performance and low computational complexity. Because the proposed model has low computational complexity, the proposed model can be employed to process multimedia in real time.

References

[1] J Braun, D Sagi. *Vision outside the focus of attention*. *Percept. Psychophys*. 1990; 48(1): 45-48.
[2] ND Bruce, JK Tsotsos. Saliency based on information maximization. *Adv. Neural Inf. Process. Syst*. 2006; 18: 5-162.
[3] D Gao, N Vasconcelos. *Bottom-up saliency is a discriminant process*. Proc. IEEE Int. Conf. Computer Vision. 2007.
[4] V Gopalakrishnan, Y Hu, D Rajan. Salient region detection by modeling distributions of color and orientation. *IEEE Trans. Multi-media*. 2009; 11(5): 892-905.

[5] C Guo, Q Ma, L Zhang. *Spatio-temporal saliency detectionusing phase spectrum of quaternion Fourier transform*. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition. 2008.
[6] J Harel, C Koch, P Perona. Graph-based visual saliency. *Adv.Neural Inf. Process. Syst*. 2006;19:545-552.
[7] Dongjian He, Yongmei Zhang, Huaibo Song. A Novel Saliency Map Extraction Method Based on Improved Itti's Model. IEEE *Computer Society*. 2012; 3: 323-327.
[8] X Hou, L Zhang. *Saliency detection: A spectral residual approach.* Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition. 2007.
[9] LM Hurvich, D Jameson. An opponent-process theory of color vision. *Psychological Review*. 1957; 64(6): 384-404.
[10] L Itti, C Koch, E Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 1998; 20(11): 1254-1259.
[11] L Itti. Models of bottom-up and top-down visual attention. Ph.D. dissertation, Dept. Comput. Neural Syst.. California Inst. Technol., Pasadena. 2000.
[12] C Kanan, M Tong, L Zhang, G Cottrell. SUN: *Top-down saliency using natural statistics. Visual Cognit.*, 2009; 17(6): 979-1003.
[13] Z Lu, W Lin, X Yang, E Ong, S Yao. Modeling visual attention's modulatory aftereffects on visual sensitivity and quality evaluation. *IEEE Trans. Image Process*. 2005; 14(11): 1928- 1942.
[14] T Liu, J Sun, N Zheng, X Tang, HY Shum. *Learning to detect a salient object*. Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition. 2007.
[15] A Torralba, A Oliva, MS Castelhano, JM Henderson. Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychol. Rev*., 2006; 113(4): 766-786.
[16] A Treisman, G Gelade. A feature-integration theory of attention. *Cognit. Psychol*., 1980;1 2(1): 97-136.