EXTRACTING REGIONS OF ATTENTION BY IMITATING THE HUMAN VISUAL SYSTEM

Fei Qi, Jinjian Wu and *Guangming Shi*

School of Electronic Engineering, Xidian University
Xi'an, Shaanxi, 710071, P. R. China

ABSTRACT

Detecting and segmenting out the regions of interest (ROIs) is one of the foundations in image processing and analysis. Because the final information sink of images is human, for segmenting out the ROIs effectively, we need to study human visual system (HVS) and imitate the behaviors when human viewing a scene. Researchers have found serveral factors which affect human attentions by studying eye movements when one views an image. In this paper, a method is proposed to detect the ROIs automatically based on HVS. In the proposed algorithm, the properties of pixels such as the contrast, location and edges are analyzed, and the pixels are enhanced according to the sensitivity of HVS. Then these factors are combined to a salient map, which classifies each pixel of the image in relation to its perceptual importance. Finally, the ROIs are segmented according to the salient map. This algorithm is easy to work, and can segment the objects from complex background efficiently.

Index Terms— region of interest, character of image, local contrast, global contrast, salient map

1. INTRODUCTION

In image processing and computer vision, it is an important problem to segment out regions of interest (ROIs) from images automatically. The most difficult work is how to detect ROIs based on human visual attention, though it's simple for human beings. Researchers studied the movements of human eyes and found characters of the image which affect the response of human visual attention system. These characters can be classified into low level and high level factors [1, 2].

Researchers have worked on how humans percept visual information for decades and several methods for extracting ROIs based on HVS are proposed. In the early years, most methods are based on image segmentation working on one characteristic of the image [3]. These methods are ineffective to images which have complex background. Osberger and Maeder [4] introduce an algorithm which simulate human visual attention system. But the success of this algorithm is limited to the segmentation which segments the image into small regions for characters analyzing. Laurent Itti *et al*. [5] use three features, color, orientation and intensity, to construct the salient map which can locate the salient objects. In Itti's model, they use the Gaussian pyramid to analysis the texture of the image. This algorithm can locate the salient object effectively. But with the down-sampling operator, it loses some details of the salient objects, such as boundaries of the objects. SooYeong Kwak *et al*. [6] and Ko Byoung Chul *et al*. [7] proposed algorithms using attention window (AW) to detect the salient objects. Both of these algorithms use Itti's model to construct the salient map. When there are multi-objects separated in the image, the AW will not work(see figure 5 and 6 for example).

We introduce a novel algorithm which directly analyzes individual pixel. We analyze three factors, i.e. contrast, edge and location, to get the characteristic charts of the image. Then we combine these characteristic charts to form a salient map and segment the ROIs based on it. The main contribution of our algorithm is: first, it is easy to work and can detect multi-objects in an image; second, it can locate the salient objects and get the boundaries of the salient objects effectively. We test a lot of images, our algorithm is hopeful. Our algorithm have a higher accuracy and a broader application scope, and it can be used to automatically segment the ROIs effectively in various images with complex background.

In section 2, factors which affect the human visual attention are explicated in detail. Modeling the salient map and segmenting the ROIs are introduced in section 3. The experimental results are showed in section 4. In section 5, conclusions are drawn.

2. THE HUMAN VISUAL ATTENTION SYSTEM

2.1. Factors influencing eye movements and attention

One of the main functions of human visual attention system is to decrease the uncertainty when watching the scene. Selective visual attention plays an important role in visual attention operation. It reduces the computation load of vision processing and raises the speed of vision processing. Researchers have found low level and high level factors which affect human visual attention system.

This work is supported by National High Technology Research and Development Program of China under Grant No. 2007AA01Z307, and National Natural Science Foundation of China under Grant No. 60736043.

The low factors generally can quickly guide human visual attention system and feed forward some information of images. Some low factors have been found as follows:

1) Contrast [1]. Researchers indicate that human visual attention system converts the luminance of the image into contrast at an early stage of processing. Pixels stand out from surrounding will be more attractive, for example, while viewing a lawn, a bright flower in it will certainly attract our attention.

2) Shape [2]. People pay more attention to those objects whose shapes are long and thin, rather than objects which have the same area and contrast. Edge reflects the shape of object, so pixels in the edge attract more attention than others.

3) Other low level factors that have been found which influence human visual attention include size, color, motion, brightness, location and so on [4, 1, 2].

The high level factors involve some feedback process from memory and may involve template matching. Researchers have found some high level factors as follows:

1) Context [2]. When viewing an image, the brain will control the movement of the visual attention focus by dealing with the context of the image and find the ROIs according to the feedback information.

2) Location [4]. Eye movement experiments have shown that observers firstly focus on the center 25% of a scene to find objects, namely, we pay more attention to the center of images.

3) Foreground/background [4]. Visual attention experiments indicate that human pay more attention to the foreground to find the objects, and it may be much more likely to neglect objects at the background.

4) People [1]. Human visual attention system is more familiar with information about human body and is very sensitive to it. Human's face, arm and other parts in images attract more viewer attention when viewing images. It's a very salient high level factor.

2.2. Choosing factors influencing visual attention system

We only analyze three characters, i.e. intensity contrast, object edge and location, to form the salient map of the image and segment the ROIs based on the salient map. The model is shown as fig. 1.

In all of these factors which influence the human visual attention system, some may be more salient than others. Such as the motion factor, it's a very sensitive character. But what we deal with in this paper is static picture, so we do not take it into account. The factor of 'people' is also very sensitive to human visual attention system, but it is too complex to be analyzed. We also have not considered some factors which have relation with region for what we analyze here is individual pixel.

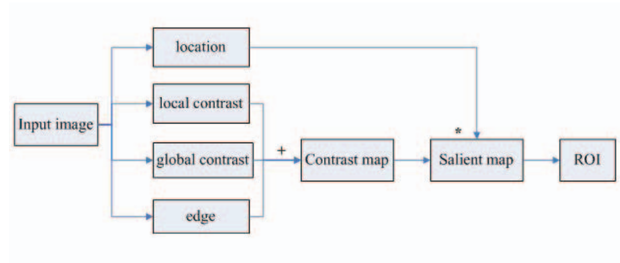

Fig. 1. Block diagram for ROIs detecting

3. ALGORITHM FOR THE ROIS SEGMENTATION

3.1. Algorithm for modeling salient map

To form a salient map, we need to analyze some characters to get the importance of each pixel. The basic operations of our algorithm are shown as fig. 1.

First, we calculate the intensity contrast. If an object stands out from its surrounding, it will certainly attract our attention. It means that if pixels are different from its surrounding or different from the global image, they may be attractive. Every pixel's intensity contrast to the global image is calculated as follow [2]:

$$G_{\text{contrast}}(x, y) = \frac{|L_m(x, y) - L_M|}{L_m(x, y) + L_M} \tag{1}$$

where $L_m(x, y)$ is the mean gray within a 7×7 surrounding of the center pixel, L_M is the mean gray of the whole image.

The Differential of Gaussian (DoG) function indicates the different between the center to its surrounding. We can use the DoG function to calculate the contrast between the pixel to its surrounding.

$$I(\sigma_i) = I \otimes G(\sigma_i) \tag{2}$$

$$F_{\text{contrast}}(x, y; \sigma_1, \sigma_2) = |I(x, y; \sigma_1) - I(x, y; \sigma_2)| \tag{3}$$

where I is the input image, $G(\sigma)$ is the Gaussian function.

We can get the local intensity contrast from equation 3. The effect of the result is depending on the value of σ_1 and σ_2. In this paper, we set $\sigma_1 = 1$. There are some relationship between the size of the objects and the value of σ_2. When the value of σ_2 is small, this algorithm detects small objects in the image more effectively. And with a big value of σ_2, it detects big objects more effectively. From experiment we know that when the value of σ_2 is about half of the object size, it can detect the object much well. Human visual attention system is not sensitive to small regions (less than 100 pixels). So we set $\sigma_2 = 5$ (with this value, it can detects small objects well) or σ_2 equal to a quarter of the image size (with this value, it can detects big objects better) to form two local contrast maps.

We combine the two maps with global non-linear normalization followed by summation, denoted by $N(\cdot)$, into a local contrast map [3]. This normalization operator consists as follows:

- Normalize the value of the images to a fixed range.
- Find the global maximum value(M) and average value of all the local maximum values($\bar{m}$).
- Globally multiplying the map by$(M - \bar{m})^2$.

This normalization operator can globally promote maps which have a small number of strong peaks, and globally suppress maps which contain numerous comparable peaks.

The local contrast map is constructing as following:

$$L_{\text{contrast}}(x, y) = \sum_{i=2}^{3} N\left(F_{\text{contrast}}(x, y; \sigma_1, \sigma_i)\right). \qquad (4)$$

Edge contains massive information of the image. It is a primary factor of understanding objects in the image. We use the Sobel operator to segment the edge information.

At last, we combine the global contrast map, local contrast map and the edge map to form the contrast map(CM) of the image.

3.2. Algorithm for modeling weighting matrix

We choose the high level factors, location, as the weighting factor. As we know, people pay more attention on the center of the image. We found an equation which suits well with the location character known to influence attention from our studying. Each pixel's weight of location is calculated as follow:

$$W_{\text{center}}(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \text{center}, \\ \frac{1+\cos^2 \frac{\pi r}{2R}}{2} & \text{else.} \end{cases} \qquad (5)$$

where "center" is the central $\frac{1}{4} \times \frac{1}{4}$ of the image, r is the distance from the pixel (x, y) to the central region, and R is the distance from the edge to the central region.

3.3. Salient map and ROIs regions

Now we have acquired the contrast map and the weight of each pixel the of the image. We define the salient map as follow:

$$SM(x, y) = CM(x, y)W_{\text{center}}(x, y) \qquad (6)$$

In the salient map, each pixel's luminance indicates the importance to human visual attention. In other words, the pixel with higher luminance attracts more attention. So we set a suitable threshold to segment pixels which have higher luminance. In our model, the salient map has almost indicated the boundaries of the objects. So after a series of operators, we can segment the ROIs of the image effectively.

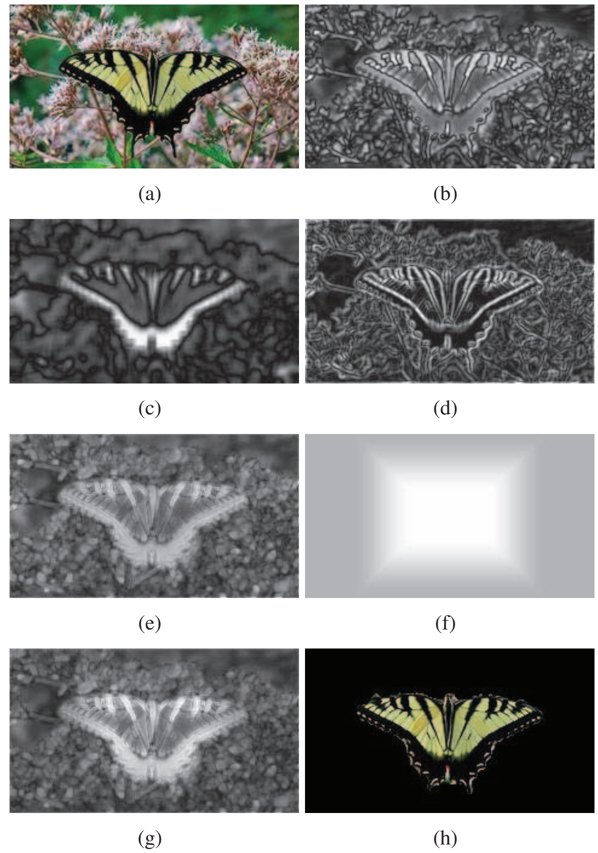

Fig. 2. The ROIs segment of butterfly. (a) original image, (b) local contrast, (c) global contrast, (d)edge , (e) contrast map, (f) location map, (g) salient map, (h) ROIs

4. EXPERIMENTS AND DISCUSSION

We analyze a wide variety of images and from the results we can see that the salient map can indicate the ROIs effectively.

Fig. 2 demonstrates the processes how we segment the ROIs of an image. Fig. 2(a) is the original image(butterfly). Fig. 2(b) demonstrates the difference between pixels from its surrounding. Fig. 2(c) is the intensity contrast of pixels to the global image. Fig. 2 (d) is the edge map. Fig. 2(e) is the contrast map which is form by the combination of fig. 2(b), (c), (d). Fig. 2(f) is the location feature map. Fig. 2(g) is the salient map. This salient map is very different from Itti's model. We can just see the rough locations of the salient objects in Itti's model. But in our model, we can see the salient object's(butterfly) location and its boundary clearly from the salient map. Fig. 2(h) demonstrates a satisfying result of the ROIs segmentation. The butterfly in the image is segmented clearly from the background.

Figure 3, 4 and 5 show some results of ROIs segment by using this algorithm. All of these salient maps can demonstrate the salient objects locations and boundaries very clearly. In figure 3 and figure 4, the salient objects are locating in the

(a) (b)

Fig. 3. Skating boy. (a) original image, (b) salient map

(a) (b)

Fig. 4. Woman. (a) original image, (b) salient map

center of the image, our algorithm detects the salient objects effectively from the background. In figure 5, there are two salient objects and both of them are not locating in the center of the image. From the fig 2(b), the salient map, we can see that it indicate the objects well. We can segment the ROIs effectively based on these salient maps. Figure 6 shows an image with more complex background and multi-objects and its salient map.

The algorithm we used to segment the ROIs combines 3 factors which affect human visual attention, and it is based on individual pixel. Though the calculation here is simple but effective, it can segment the ROIs successfully.

5. CONCLUSION

In this paper, we introduce a novel method to segment the ROIs automatically. The algorithm is based on some productions from which scientists research human visual attention system while viewing scenes. This algorithm has been designed in a flexible manner which can accommodate to modifications and be calculated easily. It's efficient to both simple and complex images.

However, many areas for improvement still remain. Our method uses only intensity feature to form the contrast map, we do not take the factor of color into account. When forming the salient map, we just take four factors into account. While the objects have a slight contrast from their surrounding, this algorithm will be invalidation. We are working on some other factors to deal with this problem.

(a) (b)

Fig. 5. Dog. (a) original image, (b) salient map

(a) (b)

Fig. 6. Girl. (a) original image, (b) salient map

6. REFERENCES

[1] J. Findlay, "The visual stimulus for saccadic eye movement in human observers," *Perception*, vol. 9, pp. 7–21, 1980.

[2] J. Senders, "Distribution of attention in static and dynamic scenes," in *Proceedings SPIE*, San Jose, Feb 1997, vol. 3016, pp. 186–194.

[3] Claudio M. Privitera and Lawrence W. Stark, "Algorithms for defining visual region-of -interesting: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–982, 2000.

[4] Wilfried Osberger and Anthony J. Maeder, "Automatic identification of perceptually important regions in an image," in *Proc. Int'l Conf. Pattern Recognition*, 1998, pp. 17–20.

[5] Itti laurent, Koch Christof, and Niebur Ernst, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, 1998.

[6] SooYeong Kwak, ByoungChul Ko, and Hyeran Byun, "Automatic salient-object extraction using the contrast map and salient points," in *Advances in Multimedia Information Processing — PCM 2004*. 2004, vol. 3332 of *LNCS*, pp. 138–145, Springer Berlin.

[7] Ko Byoung Chul and Nam Jae-Yeal, "Automatic object-of-interest segmentation from natural images," in *Proc. Int'l Conf Pattern Recognition*, 2006, pp. 45–48.