Salient Region Detection based on Frequency-tuning and Region Contrast

Fu Li-hua and Guo Liang
College of Computer Science
Beijing University of Technology
Beijing, China
sxlfguoliang@126.com, fulh@bjut.edu.cn

***Abstract*—Frequency-tuned saliency detection analyzes image saliency from the perspective of frequency domain and fully combines image segmentation method, which outputs well-defined boundaries of salient objects. However, the method ignores spatial relationships across image parts. This paper proposes an improved saliency detection method on the basis of the frequency-tuned method. In this method, we first segment the input image into regions and then analyze the image from the frequency domain. After that, we preprocess it using Gaussian filter to eliminate noise and coding artifacts. For each region, we can get saliency map in region-level based on region contrast. Finally, salient regions are selected by "winner-take-all" (WTA) neural network and Inhibition of Return(IOR) mechanism. The proposed salient region detection algorithm combines the virtues of frequency-tuning and region contrast. The experimental results show the feasibility and validity of this algorithm.**

Keywords-Image retrieval; Semantic gap; Saliency detection; Region-level saliency map; Visual attention

I. INTRODUCTION

With the rapid development of digital image technology, image data increases exponentially. How to effectively and rapidly process image data has been becoming a hot topic of image processing field[1]. CBIR(Content-Based Image Retrieval) is one focus in the research field[2]. CBIR can be divided mainly into three steps: index construction based on low-level features, distance calculation and the query. CBIR considers image content and has strong objectivity. But it does not necessarily reflect or match high-level semantic information, that is to say, there is a gap between low-level features and high-level semantic namely Semantic Gap[3].

The involved techniques to fill the semantic gap include: Relevance Feedback[4], Image Segmentation[5] and Saliency Detection[6]. Facing a complex scene, human can routinely and effortlessly judge the importance of image regions, and focus attention on important parts. Image retrieval technique based on the visual saliency selects critical region which can cause the most user interest and most closely represents the image content simultaneously to describe the image. Saliency detection can fill the semantic gap to some extent, which opens up many possibilities for image understanding.

There are two kinds of saliency detection models: fast, bottom-up ones and relatively slow, top-down ones. This article mainly studies the slow, top-down ones. Recently many computational models have been proposed for saliency detection. These models can broadly be categorized as biologically motivated ones, purely computational ones or their combination.

The method put forward by Itti et al.(IT method for short)[6] is one of the classic saliency detection algorithms based on the biological plausible architecture given by Koch and Ullman[7]. They defined image saliency using central-surrounded differences across multi-scale image features. IT method produces saliency map which is just $1/256^{th}$ the original image size in pixels, thus the salient region obtained does not definitely match the region human has interest in. Moreover, the high calculation complexity is not optimistic.

The Frequency Tuning Method presented by Achanta et al.(FT method for short)[8] is a kind of models based on purely computation. FT method defines pixel saliency using a pixel's color difference from the average image color. The method can output the saliency map with the same size of the original image. It has better performance in preserving all the information of frequency domain of the image. However, FT method only considers first order average color and ignores the influence on the saliency detection from spatial relationships across image parts and that is the key defect [9].

II. SALIENCY DETECTION MODEL

A. IT Method

Itti et al. base their method on the biological principles and this is one of the most classical saliency detection models. The basic idea of their method is described as below: First, the low-level features of the colors, intensity and orientations from the input image are extracted. Each feature map is obtained by the center-surround differences after nonlinear normalization, and then the total saliency map is given by the linear combination of the feature maps. Finally, the visual attention shift is realized by the WTA neural network and the IOR mechanism on this basis[10].

IT method mainly has four steps: visual feature extraction, calculation of the feature map, the combining operations of the feature maps and the visual attention shift. The structure is shown in *Fig. 1*.

B. FT Method

Achanta et al. relied on frequency domain processing to compute saliency. They analyzed the spatial frequency content from the original image. Let ω_{lc} be the low frequency cut-off value and ω_{hc} be the high frequency cut-off value. In

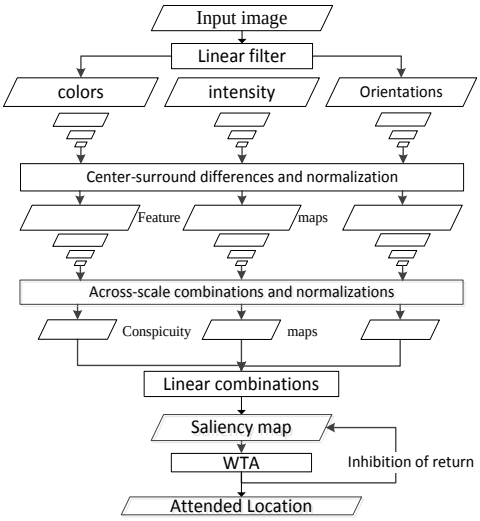


Figure 1. Itti's visual attention model

order to highlight large salient objects, ω_{lc} has to be low. This also helps highlight salient objects uniformly. To have well defined boundaries, ω_{hc} has to be high. However, to avoid noise, texture patterns and coding artifacts, the highest frequencies need to be disregarded.

For getting a saliency map containing a wide range of frequencies, they combined the outputs of several *DoG* (Difference of Gaussians) band pass filters with contiguous $[\omega_{lc}, \omega_{hc}]$ pass bands.

Thus, the saliency of each pixel p is calculated as:

$$S(p) = \left\| I_{\mu} - I_{\omega hc}(p) \right\| \tag{1}$$

where I_{μ} is the average feature of the image, $I_{\omega hc}(p)$ is the feature of pixel p in L^*a^*b color space after the image is smoothed by Gaussian method, and $\|\cdot\|$ is the L_2 norm.

III. OUR WORK

As mentioned above, IT method has not only the high calculation complex but also the low resolution of saliency map. Sometimes the salient objects cannot be detected. Based on image frequency, FT method can avoid the influence on saliency detection from the disturbance factors such as texture, noise and coding artifacts.

Considering IT method and FT method, an obvious fact could be found that the utilization of edge features in the FT method is more completed than IT method. IT method simply obtains salient regions on the basis of the saliency map obtained by detection. The edge features are obviously ignored so that the final saliency regions often present a series of local fragments. As a comparison, FT method makes full use of the results of image segmentation to select the regions according to the saliency. Therefore, the better edge information of the salient object is acquired using the FT method.

However, from another point of view, FT method inevitably has two weaknesses:

(1) FT method omits the influence from the spatial relationship across the regions of the image which plays an important role in human visual attention. The high contrast between the adjacent regions leads to more attention than that between the distant ones.

(2) When taking the spatial relationship into consideration, the judgment of the saliency on pixel level in FT method undoubtedly results to the large cost of calculation.

This article focuses on promoting FT method by considering region-level saliency. There are two steps in this method: calculation of the region-level saliency map and visual attention shift.

A. Calculation of Region-level Saliency Map

FT method analyzes image from the aspect of frequency. Equation (1) shows the process for calculating the saliency value of each pixel. That is to say, the contrast between each pixel and the average information of image after gaussian smoothing acts as the saliency of each pixel. There are two major deficiencies about the FT method. Therefore, the procedure of our method can be described as follows: First, segment the image adopting the group-based segmentation algorithm and simultaneously get the image after Gaussian smoothing. Then, for each region r_k, determine the saliency by the calculation of contrasts between other image regions and r_k. The algorithm is given in *Fig. 2*.

In this article, the contrast $S(r_k)$ of region r_k is defined as follow:

$$S(r_k) = \sum_{r_k \neq r_i} \exp\left(-D_s(r_k, r_i)/\sigma_s^2\right) \omega(r_i) D_r(r_k, r_i) \tag{2}$$

where $D_r(r_k, r_i)$ is the distance between region r_k and r_i in the L^*a^*b color space after the Gaussian method is used to smooth the image, and $D_s(r_k, r_i)$ is the spatial distance. σ_s controls the spatial weight and $\omega(r_i)$ is the number of pixels in region r_i.

The algorithm calculates the distance $D_r(r_k, r_i)$ between r_1 and r_2 as:

$$D_r(r_1, r_2) = \sum_{i=1}^{m} \sum_{j=1}^{n} D(I_{1,i}, I_{2,j}) \tag{3}$$

where I is the feature value in L^*a^*b color space of the image after the Gaussian smoothing and region r_1 and r_2 respectively have m pixels and n pixels.

In (2), we overcome the lack of spatial relationship in FT method by introducing spatial distance $D_s(r_k, r_i)$ as a weight. And the second weight $\omega(r_i)$ emphasizes the contrast of regions with more large areas to make some error points eliminated.

Region Contrast Method (RC method for short) defines saliency similar to our method which is one of the saliency

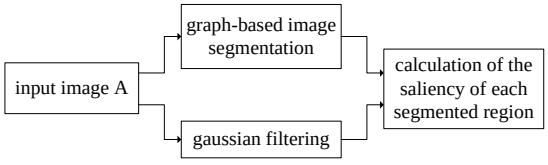


Figure 2. Calculation of the saliency map in the region-level

detection algorithms with best performance. However, we define the region-level saliency $S(r_k)$ from the frequency domain instead of the contrast of each region calculated by pixels in the initial image adopted by the RC method. From the analysis above, it is easy to find that some noise is involved in the original input image. The degree of the image variation is characterized by frequency domain of the image which also indicates the distribution of the gradient of image. For the purpose to reduce the interference of the detail information such as texture, the *DoG* band pass filtering is used to eliminate the high-frequency noise[8]. Thus, we employ the distance among regions in *L*a*b* color space after the Gaussian smoothing to calculate the region contrast.

B. Visual Attention Shift

IT method uses the WTA neural networks and the IOR mechanism to extract the regions of interest containing the salient object successively. IT method circles the salient regions of the input image with the closed curve according to the salience value successively. But these regions are irregular and not easy for the post-processing. So, in our method, the visual attention shift is based on the saliency map in region level. We refer to the visual attention shift in IT method to extract the salient rectangle regions of the multiple-target image one by one. In more detail, this visual attention shift approach can be realized as the following steps shown in *Fig. 3*:

(1) Calculate the average saliency S_{mean} of the region-level saliency map.

(2) Filter the segmented regions with the strategy $S>2S_{mean}$, and get the binary image containing multiple salient objects.

(3) Combine the saliency map and the binary image to correct the saliency map.

(4) Obtain the current region with the most interest based on the WTA neural network.

(5) Extract the region of interest in the form of rectangle whose horizontal edges are respectively the horizontal lines in which the two pixels among all the pixels with the maximum ordinate and the minimum ordinate are and whose vertical edges are respectively the vertical lines in which the two pixels among all the pixels with the maximum abscissa and the minimum abscissa are.

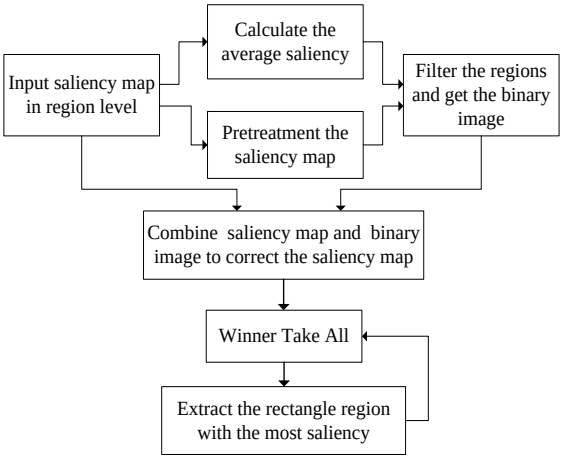


Figure 3. Visual attention shift

(6) Avoid the secondary treatment to the past salient regions by the IOR in vision research.

IV. Experiments and Discussions

For the purpose to verify the validity of the method, two experiments are designed for all kinds of images: the experiment on the calculation of saliency map and the experiment on visual attention shift. Both the experiments realize the algorithm in VS 2010 programming environment in the computer with the Pentium IV 3.00GHz CPU, 512MB memory and Windows XP operating system.

A. Calculation of Saliency Map

We utilized the public test set provided by Achanta et al.[9] to testify the validity of the calculation of saliency map. The method was compared with the IT method, FT method, HC method and RC method and we derived the good result as is shown in *Fig. 4*.

From the *Fig. 4*, some conclusions are observed: IT method obtained the definition of the saliency of image from the biology. But the features extracted from the model were partial so that the salient regions detected were often the smaller regions with the high contrast. For instance, it can get the edge of some big objects while it had an unsatisfactory performance inside the region. FT method defined the saliency of image from the frequency domain. It obtained the frequency domain information about the image by a summation over *DoG* filter and the saliency regions were well reflected. But the spatial relationship between different parts of image was ignored. HC method calculated the saliency by the feature difference of the pixels, but it omitted the spatial relationship among different parts of image. RC method segmented the image into small regions and then calculated the global contrast in region units. The factors such as colors were taken into consideration. However, RC method adopted the original input image to calculate the contrast of regions. It was not analyzed from the frequency domain so that the texture, noise and coding artifacts still had the large influence on the saliency detection. Our method combined the FT method and the RC method to detect the saliency map, which considered both the information from the frequency domain and the influence on saliency from the spatial relationship. And we finally obtained the good results of the saliency detection.

B. Visual Attention Shift

In the second experiment, we attempted to do the visual attention shift experiment on the multiple-target images. The visual attention shift experiment in this article is mainly based on the saliency map in region-level. The visual attention shift mechanism in the IT method was referred to extract each rectangle region in the multiple-target image successively.

Figure 5 shows that the visual attention shift on the saliency map acquired by IT method only gave the approximate regions of the salient objects while the attention

shift on the region-level saliency map by our method can indicate the rectangle regions containing the salient objects. In conclusion, the saliency map based on IT method only told the approximate scale of the salient regions so that its results of visual attention shift did not precisely give the description of the edge of the salient object. By contrast, our effect of visual attention shift well contained the whole salience object in that the saliency map obtained by our method took frequency information of image into consideration and the image segmentation thought over the edge information of the objects in image.

V. CONCLUSIONS

This article combines the advantage of FT method and RC method. A salient region detection algorithm taking frequency information and spatial relationship into consideration is finally proposed. The experiment on the calculation of saliency map indicates that this algorithm can get the saliency map with more edge information. And the experiment on the visual attention shift gains the rectangle regions containing the whole salient object. Our salient region detection method lays the foundation for implementing image retrieval and image compression which are based on the salient regions.

ACKNOWLEDGMENT

This work is supported by the Beijing Municipal Education Commission Foundation (No. 007000546311501).

REFERENCES

[1] M. Flickner, H. Sawhney, W. Niblaek, et al. Query by image and video content: the QBIC System. IEEE Computer, 1995, 28(9): 23-32.

[2] S.K. Chang, A. Hsu. Image information systems: where do we go from here? IEEE Transaction on Knowledge and Data Engineering, 1992, 5(5): 431-442.

[3] A.W.M. Smeulders, M. Worring, S. Santini. Content-based image retrieval at the end of the early years. Pattern Analysis and Machine Intelligence, 2000: 1349-1380.

[4] Y. Lv, C. Zhai. Positional relevance model for pseudo-relevance feedback. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval, 2010: 579-586.

[5] M. Frucci, G. Baja. From segmentation to binarization of gray-level images. Journal of Pattern Recognition Research. 2008, 3(1): 1-13.

[6] L. Itti, C. Koch, E. Niebur. A mode of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

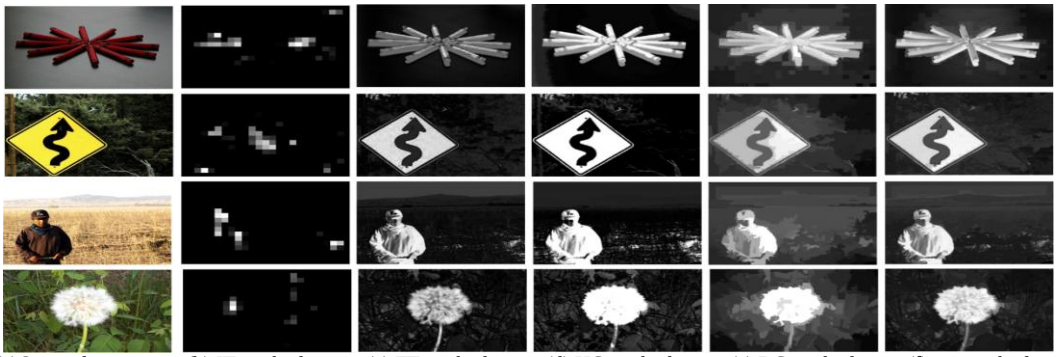
[7] C. Koch, S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology, 1985, 4(4): 219-227.

[8] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk. Frequency-tuned salient region detection. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1597-1604.

[9] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, S. Hu. Global contrast based salient region detection. In 2011 IEEE Conference on Computer Vision and Pattern Recognition, 2011: 409-416.

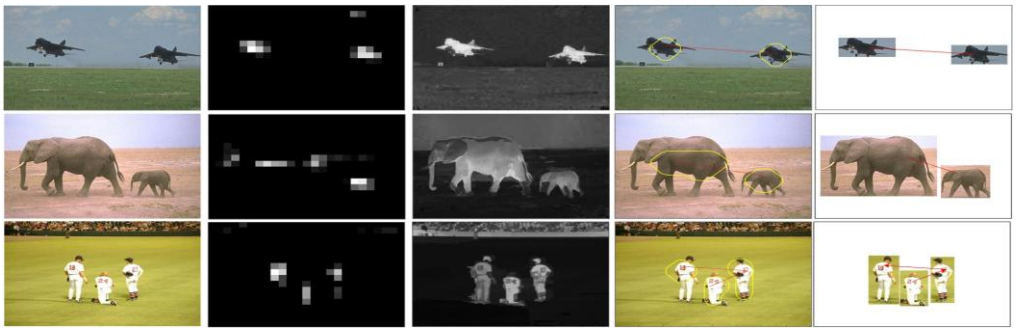
[10] L. Itti, C. Koch. Computational modeling of visual attention. Nature Reviews Neuroscience, 2001, 2(3): 194-203.

[11] P. Felzenszwalb, D. Huttenlocher. Efficient graph-based image segmentation. IJCV, 2004, 59(2):167-181.



(a)Original image (b) IT method (c) FT method (d) HC method (e) RC method (f) our method

Figure 4. Visual attention shift



(a) Original image (b) IT method (c) our method (d) attention shift by IT (e) by our method

Figure 5. Visual attention shift