Saliency detection based on the fusion of spatial and frequency domain analysis

Yongcai Pan*, Yuwei Zhang, Yanxia Wei, Qingzheng Liu
School of Electrical and information Engineering
Guangxi University of Science and Technology
Liuzhou, China
The author's e-mail:panyongcai@163.com

Abstract—**Saliency detection is a challenging direction in the field of computer vision. To get the more accurate salient target, we propose a visual saliency detection method based on the fusion of spatial and frequency domain analysis. Firstly, the image is smoothed by a Gaussian filter and then segmented by the superpixel segmentation of SLIC. The spatial saliency map is calculated by the color distance. Secondly, the saliency map of frequency domain is obtained by hypercomplex Fourier transform. Finally, the point multiplication operation is adopted to merge the spatial and the frequency domain saliency map. Compared with the others methods, the model we proposed can well retain the complete contour, highlight the target and suppress the background.**[1]

Keywords- saliency detection;SLIC; HFT; frequency domain

I. Introduction

Visual attention mechanism plays an important regulatory role in the human visual system, which enables the human eye to acquire important information, transfer to and maintain the attention to this information efficiently. Saliency detection is the most important initial step in visual attention mechanism. It can quickly extract the regions of human visual interest from environmental background. Image saliency detection is widely used in image segmentation, target detection, image retrieval, image compression and other fields. Saliency detection is a low-level visual processing task in human visual system and also a preprocessed process. Effective saliency detection is of great significance and wide applicable value for computer vision tasks of higher level such as image analysis and recognition and visual application [1-2]. After more than 20 years of development, the saliency detection has obtained some achievements, but compared with the sophisticated human visual system, the present saliency detection model is still far behind, the accuracy is low and the detection efficiency needs to be improved, and has a great promotion space. Saliency detection will continue to be a hot research topic in computer vision and artificial intelligence in a long period of time in the future.

Saliency detection model can be divided into two types of spatial domain and frequency domain according to the different treatment. The spatial domain models process pixels directly, transform and filter the color value, grey value, then get the salient area. The frequency domain models transform the image into the frequency domain firstly, and then deal with transformation and filtering operation based on frequency variable, get the target area through inverse transform lastly. The research results of saliency detection are briefly described according to those two models.

In the field of saliency detection, the earliest model was proposed by Itti [3] in 1998, which is the first model that can be implemented by computational method. The image saliency map can be obtained based on the difference between the centre and the edge by using the feature map of image direction, color and intensity. In 2006, Harel optimized the Itti model, based on the idea of graph theory, he simulated the Markov random field to obtain the saliency map, namely GBVS model [4]. However, the salient areas detected by these methods cannot highlight the target well and remove the background completely. In 2006, Bruce and Tsotsos proposed an information maximization visual attention model (AIM) based on the principle of information maximization [5]. In 2008, Achanta proposed an AC algorithm [6], which is a simple and effective computing model, but this method ignores the important mechanism of biological vision. In 2011, Cheng proposed a global contrast HC algorithm [7], which defined the salient value as the difference between pixel colors by analyzing the global pixel value of the image, and obtained the computing speed by simplifying the color component. The salient target was relatively uniform, but it was not ideal for the image effect of complex texture background.

Compared with the saliency detection in the spatial domain, the detection in the frequency domain has the advantages of simpler model, faster calculation speed, simpler and adjustable parameters, etc. In addition, the frequency domain model conforms to the perception theory of human eyes. The earliest frequency domain model was the SR model put forward by Hou in 2007 [8]. This model believes that the image information is composed of non-significant part and significant part, and the significant part corresponds to the information in the spectrum residual of the image. The SR algorithm is simple and fast. In 2008, Guo only used the phase spectrum of Fourier transform to obtain saliency results, which was called PFT model [9], but ignoring the amplitude spectrum. Meanwhile, the PFT method is extended to PQFT for color images. In 2012, Hou proposed an image descriptor IS model [10], which only retained the symbol information in DCT for saliency detection. This method is not only relatively simple but also requires less

[1] This work is supported by Research Program of Science at Universities of GuangXi Autonomous Region(2018KY0323, 2017KY0358), Key Laboratory of Industrial Process Intelligent Control Technology of Guangxi Higher Education Institutes(IPICT-2016-07, IPICT-2016-08 and IPICT-2016-12).

978-1-7281-3584-7/19/$31.00 ©2019 IEEE

computation. Schauerte extended DCT algorithm to QDCT algorithm [11]. In 2013, Li proposed an HFT detection algorithm. The intensity feature and color feature used in this method are refer to the visual model mechanism of human eyes. HFT model use multi-scale filters in the frequency domain, and use the principle of the minimum entropy of image information to obtain the saliency map [12], then get the ideal effect.

But each method is not ideal, whether in spatial or frequency domain. There is still a large distance to the truth map. Observed that the method based on superpixel segmentation can retain the complete contour, while the method based on hypercomplex Fourier transform can highlight the target and suppress the background. In this paper, in order to make the saliency detection not only highlight the target, suppress the background, but also have a clear outline, we fuse the saliency maps extracted from the previous two methods.

II. Spatial saliency detection

A. Gaussian Smoothing

Gaussian filter is a kind of linear filter, which can effectively suppress noise and smooth the images. Its working principle is similar to the mean filter, which takes the mean value of the pixels in the filter window as the output. However, the coefficient of the window template is different from that of the mean filter. The template coefficient of Gaussian filter decreases with the increase of the distance from the template centre. Therefore, the Gaussian filter is less fuzzy than the mean filter. The two-dimensional Gaussian smoothing formula is as follows.

$$\mathrm{G(x,y)} = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

After smoothing, the sharp parts in the original image can be removed, which is conducive to the saliency detection based on spatial contrast and ensures the accuracy of the saliency of the target. Fig. 1 shows the result of Gaussian smoothing.

B. Superpixel Segmentation

SLIC (Simple Linear Iterative Clustering) is a simple segmentation algorithm proposed by Achanta [13]. This algorithm firstly converts color image into CIELAB image space, and uses the combination of brightness component, two color components and x and y coordinates in LAB color space to carry out k-means clustering for five-dimensional space vector. The algorithm steps are as follows.

- Set the number of superpixel segmentation, and calculate the number of pixel points contained in each superpixel, and initialize the cluster center C_k= [l_k, a_k, b_k, x_k, y_k] on the grid nodes with an interval of S;
- Select the pixel with the smallest gradient in the 3×3 neighborhood as the new clustering centre;
- Repeat the above steps;

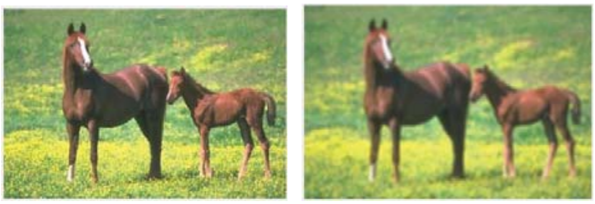

Figure 1. The result of Gaussian smoothing. From left to right are the original image and the smoothed image respectively.

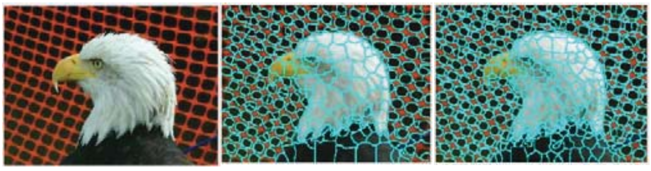

Figure 2. The result of SLIC. From left to right are the original image, the segmentation result whose blocks number is 300 and 600 respectively.

- For each clustering centre, pixel points are allocated in the $2S\times2S$ region of the clustering centre according to the distance;
- Recalculate the clustering centre and re-cluster, and calculate the distance of the two clustering centres;
- Repeat the iterative process, the clustering end if the distance between the two clustering centres is less than the default threshold.

The result of SLIC superpixel segmentation is shown in the Fig. 2. The blocks number has certain influence on segmentation effect and running speed. As can be seen from the figure, the target and background are divided into different parts after superpixel segmentation. The dividing line of the contour of the target is very clear. If it is applied to the saliency detection, a more accurate boundary can be obtained.

C. Saliency Detection Based on SLIC

The saliency of a pixel can be represented by the color distance. Compared with the general pixel-by-pixel processing of the image, after the superpixel segmentation, the image blocks composed of several pixels are processed in the subsequent processing, which greatly reduces the time-consuming and complexity of the algorithm.

The space distance d_{xy} and the color distance d_{lab} in LAB space are calculated as follows.

$$d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \tag{2}$$

$$d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} \tag{3}$$

The comprehensive distance can be calculated with the space distance and the color distance by an adjustment coefficient as shown in (4), in this paper, $w=0.1$.

$$D_s = \sqrt{w * \left(d_{xy}\right)^2 + (d_{lab})^2} \tag{4}$$

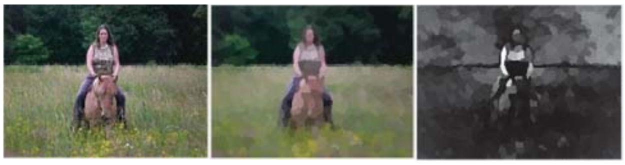

Figure 3. The result of spatial saliency detection. From left to right are the original image, the color filling image, SLIC Saliency map.

The saliency of one pixel is the sum of its distance from other pixels as shown in (5).

$$S(i) = \sum_{j=1}^{n} D_S(i,j) \tag{5}$$

The saliency calculation after segmentation is equivalent to the saliency of the segmentation block and the weight of the position. As can be seen from the Fig. 3, the saliency result can highlight the target, keep the original target and also have obvious boundaries, but there are still many background elements in the figure that are not suppressed.

III. Frequency domain saliency detection

A. Hypercomplex Fourier Transform（HFT）

As a mathematical tool, Fourier transform has been widely used in various fields of industry, especially in the field of signal processing. Based on the Fourier transform, the researchers proposed the hypercomplex Fourier transform, which can be used in the color images processing, such as RGB images. The color information of the RGB image is stored in each channel, so the quaternion can be used to represent the color image and realize mathematical operation, while the hypercomplex Fourier transform can transform the quaternion into the frequency domain for convenient processing. Equation (6) is the form of hypercomplex matrix.

$$f(n,m) = a + bi + cj + dk \tag{6}$$

where, i , j, k are imaginary units, and satisfy $ij = k$, $jk = i$, $ki = j$, $i^2 = j^2 = k^2 = -1$. In the discrete case, the (6) trans to (7) through the hypercomplex Fourier transform.

$$F_H[u,v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{\mu 2\pi\left(\frac{mu}{M}+\frac{nv}{N}\right)} f(n,m) \tag{7}$$

where, μ is a unit pure quaternion, and $\mu^2 = -1$. After the quaternion is transformed by the hypercomplex Fourier transform, the inverse transformation as shown in (8) can be used to realize the transformation from the frequency domain to the original domain.

$$f(n,m) = \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{\mu 2\pi\left(\frac{mv}{M}+\frac{nu}{N}\right)} F_H[u,v] \tag{8}$$

B. The Hypercomplex Form of Image

The saliency detection task of image is based on image features, so the quaternion form of combined image features can be used to represent the image, as shown in (9).

$$f(n,m) = w_1 f_1 + w_2 f_2 i + w_3 f_3 j + w_4 f_4 k \tag{9}$$

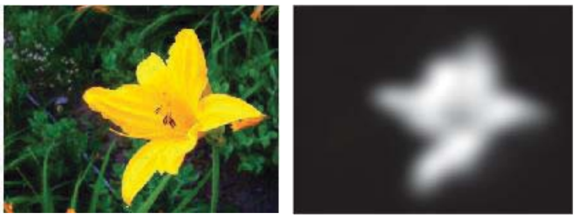

Figure 4. The result of HFT. From left to right are the original image, the saliency map.

where, w_1-w_4 is the weight of the features, $f_1 - f_4$ is the feature, f_1 is the motion feature, f_2 is the intensity feature, and f_3, f_4 are the color features. When processing the static image, f_1=0, other features are calculated as follows: $f_2 = I = (r+g+b)/3$, $f_3 = RG = R - G$, $f_4 = BY = B - Y$. $R = r - (g+b)/2$, $G = g - (r+b)/2$, $B = b - (r+g)/2$, $Y = (r+g)/2 - |r-g|/2 - b$, where, *r, g* , *b* respectively represent the red, green and blue channels of the input color image. When calculating, $w_1 = 0$, $w_2 = 0.5$, $w_3 = 0.25$, $w_4 = 0.25$. The polar coordinate form of the Fourier transform of the quaternion image is (10).

$$F_H[u,v] = \|F_H[u,v]\| e^{\mu\Phi(u,v)} \tag{10}$$

where, $\|\cdot\|$ represents the modulus of each element of the hypercomplex matrix. $F_H[u,v]$ is the hypercomplex transform of $f(n,m)$.

C. Saliency Map of HFT

After the 1-D continuous signal transformed by Fourier, it can be seen from the spectrum that the non-salient information has obvious spikes, and the salient information can be obtained by inhibiting the spikes after the inverse transform. Similarly, the quaternion form of the image combined features can be transformed into a hypercomplex, and the non-salient information of the image can be removed by the Gaussian function filter and the salient information can be retained. Here, different scales are used for filtering the amplitude spectrum, and the Gaussian kernel function is (11).

$$g(u,v;k) = \frac{1}{\sqrt{2\pi}2^{k-1}t_0} e^{-(u^2+v^2)/(2^{2k-1}t_0^2)} \tag{11}$$

where, t_0=0.5. The k is the spatial scale parameter, *k=1, 2..., K*, and *K* is determined by the size of the image, the image saliency map can be reconstructed using the amplitude spectrum filtered and phase spectrum. The saliency map $\{s_k\}$ corresponding to different spatial scales are

$$s_k = g * \left\| F_H^{-1}\{\Lambda_K(u,v)\mathrm{e}^{\chi P(u,v)}\} \right\|^2 \tag{12}$$

where, g is a fixed scale Gaussian kernel function. And the saliency map is chosen by the entropy minimum algorithm.

Fig. 4 shows the result of HFT. By repetitious experiments, results analysis and contrast, the saliency map obtained using the hypercomplex transformation can detect the objects of different sizes at the same time, and the effect is better. The background is suppressed while the target is highlighted, but the boundary is not clear.

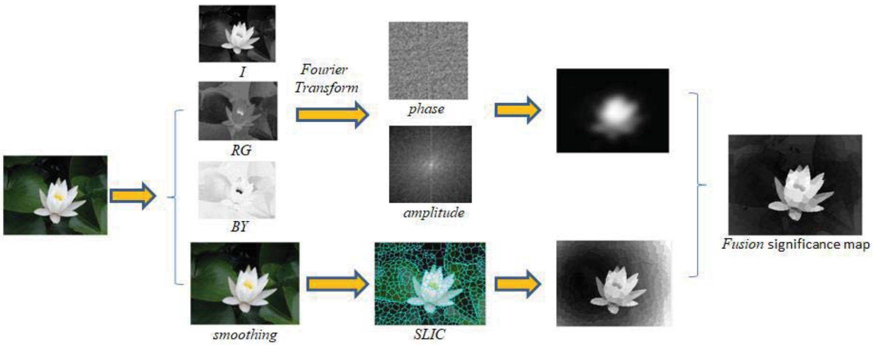

Figure 5. The flow chart and the results of our algorithm.

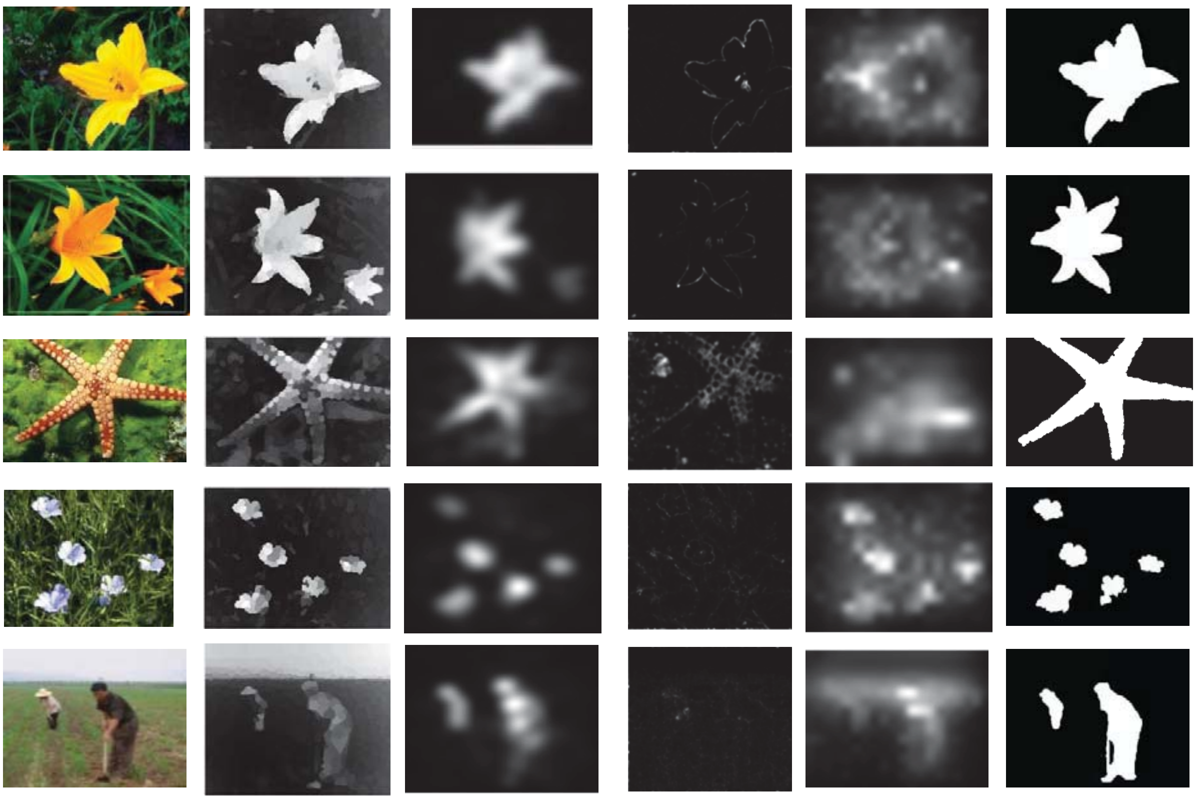

Figure 6. The results of different models. From left to right are the original image, the result of our model, HFT model, SR model, GBVS model, the truth saliency map.

IV. Fusion Saliency Map of Different Algorithms

The comparation between the two methods of saliency detection proved that their advantages and disadvantages are complementary. The method based on superpixel segmentation can retain the complete contour, while the method based on hypercomplex Fourier transform can highlight the target and suppress the background. In order to make the saliency detection not only highlight the target, suppress the background, but also have a clear outline, we use both saliency map of the tow methods. The specific flow chart and the results of the algorithm fusion is shown in Fig. 5. The saliency maps obtained based on superpixel segmentation and hyper-complex Fourier transformation are fused by point multiplication operation. The fusion expression is as follows.

$$S = S_{SLIC} .* S_{HFT} \tag{13}$$

The detection effect after fusion is better than that of both the two methods. Meanwhile, the results of the algorithm in this paper are compared with other models, as shown in Fig. 6. For multiple targets or the targets with different sizes, our proposed model has good detection results. When the gray value of the target does not differ much from the background, the detection effect needs to be improved, such as the last row.

V. Conclusion

Saliency detection is a low-level visual processing task, which plays an important role in high-level computer vision tasks such as image analysis and understanding, recognition and visual application. In this paper, a visual saliency detection method based on the fusion of spatial and frequency domain is used for saliency detection. Compared with other traditional models, the proposed model can highlight the target, suppress the background, and the results obtained are more accurate and stable, but also have clear boundaries. In the future, we will strive to improve the model, so that it can also have a better detection effect in less contrast environment.

References

[1] P. SHARMA, “Evaluating visual saliency algorithms: past, present and future,” Journal of Imaging Science and Technology, 2015, 59(5) : 50501-1-50501-17.

[2] M. MANCAS, V. P. FERRERA, N. RICHE, “The future of attention models: information seeking and self-awareness,” From Human Attention to Computational Attention, Springer Series in Cognitive and Neural Systems 10. New York: Springer, 2016: 447-459.

[3] L. ITTI, C. KOCH, E. NIEBUR, “A model of saliency based visual attention for rapid scene analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.

[4] B. SCHOLKOPF, J. PLATT, T. HOFMANN, “Graph-based visual saliency,” Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. Piscataway: IEEE, 2007: 545-552.

[5] N. D. B. Bruce, J. K.Tsotsos, “Saliency Based on Information Maximization,” Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada]. MIT Press, 2005.

[6] R. ACHANTA, S. HEMAMI, F. ESTRADA, “Frequency-tuned salient region detection,” Proceedings of the 2009 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1597-1604.

[7] M. M. CHENG, G. X. ZHANG, N. J. MITRA, “Global contrast salient region detection,” Proceedings of the 2011 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2011: 406-416.

[8] X. D. HOU, L. Q. ZHANG, “Saliency detection: An spectral residual approach,” Proceedings of the 2007 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2007:1-8.

[9] C. L. GUO, Q. MA, L. M. ZHANG, “Spatio-temporal saliency detection using phase spectrum quaternion Fourier transform,” Proceedings of the 2008 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.

[10] X. D. HOU, J. HAREL, C. KOCH, “Image signature: highlighting sparse salient region,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194 -201.

[11] B. SCHAUERTE, R. STIEFELHALGAN, “Predicting human gaze using quaternion DCT image signature saliency and face detection,” Proceedings of the 2012 IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE, 2012: 137-144.

[12] J. LI, M. D. LEVIN, X. AN, “Visual saliency based on scale-space analysis in the frequency domain,” IEEE Transactions on Pattern Recognition and Machine Intelligence, 2013, 35(4): 996-1010.

[13] R. Achanta, A. Shaji, K. Smith, “SLIC Superpixels Compared to State-of-the-Art Superpixel Methods,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11):2274-2282.