An improved saliency detection method based on hypercomplex Fourier transform

Yongcai Pan¹, Yuwei Zhang¹, Qingzheng Liu¹

¹ School of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou China, 545006

The author’s e-mail address: panyongcai@163.com

Abstract. Saliency detection is a challenging direction in the field of computer vision. It can help us quickly obtain important information of image and video, and plays an important role in image compression, image segmentation, target recognition and other applications. Based on the Hypercomplex Fourier Transform (HFT) model, median filter and threshold segmentation are employed to obtain the image salient region, which solves the problem that the boundary is not clear and multiple targets cannot be separated, and achieves more accurate and stable detection effect.

1. 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, it 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 center 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 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.

Observed that the saliency map of HFT model is vague and there is no clear boundary. For this matter, we calculate the salient area using classifications of the single-objective image and multi-objective image in this paper. Global threshold method is used in single-objective image processing, and the multi-objective images are dealt with median filters and the iterative threshold algorithm latterly, then the final saliency map is obtained.

2. Hypercomplex Fourier transform (HFT) model

2.1. Hypercomplex Fourier transform

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 (1) is the form of hypercomplex matrix.

\[ f(n, m) = a + bi + cj + dk \]  (1)

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 equation (1) trans to equation (2) through the hypercomplex Fourier transform.

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

where \( \mu \) 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 equation (3) can be used to realize the transformation from the frequency domain to the original domain.
\[ f(n, m) = \frac{1}{\sqrt{MN}} \sum_{n=0}^{M-1} \sum_{m=0}^{N-1} e^{\mu 2\pi \left( \frac{m v}{MN} + \frac{n u}{MN} \right)} F_H[u, v] \] (3)

2.2. 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 equation (4).

\[ f(n, m) = w_1 f_1 + w_2 f_2 i + w_3 f_3 j + w_4 f_4 k \] (4)

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:

\[ f_2 = l = (r + g + b)/3 \] (5)
\[ f_3 = RG = R - G \] (6)
\[ f_4 = BY = B - Y \] (7)
\[ R = r - (g + b)/2 \] (8)
\[ G = g - (r + b)/2 \] (9)
\[ B = b - (r + g)/2 \] (10)
\[ Y = (r + g)/2 - |r - g|/2 - b \] (11)

where \( r, g, b \) respectively represent the red, green and blue channels of the input color image. When calculating, set the weight \( w_1 = 0, w_2 = 0.5, w_3 = 0.25, w_4 = 0.25 \). The polar coordinate form of the Fourier transform of the quaternion image is

\[ F_H[u, v] = \|F_H[u, v]\| e^{\mu \Phi(u, v)} \] (12)

where, \( \|\| \) represents the modulus of each element of the hypercomplex matrix. \( F_H[u, v] \) is the hypercomplex transform of \( f(n, m) \). The amplitude spectrum \( A(u, v) \), phase spectrum \( P(u, v) \) and pure quaternion matrix \( \chi(u, v) \) can be calculated by the following equation after the transform.

\[ A(u, v) = \|F_H[u, v]\| \] (13)
\[ P(u, v) = \Phi(u, v) = \tan^{-1} \frac{\|v(F(u, v))\|}{S(F(u, v))} \] (14)
\[ \chi(u, v) = \mu(u, v) = \frac{v(F(u, v))}{\|v(F(u, v))\|} \] (15)

2.3. Multi-scale filtering

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, and the Gaussian kernel function is

\[ g(u, v; k) = \frac{1}{\sqrt{2\pi} z^{k-1} t_0} e^{-\left( u^2 + v^2 \right)/(2 z^{k-1} t_0^2)} \] (16)

where \( t_0 = 0.5 \). After filtering, the results of filtering at different scales constitute spectral scale space, which can be expressed as follows

\[ \Lambda(u, v; k) = (g(\cdot, \cdot; k) \ast A)(u, v) \] (17)
where $k$ is the spatial scale parameter, $k=1, 2..., K$, and $K$ is determined by the size of the image, $K = |lbmin[H, W]| + 1$, $H, W$ are the height and width of the image, respectively.

### 2.4. Calculation of saliency map

Gaussian kernel is used to filter the amplitude spectrum after the hypercomplex transform. The phase spectrum is not changed, and 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

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

(18)

where $g$ is a fixed scale Gaussian kernel function. We use the entropy criterion to select the most appropriate saliency map from the $k$ saliency maps, that is, the entropy minimum is the most appropriate.

$$k_p = \text{argmin}\{\lambda_k^{-1}H_{2D}(s_k)\}$$

(19)

Where $\lambda_k = \sum\sum k(n, m)N(s_k(n, m)), k(n, m)$ is a 2D centered Gaussian mask of the same size as $S$ and $\sum\sum k(n, m) = 1$. $H_{2D}(x)$ is the entropy calculated by convoluting a low-pass Gaussian function $g$ and 2-D signal $x$: $H_{2D}(x) = H\{g \ast x\}$.

### 3. The post process of the saliency map and the result

#### 3.1. Using the global automatic threshold

It is observed that in the HFT algorithm, the saliency map obtained by the minimum entropy method is relatively fuzzy, with no clear boundary, and there is a certain gap with the truth map, so this paper we use binarization method for subsequent processing.

On the basis of the hypercomplex Fourier transform, the automatic threshold method (called HFT-P1 in this paper) is used for processing, and the results are shown in Figure 1. As can be seen from the experimental results in Figure 1, the saliency map processed by the automatic threshold method can truly highlight the salient target and present a clear boundary. However, the contour of the saliency map is fuzzy, and there is a certain gap between it and the actual target. For example, the target cannot be distinguished out of the saliency map in the image of bicycle riding. In addition, for images with multiple targets, the detection results is not ideal and the targets cannot be separated. This is because the automatic threshold method cannot automatically recognize fuzzy boundaries.
3.2. Using median filtering and iterative threshold methods

Because the threshold is determined by the image histogram in the global threshold method, there is certain limitation for different images. Faced the problem that the global automatic threshold method is not ideal to multi-objective saliency detection, we use median filter and the iterative threshold method to improve saliency detection based on the HFT model.

For the selection of the scale of median filter, HFT image saliency detection is processing in 128*128 resolution. In the experiment, we choose four groups scales of median filters, respectively 3*3, 5*5, 7*7, 9*9. After a number of experiments and analysis, the most appropriate scale is the 5*5 template. In the post-processing, median filtering is used by the 5*5 template, and is shown as the equation (20).

$$m_k = \text{medfilt}\{\|F^{-1}_H\{\Lambda_K(u, v)e^{\theta(u,v)}\}\|^2 \}$$

(20)

After median filtering, the salient region of the image becomes more uniform partly. Subsequently, in order to obtain a better detection result, iterative threshold method is used for processing in this paper. The optimal threshold can be obtained by iterative method, which has certain adaptability.

By processing by the median filtering and iterative threshold method (called HFT-P2 in this paper), the detection result is relatively accurate and stable, as shown in figure 2. The experimental results show that the proposed method can detect the salient target of the image well, and clearly distinguish the salient region of single target and multi-target. Meanwhile, the results compared with other models are shown in figure 3.
4. 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, based on the hypercomplex Fourier transform model, the method of median filtering and iterative threshold segmentation is used for saliency detection. Compared with other traditional models, the results obtained have clear boundaries and the multiple targets can be well separated, and the detection effect is more accurate and stable. However, there is a certain distance between the results in this paper and the truth value graph, and the model still has a large space for improvement. In the later research, the saliency detection will be carried out by combining the task and the image itself.

Acknowledgments
This work is supported by Research Program of science at Universities of Guang Xi Autonomous Region(2018KY0323),Key Laboratory of Industrial Process Intelligent Control Technology of Guangxi Higher Education Institutes(IPICT-2016-07, IPICT-2016-08 and IPICT-2016-12 ).

Reference
[1] SHARMA P. Evaluating visual saliency algorithms: past, present and future[J]. Journal of Imaging Science and Technology, 2015, 59(5): 50501-1-50501-17.
[2] MANCAS M, FERRERA V P, RICHE N. The future of attention models: information seeking and self-awareness[C]//From Human Attention to Computational Attention, Springer Series in Cognitive and Neural Systems 10. New York: Springer, 2016: 447-459.
[3] ITTI L, KOCH C, NIEBUR E. A model of saliency based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.
[4] SCHOLKOPF B, PLATT J, HOFMANN T. Graph-based visual saliency[C]//Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. Piscataway: IEEE, 2007: 545-552.
[5] Bruce N D B, Tsotsos J K. Saliency Based on Information Maximization[C]// 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] ACHANTA R, HEMAMI S, ESTRADA F. Frequency-tuned salient region detection[C] //Proceedings of the 2009 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1597-1604.
[7] CHENG M M, ZHANG G X, MITRA N J. Global contrast salient region detection[C] //Proceedings of the 2011 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway, IEEE, 2011: 406-416.
[8] HOU X D, ZHANG L Q. Saliency detection: An spectral residual approach[C]// Proceedings of the 2007 IEEE International Conference on Computer Vision and Pattern Recognition Piscataway: IEEE, 2007: 1-8.
[9] GUO C L, MA Q, ZHANG L M. Spatio-temporal saliency detection using phase spectrum quaternion Fourier transform [C] //Proceedings of the 2008 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1-8.
[10] HOU X D, HAREL J, KOCH C. Image signature: highlighting sparse salient region[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194 -201.
[11] SCHAUERTE B, STIEFELHALGAN R. Predicting human gaze using quaternion DCT image signature saliency and face detection[C]//Proceedings of the 2012 IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE, 2012: 137-144.
[12] LI J, LEVIN M D, AN X. Visual saliency based on scale-space analysis in the frequency domain[J]. IEEE Transactions on Pattern Recognition and Machine Intelligence, 2013, 35(4): 996-1010.