An Improved Ft Algorithm Based on Center Enhancement and Eigenvalue Normalization

Qi Zhang, Baolong Guo*, Zuming Chen and Cheng Li

Institute of Intelligent Control and Image Engineering, Xidian University, School of Aerospace science and technology, No. 2 South Taibai Road, Xi’an, Shaanxi, China.
Email: blguo@xidian.edu.cn

Abstract. The Frequency-tuned method (FT) detection does not work very well when the saliency maps is large and the background environment is complex. Aiming to solve the above problems, this paper improves the traditional FT method, which introduces centre enhancement and eigenvalue normalization. The improvements are implemented to three aspects: enhancing the central area, normalizing the LAB colour feature values, and weighting the LAB three-channel information. Experimental results show that the algorithm is superior to the origin FT in the performance of significant detection, accuracy, recall and other performance.

1. Introduction
The research on visual saliency detection is mainly based on the visual attention mechanism of human visual system. saliency target detection is widely used in applications such as image compression, image segmentation, target detection and other fields. There are two ways to distinguish the saliency method according to different standards. According to the internal factors of visual attention, there are two types of saliency detection methods. One is the bottom-up method driven by data in the scene and non-actively aware. The other is the task-driven, subjective, top-down approach. According to the differences in detection results and application areas, there are two types of saliency detection methods. One is based on saliency points. The other type is based on saliency objects.

Itti and Koch [1] proposed a saliency detection model for multi-scale centre-peripheral differences based on three features of intensity, colour, and direction. At the beginning, the model obtains the feature maps of these three features. Then, it fuses these three types of features to obtain the final saliency map. The final saliency map has become the standard for the bottom-up visual attention model. Hou [2] used spectrum analysis method to detect saliency map as known as spectral residual (SR). Then Guo [3] proposed the phase spectrum Fourier transform method. Achanta [4] proposed the Frequency-tuned algorithm to detect saliency map. Li [5] proposed Hypercom-plex Fourier transform algorithm (HFT). MM Cheng [6] extensively compare, qualitatively and quantitatively, state-of-the-art models over seven challenging data sets for the purpose of benchmarking salient object detection and segmentation methods.

The frequency-tuned method (FT) detection is not ideal when the saliency maps is large and the background environment is complex. Aiming to solve the above problems, this paper improves the traditional FT method, which introduces centre enhancement and eigenvalue normalization. The improvements are implemented to three aspects: enhancing the central area, normalizing the LAB color feature values, and weighting the LAB three-channel information.

2. Frequency-Tuned Algorithm
Achanta set the following requirements for a saliency detection.
• Emphasize the largest salient objects.
• Uniformly highlight whole salient regions.
• Establish well-defined boundaries of salient objects.
• Disregard high frequencies arising from texture, noise and blocking artifacts.
• Efficiently output full resolution saliency maps.

Based on the above five requirements, Achanta proposed the Frequency-tuned algorithm. The input image is a RGB image $I$ with size of $M \times N \times 3$. The algorithm is implemented as follows:

Step 1) Smooth the image $I$ to obtain $I_g$ through Gaussian filter, which can eliminate the noise and reduce errors

$$I_g(i, j) = I(i, j) \times G$$  \hspace{1cm} (1)

where $(i, j)$ is the pixel coordinates, “$\times$” indicates convolution operation, $G$ denotes a Gaussian smoothing filter, and the scale of $G$ is $3 \times 3$ or $5 \times 5$.

Step 2) Convert $I_g$ from RGB colour space to Lab for intensity $L$ as well as colour components $a$ and $b$, whose average value $L_\mu$, $a_\mu$ and $b_\mu$ are defined as

$$I_\mu = \begin{bmatrix} L_\mu \\ a_\mu \\ b_\mu \end{bmatrix} = \frac{1}{M \times N} \begin{bmatrix} \sum_{i=1}^{M} \sum_{j=1}^{N} L(i, j) \\ \sum_{i=1}^{M} \sum_{j=1}^{N} a(i, j) \\ \sum_{i=1}^{M} \sum_{j=1}^{N} b(i, j) \end{bmatrix}$$ \hspace{1cm} (2)

Step 3) Define the salience value at each pixel point: $S_L(i, j)$, $S_a(i, j)$, $S_b(i, j)$:

$$S_L(i, j) = [L(i, j) - L_\mu]^2$$
$$S_a(i, j) = [a(i, j) - a_\mu]^2$$
$$S_b(i, j) = [b(i, j) - b_\mu]^2$$ \hspace{1cm} (3)

where $S_L(i, j)$ is the distance between the brightness feature map and the average of intensity. $S_a(i, j)$ and $S_b(i, j)$ are the distances between the colour features and the average of colour feature.

Step 4) Define the saliency value of image $I$ at pixel $(i, j)$ as $S(i, j)$:

$$S(i, j) = ||I_g(i, j) - I_\mu|| = S_L(i, j) + S_a(i, j) + S_b(i, j)$$ \hspace{1cm} (4)

where $|| ||$ represents the Euclidean distance

3. **REFT Algorithm**

Frequency-tuned detection algorithms are not ideal in complex backgrounds and various large saliency map of an image. The REFT algorithm improves the original FT algorithm.

3.1. Normalizing Intensity and Color Feature

In the Lab colour space, the value of $L(i, j)$ is from 0 to 100, while the value of $a(i, j)$ and $b(i, j)$ are from -128 to 127. In the original FT algorithm, $I_g$ is converted from the RGB colour space to the Lab. And the values of the three feature maps are accumulated after being processed. However, the
range of the three colour components \( L, a, b \) are unequal, which may cause the effect of small components to be suppressed. Therefore, the method of Yu [12] is used to normalize the feature maps before being accumulated.

\[
S_{nor_L} = \frac{S_L}{\max(S_L)} \\
S_{nor_a} = \frac{S_a}{\max(S_a)} \\
S_{nor_b} = \frac{S_b}{\max(S_b)}
\]  

(5)

Redefine the salience value of image \( I \) at the pixel \((i, j)\):

\[
S_{nor}(i, j) = \frac{1}{3} (S_{nor_L}(i, j) + S_{nor_a}(i, j) + S_{nor_b}(i, j))
\]

(6)

3.2. Centre Highlights Prominent Goals

An important role of human visual attention is to better interpret content by focusing on certain areas of content. Selective visual attention plays a very important role in the visual system. This mechanism can reduce the complexity of visual information and speed up visual processing capabilities. Based on the visual attention mechanism of the human visual system, the research of visual saliency detection is mainly to detect the most attractive area or object in the scene. According to eye movement experiments, scientists point out that the human eye is more inclined to focus on the centre of the image and find the target in the area.

Since humans pay more attention to the visual centre region, this paper assign a weight value to an image based on the position information of the pixel. According to the position information of each pixel in the image, the position weight is defined as:

\[
W(i, j) = (1 - \frac{|X_0 - i|^3}{X_0^3}) \times (1 - \frac{|Y_0 - j|^3}{Y_0^3})
\]

(7)

where \((i, j)\) represents the pixel coordinates, \((X_0, Y_0)\) is the centre of the image, and \(\|\|\) indicates the absolute value.

Combine the normalized result with the position weight map:

\[
S_f(i, j) = S_{nor}(i, j) \times W(i, j)
\]

(8)

4. Experiment

This paper compare the algorithms with ten other saliency detection algorithms. These ten significant detection algorithms are GC[7], HC[8], GU, SWD[9], SR[2], SeR[10], CA[11], SIM[12], SUN[13], and FT[4]. In order to get reliable result, we use MM Cheng’s benchmark to make experiments on the HKU-IS dataset with a same segmentation algorithm. The experiment result is showing in Figure 1:
4.1. Evaluation
We use AUC, F-Measure, Recall, and precision to evaluate the consistency between the prediction results and the manually marked data. We use $PS$ to represent the saliency map, and use $GT$ to represent the manually labeled binomial significant target, and use $|M|$ to represent the number of non-zero entries in the mask.

4.2. Precision-Recall (PR)
We can convert the saliency map $PS$ into a binary mask $PM$ and compare it with $GT$ to calculate Precision and Recall.

$$\text{Recall} = \frac{|PM \cap GT|}{|GT|}, \text{Precision} = \frac{|PM \cap GT|}{|PM|}$$

(9)

In general, there are three ways to get the binary image $PM$. The first one was proposed by Achanta:

$$\text{threshold} = \frac{2}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} PS(i, j)$$

(10)

where $M$ and $N$ respectively represent the height and width of an image. $PS(i, j)$ is the saliency value of the pixel at position $(i, j)$.

The second method is to use the Saliency Cut algorithm [14], which uses a series of relatively loose thresholds to generate a PM image. This method has higher recall but lower precision. The third method is segmenting the image with a threshold, which is changed from 0 to 255. Each threshold will get a corresponding precision/recall. Then we can get a curve named precision-recall (PR). From the precision-recall (PR) curve shown in Figure 2, we can see that this algorithm is not only 10% better than the original algorithm, but also superior to other algorithms, such as SUN and HC.
4.3. Receiver Operating Characteristics (ROC) Curve

Except the precision-recall (PR) curve, we can also get the Receiver operating characteristics (ROC) curve. Receiver operating characteristics (ROC) curves reflect the relativity of the false positive rate (FPR) and true positive rate (TPR) parameters as the threshold changes from 0 to 255:

$$TPR = \frac{|PM \cap GT|}{|GT|}, \quad FPR = \frac{|PM \cap \neg GT|}{|\neg GT|}$$

(11)

from the ROC curve, which is shown in Figure 3, we can see that our algorithm is superior to the other eight algorithms except GU and CA. It takes 34s to process an image with the CA method, while only takes 0.5s with this improved method.
4.4. F-measure
Achanta proposed a method to calculate F-measure:

\[ F_\beta = \frac{(1 + \beta^2) \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}} \]  \hspace{1cm} (12)

where \( \beta \) is set to 0.3. We use two different methods to obtain their F-measure. As is shown in Figure 4, the REFT algorithm is not only 20% better than the original algorithm, but also better than the other nine algorithms except CA. Although ca performs better, the CA algorithm is 20 times slower than our algorithm.

![Figure 4. F-measure of our method REFT and the other ten methods.](image)

![Figure 5. AUC of our method REFT and the other ten methods.](image)

4.5. Area under ROC Curve (AUC) Score
The ROC curve is a two-dimensional curve that reflects the performance of the model. The AUC of a perfect model will be one. From the AUC comparison chart of Figure 5, we can see that our algorithm RETT is better than the other eight algorithms except CA and SWD and improves by 10% compared to original algorithm in AUC.
5. Conclusions
This paper improves the traditional FT method, which introduces centre enhancement and eigenvalue normalization. The improvements are implemented to three aspects: enhancing the central area, normalizing the LAB colour feature values, and weighting the LAB three-channel information. Compared with the FT algorithm, our algorithm has obvious advantages in terms of significant visual effects, Precision-recall (PR), ROC, F-measure, and AUC. Compared with other nine algorithms, this method also has considerable advantages.

6. Acknowledgments
This work is supported by the National Natural Science Foundation of China (61571346). The research is also supported by the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University.

7. References
[1] L. Itti, C. Koch, and E. Niebur. A Model of Saliency-based Visual Attention for Rapid Scene Analysis [J]. IEEE TPAMI, 1998, 20(11):1254 – 1259.
[2] Hou X, Zhang L. Saliency detection: a spectral residual approach. In: Proceedings of the 2007 Computer Vision and Pattern Recognition. Minneapolis, MN: IEEE, 2007. 1-8
[3] Guo C L, Ma Q, Zhang L M. Spatio-temporal Saliency detection using phase spectrum of quaternion Fourier transform. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, AK:IEEE, 2008. 1-8
[4] Achanta R, Hemami S, Estrada F, Susstrunk S. Frequency-tuned salient region detection. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL: IEEE, 2009. 1597-1604
[5] Li J, Levine M D, An X, Xu X, He H. Visual saliency based on scale-space analysis in the frequency domain. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35 (4): 996-1010
[6] Borji A, Cheng M M, Jiang H, et al. Salient Object Detection: A Benchmark [J]. IEEE Transactions on Image Processing, 2015, 24 (12): 5706 - 5722.
[7] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, “Efficient salient region detection with soft image abstraction,” in ICCV, 2013, pp. 1529 – 1536.
[8] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, “Global contrast based salient region detection,” IEEE TPAMI (CVPR 2011), 2014.
[9] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, “Visual saliency detection by spatially weighted dissimilarity,” in CVPR, 2011, pp.473–480.
[10] H. J. Seo and P. Milanfar, “Static and space-time visual saliency detection by self-resemblance,” J. Vision, 2009.
[11] S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” IEEE TPAMI, vol. 34, no. 10, 2012.
[12] N. Murray, M. Vanrell, X. Otazu, and C. A. Parraga, “Saliency estimation using a non-parametric low-level vision model,” in CVPR, 2011, pp. 433 – 440.
[13] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, “Sun: A bayesian framework for saliency using natural statistics,” J. Vision, vol. 8, no. 7, pp. 32, 1 – 20, 2008.