

FROM QUATERNION TO OCTONION: FEATURE-BASED IMAGE SALIENCY DETECTION

Hong-Yun Gao and Kin-Man Lam

Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hong Kong, China

ABSTRACT

A novel computational model for detecting salient regions in color images is proposed by utilizing early visual features and performing spectral normalization in the octonion algebra framework, which can accommodate more feature channels than quaternions can. Firstly, feature maps based on edge intensity, the black-white, red-green, and blue-yellow color opponents, as well as the Gabor features with four directions, are incorporated into the eight channels of the octonion image. Then, spectral normalization is achieved by preserving the phase information of the octonion image. Finally, saliency maps are generated at different scales using Gaussian pyramids, and are combined to form the final saliency map. The integration of frequency normalization into the octonion image and saliency-map pyramids exploits the benefits from both the spectral domain and the spatial domain. Experimental results on the MSRA dataset demonstrate that our proposed method outperforms five existing saliency detection models.

Index Terms—Saliency detection, octonion, feature map, spectral normalization, Gaussian pyramids

1. INTRODUCTION

Visual attention is a remarkable mechanism of the human visual system (HVS) which can extract salient and important information from natural scenes efficiently. According to some previous research [1] and [2], the visual-attention mechanism can be divided into two categories, namely bottom-up and top-down. The bottom-up mechanism is fast, simple, and task-independent, in which early features are processed. In contrast, the top-down mechanism involves high-level processing, which may require prior knowledge and memory. Both mechanisms have been exploited to generate saliency maps for detecting salient regions. In the existing bottom-up models, the saliency at a given pixel position is determined by computing the pixel's dissimilarity from its neighboring pixels. "Dissimilarity" can be defined in many ways, e.g. center-surround contrast [3], self-information [4], and even spectral residual [5]. Saliency maps are useful in many applications, such as object detection and image compression. The object detection task

can be significantly simplified by first computing a saliency map for the image under consideration; only the salient regions are analyzed further for object detection. Image compression can also benefit from the saliency map, which divides an image into salient and non-salient regions. For the non-salient regions, the compression ratio can be greatly increased, since people tend to neglect these regions.

One of the earliest computational models was proposed by Itti *et al.* [3]. Their algorithm is based on center-surround contrast, which is implemented as the difference between the fine and coarse scales in the intensity, color and orientation maps. Three conspicuity maps are hence generated by the summation of the feature maps in intensity, color and orientation, respectively, and are combined to form the final saliency map. Ma [6] simulated human perception and proposed a saliency model based on local contrast analysis. Hou [5] firstly proposed a spectral saliency model, namely the spectral residual model, where the difference between the perceived log-spectrum and the characteristic log-spectrum of natural images is extracted. The final saliency map is obtained by transforming the spectral residual back into the spatial domain. Guo *et al.* [7] proposed that phase is the key in obtaining the saliency map, and introduced the use of quaternion image to detect salient regions by preserving phase information only. Achanta *et al.* [8] proposed to use a DoG (difference of Gaussian) filter to eliminate redundant information, and output full-resolution saliency maps with well-defined boundaries of salient objects. Liu *et al.* [9] proposed a new learning algorithm based on the conditional random field to effectively combine features, such as multi-scale contrast, center-surround histogram, and color spatial distribution, for salient object detection.

In this paper, we propose a novel approach for visual-saliency detection. In our approach, spectral normalization is performed in the octonion framework, which can provide at most eight channels – doubling the number of the quaternion. In order to highlight the important salient regions, and attenuate the insignificant salient regions, Gaussian pyramids are used to enhance the saliency maps after spectral normalization.

The rest of the paper is organized as follows. Section 2 introduces the basic concept of octonion, and then presents the proposed model in detail. Experimental results and the

evaluation of the proposed model are given in Section 3. Finally, the conclusion is provided in Section 4.

2. PROPOSED FRAMEWORK

2.1. Basic octonion properties

Among the normed algebra in mathematics, the octonions O are the largest such group, with the other three being the real numbers R , the complex numbers C , and the quaternions H . Octonions, which were discovered by Graves and Cayley [10] independently, have eight dimensions, thus doubling the number of the quaternions, and can be defined as follows:

$$o = x_0e_0 + x_1e_1 + x_2e_2 + x_3e_3 + x_4e_4 + x_5e_5 + x_6e_6 + x_7e_7, \quad (1)$$

where $\{e_0, e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$ are unit octonions that can be treated as perpendicular axes in the eight-dimensional space. $\{x_0, x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$ are real numbers.

The addition and subtraction of octonions are done by adding and subtracting corresponding terms. However, multiplication is complex for being neither commutative nor associative, i.e. $e_ie_j = -e_je_i \neq e_je_i$, if i, j are distinct and non-zero, and $(e_ie_j)e_k = -e_i(e_je_k) \neq e_i(e_je_k)$, if i, j, k are distinct and non-zero. The multiplication rule is shown in Table 1.

\times	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_0	e_0	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_1	e_1	$-e_0$	e_3	$-e_2$	e_5	$-e_4$	$-e_7$	e_6
e_2	e_2	$-e_3$	$-e_0$	e_1	e_6	e_7	$-e_4$	$-e_5$
e_3	e_3	e_2	$-e_1$	$-e_0$	e_7	$-e_6$	e_5	$-e_4$
e_4	e_4	$-e_5$	$-e_6$	$-e_7$	$-e_0$	e_1	e_2	e_3
e_5	e_5	e_4	$-e_7$	e_6	$-e_1$	$-e_0$	$-e_3$	e_2
e_6	e_6	e_7	e_4	$-e_5$	$-e_2$	e_3	$-e_0$	$-e_1$
e_7	e_7	$-e_6$	e_5	e_4	$-e_3$	$-e_2$	e_1	$-e_0$

Table 1. Multiplication rules for the unit octonions.

The conjugate of an octonion is defined as follows:

$$o^* = x_0e_0 - x_1e_1 - x_2e_2 - x_3e_3 - x_4e_4 - x_5e_5 - x_6e_6 - x_7e_7. \quad (2)$$

The product of an octonion with its conjugate is always a non-negative number as follows:

$$o^*o = x_0^2 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2. \quad (3)$$

Using the definition above, the norm is consistent with the standard Euclidean norm, and is defined as follows:

$$\|o\| = \sqrt{o^*o}. \quad (4)$$

2.2. Feature map and octonion image construction

The saliency maps of color images are usually generated by processing the different color channels separately, and combining the results from the respective channels to form the final saliency map. However, these channels are in fact correlated [11]. In general, color images have 3 channels, such as *RGB*, *YUV* or *Lab*. Instead of using the *RGB* color

space as the feature maps, an image is transformed into the *CIE Lab* color space. The transformation decouples the luminance channel from the two color-carrying channels. The two color-carrying channels are red-green (RG) and blue-yellow (BY) color opponents, which are consistent with [12]. We retain the two color-opponents maps as the color feature maps. Besides, we use the luminance channel of the *YUV* color space as the black-white (KW) color-opponents feature map. To better discriminate between different objects, we propose using edges to form another early feature map. Itti *et al.* [3] also incorporated the orientation feature maps in their framework, to approximate the receptive-field sensitivity profile of orientation-selective neurons in the primary visual cortex [13]. The orientation information is extracted using a Gabor filter, which has a simplified form as follows:

$$g(x, y; \sigma, \theta, \omega, \varphi) = \exp\left(-\frac{x'^2 + y'^2}{2\sigma^2}\right) \exp(i(\omega x' + \varphi)), \quad (5)$$

where $x' = x \cos \theta + y \sin \theta$ and $y' = -x \sin \theta + y \cos \theta$. In (5), σ is the standard deviation of the Gaussian envelope, θ is the orientation selectivity of the Gabor filter, ω is the frequency of the sinusoidal factor, and φ is the phase offset. For simplicity, we select only four orientations, namely the horizontal, vertical, and two diagonal orientations. Hence, the orientations for the Gabor filter are: $\theta \in \{0^\circ, 90^\circ, 45^\circ, 135^\circ\}$.

After selecting these eight early feature maps, they can be integrated to form an octonion image. The constructed octonion image is as follows:

$$o(x, y) = e(x, y)e_0 + kw(x, y)e_1 + rg(x, y)e_2 + by(x, y)e_3 + o_0(x, y)e_4 + o_{90}(x, y)e_5 + o_{45}(x, y)e_6 + o_{135}(x, y)e_7, \quad (6)$$

where $e(x, y)$ is the edge map, and $kw(x, y)$, $rg(x, y)$ and $by(x, y)$ are the black-white, red-green, and blue-yellow color-opponents maps, respectively. $o_0(x, y)$, $o_{90}(x, y)$, $o_{45}(x, y)$, $o_{135}(x, y)$ are the four orientation maps for 0° , 90° , 45° , 135° , respectively.

2.3. Fourier transform of the octonion image

After the construction of the octonion image, we need to perform Fourier transform of the octonion image. Since there is no existing clearly-defined Fourier transform for octonion images, we propose extending the quaternion Fourier transform on color images, as proposed by Ell and Sangwine [14], to implement the octonion Fourier transform. According to [14], the general quaternion Fourier transform is defined as follows:

$$F[u, v] = \frac{1}{\sqrt{MN}} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} e^{-\mu 2\pi((mv/M) + (nu/N))} f(n, m), \quad (7)$$

where (n, m) and (u, v) are locations in the spatial and the frequency domain, respectively. M and N are the height and width of the image $f(n, m)$, respectively; μ is any unit pure quaternion.

In order to implement the octonion Fourier transform with the quaternion Fourier transform, we can express an octonion image in the symplectic form as follows:

$$o(x, y) = q_1(x, y) + q_2(x, y)e_4, \quad (8)$$

where $q_1(x, y)$ and $q_2(x, y)$ are the simplex part and the perplex part of the octonion image $o(x, y)$, respectively, defined as follows:

$$q_1(x, y) = e(x, y)e_0 + kw(x, y)e_1 + rg(x, y)e_2 + by(x, y)e_3, \quad (9)$$

$$q_2(x, y) = o_0(x, y)e_0 + o_{90}(x, y)e_1 + o_{45}(x, y)e_2 + o_{135}(x, y)e_3. \quad (10)$$

Sometimes, the quaternions can subject to basis transform, hence we can use the generalized complex operator in expressing the quaternion image as follows:

$$q_1(x, y) = e(x, y) + kw(x, y)\mu_1 + rg(x, y)\mu_2 + by(x, y)\mu_3, \quad (11)$$

$$q_2(x, y) = o_0(x, y) + o_{90}(x, y)\mu_1 + o_{45}(x, y)\mu_2 + o_{135}(x, y)\mu_3, \quad (12)$$

where $\mu_i (i=1,2,3)$ are three arbitrary pure quaternions such that they are perpendicular to each other, and $\mu_3 = \mu_1\mu_2$. Taking the quaternion Fourier transform of (8), we obtain:

$$O(u, v) = Q_1(u, v) + Q_2(u, v)e_4, \quad (13)$$

where $Q_1(u, v)$ and $Q_2(u, v)$ are the quaternion Fourier transform that can be calculated by (7).

As indicated in [14], a quaternion image can be expressed in the polar form. Hence, $Q_i(u, v)$, where $i = 1, 2$, can be expressed as follows:

$$Q_i(u, v) = |Q_i(u, v)|e^{i\mu\varphi_i(u, v)}, \quad (14)$$

where μ and φ are referred to as the eigenaxis and eigenangle of the quaternion, respectively.

Oppenheim and Lim [15] demonstrated the importance of phase information in signal reconstruction in their early work. We follow [7] and [11] by maintaining the magnitude of the spectral domain to unity and keeping the phase information unchanged. By transforming back to the spatial domain, we can obtain the spectral-normalized quaternion image $q'(x, y)$ to form the saliency map as follows:

$$Q'(u, v) = e^{i\mu\varphi(u, v)}, \quad (15)$$

$$q'(x, y) = \frac{1}{\sqrt{MN}} \sum_{v=0}^{M-1} \sum_{u=0}^{N-1} e^{-i\mu_1 2\pi((mv/M) + (nu/N))} Q'(u, v), \quad (16)$$

where $Q'(u, v)$ is the normalized frequency representation.

The final saliency map can be calculated by squaring the eight normalized channels and summing together, as follows:

$$SM = A_0^2 + A_1^2 + A_2^2 + A_3^2 + A_4^2 + A_5^2 + A_6^2 + A_7^2, \quad (17)$$

where SM is the final saliency map; $A_0, A_1, A_2, A_3, A_4, A_5, A_6$, and A_7 are the eight normalized channels, respectively.

2.4. Gaussian pyramid

Gaussian pyramids have been used by Itti *et al.* [3] and Liu [16] to generate feature maps. We follow the analysis by Liu [16] in that the number of Gaussian pyramid levels is defined as: $n = \log_2(\min(w, h)/10)$, where w and h are the width and height of the image, respectively. Firstly, n scales of the original image are created using the Gaussian pyramid, which progressively low-pass and down-sample the images into half-sized counterparts. Then, the images at different scales are individually subject to spectral normalization within the octonion framework, as described in the previous sub-section. Finally, those saliency maps with different scales are rescaled using bilinear interpolation to the original size, and are summed together to form the final saliency map. The reason for using Gaussian pyramids to enhance saliency detection is straightforward: Gaussian pyramids accentuate the important salient regions, while attenuating the trivial salient regions. In the original image, many locations may be salient as rather different compared with the neighbors, but some of those locations are trivial when viewed at a finer scale. Gaussian pyramids can thus be used to eliminate the trivial salient regions by producing their low-passed and down-sampled counterparts.

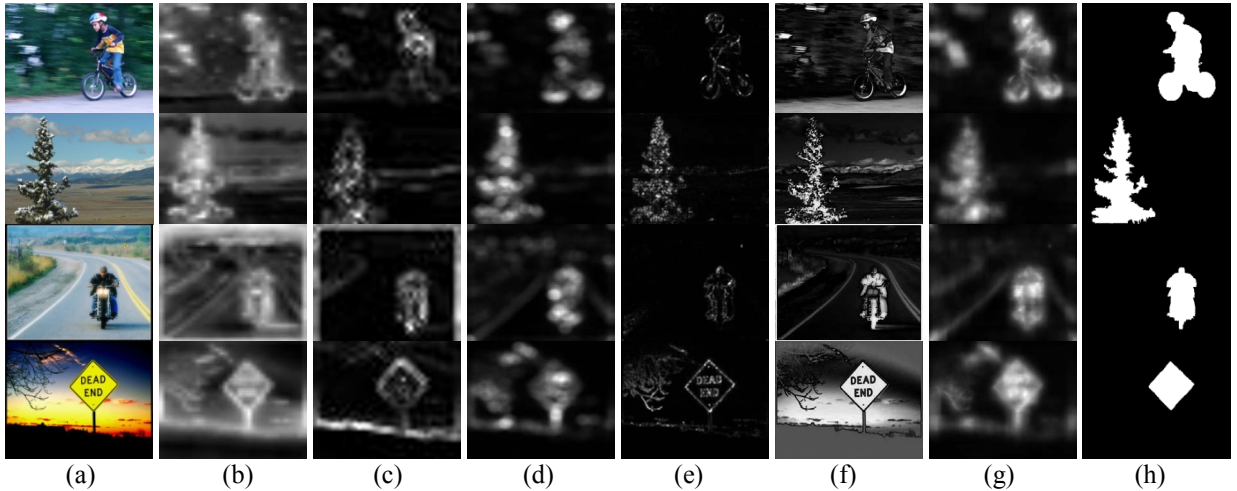


Fig. 1. Comparison of saliency maps generated by different methods on the MSRA dataset [9]: (a) Original images, (b) IT [3], (c) MZ [6], (d) SR [5], (e) PQFT [7], (f) FTS [8], (g) The proposed method, and (h) Ground-truth images from [8].

3. EXPERIMENTAL RESULTS

In this work, the MSRA dataset [9], which contains 5,000 color images as well as the ground-truths for all the images, is used to evaluate the performance of the proposed method. Achanta *et al.* [8] refined the ground-truths more accurately by segmenting salient objects entirely, instead of depicting the objects with rectangular boxes only. As mentioned in Section 1, we compare our method with IT [3], MZ [6], SR [5], PQFT [7] and FTS [8]. Fig. 1 shows some sample images for comparison. The experiment results demonstrate the excellent performance of our proposed model, since it can locate true salient objects well.

We evaluate the performance of our proposed method quantitatively in terms of the Precision, Recall and F-measure, proposed by Liu *et al.* [9]. The Precision, Recall and F-measure are defined as follows:

$$\text{Precision} = \frac{\sum_{x,y} s(x,y)g(x,y)}{\sum_{x,y} s(x,y)}, \quad (18)$$

$$\text{Recall} = \frac{\sum_{x,y} s(x,y)g(x,y)}{\sum_{x,y} g(x,y)}, \quad (19)$$

$$F = \frac{(1 + \alpha) \times \text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + \text{Recall}}, \quad (20)$$

where $s(x, y)$ and $g(x, y)$ are the values of the saliency mask and the ground-truth image at location (x, y) , respectively. α is a parameter to determine the relative importance of Precision and Recall, and chosen as 0.3 in our experiments. As shown in Fig. 2, our proposed method outperforms the other methods on the Recall and F-measure values, while the Precision value of our model is slightly lower than FTS [8]. However, as the F-measure is an overall performance indicator, our method shows a better performance compared with other methods according to this benchmark.

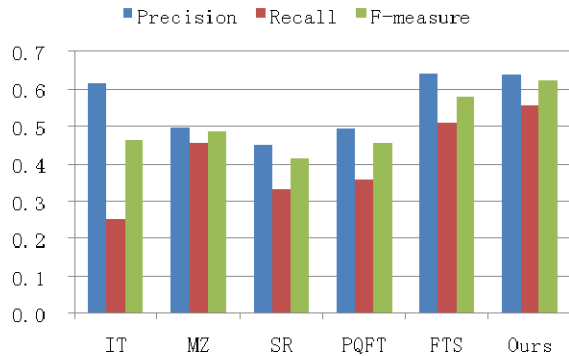


Fig. 2. The Precision, Recall and F-measure of the different saliency detection methods.

We also evaluate the performance of the proposed method using the ROC curve [17]. The percentage of the target points in the ground-truths that fall into the salient

points in the saliency map is the True Positive Rate. The percentage of the background points that fall into the salient points in the saliency map is the False Positive Rate. The overall performance can be reflected by the area under the ROC curve (AUC). The ROC curve of these methods is shown in Fig. 3. The AUC of these methods are tabulated in Table 2. From Fig. 3 and Table 2, it can be seen that our method has the largest AUC, and thus achieves the best performance when compared to the other methods.

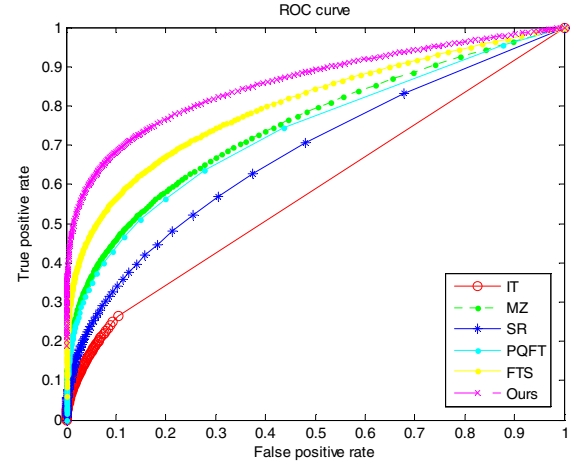


Fig. 3. ROC curve of different saliency detection methods.

Model	ROC area	Model	ROC area
IT[3]	0.6301	PQFT[7]	0.7376
MZ[6]	0.7481	FTS[8]	0.8007
SR[5]	0.6789	Ours	0.8563

Table 2. The AUC of different saliency detection methods.

4. CONCLUSION

In this paper, we propose a novel computational model for saliency detection, based on spectral normalization in the framework of octonion algebra. Various feature maps are generated based on the *Lab* and *YUV* color space, edges, and Gabor features. By incorporating the eight feature maps into an octonion image, we can perform Fourier transform on the octonion image as a holistic framework with the extension from quaternion. Spectral normalization is then achieved on the octonion image to generate the saliency map. Furthermore, Gaussian pyramids are used to enhance the performance by combining several saliency maps of different scales into the final saliency map. The experiment results on the MSRA [9] dataset demonstrate the excellent performance of our method when compared to the other five existing saliency detection algorithms.

5. ACKNOWLEDGEMENT

This work is supported by an internal grant from the Hong Kong Polytechnic University (Grant No. G-YN21).

6. REFERENCES

- [1] A. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, no. 1, pp. 97-136, 1980.
- [2] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219-227, 1985.
- [3] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.
- [4] N. Bruce and J. Tsotsos, "Saliency based on information maximization," in *Advances in Neural Information Processing Systems*, 2005.
- [5] X. Hou and L. Zhang, "Saliency Detection: A Spectral Residual Approach," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [6] Y-F. Ma and H-J. Zhang, "Contrast-based Image Attention Analysis by Using Fuzzy Growing," in *ACM International conference on Multimedia*, 2003.
- [7] C. Guo, L. Ma and L. Zhang, "Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [8] R. Achanta, S. Hemami, F. Estrada and S. Susstrunk, "Frequency-tuned salient region detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [9] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and H-Y. Shum, "Learning to Detect a Salient Object," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353-367, 2011.
- [10] Arthur Carlay, "On Jacobi's elliptic functions, in reply to the Rev. B. Bronwin; and on quaternions", *Philosophical Magazine*, vol. 26, pp. 208-211, 1845.
- [11] P. Bian and L. Zhang, "Biological Plausibility of Spectral Domain Approach for Spatiotemporal Visual Saliency," in *International Conference on Neural Information Processing*, 2008.
- [12] S. Engel, X. Zhang, and B. Wandell, "Colour Tuning in Human Visual Cortex Measured With Functional Magnetic Resonance Imaging," *Nature*, vol. 388, no. 6637, pp. 68-71, 1997.
- [13] A. Leventhal, *The Neural Basis of Visual Function: Vision and Visual Dysfunction*, vol.4, *CRC Press*, 1991.
- [14] T. Ell and S. Sangwine, "Hypercomplex Fourier Transform of color images," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 22-35, 2007.
- [15] A. Oppenheim and J. Lim, "The Importance of Phase in Signals," in *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529-541, 1981.
- [16] F. Liu and M. Gleicher, "Region Enhanced Scale-invariant Saliency Detection," in *IEEE International Conference on Multimedia and Expo*, 2006.
- [17] B.W. Tatler, R.J. Baddeley and I.D. Gilchrist, "Visual Correlates of Fixation Selection: Effects of Scale and Time," *Vision Research*, vol. 45, no. 5, pp. 643-659, 2005.