FqSD: Full-Quaternion Saliency Detection in Images

Reynolds León Guerra$^{(\boxtimes)}$, Edel B. García Reyes,
Annette M. González Quevedo, and Heydi Méndez Vázquez

Advanced Technologies Application Center (CENATAV), Havana, Cuba
`rleon@cenatav.co.cu`

Abstract. Saliency detection aims to segment in two groups the pixels of an image, in important or less important visual information. Important information can be used to detect objects with some semantics for tasks in computer vision. In this paper, we develop a saliency detection method using full-quaternion. The proposed method makes a combination between local and global approaches. Local features are obtained at patches level using a Module Local Binary Patterns and the comparison between feature vectors (module and phase). The salient object is obtained by a weighted combination of salient maps where a function of center-bias and refinement is applied. To verify the effectiveness of our method, this is validated using the mean absolute error metric in the ECSSD-1000 and DUT-OMRON datasets and compared with others state of the art algorithms. A statistical analysis is applied to the results to obtain the statistical significance among color spaces with the method of Wilcoxon signed test. The results show that the HSV color space is better in effectiveness than others.

Keywords: Full-quaternion · Salient map · Color · Feature · Object

1 Introduction

Saliency detection task aims to detect objects in images that are visually salient for persons [1]. When a person looks at a scene it has the tendency to focus on relevant information and reject the redundant information [2]. This can be useful to make different tasks in computer vision, for example, visual classification [3], image retrieval [4], person re-identification [5] and so on. In general, there are two approaches to search salient objects in images, **bottom up** and **top down**. The baseline in the first approach is to get the features less frequent in the scene and consider them as salient regions. **Top down** is used to look for salient regions with some prior knowledge about the objects.

In recent years, different saliency detection methods have been developed. Erdem and Erdem [6] use region covariance descriptor with features of color, orientation and spatial information, extracted at patches level to compare among

themselves. Hu et al. [7] proposed a method where local and global features are combined in a final map and all pixels are weighted according to their distance to center of the image. Liu and Hu [8] make a combination among maps obtained using Quaternion Fast Fourier Transform (QFFT) to look for the optimum. Yu et al. [9] proposed to use the global contrast of color to obtain a salient map grouping the background pixels, considering that these are similar. Wang et al. [10] based their work on two neural networks to learn local and global features to obtain the salient maps and later these are combined in a single map. Rajankar and Kolekar [11] applied a scale reduction using interpolation of the Fourier coefficients in quaternion space getting the salient map.

Colors and contrasts are features very important to obtain salient maps. However, in the works using this features the correlation among them is not considered at different levels with different approaches. To solve this, we propose a saliency detection method **FqSD** where unlike other works, we link the spatial and frequency information using quaternions preserving the correlation among colors and contrast. Contributions of this paper are summarized as follows: First, a combination of local and global features in quaternion space at different scales is done. Second, a comparative study is developed where we obtain the best color space to use the proposed method.

This paper is organized as follows. In Sect. 2 we explain our approach. In Sect. 3 experimental results are presented and analyzed. Finally the conclusions are set out in Sect. 4.

2   Our Approach

The data input to method **FqSD** is an image represented in any color space. In step 2, a Gaussian pyramid reduction is applied to get several images with different resolutions where a lot of less important information is eliminated. After, in step 3 each image is processed to build salient maps using local and global approaches in the spatial and frequency domain using full-quaternions (combining values of contrast and color channels). Also previous images are merged into a single image per method. As there are three salient maps, another merge step is required (step 4), which is done by means of a weighted sum of maps. In step 5, two functions (center-bias and refinement) are applied to finally obtain the salient object (output).

2.1   Multiple Resolution

Generally, salient objects are invariant when a scale transformation is applied to the image. But, the information of not-salient objects is lost during the change of resolution e.g. background information in the image. As explained before, we used a Gaussian pyramid by reduction [12] in step 2, where four images are obtained and we preserve the original image. As original images have different sizes, these are normalized to get a standard size.
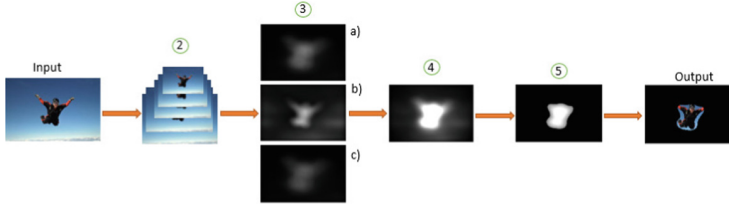
Fig. 1. Show the different steps in the proposed method **FqSD**.

2.2   Local and Global Salient Maps

After obtaining the images in step 2, these are transformed from original space to full-quaternion space, where the image has four channels. First channel is the clear-dark contrast effect (for more detail see [13]) and the other channels are values of a color space, for example, Red-Green-Blue **(RGB)**. We develop three approaches to obtain the salient maps (Local Salient Map **(LSM)**, Module Local Binary Pattern Salient Map **(MLBP)** and Quaternion Fast Fourier Transform to global Salient Map **(QFFT)**).

For a better understanding of the proposed method, firstly several properties of quaternion algebras are explained. Quaternion is a hypecomplex number given by Hamilton [14] and is denoted with letter $\mathcal{H}$. If $q \in \mathcal{H}$, this is represented as follows:

$$\{q = t + xi + yj + zk | (t, x, y, z) \in \mathcal{R}\} \tag{1}$$

Where the complex operators $\mathbf{i}, \mathbf{j}, \mathbf{k}$ have the next rules $\{i^2 = j^2 = k^2 = ijk = -1, ij = k = -ji, ki = j = -ik, jk = i = -kj\}$. It is clear that the multiplication between quaternions is not commutative. If $t = 0$, is a pure quaternion and $q = xi + yj + zk$, if $t \neq 0$ is a full-quaternion. The module and the phase are:

$$|q| = (t^2 + x^2 + y^2 + z^2)^{1/2} \tag{2}$$

$$\phi = \tan^{-1}\left(\frac{(x^2 + y^2 + z^2)^{1/2}}{t}\right) \tag{3}$$

LSM: Images are divided in little patches and in each one of them two feature vector are obtained. The first vector is associated to each full-quaternion in the patch and their elements are the module and phase (see Eqs. (2) and (3)). The second feature vector has the average of the module and phase of all the full-quaternions in the patch. To obtain a salient map in each patch an euclidean distance is applied between the first and second feature vector. The full-quaternions with high value have the highest probability to be different from their neighbors (see Fig. 1, 3.(a)).

MLBP: In this approach, each full-quaternion is codified using Local Binary Pattern (LBP). We extend the LBP to the full-quaternion using the module

because is sensitive to the color changes. See following equation:

$$mLBP_{S_i} \sum_{J=0}^{p-1} h(s_j - s_i)2^j, h(r) = \begin{cases} 0 \text{ if } r \geq 0 \\ 1 \text{ if } r < 0 \end{cases} \tag{4}$$

Where S is a 3×3 windows, $p \in S$, s_i and s_j module of full-quaternion analysis and its neighborhood respectively. Here, the salient map is obtained equal that the **LSM** method, but using the value of the modules (see Fig. 1, 3.(c)).

QFFT: Quaternion Fast Fourier Transform is used to build salient map in a global way [8]. Spectral module is modified using a low-pass filter, where a stable color region is obtained with the Inverse Fourier Transform as follows.

$$F(p, s) = S \sum_{m=0}^{m-1} \sum_{n=0}^{n-1} e^{-\mu 2\pi(\frac{pm}{M}) + (\frac{sn}{N})} f(m, n) \tag{5}$$

$$f(m, n) = |S \sum_{p=0}^{p-1} \sum_{s=0}^{s-1} e^{\mu 2\pi(\frac{pm}{M}) + (\frac{sn}{N})} (exp(\Upsilon + (\Lambda(p, s) \circ \Gamma)))|^2 \tag{6}$$

$$\Lambda = \frac{Vq}{|Vq|}, \Lambda \in F(p, s) \tag{7}$$

Where, $S = \sqrt{\frac{1}{MN}}$, the Eqs. (5) and (6) are the direct and inverse Quaternion Fourier Transforms, μ is a unit pure quaternion (module is aqual to 1), p y s are frequency coefficients, m y n are spatial coordinates of the image. Λ y Γ are the eigenaxis and phase, Υ is the spectral module modified to the filter and $\circ$ is the Hadamard product (see Fig. 1, 3.(b)).

2.3   Image Fusion

To show the salient map obtained in step 3, it is necessary to make a fusion among the images obtained by each level in the Gaussian pyramid and process these images (as explained in the Sect. 2.2). Finally, a single image is obtained by each method. We developed the Eq. (8) taking into account that in level 2 there is a trend to keep high values of salient points. Nonetheless, in other levels the values are high or low according to the scene.

$$SM = max(Map_0, Map_1, Map_3, Map_4) + Map_2 \tag{8}$$

Where, SM, may be LSM, mLBP-SM or QFFT-SM. $Map\{.\}$ is the map obtained in different levels. The next step is to fuse the three salient maps obtained previously with a weight value to each map as follow:

$$MSsingle = \alpha(\psi * LSM) + \beta log((\psi * MLBP) + 1) + \delta(\psi * QFFT) \tag{9}$$

Where, ψ is a radial filter, $\{\alpha, \beta, \delta\}$ are parameters of weight for the maps and $\{*\}$ convolution product (see Fig. 1, image 4).

2.4   Refinement

A refinement step is needed to improve the salient map obtained in the last step. Center-bias here is used to give a weight to each element of a salient map [15]. Theory of center-bias is based on the way that images are captured, in general the salient object is localized in the center of the image. Hence, Eq. 10 is used to weight the values according to the distance of a pixel to the center of the image.

$$MS_{cb} = MSsingle_{(m,n)}(1 - d), d = \sqrt{(m - \upsilon)^2 + (n - \rho)^2}/(\upsilon^2 + \rho^2) \quad (10)$$

Where, υ and ρ are the center coordinates of the image.

Generally speaking, in practical tasks we only need to know the values associated with salient objects. To solve this, unlike other works, an adaptive threshold is applied to eliminate values that are far from the interest objects as follows.

$$MS_{threshold} = \begin{cases} 0 \text{ if } r \leq 0 \\ \omega \text{ if } r > 0 \end{cases} \quad (11)$$

Where, ω is the value of the $MS_{cb}(m, n)$, r is the threshold (sum of the average with the standard deviation in MS_{cb}).

Other interesting detail of salient maps is that a salient object can have different parts with several probability of be observed. Therefore, it is done a refinement as follow:

$$MS_{final} = log(|MS_{threshold}| + 1) \quad (12)$$

3   Experimental Results

Our aim is to validate the performance of the developed method using different color spaces and performing a comparison with other state-of-the-art methods. The datasets used are ECSSD-1000 and DUT-OMROM. ECSSD-1000 [16]: has 1000 images (with their salient mask respectively) where there are from 1 to n salient objects. Salient objects have been labeled by five persons to obtain the mask or ground truth (GT). DUT-OMROM [17]: contains 5168 images with complex scene and their masks respectively. We performed experiments using 4 color spaces: RGB, HSV (Hue-Saturation-Value), Lab (lightness, color opponents green-red and blue-yellow) and Ycbcr (Y is the Luma component and CB and CR are the blue-difference and red-difference chroma components). Our best parameter configuration is: $\delta = 1.4$, $\beta = 1.3$, $\delta = 0.3$, $\mu = (i, j, k)/\sqrt{3}$. As metric of evaluation we used Mean Absolute Error (MAE), see Eq. 13. The Wilcoxon signed test is used to know the statistical significance than there are among the result obtained in the different color spaces. In this test, a value of 1 means significance (the results are not casual), if the value is 0 the results are doubtful.

$$MAE = \frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} |MS_{final}(m, n) - GT(m, n)| \quad (13)$$

Table 1. Different results among FqSD and other methods in terms of MAE.

Methods	ECSSD-1000	DUT-OMROM
FqSD(HSV)	**0.2002**	**0.1551**
FqSD(RGB)	0.2102	0.1718
FqSD(Lab)	0.2198	0.1551
FqSD(Ycbcr)	0.2241	0.1552
IT (1998) [18]	0.2730	0.1980
Wco (2014) [19]	0.2250	-
LPS (2015) [20]	0.2370	-
Jiang (2017)[21]	0.2430	-
BL (2015) [22]	0.2620	0.2390
CPMC-GBVS (2014) [23]	-	0.2345
HDCT (2014) [24]	-	0.1669
BSCA (2015) [25]	-	0.1910
XIA (2017) [26]	-	0.1632
BSP (2018) [27]	0.2200	-

Table 2. Statistical significance among the different color space vs HSV.

Color spaces	ECSSD-1000	DUT-OMROM
RGB	1	1
Lab	1	0
Ycbcr	1	0

We can observe in Table 1 that our method using the HSV color space got the best results in both datasets (see Fig. 2). However, the results with Wilcoxon signed test show difference in the dataset. For the images analyzed in ECSSD-1000 represented in other color spaces different than HSV there is statistical significance (see Table 2) and the results are better than the algorithms of state of the art. On the other hand, the results in the DUT-OMROM dataset have zero value in the Wilcoxon signed test. This result is associated with the characteristic present in dataset, where the patterns in different images are repeated with high frequency along the dataset (the variance among image data is small).

The advantage of HSV color space over Lab, Ycbcr and RGB, is because there is a correlation among different features represented by full-quaternions, where the four features (Hue, Saturation, Value, Clear-Dark Contrast) are combined linearly and processed as a single element. Moreover, the combination between local and global features allows highlighting regions of interest that could be ignored with a simple analysis (local or global). Center-bias and refinement act as a function of adjustment delineating better the contour of the salient objects.

Fig. 2. Different salient objects with the proposed method **FqSD(HSV)** in ECSSD-1000 and DUT-OMROM dataset. From top to bottom the rows are, original image, ground truth and salient objects respectively.

4   Conclusions

Experimental results show good performance using the HSV color represented by means of full-quaternions. The integration of the local and global salient maps to look for features that are less frequent in images allows improving the results analyzed with the Mean Absolute Error. The Wilcoxon signed test showed that little variety in the images in a dataset can give untrustworthy results. In future works, we plan to develop a deep learning-based method with neural network using full-quaternions to learn parameters in front of the complexity of different scenes that appear in the real world.

References

1. Wang, X., et al.: Edge preserving and multi-scale contextual neural network for salient object detection. IEEE Trans. Image Process. **27**(1), 121–134 (2018)
2. Aytekin, C., Iosifidis, A., Gabbouj, M.: Probabilistic saliency estimation. Pattern Recognit. **74**, 359–372 (2018)
3. Murabito, F., et al.: Top-down saliency detection driven by visual classification. arXiv preprint arXiv:1709.05307 (2017)
4. Li, S., Mathews, P.: Can image retrieval help visual saliency detection? arXiv preprint arXiv:1709.08172 (2017)
5. Zhu, F., et al.: A novel two-stream saliency image fusion CNN architecture for person re-identification. Multimed. Syst., 1–14 (2017)
6. Erdem, E., Erdem, A.: Visual saliency estimation by nonlinearly integrating features using region covariances. J. Vis. **13**(4), 11–11 (2013)
7. Hu, D., et al.: Saliency region detection via local and global (2015)
8. Liu, S., Hu, J.: Visual saliency based on frequency domain analysis and spatial information. Multimed. Tools Appl. **75**(23), 16699–16711 (2016)
9. Yu, C., Zhang, W., Wang, C.: A saliency detection method based on global contrast. Int. J. Signal Process. Image Process. Pattern Recognit. **8**(7), 111–122 (2015)
10. Wang, L., et al.: Deep networks for saliency detection via local estimation and global search. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3183–3192. IEEE (2015)
11. Rajankar, O.S., Kolekar, U.D.: Scale space reduction with interpolation to speed up visual saliency detection. Int. J. Image Graph. Signal Process. **7**(8), 58 (2015)

12. Burt, P.J., Adelson, E.H.: The Laplacian pyramid as a compact image code. In: Readings in Computer Vision, pp. 671–679 (1987)

13. Guerra, R.L., García Reyes, E.B., Mata, F.J.S.: Full-quaternion color correction in images for person re-identification. In: Mendoza, M., Velastín, S. (eds.) CIARP 2017. LNCS, vol. 10657, pp. 339–346. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75193-1_41

14. Morais, J.P., Georgiev, S., Sprößig, W.: Real Quaternionic Calculus Handbook. Springer, Basel (2014). https://doi.org/10.1007/978-3-0348-0622-0

15. Buso, V., Benois-Pineau, J., Domenger, J.-P.: Geometrical cues in visual saliency models for active object recognition in egocentric videos. Multimed. Tools Appl. **74**(22), 10077–10095 (2015)

16. Yan, Q., et al.: Hierarchical saliency detection. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1155–1162. IEEE (2013)

17. Yang, C., et al.: Saliency detection via graph-based manifold ranking. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3166–3173. IEEE (2013)

18. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)

19. Zhu, W., et al.: Saliency optimization from robust background detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821 (2014)

20. Li, H., et al.: Inner and inter label propagation: salient object detection in the wild. IEEE Trans. Image Process. **24**, 3176–3186 (2015)

21. Jiang, L., Zhong, H., Lin, X.: Saliency detection via boundary prior and center prior. Int. Robot. Autom. J. **2**(4), 00027 (2017). https://doi.org/10.15406/iratj.2017.02.00027

22. Tong, N., et al.: Salient object detection via bootstrap learning. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1884–1892. IEEE (2015)

23. Li, Y., et al.: The secrets of salient object segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 280–287. IEEE (2014)

24. Kim, J., et al.: Salient region detection via high-dimensional color transform. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 883–890 (2014)

25. Qin, Y., et al.: Saliency detection via cellular automata. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 110–119. IEEE (2015)

26. Xia, C., Zhang, H., Gao, X.: Saliency detection by aggregating complementary background template with optimization framework. arXiv preprint arXiv:1706.04285 (2017)

27. Wang, H., et al.: Saliency region detection method based on background and spatial position. Int. J. Pattern Recognit. Artif. Intell. **32**(07), 1850024 (2018)