 SL, vol.7, no.1, pp.45-57, 2013

Multiscale Phase Spectrum Based Salient Object Detection

Deepak Singh[1] **and Sukadev Meher**[1]

Abstract: Automatic image segmentation is emerging field in image processing research domain. In a visual scene, the objects which are different from their surroundings get more visual importance and get high gaze attention of the viewer. Saliency detection predicts significantly important regions in a scene, which should be considered for further processing according to specific applications. There are several applications where saliency detection is used as core modules such as object based surveillance, content adaptive data delivery for low data rate systems, automatic foveation system. There are two different approaches to predict the viewer's gaze in a visual field: topdown approach and bottom-up approach. Top-down approach is target driven while the bottom-up method is independent of target and stimuli driven. Hardware based eye tracker devices are also commercially available but the cost is comparatively very high. In this paper, an efficient multiscale phase spectrum based salient object detection method is proposed. It is observed that a fixed scale of the original image may not predict properly the salient objects. Saliency predicted in one resolution may not predict the same fixation region on another resolution. It is proposed to apply saliency detection to multiple scales of the original image. Saliency is detected using phase spectrum of Fourier transform as positional information is contained in the phase spectrum while amplitude spectrum contains the presence of frequency components. The proposed method performs much better than other previous methods and predicts more precisely salient objects. Simulation results of six state-of-art techniques for salient object detection are analyzed along with the ground truth and compared against the proposed method. The performance of the proposed method is measured on the basis of objective and subjective analysis. The simulation verifies that the proposed method is suitable candidate for prediction of salient objects.

Keywords: Foveated imaging, salient object detection, object based segmentation, computer vision.

[1] Department of Electronics & Communication National Institute of Technology Rourkela 769008, Odisha, India

1 Introduction

Salient object detection of a visual scene is useful for many applications like object based segmentation, region based adaptive compression, object recognition and computer vision. Saliency detection is a method which is used for determining visually important regions within a scene. Task dependent and task independent methods (also known as topdown and bottom-up methods respectively) are two different approaches to determine the salient object in vision analysis. The task dependent method is top down, computational aggressive and slower in processing while task independent method is bottom up approach, scene dependent, saliency driven and responds quickly [Niebur and Koch (1998)].

Most of the models of saliency detections are biological models based and inspired by human visual system. They work on a bottom-up computational method. Various lowlevel visual features such as intensity, color, orientation, texture and motion are extracted from the image at fixed scale or images at multiple scales either by using Gaussian or laplacian pyramids to determine saliency [Adelson, Anderson, Bergen, Burt, and Ogden (1984a)]. After a saliency map is computed from each of these features, they are normalized and summed up in a linear or non-linear fashion to form a master saliency map that represents the saliency of each pixel of original image. Flow chart shown in Fig. 1 describes the steps to follow for salient object based segmentation. Bottom up saliency model is much popular than top down approach. Many bottoms up models are based on center-surround contrast method which simulates visual receptive fields model. Although there are others models too which are not related to biological vision system, in spite of that they exploit image properties by mathematical equations.

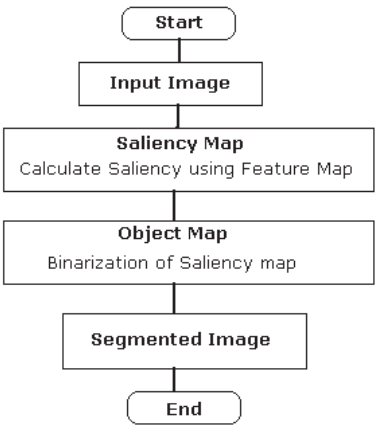


Figure 1: Flowchart of Salient Object Based Segmentation

Itti et al. (BM) proposed a bottom-up human vision attention model and which became the base of a system called Neuromorphic Vision C++ Toolkit (NVT) [Itti, Koch, and Niebur (1998)]. After that, based on Rensinks theory [Rensink (2000)], Walther created the most useful commercial product for saliency is SaliencyToolBox (STB) [Walther and Koch (2006)]. Harel et al. (GB) used graph for saliency detection [Harel, Koch, and Perona (2006)]. They first form the activation map based on particular feature and normalize it to generate the saliency map. Hou et al. (SR) proposed a model which is independent of features or biological system. It determines saliency map by analyzing the log spectrum of an input image [Hou and Zhang (2007)]. It is also shown that this is very fast approach for saliency detection. Ma et al. (CB) generates the saliency map based on center surround scheme by contrast analysis [Ma and Zhang (2003)]. The theory behind this is; contrast is the most important feature which directs the human visual attention than any other feature like color, texture or orientation. Achanta et al. (FT) generates the full resolution saliency map unlike Itti and CB; by preserving more frequency content by exploiting feature of color and luminance [Achanta, Hemami, Estrada, and Ssstrunk (2009)]. An example of region based segmentation using saliency is shown in Fig. 2.

(a) Input Image (b) Saliency Map (c) Ground Truth (d) Segmented Image

Figure 2: Example of Region based segmentation

Figure 3: Example of Gaussian Pyramids

In this paper, we have introduced an efficient salient object detection technique. Our approach is inspired by phase spectrum of Fourier transform (PFT) based saliency detection technique proposed by Guo, Qi, and Zhang (2008). But they have calculated saliency only for one scale of resolution of image and there is huge chance of missing information in one scale to another scale. Our method take advantage of multi-scale saliency maps over PFT and gives better results as compared to PFT. This method is simple and fast. This paper is organized as follows. Background concepts of phase spectrum based saliency is discussed in detail in section 2. Proposed model is explained in detail to section 3. The experimental results of our method and comparative performance of other techniques against our proposed method are shown in section 4. Finally conclusion is given in section 5.

2 Background Concepts

2.1 Gaussian pyramids

In saliency detection, image resolution plays major role. As we know that any scene in the real world may contain one or more objects of various sizes. And objects can be placed at difference in distance from viewers direction of gaze. So due to variations in distance, orientations and viewers view angles between objects, any algorithm applied to image for vision analysis will not correctly work for all objects or features. Analysis of one scale may not have information at another scale or resolution [Adelson, Anderson, Bergen, Burt, and Ogden (1984b)]. So to determine saliency more accurately we have proposed multiscale salient object analysis.

In pyramid architecture, the original image is decomposed into sets of lowpass or bandpass pyramids which is known as Gaussian pyramid and Laplacian pyramids respectively. The Gaussian pyramid is obtained by first smoothing or blurring the image with a Gaussian smoothing filter 'kernel' and then sub sampled the smoothed image by a reducing factor of two to both the horizontal and vertical direction. After that same process iteratively repeat for further levels.

Let original image is represented as $I(x,y)$. The Gaussian pyramids are obtained iteratively as:

$$G_0 = I(x,y), \text{ for base level,} l = 0 \tag{1}$$

$$G_l = \sum_{i=-2}^{2} \sum_{j=-2}^{2} h(i,j) G_{l-1}(2x+i, 2y+j), 1 < l < N \tag{2}$$

where $h(i,j)$ is a weighting function also known as generating kernels and this is identical to all levels. This N level pyramid representation of multiresolution images is known as Gaussian pyramids. The value of generating kernel is $[\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}]$.

Each element of pyramid represents a local average resulted by applying weighting function to image at different scales. So the Gaussian pyramid contains local averages at various scales. In section 3 we have shown that obtaining the Gaussian pyramids is one of the essential steps in our proposed method for saliency detection.

2.2 Saliency calculation using Phase Spectrum

In a visual scene, the objects which got focused attention without any prior goal is known as salient objects. Salient objects or regions those get focused attentions are distinctively different from their surroundings. Saliency at any region is decided by how different this region is from its surround in intensity, color or orientation, etc. Treisman and Gelade proposed Feature Integration Theory (FIT) [Treisman and Gelade (1980)]. According to FIT theory, any visual scene is analyzed at 'preattentive stage or early representation' by different receptors those are selectively stimulated by separable properties or dimensions such as intensity, color, orientations, direction of movements and map these dimensions in different areas of brains. Later Koch and Ullman extended it that, each feature map registers individual conspicuous locations. The combination of all these feature maps for the measure of global conspicuous location is termed as Saliency Map that represents the conspicuity for every pixel in a visual scene [Koch and Ullman (1985)].

By using Fourier transform a signal in frequency domain can be decomposed in amplitude spectrum and phase spectrum. The 2D discrete Fourier transform (DFT) of a image $I(x,y)$ of size $M \times N$ is obtained as:

$$F(u,v) = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{x=0}^{N-1} I(x,y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})} \tag{3}$$

where $u = 0,1,2,3,\ldots,M-1$, $v = 0,1,2,3,\ldots,N-1$ and $j = \sqrt{-1}$.

The Fourier amplitude spectrum and phase spectrum are defined as:

$$|F(u,v)| = \sqrt[\frac{1}{2}]{(\Re(F(u,v)))^2 + \Im(F(u,v)))^2} \tag{4}$$

$$\theta(u,v) = \arctan\left[\frac{\Im(F(u,v))}{\Re(F(u,v))}\right] \tag{5}$$

where $\Re(F(u,v))$ represents real part and $\Im(F(u,v))$ is representing the imaginary part.

Amplitude spectrum represents how much of each frequency component is present, while phase spectrum shows where these frequency components are present in the image. It is shown by extensive experiments that amplitude spectrum is not unique

for individual image [van der Schaaf and van Hateren (1996)]. The phase spectrum of an image is more important that the amplitude spectrum [Oppenheim and Lim (1981)]. So for the reconstruction of image $I(x,y)$, the inverse transform of $e^{j\theta(u,v)}$ retains the resemblance to original image. Fig. 4 shows some of examples of reconstruction of images using amplitude spectrum and phase spectrum.

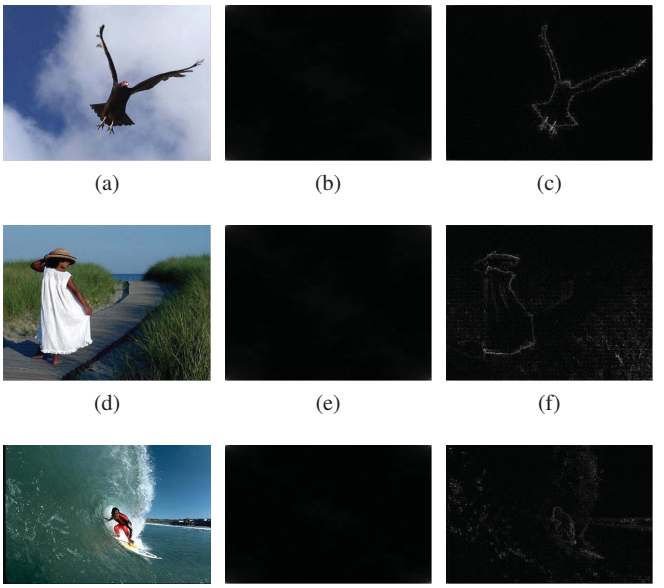


Figure 4: Examples of Images reconstruction: (a),(d),(g) Input Image, (b),(e),(h) Reconstructed with amplitude spectrum, (c),(f),(i) Reconstructed with phase spectrum

3 Proposed Model for Salient Object Detection

In our proposed algorithm, we reconstruct the image from unit magnitude spectrum and the original phase spectrum. To obtain the better feature map, we propose to blur the image first and then get the feature maps based on the phase spectrum of the blurred image. We propose to blur or smooth the image $I(x,y)$ by using Gaussian filter (g_1) with size of 3×3 and $\sigma = 0.5$.

$$\bar{I}(x,y) = \sum_{s=-1}^{1}\sum_{t=-1}^{1} g_1(s,t)I(x+s,y+t) \tag{6}$$

where $x = 0,1,2,\ldots,M-1$ and $y = 0,1,2,\ldots,N-1$ for image size of $M\times N$.

Now RGB image is converted to CIE L*a*b* color space. CIE Lab color model is perceptually uniform of color distribution and L component closely resembles human perception of intensity.

The selection of image resolution is another factor to take into account while calculating the saliency. Various resolution of an image represents the different perceptual observation of scene. So each resolution has different perception of saliency. There is equal chance of detecting unimportant objects as salient in one resolution and may miss the salient objects in another. Now as we know that PFT is applied only on a fixed scale of image to extract the salient object features. So there is definite chance of missing the salient objects in PFT. This drawback of PFT gives us scope to improve the algorithm in more precise manner by detecting the salient object in various scales, as all level of scales are equally important. We have proposed multi scale analysis for saliency detection.

To obtain multiple image resolutions for multi scale analysis we have used smoothing Gaussian kernel to generate Gaussian image pyramids. The number of levels (L) for Gaussian pyramids is calculated as:

$$\text{levels} = ||log_2(\min(\text{height}, \text{width})/10)|| \tag{7}$$

$$L = \lfloor(\text{levels}+1)/2\rfloor \tag{8}$$

where height, width are image resolution dimensions.

So L levels will give us images of different resolution ranging from $\bar{I}_0, \ldots, \bar{I}_{L-1}$, here $\bar{I}_0$ is the smoothed version of original image and $\bar{I}_{L-1}$ is the smoothest level image. Suppose original image resolution is 256×256 and $L = 3$ then we will have three images of 256×256; 128×128 and 64×64 resolutions. Now our algorithm PFT based saliency computation method is applied to each level.

$$\tilde{I}(f) = F(\bar{I}_L(x,y)) \tag{9}$$

$$A_f(f) = \Re(\tilde{I}(f)) \tag{10}$$

$$P_f(f) = \Im(\tilde{I}(f)) \tag{11}$$

$$S_L(x) = g_2(x) \times \left(F^{-1}\left[\frac{A_f(f)}{|A_f(f)|} e^{jP_f(f)}\right]^2 \right) \tag{12}$$

where F represents Fourier transform, F^{-1} is inverse Fourier transform, A_f and P_f denotes amplitude and phase of spectrum respectively. And $S_L(x)$ finally is the saliency computed for each level.

Now S_L computed for every level is summed up to take into account all the salient objects detected by every level at saliency map resolution of S_{L-1}. And finally

our algorithm generates master saliency map by taking weighted average of all the color channels.

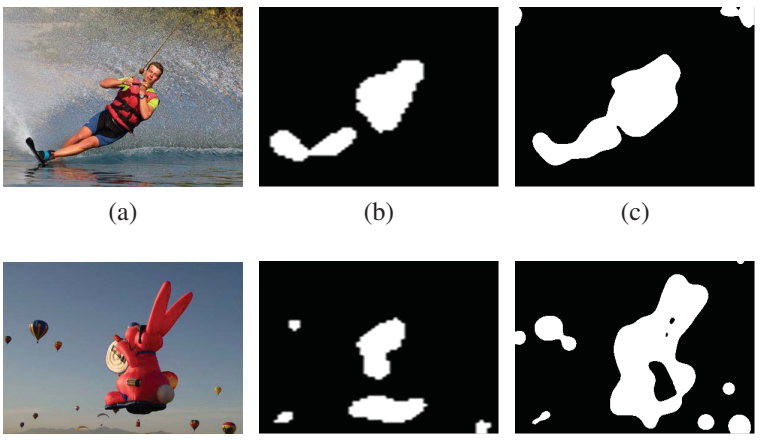
$$\bar{S} = \sum_{n=0}^{L-1} S_n \tag{13}$$

$$FS = \frac{2 \times \bar{S}(l) + \bar{S}(a) + \bar{S}(b)}{3} \tag{14}$$

Now to compute salient object based segmentation, the object map O will be computed as:

$$O = \begin{cases} 1 & \text{if } FS(x,y) > \text{Average}(FS(x,y)) \times \frac{3}{2}, \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

So here we have taken advantage of fast phase spectrum of Fourier transform based saliency detection by preserving the information at multiple scales of a scene. In the experimental results our algorithm is performing better than the PFT. Fig. 5 shows that our algorithm detects the most of the details while omitting the background while PFT based saliency detects less or some part of the salient objects.



(a) (b) (c)

(d) (e) (f)

Figure 5: Results of PFT and our proposed algorithm: (a),(d) Input Image,(b),(e) PFT and (c),(f) Our algorithm

4 Experimental Results

To analyse the results of our proposed algorithm, 100 Berkeley images are taken as test vectors. Saliency computations are done on those real images of the natural world. For comparative analysis we have tested our method against six state-of-art salient detection techniques. At first for ground truth calculation, we have asked to 4 different users to label the salient object individually, as various people have different concepts for salient objects in given image database. And the labels given by users are not same, so we have taken average of labels to decide the 'Ground Truth(G)'. Ground truth is a binary image where the average labelled pixel is represented as 1 and 0 otherwise.

$$G(x,y) = \begin{cases} 1 & \text{for } \frac{1}{4}\sum_{u=1}^{4} A_u(x,y), \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

where A_u represents labelled pixel by user uth.

The saliency maps of BM and GB are obtained by saliency detection implementations [Harel (2012)]. Other saliency maps are computed locally by implementation of above mentioned algorithms in MATLAB. The results are compared against the ground truths and our proposed algorithms.

4.1 Objective Analysis

For the objective analysis we have calculated precision (P_r), recall (R_e) and F-measure (F_α) against the ground truth [Zheshen and Baoxin (2008)]. As precision indicates fraction amount of correctly detected salient objects, while recall is the measure of fraction of ground truth salient objects detected and finally F-measure determines the weighted harmonic mean of precision and recall with a non-negative value of α. If G is the saliency map of ground truth and A is object mask of detected salient objects by any technique than:

$$P_r \frac{\sum_x g_x a_x}{\sum_x a_x}, \quad R_e = \frac{\sum_x g_x a_x}{\sum_x g_x} \tag{17}$$

And

$$F_\alpha = \frac{(1+\alpha) \times P_r \times R_e}{\alpha \times P_r + R_e}, \quad \alpha = 0.5 \tag{18}$$

For special case of $P_r = 0$ and $R_e = 0$ then $F_\alpha = 0$.

Table. 1 summarizes the performances of our proposed algorithm and six state-of-art for saliency detection. Our method performs better than BM, SR and even PFT with slightly more computation cost as comparable to PFT.

The comparative graphical analysis of average precision, average recall and average F-measure of test images for different saliency detection techniques is shown in Fig. 6. We observe that recall is much higher of our algorithm than any other techniques. It indicates our proposed method distinctively determines the salient objects in a scene. F-measure is also comparatively higher than most of the techniques. If we closely observe the analysis, FT and GB techniques are giving good performance than others but recall value of FT is less than our proposed method and GB is well known for its computational intensive and slow processing performance.

Table 1: Average Precision, average Recall and average F-measure of our proposed algorithm and other six methods including ground truth

Algorithms	P_r	R_e	F_α
Ground Truth	1.000	1.000	1.000
BM	0.693	0.134	0.259
FT	0.752	0.515	0.594
GB	0.628	0.579	0.550
CB	0.520	0.483	0.477
SR	0.487	0.249	0.337
PFT	0.486	0.313	0.368
Proposed method	*0.482*	*0.609*	*0.492*

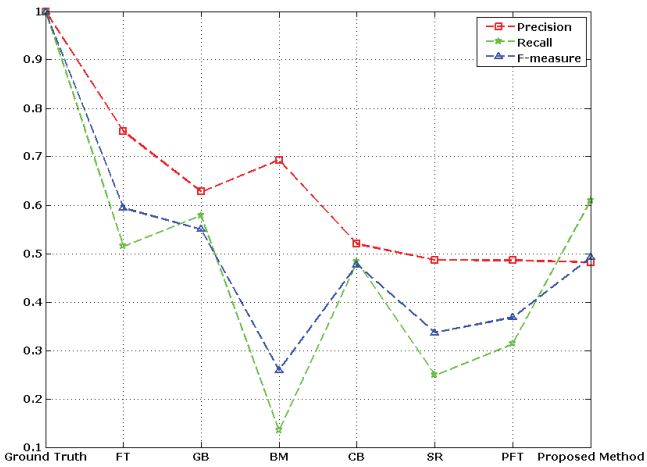


Figure 6: Performance comparison of different algorithms based on Precision, Recall and F-measure

4.2 Subjective Analysis

Fig. 7 shows the visual results of our proposed saliency detection in comparison with other six state-of-art for saliency detection. Based on visual inspection we can reason out that our proposed method is promising algorithm for salient object detections. As we see that most of the techniques detect very less fraction of ground truth salient objects in comparison to our proposed algorithm. This is also reflected in subjective analysis by higher recall value of our algorithm against others.

Figure 7: Experimental outcomes of our proposed algorithm against six state-of-art for salient object detection for subjective analysis.

5 Conclusion

In this paper, we have presented a efficient method to detect salient object in a visual scene. PFT algorithm gives good results but it operates on a fixed scale of input image, which might lead to wrong outputs. In our proposed algorithm we have incorporated the multiscale images analysis to detect the salient objects of an image using phase spectrum of Fourier transform. Saliency is calculated at each level and all saliency maps are summed up to generate final master saliency map. Intensive experimental results demonstrated that our proposed algorithm is performing much better than other six state-of-art of saliency detection. Based on objective and subjective analysis we conclude that our algorithm outperforms other methods and even performing better than PFT model. In brief, for efficient salient object detection with low complexity our proposed algorithm multiscale phase spectrum of Fourier transform based saliency detection is the promising candidate.

References

Achanta, R.; Hemami, S.; Estrada, F.; Ssstrunk, S. (2009): Frequencytuned salient region detection. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Adelson, E.; Anderson, C. H.; Bergen, J. R.; Burt, P. J.; Ogden, J. M. (1984): Pyramid method in image processing. *RCA Engineer*, vol. 29, no. 6, pp. 33–41.

Adelson, E. H.; Anderson, C. H.; Bergen, J. R.; Burt, P. J.; Ogden, J. M. (1984): Pyramid methods in image processing. *RCA Engineer*, vol. 29, no. 6, pp. 33–41.

Guo, C. L.; Qi, M.; Zhang, L. M. (2008): Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Harel, J. (2012): Saliency map algorithm : Matlab source code, 2012.

Harel, J.; Koch, C.; Perona, P. (2006): Graph-based visual saliency. In *Proc. NIPS*, pp. 545–552.

Hou, X.; Zhang, L. (2007): Saliency detection: A spectral residual approach. In *IEEE Proc. Conf. Computer Vision and Pattern Recognition*.

Itti, L.; Koch, C.; Niebur, E. (1998): A model of saliency- based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259.

Koch, C.; Ullman, S. (1985): Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology*, vol. 4, pp. 219–227.

Ma, Y.; Zhang, H. (2003): Contrast-based image attention analysis by using fuzzy growing. In *ACM International Conference on Multimedia*.

Niebur, E.; Koch, C. (1998): Computational architectures for attention. In Parasuraman, R.(Ed): *The Attentive Brain*, pp. 163–186. MIT Press, Cambridge, Mass.

Oppenheim, A. V.; Lim, J. S. (1981): The importance of phase in signals. *Proceedings of the IEEE.*, vol. 69, no. 5, pp. 529–541.

Rensink, R. (2000): Seeing, sensing, and scrutinizing. *Vision Research*, vol. 40, no. 10-12, pp. 1469–87.

Treisman, A. M.; Gelade, G. (1980): A feature-integration theory of attention. *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136.

van der Schaaf, A.; van Hateren, J. H. (1996): Modeling of the power spectra of natural images: statics and information. *Vision Research*, vol. 36, pp. 2759–2770.

Walther, D.; Koch, C. (2006): Modeling attention to salient protoobjects. *Neural Networks*, vol. 19, pp. 1395–1407.

Zheshen, W.; Baoxin, L. (2008): A two-stage approach to saliency detection in images. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 965–968.