Special Issue of International Journal of Computer Applications (0975 – 8887)
International Conference on Electronic Design and Signal Processing (ICEDSP) 2012

An Efficient Multiscale Phase Spectrum based Salient Object Detection Technique

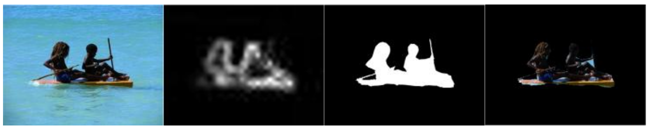
Deepak Singh
Department of Electronics & Communication
National Institute of Technology Rourkela - 769008, Odisha, India

Sukadev Meher
Department of Electronics & Communication
National Institute of Technology Rourkela - 769008, Odisha, India

ABSTRACT
Automatic image segmentation is emerging field in image processing research domain. Many researchers have developed various techniques for segmenting the interested region in an image. Saliency based image segmentation is one of the keen area of re- search. In a visual scene, the objects which are different from their surroundings get more visual importance and get high gaze attention of the viewer. There are several other applications also where saliency detection is used as core module such as object based surveillance, content adaptive data delivery for low data rate systems, automatic foveation system. In this paper, an efficient multi-scale phase spectrum based salient object detection method is pro- posed. 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 algorithm to multiple scales of the original image. As it known that, positional information is contained in the phase spectrum whereas amplitude spectrum contains the presence of frequency components, hence it is proposed to detect saliency using phase spectrum of Fourier trans- form. The proposed method performs much better than other previous methods and predicts more precisely salient objects. In experimental set-up, results of four state-of-art techniques for salient object detection are analyzed compared against the proposed method. The performance of the proposed method is measured on the basis of objective and subjective analysis.

Keywords
Foveated imaging, Salient object detection, Object based segmentation, Computer vision

1. INTRODUCTION
We ask that authors follow some simple guidelines. In essence, we ask you to make your paper look exactly like this document. The easiest way to do this is simply to download the template, and replace the content with your own material. Salient object detection of a visual scene is used in many industrial applications like object based segmentation, region based adaptive compression, object recognition and computer vision. Saliency detection is a method which is used to determine visually important regions within a scene. Task dependent and task independent methods (also known as top-down and bottom-up methods respectively) are two different approaches to define the salient object in vision analysis [7]. The task dependent method is top down; computational aggressive and slower in processing while a task independent



(a) (b) (c) (d)

Figure 1: Example of Region based segmentation (a) Input Image (b) Saliency Map (c) Ground Truth (d) Segmented output.

method is bottom up approach, scene dependent, saliency driven and responds quickly [10].

Most of the models for bottom-up saliency detections are biological models and based on human visual system (HVS). Various low- level 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 [3]. After a saliency map is computed for 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 final saliency of each pixel of the original image. An illustration of application of saliency map detection for a region based segmentation is shown in Figure 1.

Itti et al. (BM) proposed a bottom-up human vision attention model and which is the fundamental algorithm of a commercially available system called Neuromorphic Vision C++ Toolkit (NVT) [7]. After that, based on Rensink's theory [12], Walther created the most useful commercial product for saliency is SaliencyToolBox (STB)[15]. Harel et al. (GB) used graph for saliency detection [5]. They first form the activation map based on particular features 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 the saliency map by analyzing the log spectrum of an input image [6]. Ma et al. (CB) generates the saliency map based on center surround scheme by contrast analysis [9]. They believe that 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 [1].

In this paper, an efficient salient object detection technique is introduced. The phase spectrum of the Fourier transform (PFT) based saliency detection technique is proposed by Guo et al.[4]. But they have calculated saliency only for one scale of resolution of the image and it might miss information in one scale to another scale due to real world constraints. Proposed method takes 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 the phase spectrum based saliency are discussed in detail in section II. The proposed model is explained in detail in section III. The experimental results of the proposed method and comparative performance of other techniques against our proposed method are shown in section IV. The final conclusion is given in section V.

2. BACKGROUND CONCEPTS

2.1 Gaussian pyramids

In saliency detection, image resolution plays major role in deciding visual important objects. 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 the viewer's direction of gaze. So due to variations in distance, orientations and viewer's 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 [2]. Hence, multiscale salient object analysis is proposed to determine saliency more accurately.

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. Illustration of Gaussian pyramid is shown in figure 2.

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

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

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

where h(i, j) is 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 $\left[\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}\right]$. 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 III we have shown that obtaining the Gaussian pyramids is one of the essential steps in the proposed method for saliency detection.

Figure 2: Example of Gaussian Pyramids

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 the Feature Integration Theory (FIT) [13]. 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 [8].

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 × N is obtained as:

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

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

The Fourier amplitude spectrum and phase spectrum are defined as:

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

$$\theta(u, v) = tan^{-1}\left[\frac{\Im(F(u,v)}{\Re(F(u,v)))}\right] \quad (5)$$

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 [14]. The phase spectrum of an image is more important that the amplitude spectrum [11].

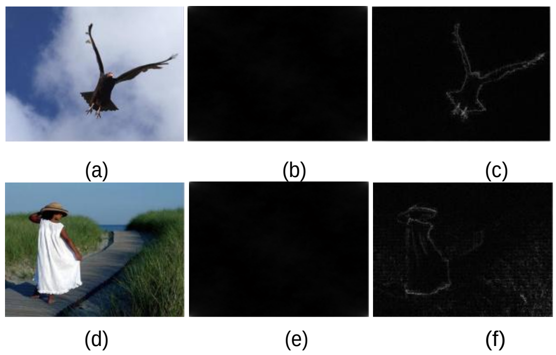


Figure 3: Examples of Images reconstruction: (d),(a) Input Image, (e),(b) Reconstructed with amplitude spectrum, (f),(c) Reconstructed with phase spectrum

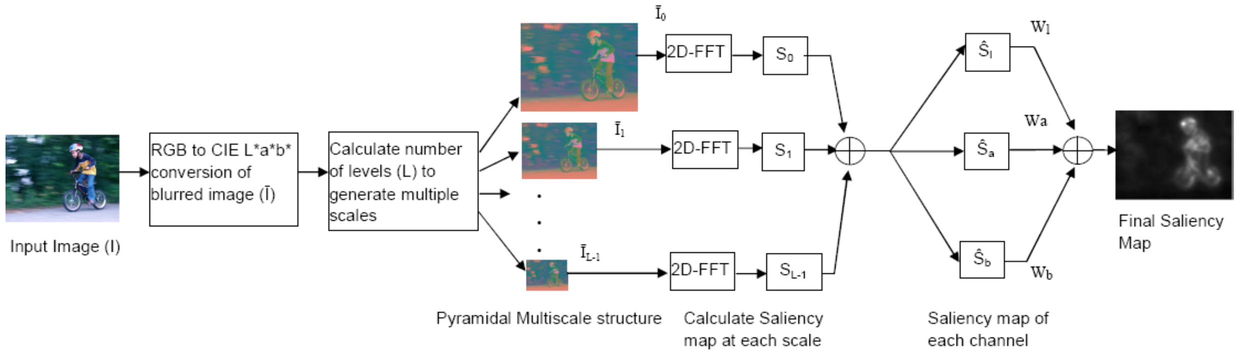


Figure 4: Flowchart of Proposed Salient Object Detection Method

So for the reconstruction of image $I(x,y)$ the inverse transform $e^{j\theta(u,v)}$of retains the resemblance to original image. Figure 3 shows some of examples of reconstruction of images using amplitude spectrum and phase spectrum.

3. PROPOSED MODEL FOR SALIENT OBJECT DETECTION

In proposed algorithm, it is proposed to reconstruct the image from unity magnitude spectrum and the original phase spectrum. To obtain the better feature map, blur the image first and then get the feature maps based on the phase spectrum of the blurred image. Blurring or smoothing of the image $I(x,y)$ by using Gaussian filter (g_1) is done as:

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

where y = 0, 1, 2, . . . , M −1 and y = 0, 1, 2, . . . , 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. And color components are having properties of color opponents same as human visual system. According to color opponent property, a color cannot have both red and green color components, or blue and yellow color components.

In determination of saliency map, selection of image resolution is very crucial. Various resolution of an image represents the different perceptual observation of scene. So each resolution has different perception of saliency. The algorithm might detect unimportant objects as salient in one resolution and may miss the salient objects in another. As PFT is applied only on a fixed scale of image, to extract the salient object features so PFT will not be able produce true salient objects. 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. To obtain multiple image resolutions for multiscale analysis, smoothing Gaussian kernel is used to generate Gaussian image pyramids. The number of levels (L) for Gaussian pyramids is calculated as:

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

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

where height, width are image resolution dimensions.

So L levels will give us images of different resolution ranging from $\bar{I}_0$... $\bar{I}_{L-1}$, 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 $256 \times 256, 128 \times 128$ and 64×64 resolutions. Now our algorithm PFT based saliency computation method is applied to each level as:

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

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

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

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

where F represents Fourier transform, F^{-1} is inverse Fourier transform, A_f and P_f denotes amplitude and phase of spectrum respectively. And finally $S_L(x)$ 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 proposed algorithm generates master saliency map by taking weighted average of all the color channels. Intensity channel (l) has given more weight than the color channels (a & b). As it is considered that, the intensity is the most important feature to decide the saliency [9]. The final saliency map will be obtained by:

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

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

The flowchart of the proposed salient object detection method is shown in figure 4. In the proposed method, we have taken advantage of fast phase spectrum of Fourier transform based saliency detection by preserving the information at multiple scales of a scene. Figure 6 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.

4. EXPERIMENTAL RESULTS

To analyze the results of the proposed algorithm, 100 Berkeley images are taken as test vectors. For comparative analysis, proposed method is tested against four state-of-art (BM[7], CB[9], SR[6] and PFT[4]) salient detection techniques. The image database along with an accurate object-contour based ground truth database and saliency maps for BM, CB and SR are obtained from [1]. Saliency maps for PFT and the proposed method are computed locally by implementing the algorithms in MATLAB version 7.10.0.499(R2010a) on 32-bit Intel(R)Core(TM)2 Duo CPU. The results are compared against the ground truths and the proposed algorithms.

4.1 Objective Analysis

For the objective analysis, precision (P_r), recall (R_e) and F- measure (F_α) are calculated against the ground truth. [16]. Precision is the fraction of retrieved documents that are relevant. is the fraction of relevant documents that are retrieved. A single measure that trades off precision versus recall is the F-measure, which is the weighted harmonic mean of precision and recall. In the present case 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:

$$P_r = \frac{\sum_x g_x a_x}{\sum_x a_x} \quad R_e = \frac{\sum_x g_x a_x}{\sum_x g_x} \tag{15}$$

And

$$F_\alpha = \frac{(1+\alpha)\times P_r \times R_e}{\alpha \times P_r + R_e} \tag{16}$$

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

Table1 summarizes the performances of proposed algorithm and four state-of-art for saliency detection. Proposed method performs better than BM, SR and even PFT.

The comparative graphical analysis of average precision, average recall and average F-measure of test images for different saliency detection techniques are shown in figure 5. It is observed that, the recall is much higher of proposed algorithm than any other techniques. It indicates proposed method distinctively determines the salient objects in a scene. F-measure is also comparatively higher than most of the techniques.

The comparative graphical analysis of average precision, average recall and average F-measure of test images for different saliency detection techniques are shown in Fig.5. It is observed that, the recall is much higher of proposed algorithm than any other tech- niques. It indicates proposed method distinctively determines the salient objects in a scene. F-measure is also comparatively higher than most of the techniques.

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

Algorithms	Pr	Re	Fα
Ground Truth	1.000	1.000	1.000
BM[7]	0.693	0.134	0.259
CB[9]	0.520	0.483	0.477
SR[6]	0.487	0.249	0.337
PFT[4]	0.486	0.313	0.368
Proposed method	0.482	0.609	0.492

4.2 Subjective Analysis

The visual results of proposed saliency detection in comparison with other four state-of-arts for saliency detection are shown in figure 6. Based on visual inspection, it can reason out that proposed method is promising algorithm for salient object detections. It is seen that most of the techniques detect very less fraction of ground truth salient objects in comparison to proposed algorithm. This is also reflected in subjective analysis by higher recall value of proposed algorithm against others.

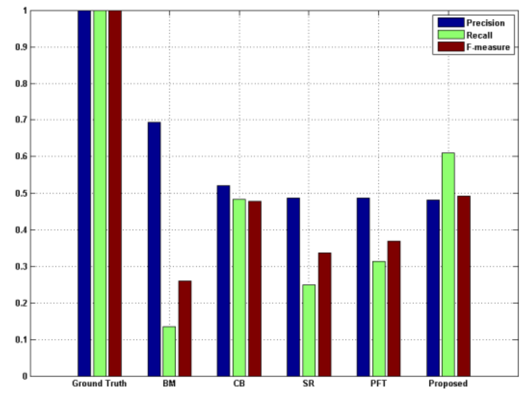


Figure 5: Performance comparison of state-of-art algorithms (BM[7], CB[9], SR[6] and PFT[4]) against the proposed method, based on Precision, Recall and F-measure.

5. CONCLUSION

In this paper, an efficient method is proposed to detect salient objects in a visual scene. PFT algorithm gives good results but it operates on a fixed scale of input image, which lead to wrong outcomes. In proposed algorithm, the multiscale images analysis method is incorporated to detect the salient objects of an image, using phase spectrum of Fourier transform. Saliency map is calculated at each level and all saliency maps are summed up to generate final master saliency map. Intensive experimental results demonstrated that the proposed algorithm is performing much better than other four state-of-art of saliency detection. However, the proposed method is computationally intensive due to multiscale analysis framework than the other aforementioned methods except BM [7] which generate 9 levels Gaussian pyramid images. Based on objective and subjective analysis, it is concluded that the proposed algorithm outperforms other methods and even performing better than PFT model. In brief, for efficient salient object detection with low complexity the proposed algorithm multiscale phase spectrum of Fourier transform based saliency detection is the promising candidate.

6. REFERENCES

[1] Radhakrishna Achanta, Sheila S. Hemami, Francisco J. Estrada, and Sabine Ssstrunk. 2009 Frequency-tuned salient region detection. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), Miami, Florida, USA, pages 1597–1604.

[2] E H Adelson, C H Anderson, J R Bergen, P J Burt, and J M Ogden. 1984. Pyramid methods in image processing. RCA Engineer, vol. 29(6):3341.

[3] P. J. Burt, E. H. Adelson, and Anderson. 1983. The laplacian pyramid as a compact image code. IEEE Transactions on Communication, vol.31(4):532–540.

[4] Chenlei Guo, Qi Ma, and LiMing Zhang. 2008. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In IEEE Computer Society

Figure 6: Experimental outcomes of our proposed algorithm against four state-of-arts for salient object detection for subjective analysis.

Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, Alaska, USA, pages 1–8.

[5] Jonathan Harel, Christof Koch, and Pietro Perona. 2006. Graph based visual saliency. In Advances in Neural Information Processing Systems 19, Proceedings of the 12th Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, pages 545–552. MIT Press.

[6] Xiaodi Hou and Liqing Zhang.2007. Saliency detection: A spectral residual approach. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), Minneapolis, Minnesota, USA.

[7] C. Niebur E. Itti, L. Koch. 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on pattern analysis and machine intelligence (PAMI), vol.20(11):1254–1259.

[8] C Koch and S Ullman. 1985. Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology, vol. 4(4):219–227.

[9] Yu-Fei Ma and HongJiang Zhang. 2003.Contrast-based image attention analysis by using fuzzy growing. In Proceedings of the 11th ACM International Conference on Multimedia, Berkeley, CA, USA, pages 374–381.

[10] E. Niebur and C. Koch. 1998. Computational architectures for attention. The Attentive Brain, pages 163–186.

[11] A V Oppenheim and J S Lim. 1981.Importance of phase in signals. Proceedings of the IEEE, vol. 69(5):529–541.

[12] R. Rensink. 2000. Seeing, sensing, and scrutinizing. Vision Research, Vol. 40, No. 10-12:1469–1487.

[13] A M Treisman and G Gelade. 1980. A feature-integration theory of attention. Cognitive Psychology, vol. 12(1):97–136.

[14] A Van Der Schaaf and J H Van Hateren. 1996. Modeling the power spectra of natural images: Statistics and information. Vision Research, vol.36:2759–2770.

[15] DirkWalther and Christof Koch. 2006. Modeling attention to salient proto-objects. Neural Networks, vol.19(9):1395–407.

[16] Zheshen Wang and Baoxin Li. 2008. A two-stage approach to saliency detectin in images. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 965–968.

.