

# PERCEPTIONAL CLEAVAGE WITH OBSESSION

P.Senthil Kumaran<sup>#1</sup>,  
P.G Scholar  
PRIST University,  
Trichy, Tamilnadu, India.

R.Sudha<sup>#2</sup>,  
Assistant professor  
PRIST University,  
Trichy, Tamilnadu, India.

B. SakthiSaravanan.<sup>#3</sup>,  
Assistant professor  
Saveetha Engineering College,  
Thandalam, Tamilnadu, Chennai, India

**Abstract**— Attention is an automatically detect visually interesting regions in images has many practical applications, especially in the design of active machine vision and automatic visual surveillance systems. An integral part of the human visual system and has been widely studied in the visual attention literature. Analysis of the statistics of image features at observers can provide insights into the mechanisms of fixation selection in humans. Using a foveated analysis framework, we studied the statistics of four low-level local image features: luminance, contrast, and band passout puts of both luminance and contrast, and discovered that image patches around human fixations had, on average, higher values of each of these features than image patches selected at random. Contrast-bandpass showed the greatest difference between human and random fixations, followed by luminance-bandpass, RMS contrast, and luminance. Using these measurements, we present a new algorithm that selects image regions as likely candidates for fixation. These regions are shown to correlate well with fixations recorded from human observers. In fact, we propose a segmentation refinement process based on such a feedback process. Finally, our experiments show the promise of the proposed method as an automatic segmentation framework for a general purpose visual system.

**KEYWORD**—Eye tracking, Fixation selection, Fixation-based segmentation, object segmentation, visual attention

## I. INTRODUCTION

The human visual system is constantly bombarded with a slew of visual data, from which it actively selects and assimilates relevant visual information in an efficient and seemingly effortless manner. Despite a large field of view, the human visual system processes only a tiny central region (the fovea) with great detail while the resolution drops rapidly towards the periphery [1]. The human (primate) visual system observes and makes sense of a dynamic scene (video) or a static scene sense of a dynamic scene (video) or static scene locations in the scene. The eye movement between consecutive fixations is called a saccade. The active nature of looking, as instantiated in the human visual system, promises to have advantages in both speed and reduced storage requirements in artificial vision systems as well. In fact, several forested vision sensor arrays have been designed and used in real-time imaging systems [1]–[3].

The next generation of efficient, forested, active vision systems [4] could potentially be applied to a diverse array of problems such as automated pictorial database query, image understanding, image quality assessment [6], automated object detection, autonomous vehicle navigation, and real-time, foveated video compression [7], [8]. The role of fixational eye movements—the involuntary eye movements—during a fixation is even more unclear. In fact, for a long time, these eye movements were believed to be just a neural tick and not useful for visual perception. However, neuroscientists have recently revived the debate about the nature of these movements and their effects on visual perception. Bottom-up approaches to gaze selection assume that eye movements are quasi-random and driven by low-level image features. They propose a computational model for human gaze selection based on image processing to accentuate certain image features that are deemed relevant for drawing gaze. The influence of certain low-level image features such as edges and areas of high curvature in drawing fixations was established as early as 1935 [10], [11]. Williams, studied the influence of color, shape, and size in visual search and concluded that, among the attributes studied, the color of the target was the most important image feature in influencing saccades.

### A. Well-Posed Problem

In computer vision literature, segmentation essentially means breaking a scene into non overlapping, compact regions where each region constitutes pixels that are bound together on the basis of some similarity or dissimilarity measure. Over the years, many different algorithms [10], have been proposed that segment an image into regions, but the definition of what is a correct or “desired” segmentation of an image (or scene) has largely been elusive to the computer vision community. In fact, in our view, the current problem definition is not well posed. let us take an example of a scene (or image) shown in Fig. 1. In this scene, consider two of the prominent objects: the tiny bull and the pair of players.

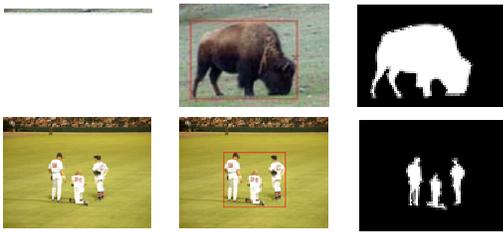


Figure 1. Segmentation of a natural scene in (a) using the Normalized Cut algorithm [36] for two different values of its input parameter



Figure 2. Examples of images used for the experiment

## II. EYE TRACKING METHODS

GAFFE is based on a gaze-attentive framework; this means that the features used for fixation selection are those that are statistically significant at recorded human gaze locations (when compared to features at randomly selected fixations). This section describes the experimental procedure that was used to record human eye movements in a natural viewing task.

### A. Stimuli and Tasks

101 static images of size 1024\*768 pixels were manually selected from a calibrated gray scale natural image database [31]. Since we were interested in developing a bottom-up framework for fixation selection, images containing man-made structures and features such as animals, faces, and other items of high-level semantic interest that could have instinctively attracted attention were omitted. Typical images are shown in Fig. 2. The stimuli were displayed on a 21-inch, gamma corrected monitor at a distance of 134 cm from the observer. The screen resolution corresponded to about 1 arc minute per pixel. Each image was displayed for 5 seconds in a fixed order for all observers. Observers were instructed to view each of the images as they desired. All observers began viewing the image stimuli from the center of the screen. Following the display of each image, observers were shown a small image patch and asked to indicate whether the image patch was from the image they just viewed or not. This task was used to encourage observers to scan the entire scene. A total of 29 (24 native) adult human volunteers participated in this study. All observers either had normal or corrected-to-normal vision.

### B. Eye Tracking

Human eye movements were recorded using an SRI Generation V Dual Purkinje eye tracker. It has an

accuracy of <10 arc minute, and a precision of ~1 arc minute. A bite bar and forehead rest was used to restrict the observer's head movements. The observer was first positioned in the eye tracker and a positive lock established onto the observer's eye

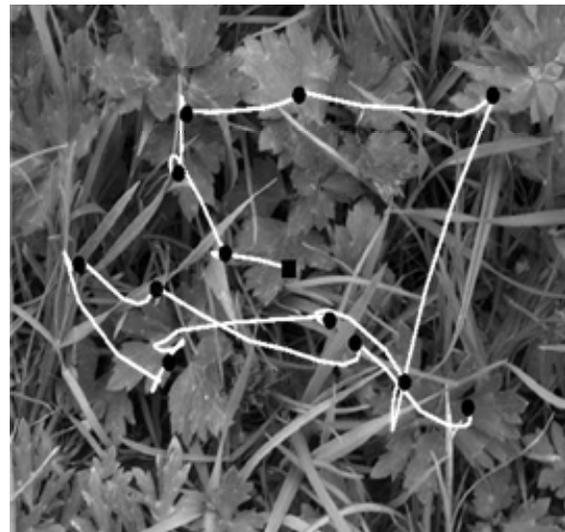


Figure 3. Example of an observer's eye movement trace superimposed on the image stimulus. The dots are the computed fixations. The square in the centre of the image is the first fixation.

A linear interpolation on a 3\*3 calibration grid was then done to establish the linear transformation between the output voltages of the eye tracker and the position of the observer's gaze on the computer display. The output of the eye tracker (horizontal and vertical eye position signals) was sampled at 200 Hz and stored for offline data analysis. This calibration routine was repeated every 10 images, and a calibration test run after every image.

### C. Image Data Acquisition

The gaze coordinates corresponding to the eye movements of the observers for each trial were divided into fixations and saccades using spatial temporal criteria derived from the known dynamic properties of human saccadic eye movements [11]. The resulting pattern of fixations for a single trial is shown by the dots in Fig. 3. The lines show the eye movement trajectories linking the fixations. As mentioned earlier, we propose a foveated framework to analyze the statistics of low-level features of image patches at the resolution at which they were encoded by the observer. To achieve this, the image was first foveated at the observer's current fixation, say  $n$ , and a patch centered at the subsequent fixation,  $n+1$ , was extracted for analysis. Thus all image patches were analyzed at the resolution

at which they were encoded *prior* to fixating the patch. We then extracted circular patches of diameters 32,64,96,160,192 pixels centered at each fixation. This corresponded to patches of diameter ranging from  $0.5^\circ$  to  $3.2^\circ$ . A consequence of using such a foveated analysis framework is that the ensemble of patches extracted around fixations contains image patches that have been blurred to different extents. Further, it is also possible that saccades of different magnitudes are driven by different features. Thus, there arises a need to perform an eccentricity-based analysis of local image features, where patches of similar blur are grouped together and the relevant image feature is analyzed separately for each blur. Tatle *ret al.*[33] have observed that the influence of image features are not uniform across saccade magnitudes and note that by ignoring the dependence of image features magnitudes, prior work in this area ([9], [10], [11]) generally tends to estimate the influence of visual features incorrectly. In our study, since we use a foveated analysis framework, we analyze patches over the range of spatial frequencies at which they were processed by the human visual system, and thus incorporate both saccade and spatial frequency dependence of image patches into our analysis.

### III. BASIC CONCEPT SEGMENTATION BY COMPOSITION

Examining the image segments of Fig.4, we note that good segments of significantly different types share a common property : Given any point within a good image segment, it is easy to compose describe its surrounding region using other chunks of the same segment (like a 'puzzle'), where as it is difficult to compose it using chunks from there mining parts of the image. This is trivially true for uniformly colored and textured segments (Fig.4.a,4.b,4.c), since each portion of the

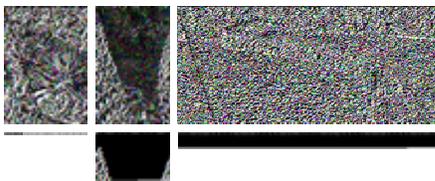


Figure 4.a, 4.b, 4.c. What is a good image segment?

segment. Portions of the same segment (the dome), but difficult to compose using chunks from there mining parts of the image (thesky). The same property carries to more complex structured segments, such as the compound segment

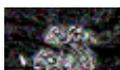


Figure 4.d

In Fig.4.d. The surrounding region of each point in the segment is easy to describe "using portions of other. The existence of several images provides 'visual evidence' that the co occurrence of different parts (orange beak, black neck, white body, etc.) is not coincidental, and all belong to a single compound segment. Similarly, one half of a complex symmetric object.

### IV. GENERATE SALIENCY DETECTION

Visual saliency is the perceptual quality that makes an object, person, or pixel stand out relative to its neighbours and thus capture our attention. Visual attention results both from fast, pre-attentive, bottom-up visual saliency of the retinal input, as well as from slower, top-down memory and volition based processing that is task-dependent [12].

The focus of this paper is the automatic detection of visually salient regions in images, which is useful in applications such as adaptive content delivery [13], adaptive region-of-interest based image compression [4], image segmentation [8, 9], object recognition [6], and content aware image resizing [2]. Our algorithm finds low-level, pre-attentive, bottom-up saliency. It is inspired by the bio- logical concept of center-surround contrast, but is not based on any biological model.

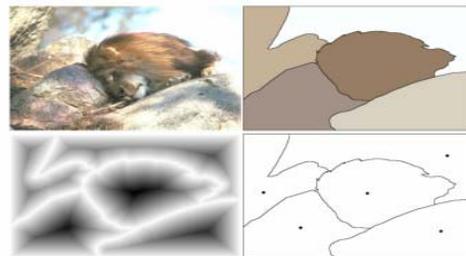


Figure 5. Original images and their saliency maps

Current methods of saliency detection generate regions that have low resolution, poorly defined borders, or are expensive to compute. Additionally, some methods produce higher saliency values at object edges instead of generating maps that uniformly cover the whole object, which results from failing to exploit all the spatial frequency content of the original image. We

analyze the spatial frequencies in the original image that are retained by five state-of-the-art techniques, and visually illustrate that these techniques primarily operate using extremely low-frequency content in the image. We introduce a frequency-tuned approach to estimate center surround contrast using color and luminance features that offers three advantages over existing methods: uniformly highlighted salient regions with well-defined boundaries, full resolution, and computational efficiency. The saliency map generated can be more effectively used in many applications, and here we present results for object segmentation. We provide an objective comparison of the accuracy of the saliency maps against five state-of-the-art methods using a ground truth of a 1000 images. Our method outperforms all of these methods in terms of precision and recall.

## V. COMPARISONS

The true usefulness of a saliency map is determined by the application. In this paper we consider the use of saliency maps in salient object segmentation. To segment a salient object, we need to binarize the saliency map such that ones (white pixels) correspond to salient object pixels while zeros (black pixels) correspond to the background.

A fixed threshold to binarize the saliency maps. In the second experiment, we perform image-adaptive binarization of saliency maps.

In order to obtain an objective comparison of segmentation results, we use a ground truth image database. We derived the database from the publicly available database used by Liu et al. [13]. This database provides bounding boxes drawn around salient regions by nine users. However, a bounding box-based ground truth is far from accurate, as also stated by Wang and Li [28]. Thus, we created an accurate object-contour based ground truth database<sup>2</sup> of 1000 images (examples in Fig. 6).

### A. Segmentation by fixed thresholding

For a given saliency map, with saliency values in the range  $[0, 255]$ , the simplest way to obtain a binary mask for the salient object is to threshold the saliency map at a threshold  $T_f$  within  $[0, 255]$ . To compare the quality of the different saliency maps, we vary this threshold from 0 to 255, and compute the precision and recall at each value of the threshold. The resulting precision versus recall curve is shown in Fig. 5. This curve provides a reliable comparison of how well various saliency maps highlight salient regions in images. It is interesting to note that Itti's method shows high accuracy for a very low recall ( $< 0.1$ ), and then the accuracy drops steeply. This is because the salient pixels from this

method fall well within salient regions and have near uniform values, but do not cover the entire salient object. Methods GB and AC have similar performance despite the fact that the latter generates full resolution maps as output. At maximum recall, all methods have the same low precision value. This happens at threshold zero, where all pixels from the saliency maps of each method are retained as positives, leading to an equal value for true and false positives for all methods.

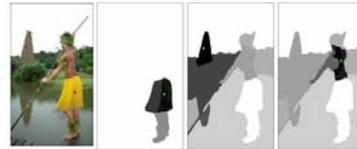


Figure 6. Ground truth examples

### B. Segmentation by adaptive thresholding

Maps generated by saliency detectors can be employed in salient object segmentation using more sophisticated

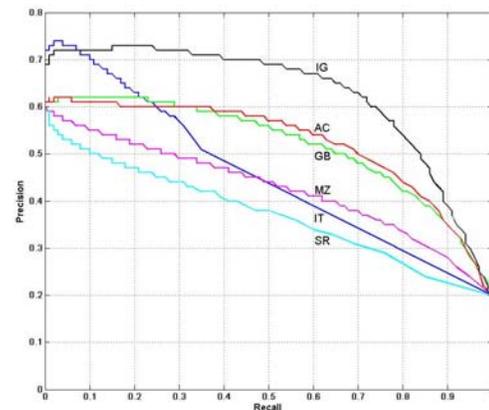


Figure 7. Precision-recall curve for naive thresholding of saliency maps.

methods than simple thresholding. Saliency maps produced by Itti's approach have been used in unsupervised object segmentation. Han et al. [9] use a Markov random field to integrate the seed values from Itti's saliency map along with low-level features of color, texture, and edges to grow the salient object regions. Ko and Nam [18] utilize a Support Vector Machine trained on image segment features to select the salient regions of interest using Itti's maps, which are then clustered to extract the salient objects. Ma and Zhang [22] use fuzzy growing on their saliency maps to confine salient regions within a rectangular region.

We use a simpler method for segmenting salient objects, which is a modified version of that presented in [1]. Their technique makes use of the intensity and color properties of the pixels along with their saliency values to segment the object. Considering the full resolution saliency map, their technique over-segments

the input image using k-means clustering and retains only those segments whose average saliency is greater than a constant threshold. The binary maps representing the salient object are thus obtained by assigning ones to pixels of chosen segments and zeroes to the rest of the pixels.

We make two improvements to this method. First, we replace the hill-climbing based k-means segmentation algorithm by the mean-shift segmentation algorithm [5], which provides better segmentation boundaries. We perform mean-shift segmentation in Lab color space. We use fixed parameters of 7, 10, 20 for  $\sigma_S$ ,  $\sigma_R$ , and  $\minRegion$ , respectively, for all the images (see [7]).

## VI. CONCLUSION

We proposed here a novel formulation of segmentation in conjunction with fixation. The framework combines static cues with motion and/or stereo to disambiguate between the internal and the boundary edges. The approach is motivated by biological vision, and it may have connections to neural models developed for the problem of border ownership in segmentation [11]. Although the framework was developed for an active observer, it applies to image databases as well, where the notion of fixation amounts to selecting an image point which becomes the center of the polar transformation. Our contribution here was to for mutate an old problem—segmentation—in a different way and show that existing computational mechanisms in the state-of-the-art computer vision are sufficient to lead us to promising automatic solutions. Our approach can be complemented in a variety of ways, for example, by introducing a multitude of cues. An interesting avenue has to do with learning models of the world. For example, if we had a model of a “horse,” we could segment the horses more correctly in Fig. 5. This interaction between low-level bottom-up processing and high-level top-down attentional processing, is a fruitful research direction.

## REFERENCES

- [1] Bagon, O. Boiman, and M. Irani, “What Is a Good Image Segment? A Unified Approach to Segment Extraction,” Proc. 10<sup>th</sup> European Conf. Computer Vision, pp. 30-44, 2008.
- [2] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, “Interactive Image Segmentation Using an Adaptive GMMRF Model,” Proc. European Conf. Computer Vision, pp. 428-441, 2004.
- [3] Y.Y. Boykov and M.P. Jolly, “Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in n-d Images,” Proc. Eighth IEEE Int’l Conf. Computer Vision, pp. 105-112, 2001.
- [4] Y.Y. Boykov and V. Kolmogorov, “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 26, no. 9, pp. 1124-1137, Sept. 2004.
- [5] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert, High Accuracy Optical Flow Estimation Based on a Theory for Warping, pp. 25-36 Springer, 2004.
- [6] N.D.B. Bruce and J.K. Tsotsos, “Saliency, Attention, and Visual Search: An Information Theoretic Approach,” J. Vision, vol. 9, no. 3, pp. 1-24, 2009.
- [7] V. Ferrari, T. Tuytelaars, and L.V. Gool, “Object Detection by

- Contour Segment Networks,” Proc. European Conf. Computer Vision, pp. 14-28, June 2006.
- [8] M. Gur, A. Beylin, and D.M. Snodderly, “Response Variability of Neurons in Primary Visual Cortex (V1) of Alert Monkeys,” J. Neuroscience, vol. 17, pp. 2914-2920, 1997
- [9] J.H.vanHateren and A.vanderSchaaf, “Independent component filters of natural images compared with simple cells in primary visual cortex,” Proc Biol Sci, vol. 265, no. 1394, pp. 359-366, Mar. 1998.
- [10] P. Reinagel and A. M. Zador, “Natural scene statistics at the centre of gaze,” Network: Computation in Neural Systems, vol. 10, no. 4, pp. 341-350, 1999.
- [11] D.J.Parkhurst and E.Niebur, “Scene contents selected by active vision,” Spatial Vision, vol. 16, no. 2, pp. 125-154, June 2003.
- [12] Y.-F. Ma and H.-J. Zhang. Contrast-based image attention analysis by using fuzzy growing. In ACM International Conference on Multimedia, 2003
- [13] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. ACM Transactions on Graphics, 26(3), 2007

## AUTHORS PROFILE

**P.Senthil Kumaran** pursuing M.Tech (CSE) in PRIST University, Tamilnadu, India. He had received the Master of Computer Applications degree from Bharathidasan University, Tiruchirappalli, Tamilnadu, India. He has more than Fifteen years of experience as Senior Lecturer. He was awarded Best Lecturer in 3 times. His area of interest are Data structures, Image Processing and Data Mining.



**R.Sudha** received her Master of Engineering in Computer Science and Engineering from Pavendar Bharathidasan College of Engineering Under Anna University, India. She is Currently working as an Assistant professor in PRIST University, Trichy. She has Published a paper titled “A System Tool for Identification of RAGAS using MIDI (Musical Instrument Digital Interface) for CMIR (Classical Music Information Retrieval)” in International Conference held in Dubai 28-30, 2009. Published a paper titled “Adaptive Location Aided Routing in Mobile ad-hoc Network (ALARM)” in a National Conference held in PSNA college, Dindigul 2006. Her area of Interest includes Multimedia (Musical Information Retrieval), Database and Networking.



**B.Sakthisaravanan** has received B.TECH from Anna University, Chennai. He also received M.Tech (IT) from Satyabama University, Tamilnadu, India. He has 3 years of experience as Lecturer. He is currently working as an Assistant professor in Saveetha Engineering College, Chennai. He has Published number of national and international conferences. His area of interest are Data Structures, Data Base Management Systems, Computer Networks and Computer Architecture.

