PREDICTING THE PERCEIVED INTEREST OF OBJECTS IN IMAGES

By

SRIVANI PINNELI

Master of Science in Electrical and Computer Engineering

Oklahoma State University

Stillwater, Oklahoma

2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2008

PREDICTING THE PERCEIVED INTEREST

OF

OBJECTS IN IMAGES

Thesis Approved:

Dr. Damon M. Chandler

Thesis Adviser

Dr. Guoliang Fan

Dr. Qi Cheng

Dr. A. Gordon Emslie

Dean of the Graduate College

ACKNOWLEDGEMENTS

First of all, I would like to thank my adviser Dr. Damon Chandler for his invaluable support throughout my Masters and research, which has been a key aid for my survival at Oklahoma State University and success in my career. I would also like to thank my lab mates at the Image coding and Analysis Lab (ICAL) for being supportive and for their critical views.

I would like to thank Dr. Guoliang Fan for his guidance in image processing and computer vision.

Lastly, I would like to thank my fiancée Sivaram Vempati for his help and support through my Masters. I would also like to thank my family for their unconditional love and support.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Motivation

EVERY image we look at has some objects that demand more attention than others. It has also been proven that under task free conditions humans tend to disregard certain objects in a scene irrespective of the observation time. So, what is it that makes certain objects more important than the others? Is it the location, color, contrast or any other high-level factor?

Consider, for example, the image shown in Figure 1.1.

Figure 1.1. An image from the Berkeley Segmentation Dataset and Benchmark image

For a human, it is quite evident that the butterfly forms the most important region in this image. So what is about this object that makes it more visually interesting than the flowers, leaves and other objects in the scene? One logical reason is the size: The butterfly is larger than the flowers and leaves in the background. Another reason is the location: The butterfly is located at the center of the image. Other factors like the butterfly's contrast and the extent to which the butterfly's presence in the foreground also contributes to our impression of visual interest.

If we are provided with an algorithm that can predict the most important region or object in an image, it would cater to many applications in the field of image processing and human vision. The ability to quantify perceived interest would also be useful in areas like unequal error protection, watermarking, variable-resolution displays, image segmentation, region based retrieval, adaptive compression, automatic image retargeting [20] and many more. For example, in image compression, the ability to quantify the perceived interest of different objects would allow us to devote more bits to the significant objects.

Here, we present an algorithm to quantify the perceived interest of objects in the images. Our algorithm follows the work of Osberger *et al.* and uses various factors (e.g., color, location, contrast, edge-strength, and blur) to determine perceived interest. However, unlike Osberger *et al.*, our approach uses a Bayesian framework based on precise likelihood functions measured explicitly via a psychophysical experiment.

Hence, the main objective of this research is to predict the importance of defining "regions of interest" in an image for improving the perceived fidelity of the image and propose an algorithm that would find these regions in an image. The "regions of interest"

or popularly known as the “ROI” can be defined as the objects or regions to which draw the attention of the viewer.

1.2 Review of Literature

Researchers have proposed algorithms to locate and quantify regions of interest in images [1]-[4]. In [1] Osberger *et al.* present a method to automatically determine the perceptual importance of different regions by combining various factors like size, location, contrast, and shape. An image is first segmented into regions. Then each factor (e.g., size) is measured and converted into a relative level of interest. The results of all the factors are then squared and summed to produce a final interest map. However, a more accurate way to measure and combine the various factors to arrive at the overall interest map still remains an open question to answer.

In [2], Itti *et al.* computed features based on linear filters and center-surround structures encoding intensity, orientation, and color to construct a saliency map that reflects areas of high attention. This model is biological in nature where the visual input is first decomposed into a set of topographic feature maps. Feature maps are then combined in a bottom-up manner to form a final saliency map which consists of locations that stand out from the surroundings. Stentiford [3] proposed a measure of visual attention that depends upon the dissimilarity of neighborhoods in an image. In [4], Stark *et al.* provide an analysis of effectiveness of various image-processing operations and clustering procedures in predicting regions-of-interest in images.

Most existing visual attention approaches are based on low-level features such as color, contrast and luminance, because the visual attention is unconsciously driven by

such low-level features. Almost all the approaches extract low-level features first and then combine them in a unique fashion to obtain the final saliency map of an image.

The approaches to determine the saliency can be broadly classified as biological models and purely computational ones. Ma and Zhang [24] propose a method that is not a biological model and is based on local contrast to generate saliency maps. Frintrop *et al* [17] use integral images in VOCUS (Visual Object Detection with a Computational Attention System) to speed up computation of center-surround differences to find salient regions using separate feature maps of color, intensity, and orientation. An improvement using integral images has been proposed by Achanta *et al* [12] where saliency maps of same size and resolution as the input image are generated using luminance and color as low-level features. This is accomplished by resizing the filter at each scale instead of the image and thus maintains the same resolution as the original image.

A coherent computational model of visual selective attention is presented by Olivier *et al* [25] who follow the work of Koch and Ullman [27]. The basic difference is the way they normalize the early visual features. The normalization is done automatically using visibility threshold. Another unsupervised extraction of visual attention objects is proposed by Han *et al* [18] *where* a generic model is used. The approach uses Markov random field (MRF) to integrate computational visual attention mechanisms with attention object growing techniques. A hierarchical selectivity for object-based visual attention is proposed by Sun *et al* [26] where integration of bottom-up and top-down attentional setting is done.

Another contribution to this field is by Jian *et al* [22] who use a large database to validate their results. In this approach, they formulate the salient object detection as an

image segmentation problem. They proposed a set of features like multi-scale contrast, center-surround histogram, and color spatial distribution to detect a salient object locally, regionally, and globally. Liu *et al* [22] presented a region enhanced scale-invariant saliency detection method. They constructed a scale-invariant saliency map by segmenting the image into regions and then enhance the saliency map with the region information. Chul Ko *et al* [16] proposed an object-of-interest (OOI) segmentation algorithm that is based on human attention and semantic region clustering. In this approach, an image is first segmented into regions and then merged as a semantic object. A support vector machine is used within an attention window (AW) which is based on the saliency map and saliency points from the image. Hu *et al* [19] proposed a subspace analysis for detecting visual attention where an image is represented in a 2D space using polar transformation of its features. A subspace estimation algorithm is proposed based on Generalized Principal Component Analysis (GPCA). Another biologically motivated visual attention system is proposed by Choi *et al* [15] using bottom-up saliency map and top-down inhibition. A training selective attention model is used which inhibits the unwanted salient area and thus only focus on an interesting area in static natural scene.

Bayesian probabilistic approaches have also been used to locate regions of interest. Luo *et al* [7] proposed a computational approach to determine the main subject in photographic images. The algorithm consists of region segmentation, perceptual growing, feature extraction, and then Bayesian probabilistic reasoning. Another application of Bayes' theorem is given by Chen *et al* [8] to discover main subjects in video; the algorithm in [8] consists of an appearance model and a motion model.

Though, a lot of work has been done in this field, there are still some major areas of concern which are addressed in our research. Most of the approaches mentioned above perform main subject detection [12, 24, and 21] whereas our algorithm can perform multi-level region of interest detection. Another common problem is the need of segmentation [22, 16] in the first stage to perform saliency detection. Due to the computational intensity of segmentation, it becomes almost impossible to use the approach for real-time applications. To encounter this problem, we have come up with a block-based approach to perform saliency detection. This approach doesn't need any segmentation and is fast enough to be used in real-time applications. We have also compared our results with the ground truth data obtained by conducting a psychophysical experiment where the subjects rate a set of 300 images from the Berkley image segmentation database.

1.3 Outline

This thesis is organized as follows: Chapter 2 describes the psychophysical experiment performed to obtain the likelihood functions. Chapter 3 presents the algorithm that utilizes various factors like location, color, luminance, contrast, edge strength, blur and then combines these factors to come up with an importance map of each object in an image. Chapter 4 gives the results obtained using our algorithm as well as a comparison with other existing algorithms. General conclusions and future work are provided in Chapter 5.

CHAPTER 2

PSYCHOPHYSICAL EXPERIMENT

The primary goal of this experiment is to obtain the likelihood functions of perceived interest of over 1100 objects in 300 images from Berkeley image segmentation database. A psychophysical experiment was performed in which subjects rated the perceived interest of 1143 objects in 300 images. For each of the 1143 objects, subjects were instructed to rate the perceived interest relative to the other objects within the image. The ratings were performed using an integer scale of 1 to 10 where 10 corresponded to greatest interest and 1 corresponded to least interest. Subjects were given unlimited time.

2.1 Methods

2.1.1 Apparatus and Subjects

Stimuli were displayed on high-resolution, ViewSonic VA912B 19-inch monitor. The display yielded minimum and maximum luminance of 2.7 and 207 cd/m^2 respectively and an overall gamma of 2.9. Stimuli were viewed binocularly through natural pupils at a distance of 46 cm under D65 lighting.

A total of fifteen subjects participated in the experiment out of which, thirteen adult subjects were naive to the purpose of the experiment, and the remaining two were the authors. The experiment was conducted prior to the development of the proposed

algorithm, and thus the psychophysical results from the authors were not biased toward the algorithm. The results from the authors were very much in agreement with those from the naïve subjects (correlation $R > 0.8$). Subjects ranged in age from 23 to 34 years. All subjects had either normal or corrected-to-normal visual acuity.

2.1.2 Stimuli and Methods

Images used in the experiment were obtained from the Berkeley Segmentation Dataset and Benchmark image database [9]. This database was chosen because its images are accompanied by human-segmented versions (averaged over at least five subjects). The database consists of 300 images out of which 200 are the training set whereas the remaining 100 form the testing set. The same set of 200 images has been used as training set for our algorithm and the remaining set of 100 images has been used to test the algorithm. The images used were of size 321×481 with 24-bit RGB pixel values.

Though the images were hand-segmented, they were not segmented into different regions. Hence, all objects in the image were given different colors to identify different objects in an image by the first author which is shown in the third column of Figure 2.1. All the 300 images were segmented into a total of 1143 objects. The average number of objects per each image was around 4. For each of the 1143 objects, subjects were instructed to rate the perceived interest relative to the other objects within the image. The ratings were performed using an integer scale of 1 to 10 where 10 corresponded to greatest interest and 1 corresponded to least interest. Subjects were given unlimited time. The ratings were converted to z-scores and then normalized to 0 to 1 which is discussed

in the next section. In general, subjects tended to agree with each other: The minimum, maximum, and mean standard deviation of the z-scores was 0, 0.94, and 0.29 respectively.

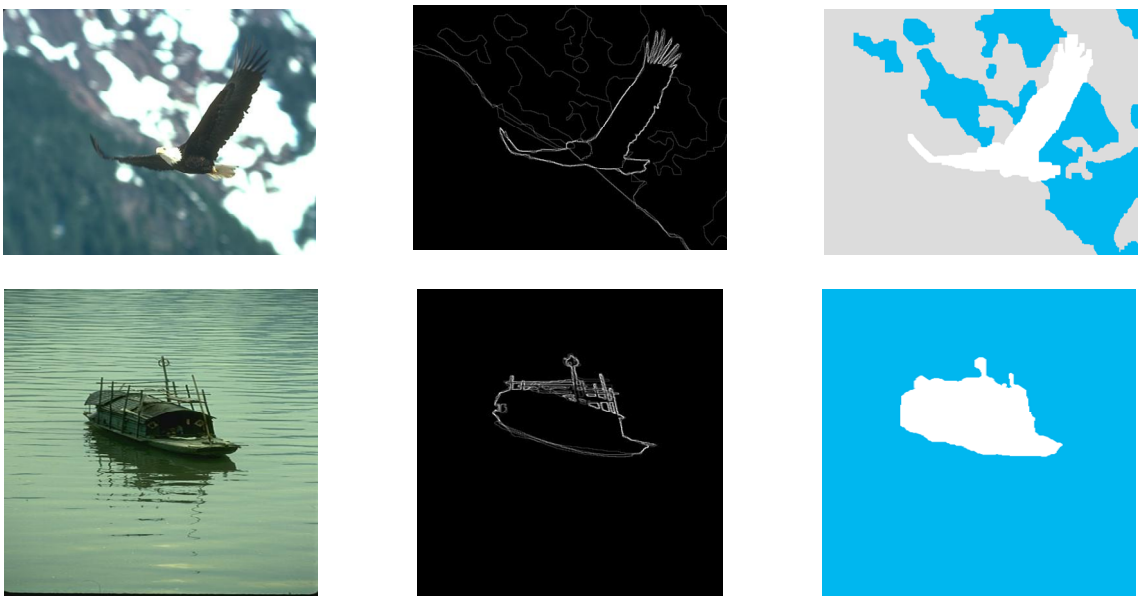


Figure 2.1: First column gives the image, second column gives the average segmented result from Berkeley image segmentation database; third column is the segmented result by first author taking the image in the second column as reference.

2.1.3 Results

Raw scores for each subject were converted to z-scores by (1) subtracting the average of all the ratings of object within the image; and then (2) dividing it with the standard deviation of all the ratings within the same image. This was done for all the subjects and then the final z-score has been obtained by taking the average of z-scores of all the subjects. The averaged z-scores were then rescaled from 0 to 1 for each image by dividing each z-score with the minimum and then dividing the result by the maximum of the z-scores within each image. Finally, the rescaled ratings were denoted by labels

"primary ROI," "secondary ROI," or "non-ROI" based on whether the score was greater than 2/3, between 1/3 and 2/3, or less than 1/3, respectively.

The graph below shows how well the subjects tended to agree with each other. In this graph, the x-axis denotes different importance levels quantized in the intervals of 0.2 and the height of each bar denotes the average rating for each range. The main point that could be taken away from this graph is the error bars, which denote the standard deviation across observers for each image.

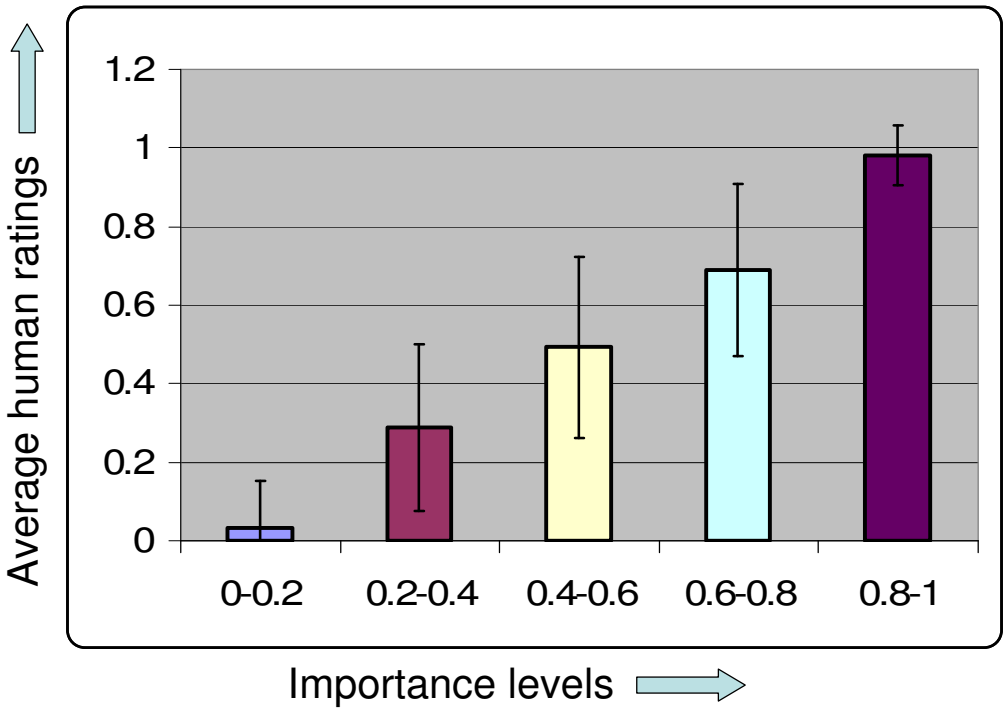


Figure 2.2: Subjective agreement of importance, where the x-axis gives the different importance levels and the y-axis denotes the average human ratings for each interval. The error bars denote the standard deviation.

What this data suggests is that with three levels we can actually see that subjects tend to agree with each other. We can notice that the error bars are least for the ranges *0 to 0.2* and *0.8 to 1.* Most discrepancy is seen in the intermediate levels. But you can also notice

that the error bars for ranges 0.2 to 0.8 tend to overlap.

After obtaining the rescaled ratings, various factors or attributes such as: location, contrast, luminance, color, edge-strength, a measure of the extent to which the object was in the foreground (discussed in Chapter 3), and blur have been measured. After this, histograms of various object attributes given that the object was rated as a primary, secondary or non-ROI have been measured.

The histograms were computed based on the results of the 200-image training set. The histograms were fitted using the distribution fitting tool from the statistics toolbox of MATLAB. The histograms were fit with either a generalized extreme value distribution or a Weibull distribution; or a Gamma distribution; the fits are shown in Figure 3.1. (1-6). These distributions were chosen as they provided a good fit to all three types (primary, secondary, and non-ROI) for each factor.

Some of the images from the Berkeley Image Segmentation Database are shown in Figure 2.3 which was used as the training set for the algorithm. Figure 2.4 gives the segmentation where each object is denoted by a different color. Figure 2.5 gives the snapshot of the excel sheet that has the ratings of different objects for each image shown in Figure 2.3. Finally, Figure 2.6 gives the ideal *interest maps* (in which brightness corresponds to perceived interest) which were generated from the ratings shown in Figure 2.5. These results demonstrate that subjects tended to rate objects having human faces or animals to be of greatest interest (primary ROI), whereas background objects such as sky and grass generally received the least interest (non-ROI).


img3.bmp


img4.bmp


img5.bmp


img7.bmp


img8.bmp


img9.bmp


img10.bmp


img11.bmp


img17.bmp


img18.bmp


img19.bmp


img21.bmp


img22.bmp


img23.bmp


img24.bmp


img25.bmp


img26.bmp


img27.bmp


img28.bmp


img29.bmp


img30.bmp


img32.bmp


img33.bmp


img35.bmp


img36.bmp


img37.bmp


img38.bmp


img39.bmp


img40.bmp


img41.bmp

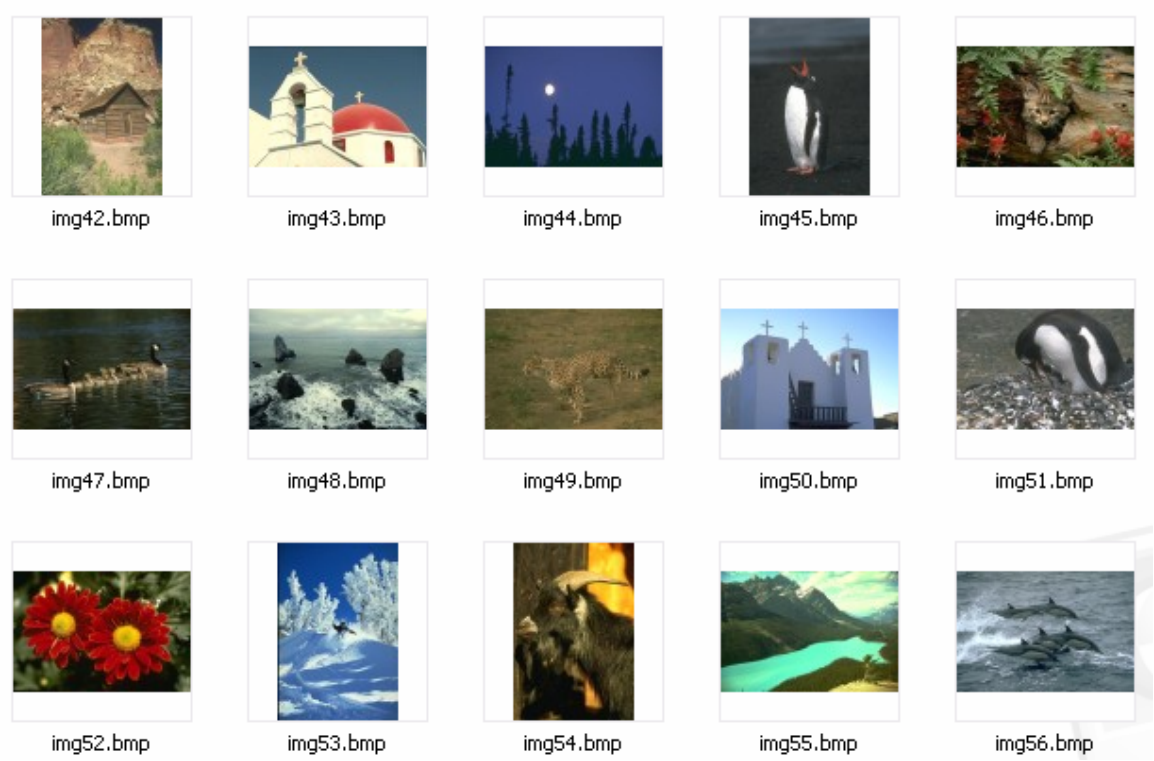
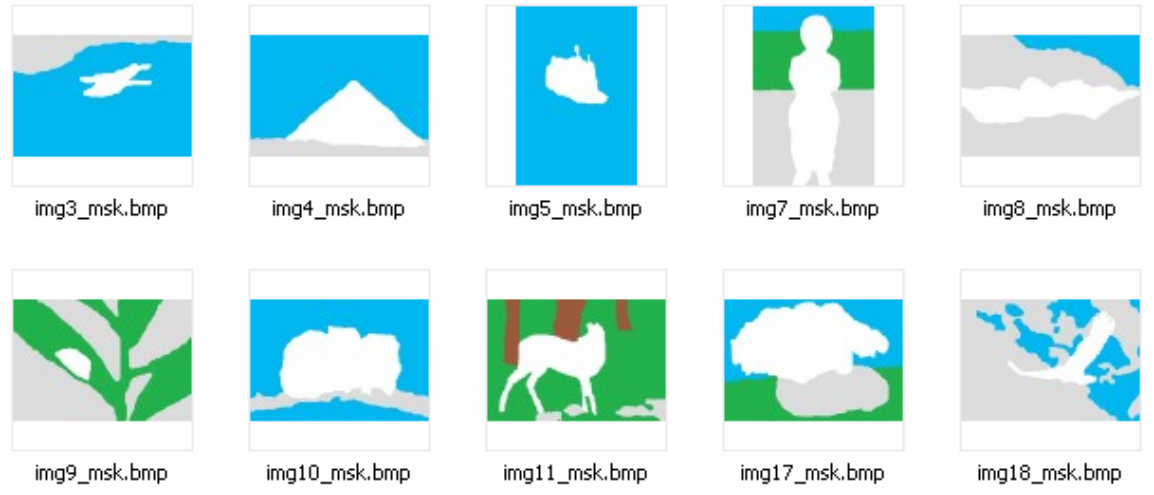


Figure 2.3: Snapshot of 50 images from the training set of 200 images of the Berkeley Segmentation dataset and Benchmark database used in Experiment.




img19_msk.bmp


img21_msk.bmp


img22_msk.bmp


img23_msk.bmp


img24_msk.bmp


img25_msk.bmp


img26_msk.bmp


img27_msk.bmp


img28_msk.bmp

img29_msk.bmp


img30_msk.bmp


img32_msk.bmp


img33_msk.bmp


img35_msk.bmp


img36_msk.bmp


img37_msk.bmp


img38_msk.bmp


img39_msk.bmp


img40_msk.bmp


img41_msk.bmp


img42_msk.bmp


img43_msk.bmp


img44_msk.bmp


img45_msk.bmp


img46_msk.bmp


img47_msk.bmp


img48_msk.bmp


img49_msk.bmp


img50_msk.bmp


img51_msk.bmp

Figure 2.4: Snapshot of the segmented images for the images shown in Figure 2.3. Each object is denoted with a different color.

no.of objects	image	object	ratings
3	img3	fish	1.00
	img3	ice	0.18
	img3	water	0.00
3	img4	pyramid	1.00
	img4	ground	0.00
	img4	sky	0.48
2	img5	boat	1.00
	img5	water	0.00
4	img7	boy	1.00
	img7	ground	0.31
	img7	wall	0.18
	img7	sky	0.00
3	img8	birds	1.00
	img8	rocks	0.56
	img8	sky	0.00
3	img9	insect	1.00
	img9	plants	0.74
	img9	back	0.00
3	img10	owls	1.00
	img10	bark	0.38
	img10	back	0.00
4	img11	deer	1.00
	img11	barks	0.27
	img11	rocks	0.00
	img11	ground	0.08

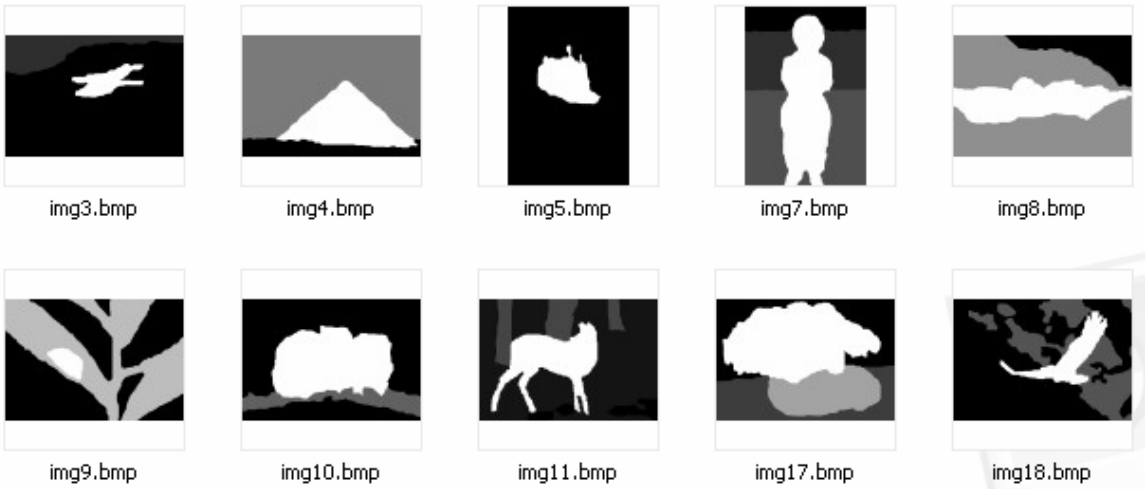
4	img17	plant	1.00
	img17	pot	0.64
	img17	ground	0.24
	img17	back	0.00
3	img18	bird	1.00
	img18	snow	0.33
	img18	back	0.00
3	img19	animal	1.00
	img19	rock	0.20
	img19	back	0.00
4	img21	animal	1.00
	img21	bark	0.15
	img21	ground	0.23
	img21	back	0.00
2	img22	building	1.00
	img22	back	0.00
2	img23	bears	1.00
	img23	back	0.00
3	img24	temple	1.00
	img24	flag	0.87
	img24	back	0.00
3	img25	horses	1.00
	img25	ground	0.30
	img25	back	0.00
3	img26	wagons	1.00
	img26	ground	0.00
	img26	sky	0.07
3	img27	woman	1.00
	img27	mesh at back	0.12
	img27	pillar	0.00

no.of objects	image	object	ratings
3	img33	man_white	1.00
	img33	man_black	0.82
	img33	back	0.00
3	img35	animal	1.00
	img35	stem	0.44
	img35	back	0.00
2	img36	owl	1.00
	img36	back	0.00
3	img37	church	1.00
	img37	back	0.18
	img37	sky	0.00
5	img38	man	0.92
	img38	iceberg	1.00
	img38	water	0.29
	img38	ship	0.63
	img38	sky	0.00
3	img39	butterfly	1.00
	img39	flowers	0.42
	img39	back	0.00
3	img40	woman	1.00
	img40	baskets	0.45
	img40	back	0.00
4	img41	animal	1.00
	img41	blue_back	0.20
	img41	grass at front	0.32
	img41	back	0.00
5	img42	hut	1.00
	img42	mountains	0.94
	img42	ground	0.25
	img42	bushes	0.48
	img42	sky	0.00
3	img43	church	0.93
	img43	dome	1.00
	img43	sky	0.00

3	img28	animal	1.00
	img28	rocks	0.70
	img28	back	0.00
2	img29	man	1.00
	img29	back	0.00
2	img30	buffalo	1.00
	img30	water	0.00
3	img32	people	1.00
	img32	moon at back	0.48
	img32	back	0.00

3	img54	animal	1.00
	img54	glowing	0.16
	img54	bark	0.00
4	img55	water	1.00
	img55	mountains	0.70
	img55	plant at front	0.00
	img55	sky	0.14
2	img56	dolphins	1.00
	img56	water	0.00

Figure 2.5: Average normalized ratings for each object for all the images shown in Figure 2.3 and Figure 2.4.




img19.bmp
img21.bmp
img22.bmp
img23.bmp
img24.bmp
img25.bmp
img26.bmp
img27.bmp
img28.bmp
img29.bmp
img30.bmp
img32.bmp
img33.bmp
img35.bmp
img36.bmp
img37.bmp
img38.bmp
img39.bmp
img40.bmp
img41.bmp
img42.bmp
img43.bmp
img44.bmp
img45.bmp
img46.bmp
img47.bmp
img48.bmp
img49.bmp
img50.bmp
img51.bmp

Figure 2.6: Human rated importance maps for all the images shown in Figure 2.3 and Figure 2.4.

CHAPTER 3

ALGORITHM

The algorithm section is broadly divided into two-sections. The first section gives the Bayesian probabilistic approach which needs segmentation and uses naïve bayes' rule of independent assumption. The second section describes the block-based segmentation approach which doesn't need segmentation and is fast-enough to be used in real-time applications.

3.1 Bayesian Probabilistic Approach

In this section, we present an algorithm which takes as input various measurable factors (attributes) of each object in an image, and then yields the relative perceived importance of each object. The algorithm operates via a Bayesian probabilistic approach using the psychophysical histograms described in the previous section.

The input to the algorithm is a segmented image that is divided into different objects. For example, the first image is an input image that is segmented into different regions denoted by different colors. The next step is to measure attributes like location, contrast, color, luminance, edge-strength, foreground/background and blur. After measuring the attributes, the raw measures are translated to likelihood importance's using the histograms generated on the training set. The choice of the object being primary, secondary or non-ROI is taken by choosing the interest which gives the maximum probability.

General flow of algorithm is given below:

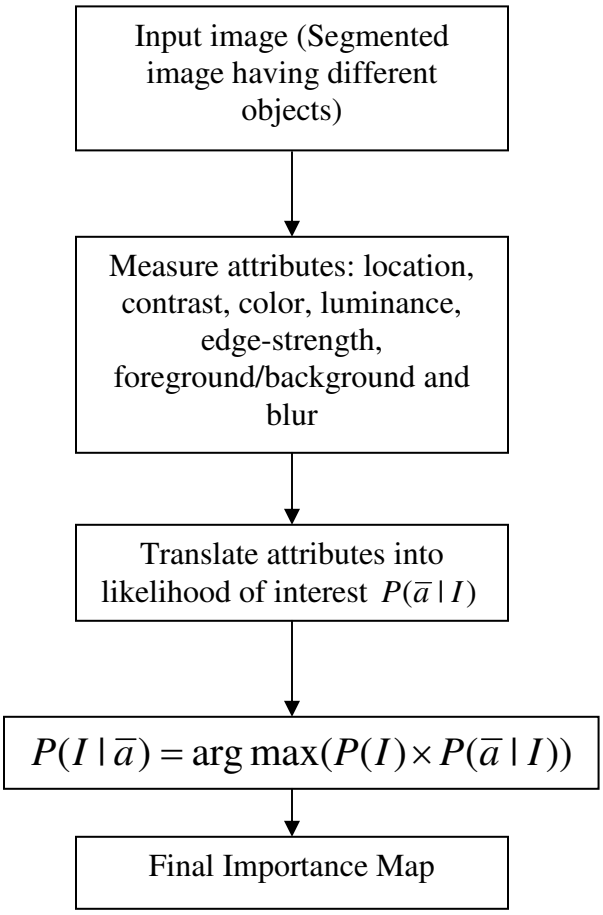


Figure 3.1: General flow of the algorithm

The following sections explain each step of the algorithm in detail.

3.1.1. Measuring Various Attributes

The various attributes involved in computing the interest maps are color, contrast, location, edge-strength, foreground/background and blur. For all 1100+ objects, each factor has been measured as follows:

3.1.1.1. Location: Usually an object toward the center of an image is more important than distant objects. The location for each object was computed as follows:

$$a_{location} = \frac{\sqrt{\left(x_c - \frac{width}{2}\right)^2 + \left(y_c - \frac{height}{2}\right)^2}}{\left(\frac{width}{2}\right)^2 + \left(\frac{height}{2}\right)^2} \times 100$$

where x_c , y_c denote the horizontal and vertical pixel coordinates of the object's centroid, respectively; and *width*, *height* denote the width and height of the full-sized image respectively.

3.1.1.2. Contrast: It is often observed that an object tends to stand out whenever it is of high luminance contrast; such an object is generally rated to be of greater interest than other, low-contrast objects. The object's contrast is measured by (1) dividing the object into $B \times B$ blocks, (2) measuring the RMS contrast of each block, and then (3) combining the per-block contrasts as follows:

$$a_{contrast} = \frac{50}{M} \sqrt{\sum_{m=1}^{M} {c_m}^2}$$

where M denotes the total number of blocks in the object, and the block size, B, is computed based on the object's size via $B = \max[4, (0.05\sqrt{N_{object}} + 0.05)]$. The quantity c_m denotes the RMS contrast of the m^{th} block, which is given by $c_m = \sigma_m/\mu_m$, where σ_m and μ_m denote the standard deviation and mean of the block luminance's, respectively. We have found this local (block-based) measure of contrast to provide a better prediction of perceived contrast than RMS contrast for natural images.

3.1.1.3. Color: It is quite obvious that colorful and bright objects draw attention. Therefore, the color distance for each object has been measured by: (1) Creating a dilated mask for each object, and then (2) measuring the Euclidean distance between the average

object color/brightness(*avg_clr/lum_object*)and the average color/brightness of the neighboring pixels (avg_clr/lum_neighbor) defined by the dilation. The distance has been computed separately for brightness ($a_{brightness}$) and color (a_{color}) in CIELAB color space, where *avg_clr_object= [a1, b1], avg_clr_neighbor= [a2 b2], avg_lum_object= [L1] and avg_clr_neighbor= [L2].*

$$a_{color} = \| avg\_clr\_object - avg\_clr\_neighbor \|$$

$$a_{brightnes} = \| avg\_lum\_object - avg\_lum\_neighbor \|$$

3.1.1.4. Edge Strength: Usually, objects with greater numbers of edges are more obvious. We obtained a measure of edge-strength (a_{edge}) for each object by: (1) applying a canny edge detector to the image, and then (2) counting the number of edge pixels for each object. The edge-strength was defined as the number of edges (N_{edges}) for each object divided by the number of pixels in each object.

$$a_{edge} = N_{edges} / N_{object}$$

3.1.1.5. Foreground/Background: To measure the extent to which an object is in the foreground, we have employed a variant of the measure specified by Osberger *et al.* [1] given by

$$a_{foreground} = 2 - \frac{2N_{object\_border}}{N_{image\_border}}$$

where N_{object_border} denotes the number of object pixels which lie within three pixels of the image's outer edges (borders), and where N_{image_border} denotes the total number of pixels in the three-pixel-wide border of the image.

3.1.1.6. Blur: Blur is another factor that has a significant influence on the importance of the object in an image. Right now, blur has been measured as a preliminary result using slope of the magnitude spectrum.

So, to measure blur, the image is first divided into 16×16 blocks with 50% overlap between the blocks. Then the slope of the magnitude spectrum is computed for each block to obtain a mask. Another mask is obtained by passing the original image through a high pass filter and then by measuring the kurtosis for each block. Kurtosis is chosen because it gives the measure of sparseness. The next step is compute the straight Euclidean distance between the two maps, and the then a median filter is applied on the resultant image where the darkness indicates the blocks that have greatest blur.

3.1.2. Histograms: Likelihood functions

In order to translate the attributes to the perceived interest, we need the histograms to compute the likelihood interests. The histograms were generated using the training set data. The histograms were fitted using the distribution fitting tool from the statistics toolbox of MATLAB. The histograms were fit with either a generalized extreme value distribution or a Weibull distribution; or a Gamma distribution; the fits are shown in Figure. 3.1. (1-6). These distributions were chosen because they provide a good fit to all three types (primary, secondary, and non-ROI) for each factor.

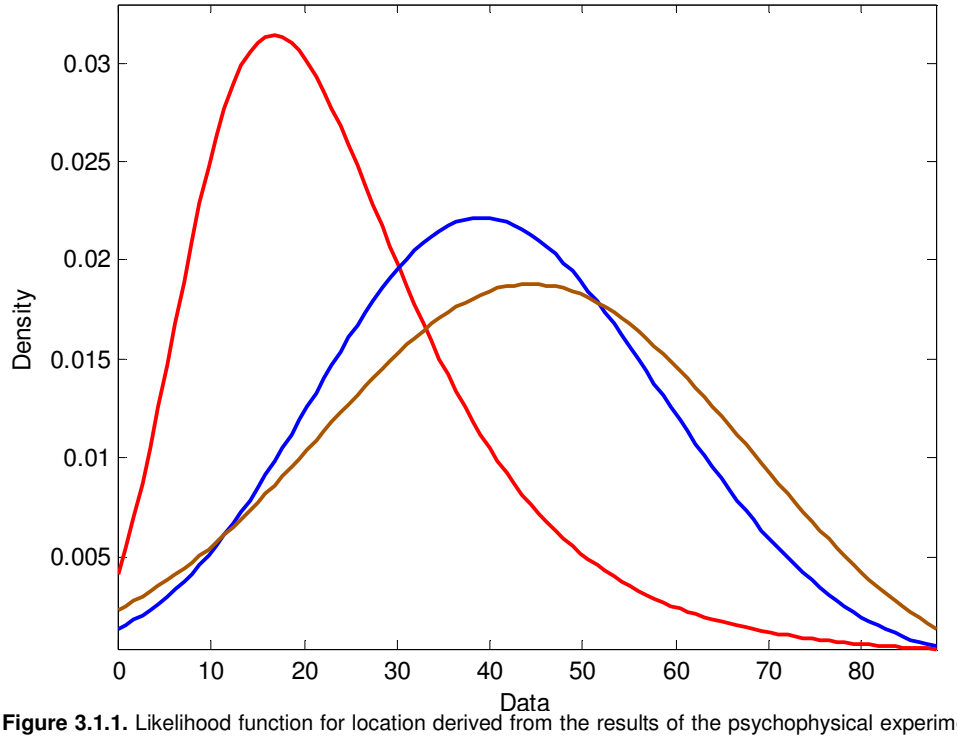


Figure 3.1.1. Likelihood function for location derived from the results of the psychophysical experiment. Red, blue and brown color curves give the histograms for location given that the object is primary, secondary or non-ROI, respectively

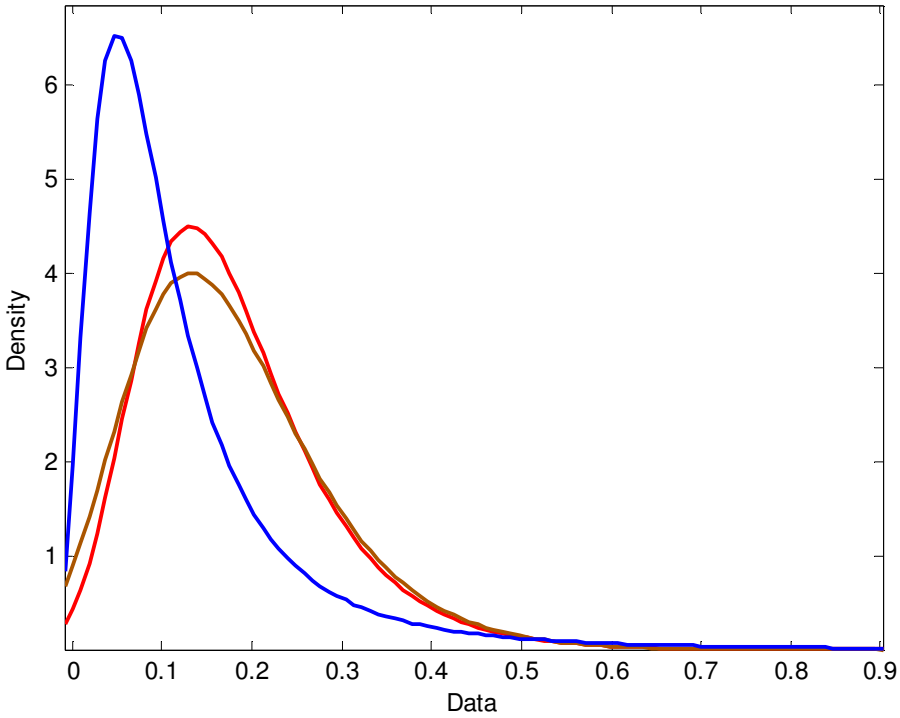


Figure 3.1.2. Likelihood function for contrast derived from the results of the psychophysical experiment. Red, blue and brown color curves give the histograms for contrast given that the object is primary, secondary or non-ROI, respectively

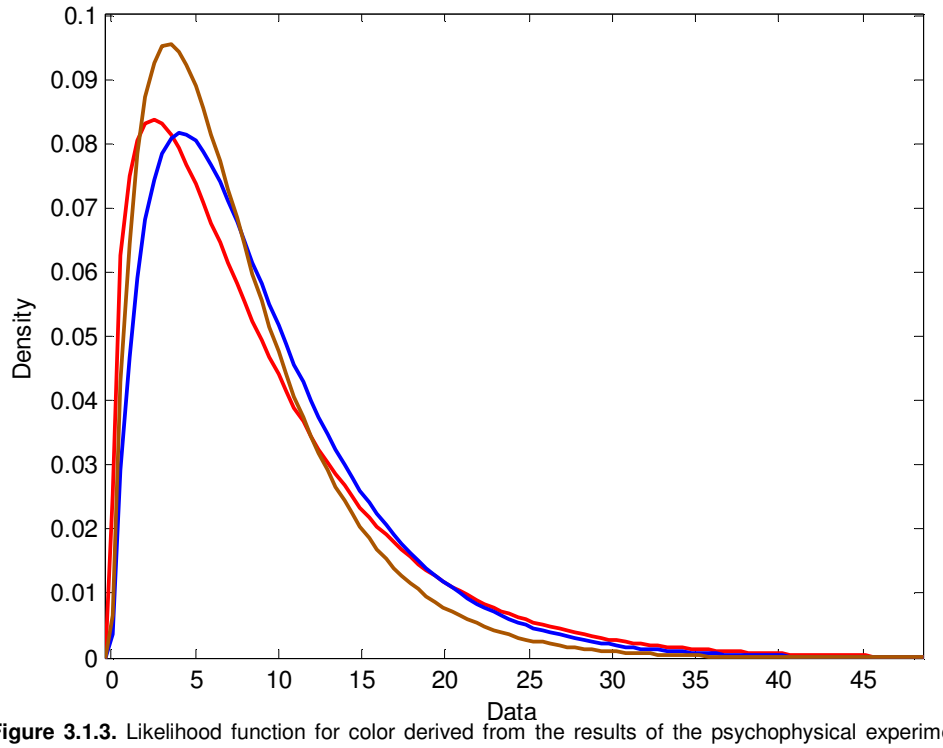


Figure 3.1.3. Likelihood function for color derived from the results of the psychophysical experiment. Red, blue and brown color curves give the histograms for color given that the object is primary, secondary or non-ROI, respectively

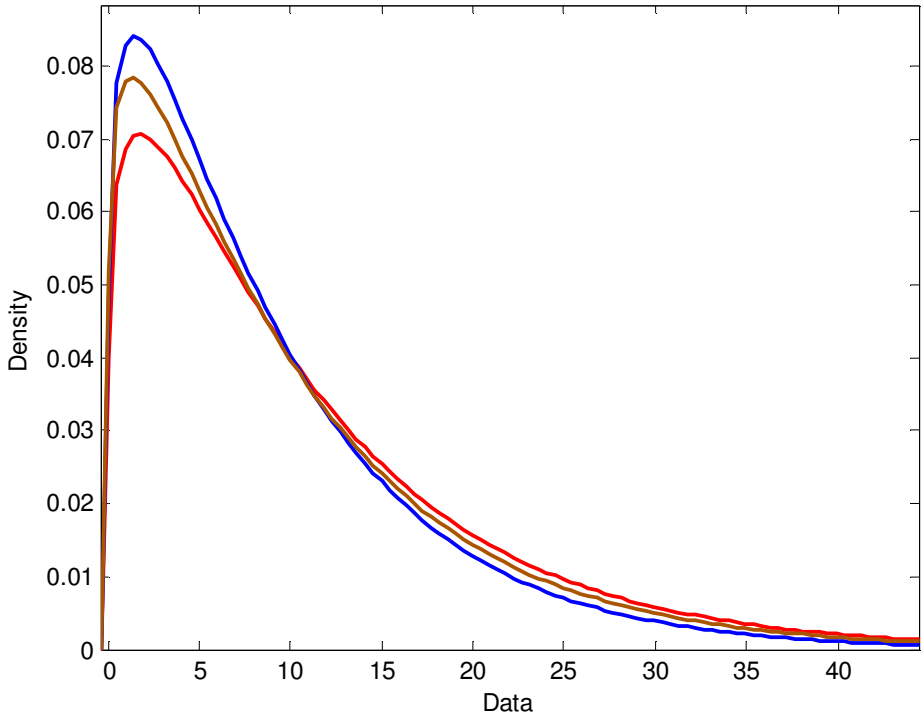


Figure 3.1.4. Likelihood function for edge-strength derived from the results of the psychophysical experiment. Red, blue and brown color curves give the histograms for luminance given that the object is primary, secondary or non-ROI, respectively

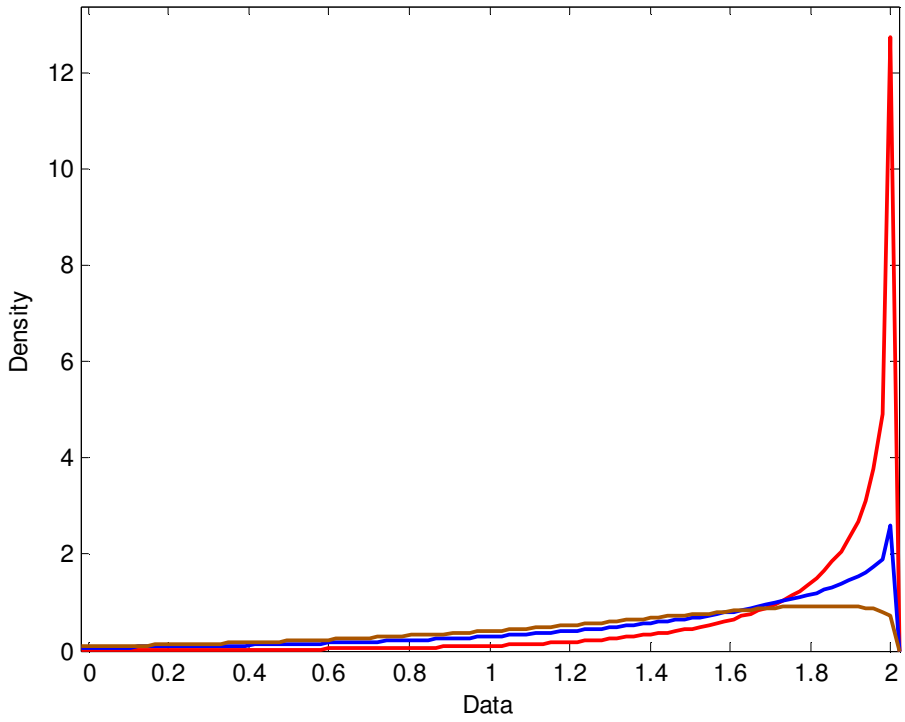


Figure 3.1.5. Likelihood function for foreground/backgorund derived from the results of the psychophysical experiment. Red, blue and brown color curves give the histograms for f/b given that the object is primary, secondary or non-ROI, respectively

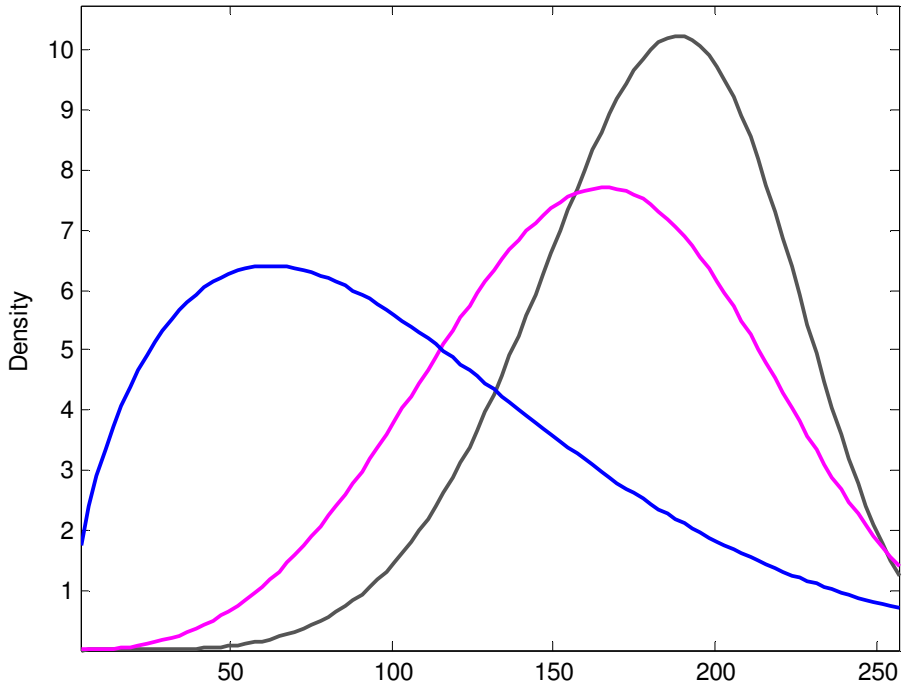


Figure 3.1.6. Likelihood function for blur derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for blur given that the object is primary, secondary or non-ROI, respectively

3.1.3. Using the Factors with Bayes' Rule

Bayes' theorem relates the conditional and marginal probability distributions of random variables. Here, we are interested in the probability of perceived interest given the attributes of location, color, contrast, edge-strength, foreground/background, and blur.

Let $\bar{a} = [a_{location}, a_{contrast}, a_{brightness}, a_{color}, a_{edge}, a_{foreground}, a_{blur}]$ denote a vector of measured attributes, and I be the perceived interest. The probability of interest given the attributes is given by

$$P(I \mid \bar{a}) = \arg_I \max(P(I) \times P(\bar{a} \mid I))$$

where $P(I)$ was assumed to be 1/3.

$P(I \mid \bar{a})$ returns the probability of interest that can take three labels: Primary, Secondary and non-ROI. The Bayesian probabilistic approach used here is based on naïve bayes. Hence we assumed statistical independence between the attributes and thus, the likelihood term can be expressed as

$$\begin{aligned} P(\bar{a} \mid I) \quad = \quad & P(a_{location} \mid I) \times P(a_{contrast} \mid I) \times \\ & P(a_{brightness} \mid I) \times P(a_{color} \mid I) \times P(a_{edge} \mid I) \times \\ & P(a_{foreground} \mid I) \times P(a_{blur} \mid I) \end{aligned}$$

where the individual probabilities were measured from the histograms generated based on the psychophysical experimental results.

The algorithm thus estimates each object to be a primary, secondary, or non-ROI by choosing the rating which yields the greatest probability given the measured attributes.

3.1.4 Algorithm Summary

In summary, given a segmented image, the algorithm performs the following steps:

i. Computes each object's factors $a_{location}$, $a_{contrast}$, $a_{luminance}$, $a_{edge\text{-}strength}$, a_{color}, $a_{foreground}$, a_{blur}.

ii. Converts the factors into perceived interest from the histograms obtained for each factor.

iii. Computes each object's total perceived interest by plugging everything into the Bayesian framework.

iv. From this Bayesian framework, the algorithm computes the probability of each object being primary, secondary or non-ROI.

v. The final outcome is the estimation that depends on the largest probability. For example, if the algorithm has highest probability with primary ROI, then the object is estimated to be a primary ROI.

3.2. Block-Based Approach

This section describes an alternative method to compute the importance of objects in the images. This approach is similar to the Bayesian probabilistic approach described in the previous section. However, the basic difference between these two approaches is the need of segmentation. Block-based approach does not need any segmentation algorithm and is fast enough to be used in real-time applications.

First of all, the image is divided into $n \times n$ blocks (where n=8, 16 or 32). After dividing the image into blocks, all the attributes are measured for each block in the image. These attributes are then converted to the perceived importances using the histograms generated

by using the training set. The basic flow chart of the block-based approach is shown below:

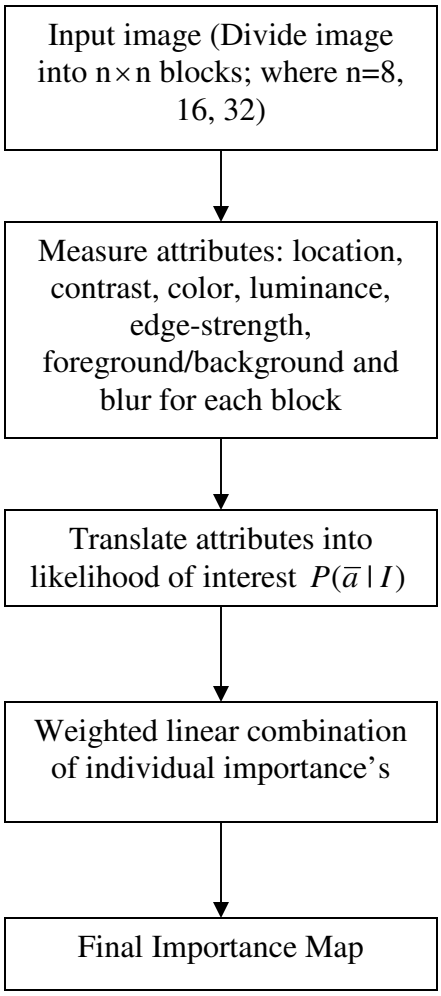


Figure 3.2. General flow-chart of the block-based approach.

Even the training set is divided into blocks and the histograms of each attribute are computed given the importance is primary, secondary or non-ROI. The histograms are fitted using the distribution fitting tool box either with weibull or generalized extreme value distribution. The histograms generated for the 200 training set are as below:

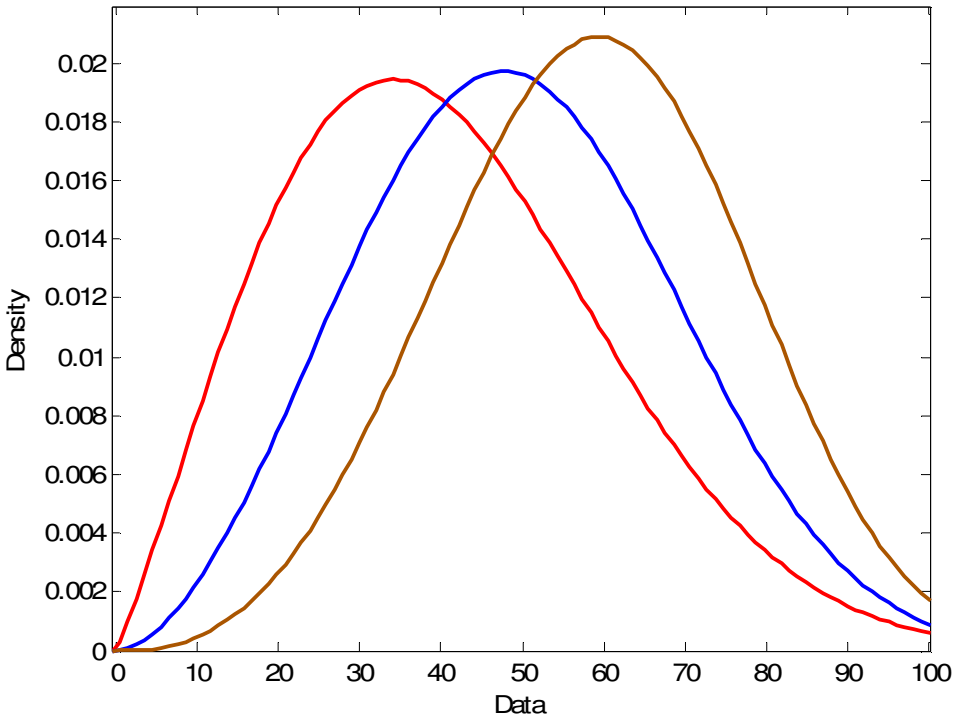


Figure 3.2.1 . Likelihood function for location derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for location given that the object is primary, secondary or non-ROI, respectively

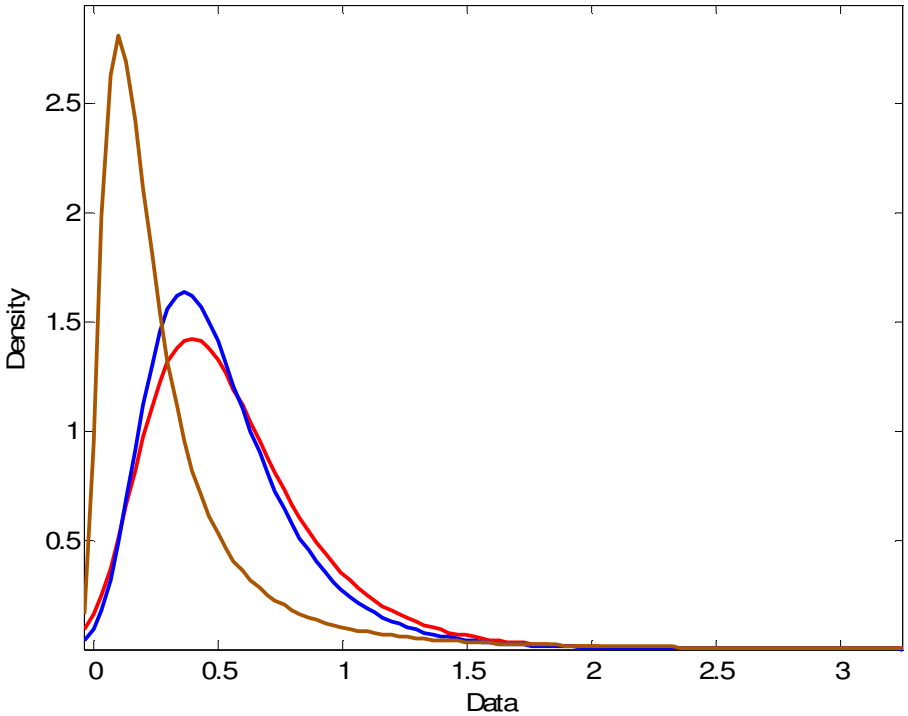


Figure 3.2.2 . Likelihood function for contrast derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for contrast given that the object is primary, secondary or non-ROI, respectively

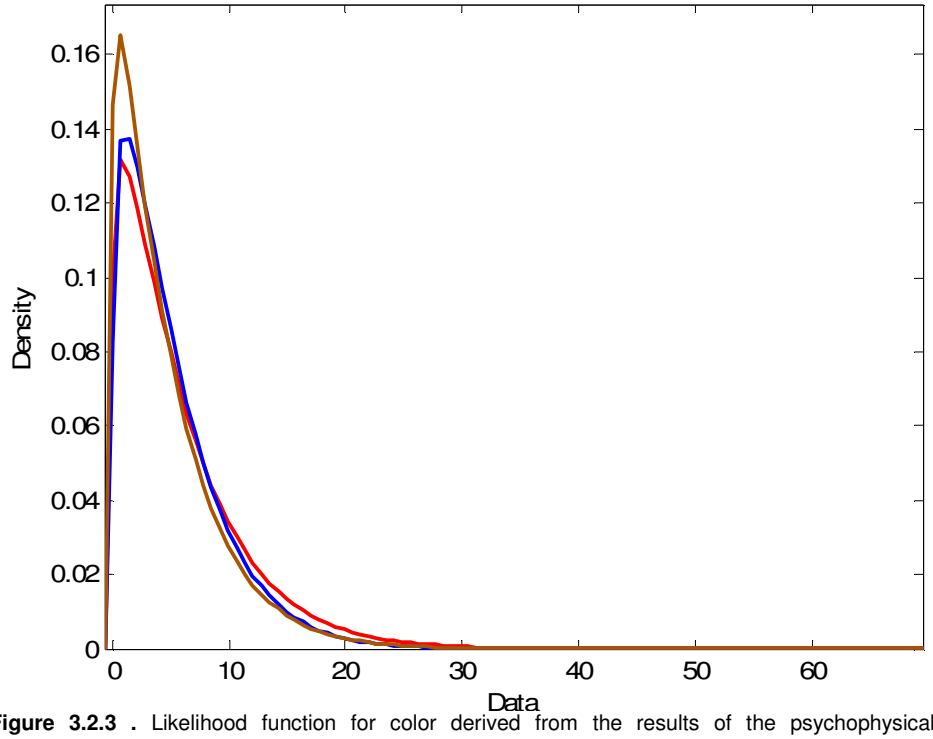


Figure 3.2.3 . Likelihood function for color derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for color given that the object is primary, secondary or non-ROI, respectively

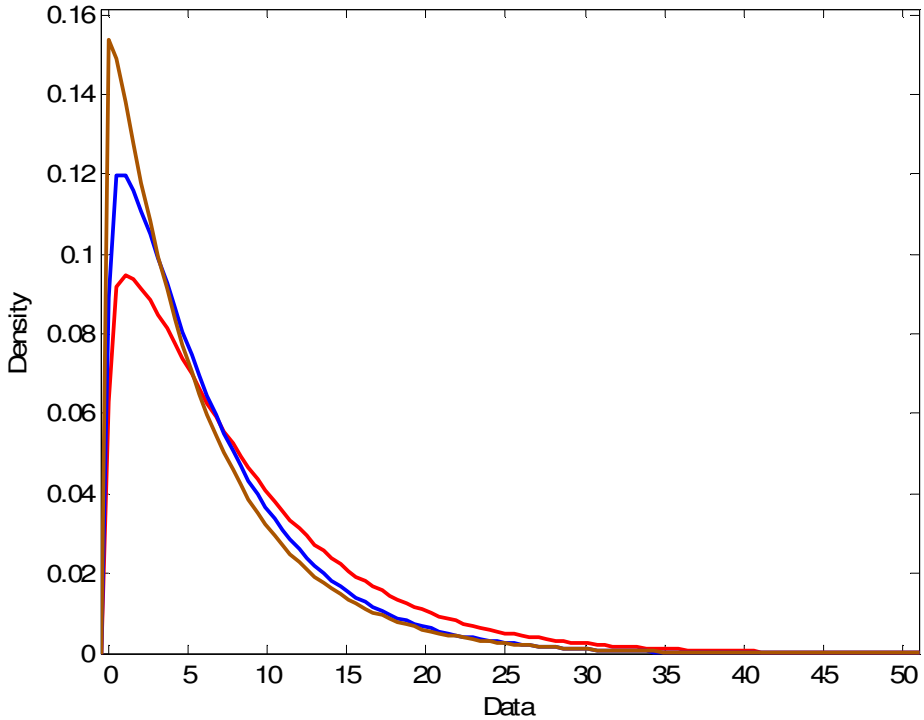


Figure 3.2.4 . Likelihood function for luminance derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for luminance given that the object is primary, secondary or non-ROI, respectively

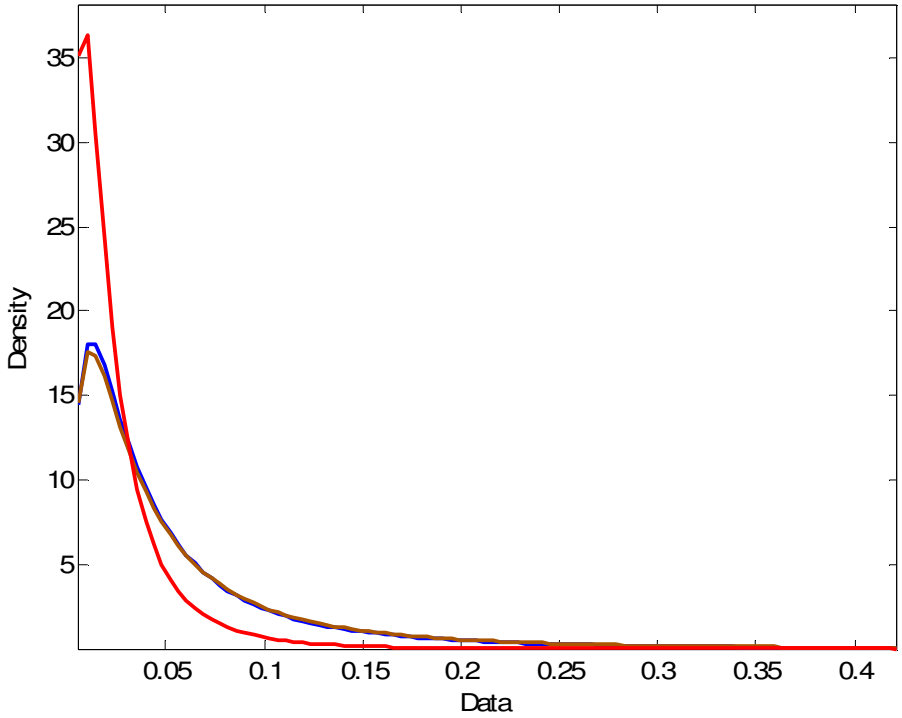


Figure 3.2.5 . Likelihood function for edge-strength derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for edge-strength given that the object is primary, secondary or non-ROI, respectively

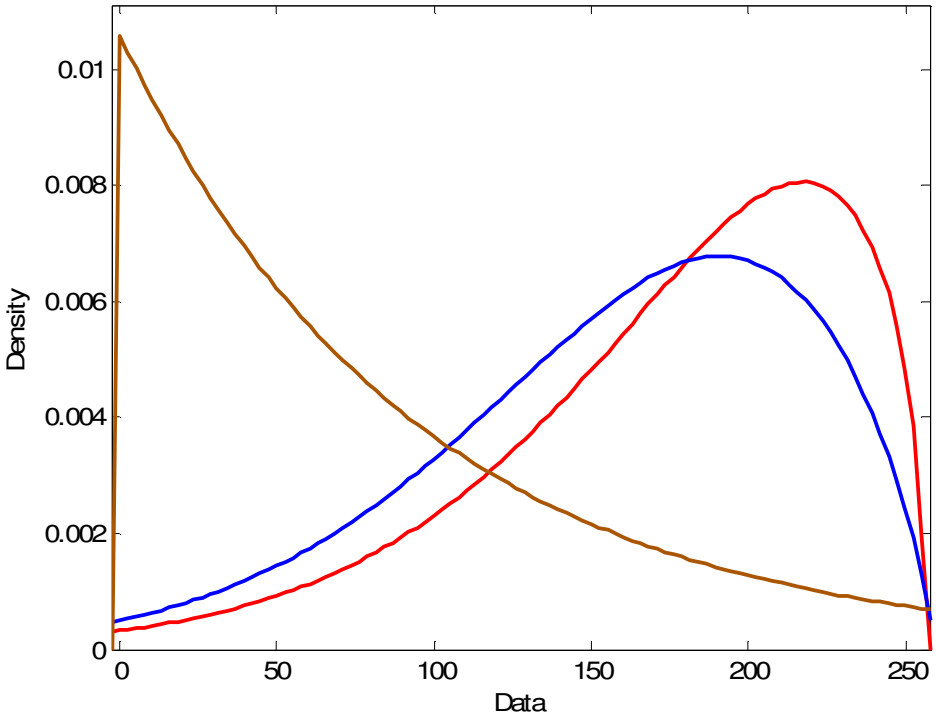


Figure 3.2.6 . Likelihood function for blur derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for blur given that the object is primary, secondary or non-ROI, respectively

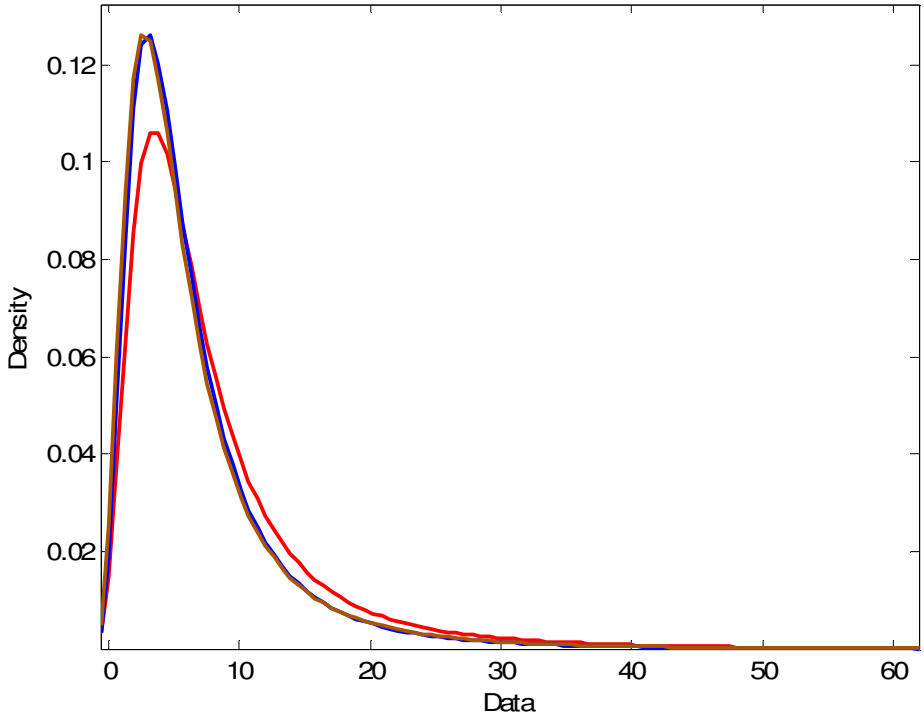


Figure 3.2.7 . Likelihood function for global color derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for global color given that the object is primary, secondary or non-ROI, respectively

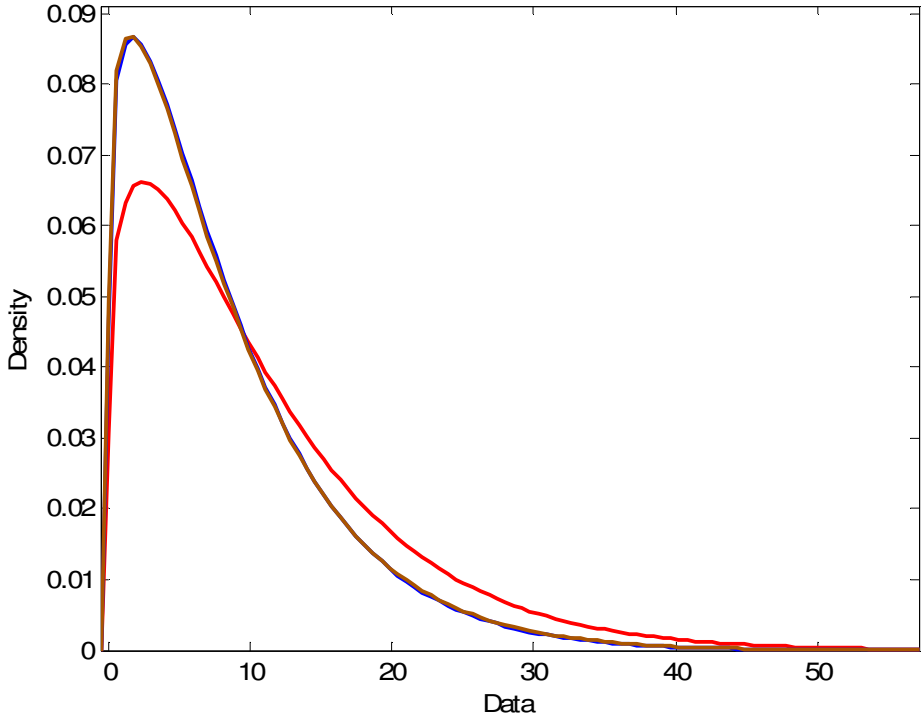


Figure 3.2.8 . Likelihood function for global luminance derived from the results of the psychophysical experiment. Pink, blue and brown color curves give the histograms for global luminance given that the object is primary, secondary or non-ROI, respectively

In Figure 3.3 the maps with raw measured factors are shown. The first image gives the raw map for location where the brightness denotes the distance from the center. Similarly for contrast, the brightness denotes greater contrast. The second row gives the maps for edge-strength and blur where the brightness for each-strength denotes more edge-strength, but the brightness in blur denotes that the block is in focus. Similarly, the brightness in the map for color distance denotes more color distance.

Again, these are just the measured factors, but these maps give a visual way of finding whether they affect the perceived importance.

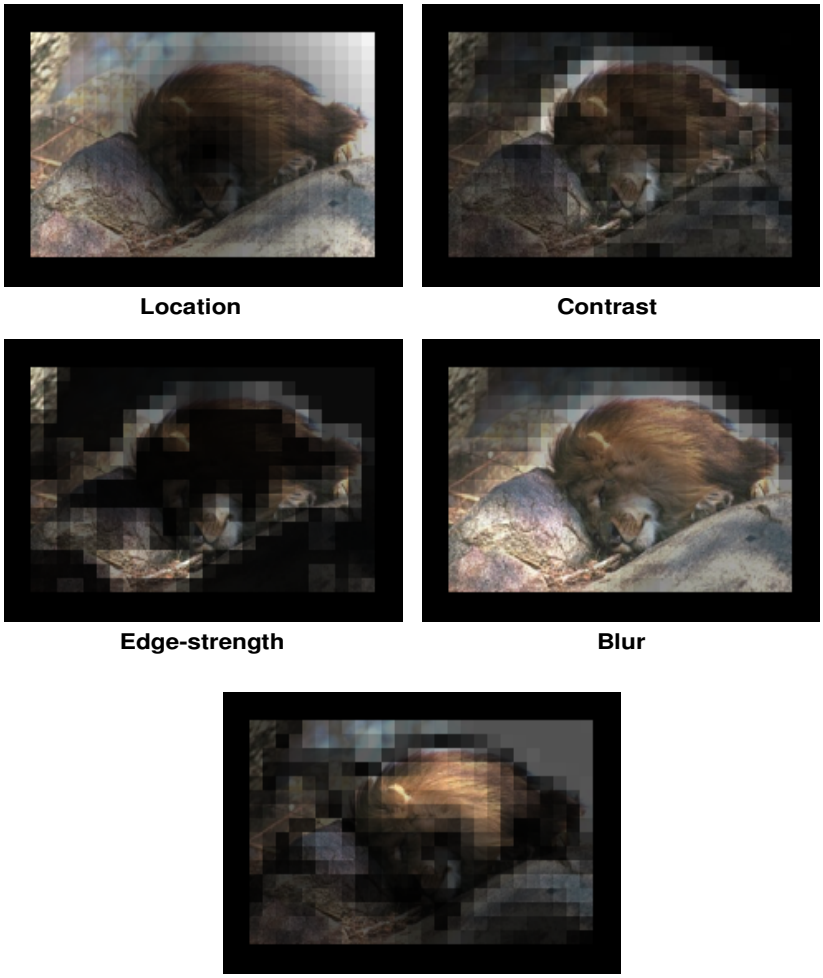


Figure 3.3. Raw maps for measured factors. First row gives the maps for location and contrast. Second row gives the maps for edge-strength and blur. The bottom row gives the map for color-distance.

The next step is to take these raw factors and translate them to importances and the way we did this is by using the histograms by considering one factor at a time. The first image in Figure 3.4 gives the importance map we would get just from location, where clear regions correspond to primary ROI, grayish regions correspond to secondary ROI and black regions denote the non-ROI. Similarly, the image on the right gives the importance map obtained by considering contrast alone. The second row gives the importance maps generated considering edge-strength and blur. And finally the bottom row gives the importance map for contrast.

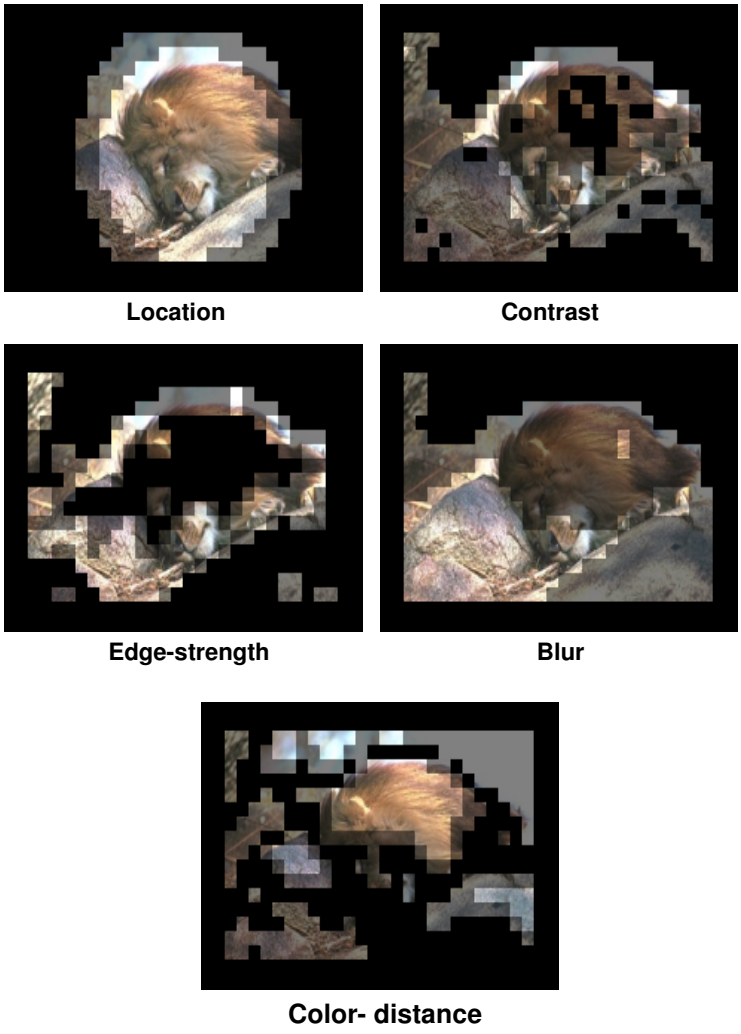


Figure 3.4. Raw maps translated to importance map for all the factors. First row gives the maps for location and contrast. Second row gives the maps for edge-strength and blur. The bottom row gives the map for color-distance.

Though both block-based and Bayesian based approaches are very similar, there are certain discrepancies.

The major difference lies in the way the histograms are generated. The histograms from the 200 training set are generated by dividing all the 200 images into blocks and then measuring the various factors. The histograms are then classified as primary, secondary and non-ROI depending on the average rating of each block. That is, each block's rating is obtained by taking the average of all the ratings (obtained from psychophysical experiment).

Some additional factors like global color distance and global luminance distance are measured for the block-based approach. A slight modification is done to the way these factors are computed. Instead of taking the color and luminance distance of each region with the corresponding neighbors, the distance is computed between each block's *color/luminance* and *the average color/ luminance* of the whole image.

$$a_{color} = \| avg\_clr\_object - avg\_clr\_image \|$$

$$a_{luminance} = \| avg\_lum\_object - avg\_lum\_image \|$$

The results with the block-based approach are shown in the results section. We observed that the block-based approach performed slightly better than the Bayesian based approach. This was due to the fact that the histograms measured for the block-based approach are dense and hence give better information. Moreover, certain factors like global color distance and global luminance distance make more sense when measured locally.

The weights are obtained using the training set of 200 images. A regression is performed

to obtain the optimum weights. The regression is performed between the algorithm's output and the results from the psychophysical experiment. The optimum weights are chosen by obtaining the maximum correlation between the algorithm's output and the ground-truth data. The weights thus obtained are given as follows:

Blur	Location	Contrast	Edge-strength	Global color distance
0.275	0.203	0.199	0.118	0.203

Table 3.1: Weights obtained from the training set for each factor.

After obtaining the weights, a weighted linear combination is performed to obtain at the final saliency map shown as follows:

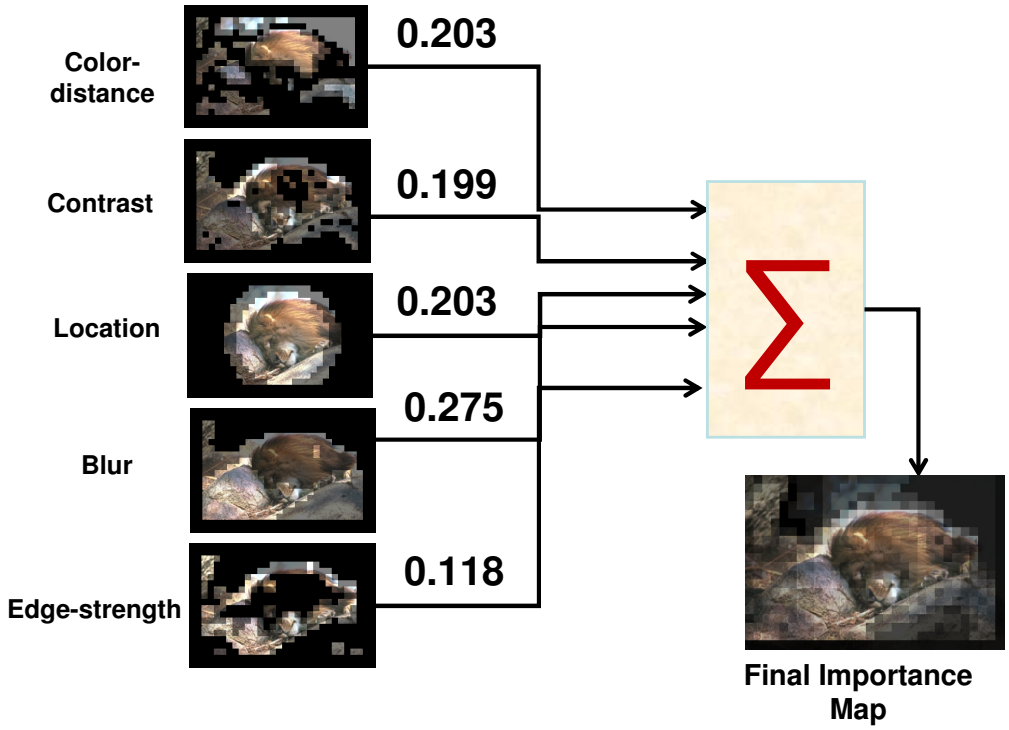


Figure 3.5: Weighted linear combination is performed to arrive at the final importance map.

CHAPTER 4

RESULTS

In this section we present the results by comparing our algorithm with the ground-truth data and also with Osberger *et al* and Itti *et al* approaches. The testing set consists of 100 images taken from the testing set of Berkeley image segmentation database as already mentioned. The images used are of size 321×481 with 24-bit RGB pixel values.

To test the efficiency of the algorithms, a comparison is made with the ground-truth data obtained from the psychophysical experiment. Since, the Bayesian probabilistic approach and Osberger *et al* approach is based on segmentation; they can easily be compared with the ground truth data. However, for the algorithms such as Itti *et al* and block-based approach, the output isn't an object level segmented importance map. Hence, to compare these approaches with the ground truth data, the work of Liu *et al* [22] has been followed where the average of the saliency points has been taken within each object of an image. That is for each object in an image, the importance is mapped back by taking the average of all the detected saliency points within that object. This could be shown in the form of demo in Figure 4.1. First of all consider the importance map generated using the human rating. Then take all the pixels corresponding to a single object for example the lions face. Then, find the same number of pixels in the Itti's or the proposed block-based map. After finding the pixels, simply compute the average of all the pixels in that region from the raw map to arrive at the final importance map. This is how the result of Itti *et al.* and the proposed block-based approach are mapped back to the segmented result.

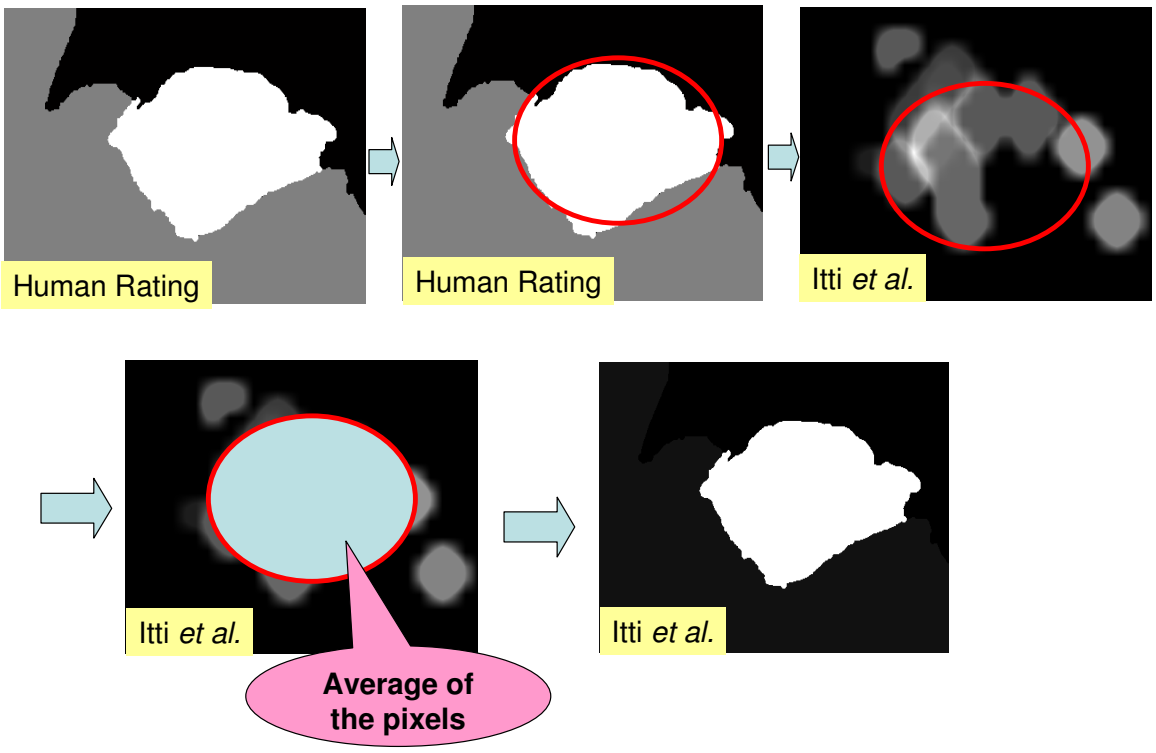


Figure 4.1: Demo showing the mapping back Itti *et al.* to the human-segmented mask.

As part of the results, we show some demonstrative images in Figure 4.2. Each image has five individual importance maps: (1) Human rated importance map (IM); (2) Importance map obtained by Osberger *et al*; (3) Importance map generated via Itti *et* al; (4) Importance map generated by proposed Bayesian probabilisitic approach and (5) Importance map generated by the block-based approach. The first row for each image is given in this order: First and foremost image is the original image. The second image is the human rated importance map which forms the ground truth data. The third image in the same row is the importance map generated using Osberger *et al* approach. The second row gives the importance map in the following order: First image is the importance map obtained using Itti *et al* which is obtained by taking the saliency map and mapping back to the segmented mask. (As explained in Figure 4.1). Second image in the second row

gives the importance map generated via the proposed Bayesian probabilistic approach and finally the third image in the second row shows the importance map generated via the block-based approach which is mapped back to the segmented image similar to the Itti *et al.*

Seven test images are presented showing the importance maps generated by all the algorithms. In all the images presented here, it is observed that almost all the algorithms predict the primary ROI perfectly. The problem arises while predicting the secondary and non-ROI in the image.

For example in the Figure 4.2.1, one can see that all the algorithms are successful in identifying the lion as the primary ROI. However, when it comes to the rocks in the foreground, which forms the next important object according to the human rating, Osberger's approach gives the sky more importance than the rocks in the foreground. Even Itti's approach gives the rocks and the sky almost the same importance. Whereas the proposed approaches do a better job finding rocks more important than the sky in the background.

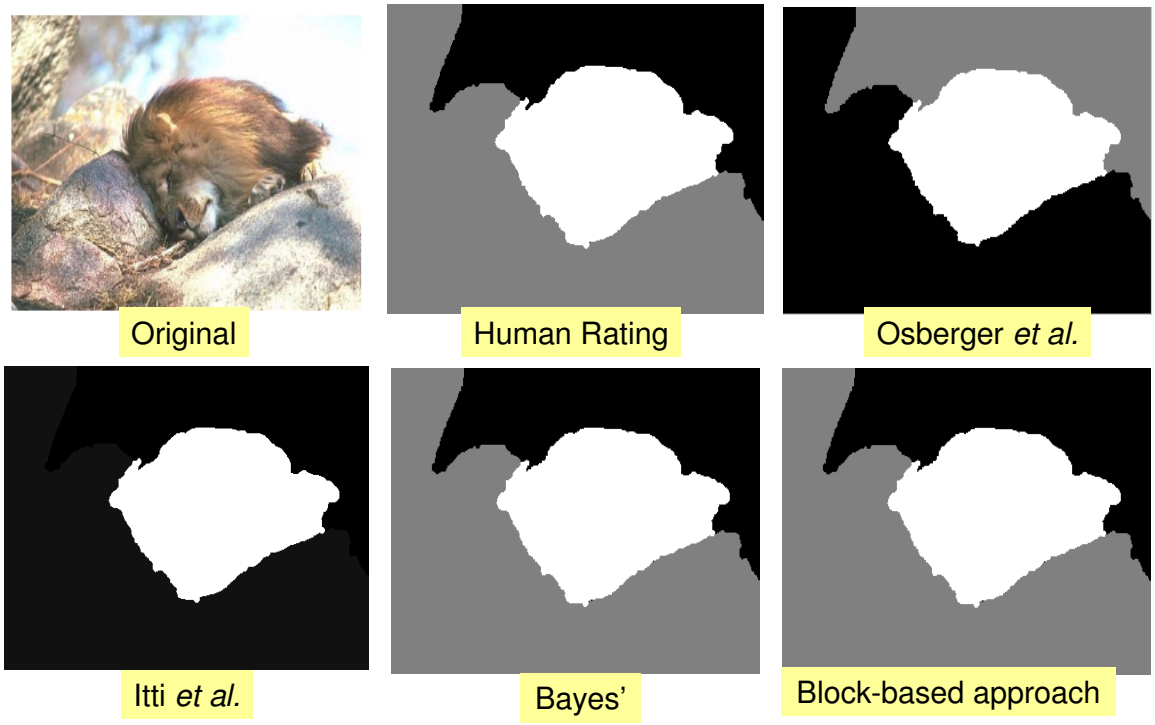


Figure 4.2.1: Results showing the importance maps generated using all the algorithms for a lion's image.

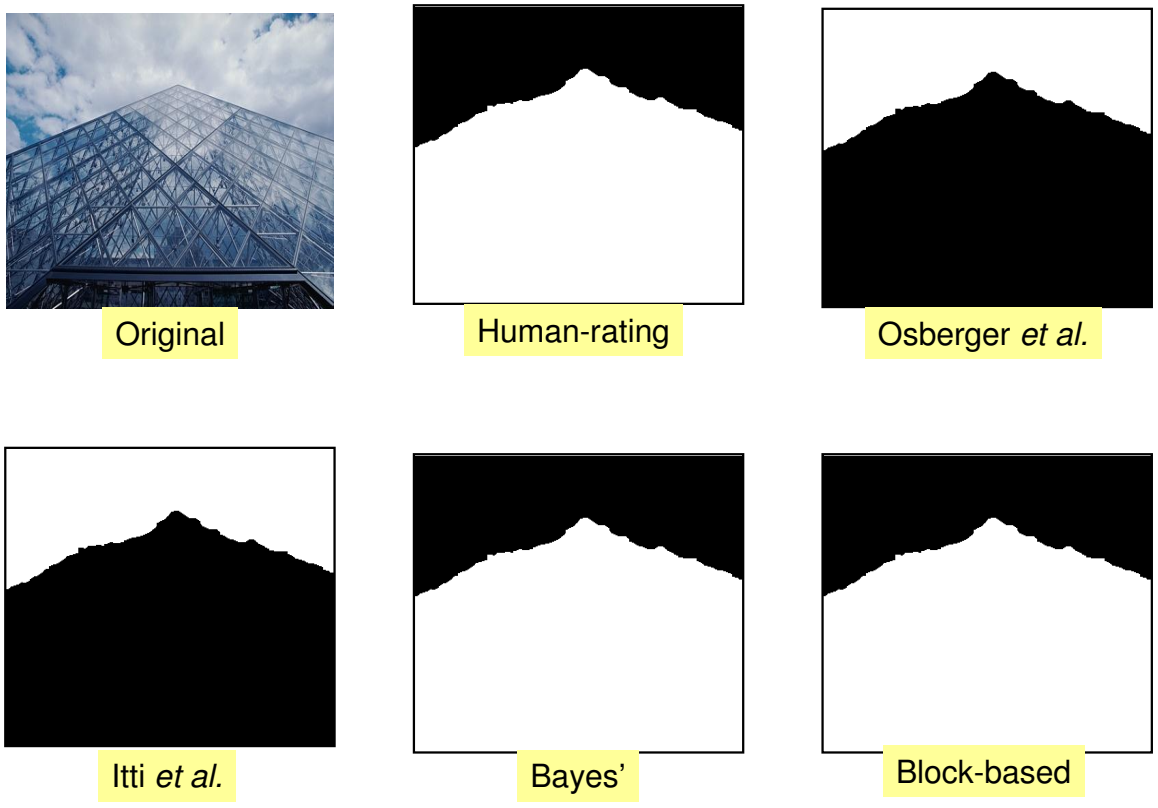


Figure 4.2.2: Results showing the importance maps generated using all the algorithms for an image of a building.

Another simple image is shown in Figure 4.2.2., where the proposed algorithm mimics the human-ratings, but both Itti and Osberger fail completely in recognizing the building as the important region.

Figure 4.2.3 gives an example of a difficult image where Osberger and the proposed approaches do a better job finding the animal as the most important region whereas Itti fails to do that. It is also observed that both Itti and Osberger give more importance to the background which is completely out of focus in the original image.

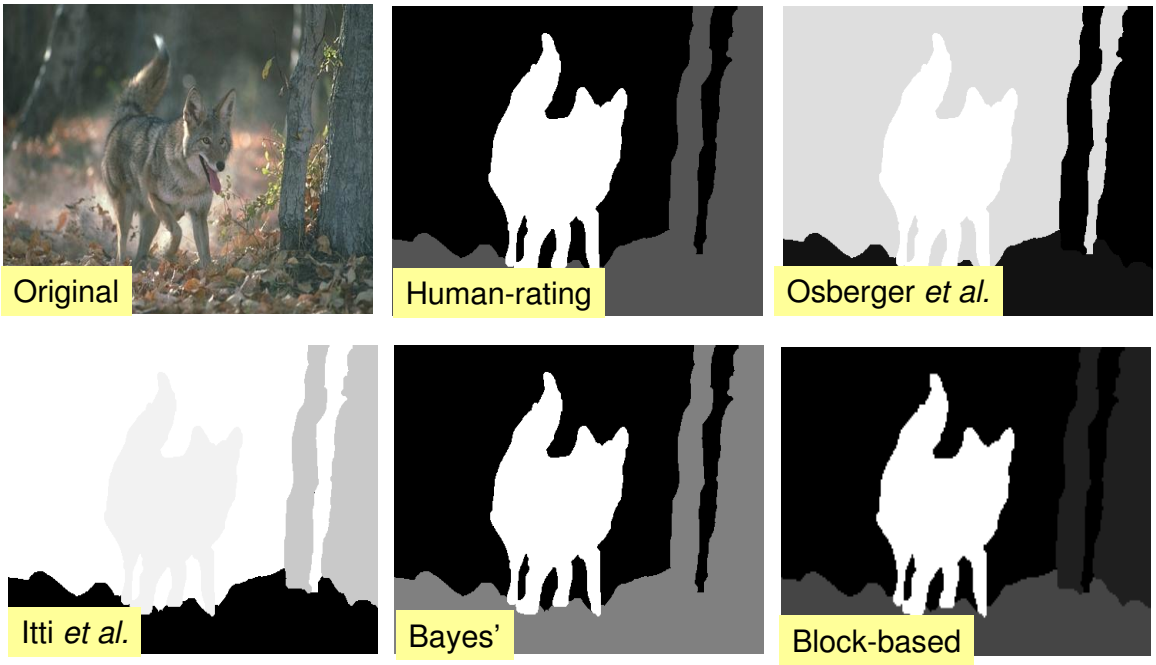


Figure 4.2.3: Results showing the importance maps generated using all the algorithms for an image of an animal.

Another difficult image is shown in Figure 4.2.4 where almost all the algorithms fail in finding the grass and the trees in the background. Our approach, assumes the trees to be more important as the trees have good contrast and color distance when compared to the sky. It is also observed that the importance map obtained via Osberger *et al.* fails in identifying the horse as the most important region.

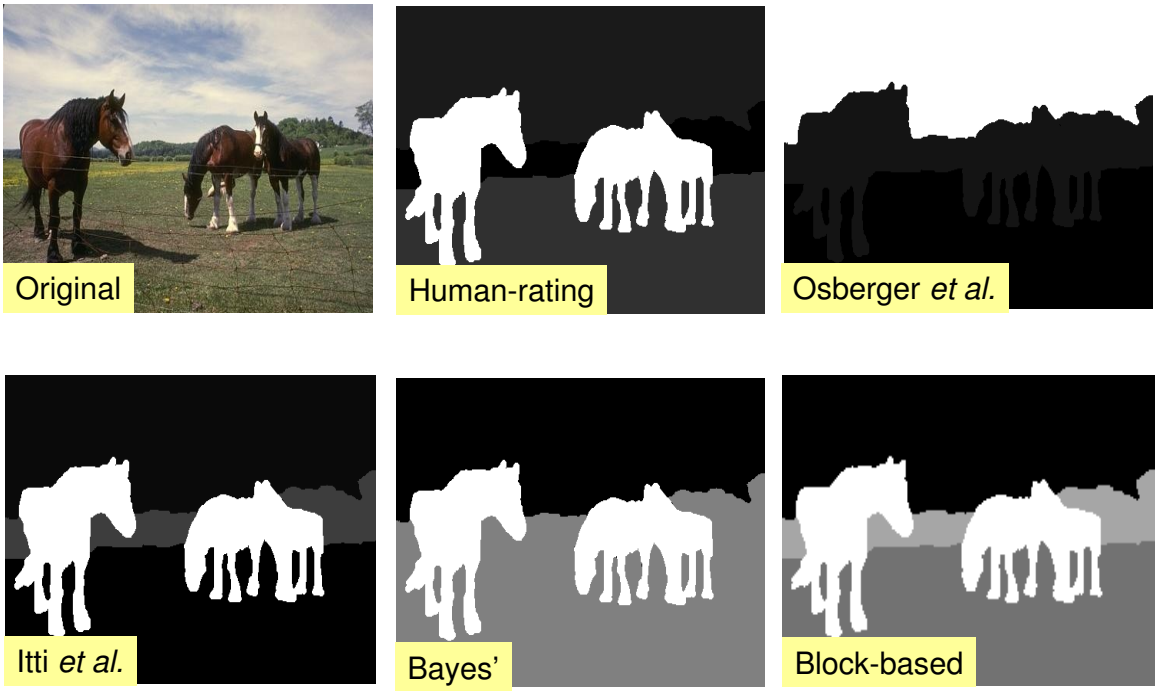


Figure 4.2.4: Results showing the importance maps generated using all the algorithms for an image of a horse.

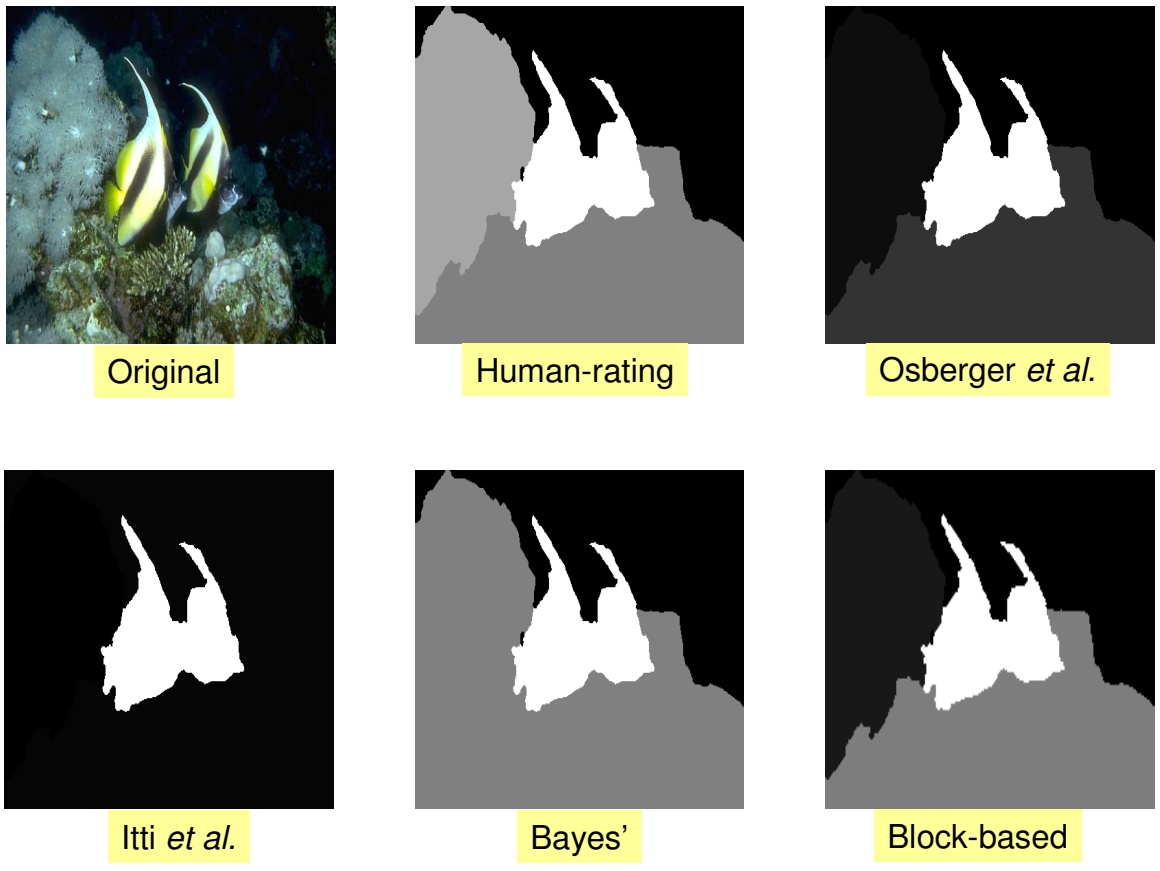


Figure 4.2.5: Results showing the importance maps generated using all the algorithms for an image of fishes.

Another example shown in Figure 4.2.5 is a difficult image to deal with because of the white corals in the scene. The corals in the scene do not have much contrast when compared to the other objects. Thus it is seen that block-based, Itti and Osberger fail to recognize the corals in the scene.

Some failure cases are shown in Figure 4.2.6 and 4.2.7. These cases tell us that there are still some hidden factors which influence the importance of objects in images and are yet to be found.

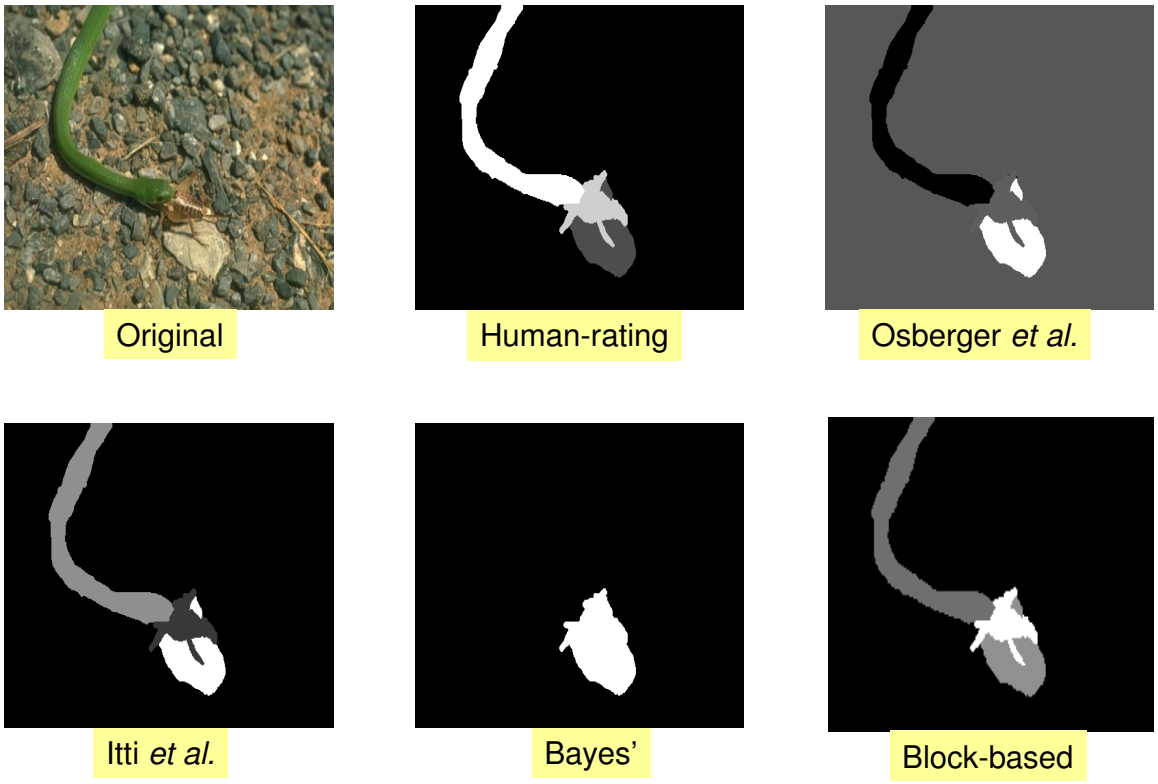


Figure 4.2.6: Failure cases showing the importance maps generated using all the algorithms for an image of a snake.

Figure 4.2.7 is an image of a snake where the proposed Bayes completely failed. Block-based does a decent job in finding the snake, but still is way away from the human rating importance map.

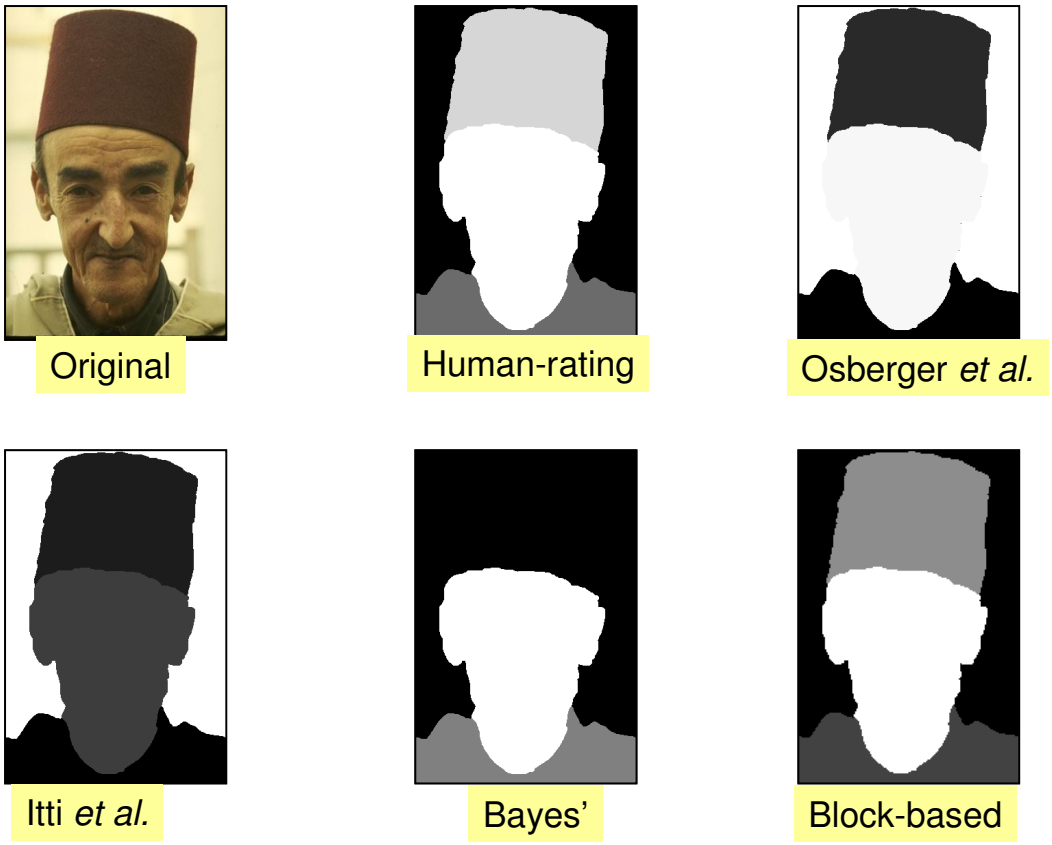


Figure 4.2.7: Failure cases showing the importance maps generated using all the algorithms for an image of a human.

Above is another image, where the proposed algorithm failed to identify the importance of the cap. However it does a decent job compared to Itti's where the face is not important and Osberger's where the background is given more importance than the face.

As mentioned earlier, all the algorithms are compared with the ground-truth data. To measure the performance of each algorithm, correlation coefficient and mean-square error is computed. Overall, our method leads to a correlation with subjective ratings of **0.707** with Bayesian probabilistic approach and **0.754** with block-based approach when applied to the 100 images from the testing set of Berkeley image segmentation database.

For the same testing set, the algorithm of Osberger *et al*. yields a correlation coefficient of **0.468**. We also compared our algorithm with Itti *et al* for same set of images. The correlation coefficient obtained using Itti *et al* is **0.512.** The correlation coefficients along with the RMSE are shown in the form of a table in Figure 4.3.

Approach	**R(Correlation coefficient)**	**RMSE(Root mean square error)**
Osberger *et al*	**0. 4683**	**0.4156**
Itti *et al*	**0.5118**	**0.4036**
Bayesian probabilistic approach	**0.7067**	**0.3312**
Block-based approach	**0.754**	**0.2639**

Table 4.1: Table showing correlation-coefficient and RMSE for all the algorithms.

We can see that the RMSE for the proposed block-based approach is the lowest. The RMSE is lower for the Bayesian probabilistic approach as well. This shows that the proposed algorithms perform better both in terms of correlation coefficient and RMSE.

CHAPTER 5

CONCLUSIONS

Here, we proposed two algorithms to estimate the perceived interest of objects in images. A psychophysical experiment has been performed to determine likelihood functions, which were then used as part of a Bayesian algorithm. Our preliminary results demonstrate that the predicted interests correlate well with human-rated interests. A correlation coefficient of 0.7 has been achieved when the algorithm is given a segmented image. On the other hand, with the original image as the input, the block-based approach gives a correlation coefficient of 0.75. This block-based approach is fast enough to be used in real-time applications. We are currently in the process of repeating the psychophysical experiment with more images containing a variety of commonplace subject matter. We are also investigating other factors that contribute to the perceived interest.

REFERENCES

[1] W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors," Proc. IEEE Int. Conf. on Image Process. **3**, pp. 414–418, 1998.

[2] L. Itti, C. Koch, and E. Neibur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell 20(11), pp. 1254–1259, 1998.

[3] F. Stentiford, "An estimator for visual attention through competitive novelty with application to image compression," Picture Coding Symposium 2001, 25-27 April 2001.

[4] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," IEEE Trans.Pattern Anal. Mach. Intell 22(9), pp. 970–982, 2000.

[5] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," Proc.R.Soc.Lond. B 265, pp. 359–366, 1998.

[6] F. Kozamernik, V. Steinmann, P. Sunna, and E. Wyckens, "SAMVIQ: A new EBU methodology for video quality evaluations in multimedia," SMPTE Motion Imaging Journal 114, pp. 152–160, 2005.

[7] J. Luo, A. Singhal, S. P. Etz, and R. T. Gray," A computational approach to determination of main subject regions in photographic images," Image and Vision Computing22 (2004) 227-241.

[8] David Liu, Datong Chen, Tsuhan Chen, "Latent layout analysis for discovering objects in images," The 18th International Conference on Pattern Recognition (ICPR), 2006.

[9] The Berkeley Segmentation Dataset and Benchmark, http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/BSDS300/html/dataset/images.html/.

[10] V. Kadiyala, S. Pinneli, E. C. Larson, and D. M. Chandler, "Quantifying the Perceived Interest of Objects in Images: Effects of Size, Location, Blur, and Contrast," *Proc. Human Vision and Electronic Imaging 2008*, San Jose, CA, January 2008.

[11] Srivani Pinneli, Damon M. Chandler, "Predicting the perceived interest of object in images", Proc. Southwest symposium on image analysis and interpretation 2008, Santa Fe, NM, March 2008.

[12] Radhakrishna Achanta, Francisco Estrada, Patricia Wils, and Sabine SÄusstrunk, "Salient region detection and segmentation", ICVS 2008, 66-75.

[13] Yuli Gao, Jianping Fan, " Automatic function selection for large scale salient object detection", MM 2006, Santa Barbara, California, USA.

[14] Li-Qun Chen, Xing Xie, Xin Fan *et al* , " A visual attention model for adapting images on small displays", Multimedia systems, 2003.

[15] Sang-Bok Choi, Sang-Woo Ban and Minho Lee, " Biologically motivated visual attention system using bottom-up saliency map and Top-down Inhibition", Neural Information Processing-Letters and Reviews, Vol 2, No. 1, January 2004.

[16] Byoung Chul Ko and Jae-Yeal Nam, " Object-of-interest image segmentation based on human attention and semantic region clustering", J. Opt. Soc. Am/ Vol. 23, No. 10/October 2006.

[17] Simone Frintrop, Maria Klodt, Erich Rome: " A real-time visual attention system using integral images", in Proc. of the 5th International Conference on Computer Vision Systems, Bielefeld, Germany, March 2007.

[18] Junwei Han, King N. Ngan, Mingjing Li, and Hong-Jiang Zhang, " Unsupervised extraction of visual attention objects in color images", IEEE transactions on circuits and systems for video technology, Vol. 16, No. 1, January 2006.

[19] Yigun Hu, Deepu Rajan and Liang-Tien Chia, " Robust subspace analysis for detecting visual attention regions in images", MM' 05, November 6-11, 2005, Singapore.

[20] Vidya Setlur et al, " Automatic Image Retargetting", Northwestern University Technical Report. NWU-CS-04-41, 2004.

[21] Jian et al, "Learning to detect a salient object", Computer vision and pattern recognition, 2007. CVPR' 07.

[22] Feng Liu, Micheal Gleicher, " Region enhanced scale-invariant saliency detection", in proceedings of ICME'2006.

[23] Jiebo Luo, Amit Singhal, Stephen P. Etz, Robert T. Gray, " A computational approach to determination of main subject regions in photographic images", Image and Vision Computing 22 (2004) 227-241.

[24] Yu-Fei Ma, Hong-Jiang Zhang, "Contrast-based Image Attention Analysis by Using Fuzzy Growing", MM' 03, November 2-8, 2003, Berkeley, California, USA.

[25] Olivier Le Meur, Patrick Le Callet, Dominique Barba, Dominique Thoreau, "A Coherent Computational Approach to Model Bottom-Up Visual Attention", IEEE transactions on pattern analysis and machine intelligence. Vol 28, No. 5, May 2008.

[26] Yaoru Sun and Robert Fisher, " Hierarchical selectivity for object-based visual attention", Lecture notes in Computer Science; Vol 2525. Proceedings of the Second International Workshop on Biologically Motivated computer vision. Pages: 427-438, 2002.

[27] C. Koch and S. Ullman, "Shifts in Selection in Visual Attention: Toward the Underlying Neural Circuitry", Human Neurobiology, vol. 4, no. 4, pp. 219-27, 1985.

VITA

Srivani Pinneli

Candidate for the Degree of

Master of Science

Thesis: PREDICTING THE PERCEIVED INTEREST OF OBJECTS IN IMAGES

Major Field: Electrical and Computer Engineering

Biographical:

Personal Data: Born in Gurazala, Andhra Pradesh, India on July 14, 1984.

Education:
Received the B.S. degree from CVR College of Engineering, Hyderabad, Andhra Pradesh, India, 2006, in Electronics and Communications Engineering

Completed the requirements for the degree of Master of Science in Electrical and Computer Engineering at Oklahoma State University, Stillwater, Oklahoma in July, 2008.

Experience: Worked as a Research Assistant in the Image Coding and Analysis Lab at Oklahoma State University for two years. Worked as a teaching assistant for Network Analysis Lab at Oklahoma State University.

Name: Srivani Pinneli Date of Degree: July, 2008

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: PREDICTING THE PERCEIVED INTEREST OF OBJECTS IN IMAGES

Pages in Study: 53 Candidate for the Degree of Master of Science

Major Field: Electrical and Computer Engineering

Scope and Method of Study: This thesis presents an algorithm designed to compute the perceived interest of objects in images based on results of a psychophysical experiment. We measured likelihood functions via a psychophysical experiment in which subjects rated the perceived visual interest of over 1100 objects in 300 images. These results were then used to determine the likelihood of perceived interest given various factors such as location, contrast, color, luminance, edge-strength and blur. These likelihood functions are used as part of a Bayesian formulation in which perceived interest is inferred based on the factors. A block-based approach is also proposed which doesn't need segmentation and is fast-enough to be used in real-time applications. Our results demonstrate that our algorithm can perform well in predicting perceived interest.

Findings and Conclusions: Our preliminary results demonstrate that the predicted interests correlate well with human-rated interests. A correlation coefficient of 0.7 has been achieved when the algorithm is given a segmented image. On the other hand, with the original image as the input, the block-based approach gives a correlation coefficient of 0.75. This block-based approach is fast enough to be used in real-time applications. We are currently in the process of repeating the psychophysical experiment with more images containing a variety of commonplace subject matter. We are also investigating other factors that contribute to the perceived interest.

ADVISER'S APPROVAL: Dr. Damon M. Chandler