Predicting the Perceived Interest of Object in Images
Srivani Pinneli and Damon M. Chandler
Image Coding and Analysis Lab, School of Electrical and Computer Engineering,
Oklahoma State University, Stillwater, OK 74078
srivanp@okstate.edu; damon.chandler@okstate.edu

Abstract
This paper presents the results of a psychophysical experiment and an associated algorithm designed to compute the perceived interest of objects in images. We measured likelihood functions via a psychophysical experiment in which subjects rated the perceived visual interest of each of 408 objects in 100 images. These results were then used to determine the likelihood of interest given various factors such as size, location, contrast, color, and edge-strength. The resulting likelihood functions are used as part of a Bayesian formulation in which perceived interest is inferred based on these factors. Results demonstrate that our algorithm can perform well in predicting perceived interest.

1. Introduction
When given an image, a human can quite easily point out the interesting or important objects in the scene. Yet, what is it that makes particular objects in an image interesting? Factors such as an object’s size, location, color, and contrast certainly contribute to our impression of interest. However, to what extent do other mid-level or high-level factors contribute to perceived interest?

The ability to automatically quantify perceived interest would have positive implications for a variety of image-processing applications. For example, in image compression, the ability to quantify the perceived interest of different objects would allow us to devote more bits to the more interesting objects. The ability to quantify perceived interest would also be useful in areas such as unequal error protection, watermarking, and variable-resolution displays. Note that perceived interest in this paper refers to images containing commonplace subject matter under general viewing conditions.

Several researchers have proposed algorithms for locating and/or quantifying regions of interest in images [1]-[4]. In [1] Osberger et al. present a method for automatically determining the perceptual importance of different regions by combining various factors such as 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 for all factors are then squared and summed to produce the final interest map. However, there remains an open question regarding the correct way to measure and combine the various factors to arrive at the overall interest map.

In [2], Itti et al. compute 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. Stentiford [3] proposed a measure of visual attention that depends upon the dissimilarity between neighborhoods in the image. In [4], Stark et al. present an analysis of the effectiveness of various image-processing operations and clustering procedures in predicting regions of interest in images.

In this paper, we present an algorithm that follows from the work of Osberger et al. and uses various factors (size, color, location, contrast, edge-strength, and category) 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.

This paper is organized as follows: Section 2 describes the psychophysical experiment performed to measure likelihood functions. The algorithm and its results are presented in Sections 3 and 4, respectively. General conclusions are provided in Section 5.

2. Psychophysical experiment

2.1 Apparatus and Subjects
Stimuli were displayed on a high-resolution, ViewSonic VA912B 19-inch monitor. The display yielded minimum and maximum luminances of respectively 2.7 and 207 cd/m² and an overall gamma of 2.9. Stimuli were viewed binocularly through natural pupils at a distance of 46 cm under D65 lighting.

Two adult subjects, naive to the purpose of the experiment, and the two authors, participated in the experiment. Subjects ranged in age from 21 to 32 years. All had either normal or corrected-to-normal vision.

2.2 Stimuli and Methods
Images used in the experiment were obtained from the Berkeley Segmentation Dataset and Benchmark image
An increase in object size. Here, the size of each object was defined as the ratio of the number of pixels in each object to the total number of pixels in the image.

### 3.1.2 Location: Usually an object toward the center of an image is more important than distant objects. We computer the location of the each object by first measuring the center of mass in both the horizontal and vertical directions. Then, the Euclidean distance from the center of mass to the center of the image was used to quantify location.

### 3.13 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. Each object’s contrast was 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:

$$\text{object contrast} = \frac{50}{M} \sqrt{\frac{\sum_{m=1}^{M} c_w^2}{M}}$$

where $M$ denotes the total number of blocks in the object, and where the block size, $B$, is computed based on the object’s size via $B = \max\{4, (0.05 \sqrt{N_{\text{object}}} + 0.05)\}$ where $N_{\text{object}}$ denotes the number of pixels in the object. The quantity $c_w$ denotes the RMS contrast of the $m^{th}$ block, which is given by $c_w = \sigma_w / \mu_w$, where $\sigma_w$ and $\mu_w$ denote the standard deviation and mean of the block luminances, respectively.

### 3.14 Color: It is quite natural that interest is drawn toward colorful and bright objects. Therefore, the color distance for each object was measured by creating a dilated mask for each object and measuring the Euclidean distance between the average object color and the average color of its neighboring regions. The distance was computed separately for brightness and color in CIELAB color space (CIE 1976, $L^*, a^*, b^*$).

### 3.15 Edge Strength: Usually, objects with greater numbers of edges are considered interesting. We measured edge-strength 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 for each object divided by the number of pixels in the object.

### 3.16 Category: A new high-level factor, category, was included. Here, each object was hand-categorized as either a Human, Animal, Object, or Background. We acknowledge that category is quite difficult to measure computationally; here, category was specified via a human. However, we also note that category is perhaps
the most important factor for obtaining proper results, and thus we feel it is important to provide results with and without category (see Section 4).

3.2 Using the factors with Bayes’ Rule

Bayes’ theorem relates the conditional and marginal probability distributions of random variables. It is expressed as follows:

\[ \text{Posterior} = \frac{\text{Prior} \times \text{Likelihood}}{\text{Constant}} \]

Here, we are interested in the probability of interest given the attributes of size, location, color, contrast, color, edge-strength, and category:

\[ P(\text{Interest} \mid \text{Attributes}) = \frac{P(\text{Attributes} \mid \text{Interest}) \times P(\text{Interest})}{\text{Constant}} \]

We assume statistical independence between the attributes; accordingly, we multiplied all the individual probabilities to get the likelihood term

\[ P(\text{Attributes} \mid \text{Interest}) = \prod P(\text{Attribute} \mid \text{Interest}), \]

where the terms \( P(\text{Attribute} \mid \text{Interest}) \) were measured from histograms generated based on the experimental results; i.e., we generated histograms for each attribute given that the object was rated as a primary, secondary or non-ROI. The histograms were fitted with either a generalized extreme value distribution or a Weibull distribution. These distributions were chosen because they provided a decent fit to all data. The fits are shown in Figure 2(a). Figure 2(b) shows the histogram for category.

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

4. Results

As part of the results, we show six demonstrative images (Figure 1) in six rows. In each row, the first image is the original image, the second image is the human-supplied ratings from the experiment, the third image is the corresponding interest map obtained via the algorithm of Osberger et al., and the fourth and fifth images show the interest map generated via our algorithm with and without category, respectively. In each result, brighter regions denote greater interest.

In Fig. 1, first row, we can see that the animals have the greatest interest. The interest decreases as we go from the trees and land to the sky in the background. Second row shows an image of a lion’s face resting on a rock with sky as the background. For this image the interest map obtained by Osberger et al works for the face, however it gives more interest to the sky than the rock in the foreground. Osberger’s approach also fails for the image in the fourth row where the background is given more importance than the animals in the foreground. In this image, the interest map from our approach with category yields better results because the size of the background is huge compared to that of the animals; category helps in classifying the objects as animals, which, based on Fig. 2(b), are more interesting than sky.

A better understanding can be obtained by looking at the two images in fifth and sixth rows. We can see in fifth row that the man is the most important object. The importance decreases when we go from the surfboard to the water to the sky in the background. Osberger’s approach considers the surfboard to be more interesting than the man. In the interest map obtained via our algorithm without category, the man and surfboard are rated primary, and water and sky are considered secondary and non-ROI, respectively. Based on the human ratings, the man is of greater interest than the surfboard; thus, a better result is obtained by using category.

A similar trend is observed in sixth row in which the animal is considered to be of equal interest as the grass in the background. Osberger’s approach fails here as it considers the sky as the most interest object. Thus, category has to be considered to get better results.

Overall, our method leads to a correlation with subjective ratings of 0.7 with the category and 0.6 without category when measured for the 100 images used in the psychophysical experiment. For the same set of images, the algorithm of Osberger et al. yields a correlation coefficient of 0.37. We acknowledge that this is a limited test suite, as these same images were used to determine the likelihood functions. We are currently in the process of repeating the psychophysical experiment with more images containing a variety of subject matter.

5. Conclusions

This paper presented a Bayesian approach to estimating the perceived interest of objects in images. A psychophysical experiment was performed to determine likelihood functions which were then used as part of a Bayesian formulation. Our results demonstrate that the predicted interests correlate well with human-rated interests. A correlation coefficient of 0.7 can be obtained when the algorithm is given a categorized image, and 0.6 without category. We are currently working to increase this correlation by adding other factors that contribute to perceived interest. We are also exploring methods of determining an object’s category.
6. References


Fig. 1. Results for five images shown in five individual rows. For each image, the original image is in the first column. The second column gives the human-rated interest map (IM) obtained via the psychophysical experiment. The third column is the interest map from Osberger et al.; the fourth and fifth columns show the interest maps obtained from our approach without category (fourth column) and with category (fifth column).

Fig. 2(a). Likelihood functions derived from the results of the psychophysical experiment. The top, middle, and bottom rows depict the histograms for all the attributes given that the objects were rated primary ROI, secondary ROI, or non-ROI, respectively. Each column depicts the probability \( P(\text{Attribute} | \text{Interest}) \) [e.g., probability of size given primary ROI].

Fig. 2(b). Histogram for category derived from psychophysical experiment given primary, secondary, and non-ROI.