A BAYESIAN APPROACH TO PREDICTING THE PERCEIVED INTEREST OF OBJECTS

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 an 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 562 objects in 150 images. These results were then used to determine the likelihood of perceived interest given various factors such as location, contrast, color, and edge-strength. These likelihood functions are used as part of a Bayesian formulation in which perceived interest is inferred based on the factors. Our results demonstrate that our algorithm can perform well in predicting perceived interest.

Index Terms—visual interest, Bayesian, region-of- interest, visual importance

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 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.

Researchers have proposed algorithms for locating and 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 location, contrast, and shape. An image is first segmented into regions, then each factor (e.g., location) 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, the correct way to measure and combine the various factors to arrive at the overall interest map remains an open question.

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

Bayesian probabilistic approaches have also been used for locating 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] for discovering main subjects in video; the algorithm in [8] consists of an appearance model and a motion model.

In this paper, we present an algorithm for quantifying the relative perceived interest (importance) of all objects in images. Our algorithm follows from the work of Osberger *et al.* and uses various factors (e.g., location, color, location, contrast, edge-strength) to determine perceived interest. The main contribution of our approach is the use of a Bayesian framework based on precise likelihood functions measured via an extensive 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. PSYCHOPHYISCAL EXPERIMENT

To obtain the likelihood of perceived interest given various factors, a psychophysical experiment was performed in which subjects rated the relative perceived interests of 562 objects in 150 images.

2.1. Apparatus and Subjects

Stimuli were displayed on a 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.

Eight adult subjects, naive to the purpose of the experiment, and the two authors[1], participated in the experiment. Subjects ranged in age from 24 to 34 years. All subjects had either normal or corrected-to-normal visual acuity.

2.2. Stimuli and Methods

Images used in the experiment were obtained from the Berkeley Segmentation Dataset and Benchmark image database. This database was chosen because its images are accompanied by human-segmented versions (averaged over at least five subjects).

[1] 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 $R > 0.8$).

One-hundred images were chosen at random and were used as a training set. Another 50 images, also chosen at random from the same database, were used for testing the algorithm. All 150 images were segmented into 562 objects in the database. The images used were of size 321×481 with 24-bit RGB pixel values.

For each of the 562 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 in which 10 corresponded to greatest interest and 1 corresponded to least interest. Subjects were given unlimited time.

2.3. Results

Raw scores for each subject were converted to z-scores; and then averaged across all subjects. The averaged z-scores were then rescaled from 0 to 1 for 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. We generated histograms for various object attributes given that the object was rated as a primary, secondary or non-ROI.

The attributes were: location, contrast, brightness, color, edge-strength, a measure of the extent to which the object was in the foreground (discussed in Section 3), and object category. The histograms were computed based on the results for the 100-image training set. The histogram for category is shown in Figure 1(a). The remaining histograms were fitted with either a generalized extreme value distribution or a Weibull distribution; the fits are shown in Figure. 1(b). These distributions were chosen because they provided a good fit to all three types (primary, secondary, non-ROI) for each factor.

Ideal *interest maps* (in which brightness corresponds to perceived interest) were generated from the ratings and are shown in the second column of Figure 2. 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). 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.

3. ALGORITHM

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.

3.1. Measuring Various Attributes

The various attributes involved in computing the interest maps are color, contrast, location, edge-strength, foreground/background and category. For all 562 objects, each factor was measured as follows:

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

3.1.2. 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 where 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.

3.1.3. Color: It is quite natural that interest is drawn toward colorful and bright objects. Therefore, the color distance for each object was measured by: (1) Creating a dilated mask for each object, and then (2) measuring the Euclidean distance between the average object color/brightness and the average color/brightness of its neighboring pixels defined by the dilation. The distance was computed separately for brightness ($a_{brightness}$) and color (a_{color}) in CIELAB color space.

3.1.4. Edge Strength: Usually, objects with greater numbers of edges are considered interesting. 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.

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

3.1.5. 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.

3.1.6. Category: A new high-level factor, category, was included ($a_{category}$). Here, each object was hand-categorized as 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. Here, we are interested in the probability of perceived interest given the attributes of location, color, contrast, edge-strength, foreground/background, and category.

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

$$P(I \mid \overline{a}) = \arg\max(P(I) \times P(\overline{a} \mid I))$$

where $P(I)$ was assumed to be 1/3. By further assuming statistical independence between the attributes, the likelihood term can be expressed as

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

where the individual probabilities were measured from histograms generated based on the experimental results.(see Figure. 1)

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.

4. RESULTS

As part of the results, we show five demonstrative images (Figure 2) in five 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 Figure 2, *top row*, we can see from the human-rated interest map that the bird has the greatest interest, whereas the snow in the background has least interest. Osberger's approach, fails for this image, the snow in the background is shown brighter than the bird at front. Figure 2, *second row*, shows an image of a man and child with the wall in the background. From the importance map obtained from Osberger's approach, we see that background is given higher importance when compared to the people in the foreground.

The difference in results with and without category demonstrates the significance of category in predicting the importance of objects in images. In general, category helps correct for the case in which the background is given more importance due to contrast and/or color. By classifying each object as Human, Animal, Object and Background, we get an improved interest map where each object is given appropriate importance. For example, in Figure 2, *bottom row,* without category, water is considered least important and thus is shown all black. But if we notice in the original image we can see that seawater has more importance when compared to the sky in the background. Thus, by classifying water as an object we assign it more importance as compared to the background. Thus, it is quite evident that the addition of category helps in obtaining better results.

Overall, our method leads to a correlation with subjective ratings of 0.8 with category and 0.7 without category when applied to the 50 images reserved for the testing set. For these images, the algorithm of Osberger *et al*. yields a correlation coefficient of 0.5. We attribute the improved performance of our approach to the likelihood functions obtained via the psychophysical experiment.

5. CONCLUSIONS

In this paper, we proposed a Bayesian probabilistic approach to estimate 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 algorithm. Our preliminary results demonstrate that the predicted interests correlate well with human-rated interests.

We are currently in the process of repeating the psychophysical experiment with more images containing a variety of commonplace subject matter, and we are investigating other factors that contribute to perceived interest.

6. 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**, 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, 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. 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 Computing 22 (2004) 227-241.
6. David Liu, Datong Chen, Tsuhan Chen, "Latent layout analysis for discovering objects in images," The 18th International Conference on Pattern Recognition (ICPR), 2006.
7. The Berkeley Segmentation Dataset and Benchmark, http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/BSDS300/html/dataset/images.html/.
8. 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.

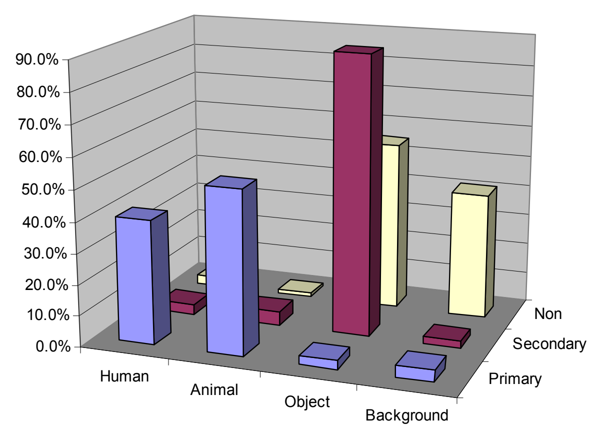

Figure 1(a). Histogram for category derived from psychophysical experiment given primary, secondary and non-ROI.

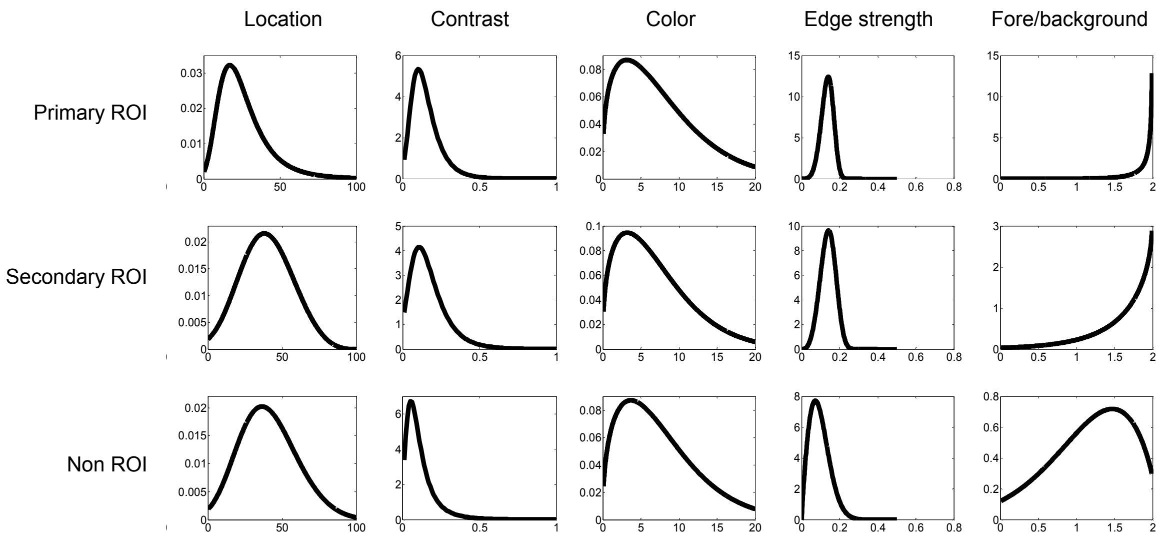

Figure 1(b). Likelihood functions derived from the results of the psychophysical experiment. Top, middle and bottom row gives the histograms for all the attributes given that the object is primary, secondary or non-ROI, respectively. Each column gives the likelihood *P*(Attribute | Interest) (e.g., likelihood of location given it is primary, etc.).

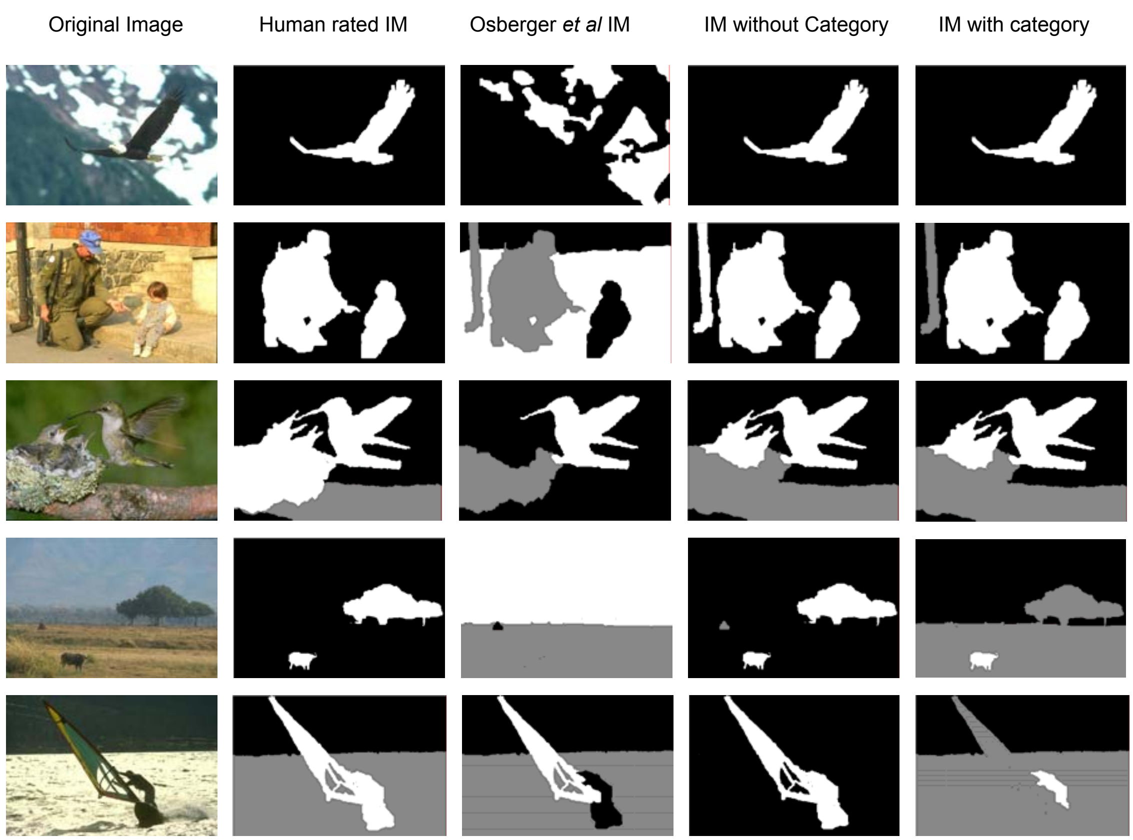

Figure 2. Results obtained using our approach with and without category for five images are shown in five individual rows. The original images are shown in the first column. The second column shows the human-rated importance map obtained via the psychophysical experiment. The third column is the importance map from Osberger *et al.*; the fourth and fifth columns show the importance maps obtained from our approach without category (fourth column) and with category (fifth column). The white, gray, and black regions denote primary, secondary, and non-ROI, respectively.