Salient Object Detection in Static Image

Akshay H. Auti
PG student, Electrical Engineering Department,VJTI Mumbai,India.

Jayalakshmi O. Chandle
Electrical Engineering Department, VJTI Mumbai, India.

Rajesh S. Ingle
PG student, Electrical Engineering Department,VJTI Mumbai,India.

***Abstract*—This model is made to identify salient objects from background of an image since we pays more attention towards the salient object in an image than its background. Here we have done binary labeling which separates salient object from its background. In this paper we have used feature extraction methods like edge detection, thresholding, multi scale contrast, center surround histogram has been used. We extract low level features based on colour, contrast and intensity of an image. After normalization and linear combination, a master map or a saliency map is computed which represents the saliency of each image pixel. Finally, the image is segmented out from the background. Feature maps are prepared using edge detection, thresholding and multiscale contrast. Saliency is computed using center surround histogram.**

Keywords—edge detection, multiscale contrast, thresholding, WTA(winner-take-all).

I. INTRODUCTION

Human eyes pay more attention towards some parts of an image than its background. Salient object is the object in an image which draws more attention towards itself. Salient object detection problem is composed as binary labeling task where the salient object and the background has been separated from each other. Applications like image cropping, video compression and adaptive image display use visual attention.

The bottom-up attention can be modeled as an integration of low level image features of different measures [1]. Koch and Ullman proposed the first computational architecture of bottom-up attention model in 1985 [2]. The bottom-up attention model proposed by Itti et al. draws great attention now a day's [3].

II. FEATURE EXTRACTION

A. Edge Detection

For smoothening of an image we use the Gaussian filter. Gaussian blur generates in an image and this is the blur in an image that causes due to application of the Gaussian function. One dimension equation of the Gaussian function is given as below

$$G(x) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \qquad (1)$$

And the two dimension equation of a Gaussian function is given as below

$$G(x,y) = \frac{1}{\sqrt{2\Pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (2)$$

Where x is the distance of horizontal axis from origin, y is the distance of the vertical axis from origin, and σ is the standard deviation of the Gaussian distribution.

Identifying points in a digital image at which the image brightness changes sharply or more formally, has discontinuities. These discontinuities are called as edge detection [4]. The various points at which image brightness changes sharply are typically organized into a set of curved line segments which are known as edges. Edge detection is a fundamental tool in an image processing, machine vision and computer vision. And this is used particularly in the areas of feature detection and feature extraction. In the ideal case the result of applying an edge detector to an image may lead to a set of connected curves that indicates the boundaries of an object. These are the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed and may therefore filter out the information which can be considered as less important, while preserving the important structural properties of an image.

Based on this one-dimensional analysis, this theory can be carried for two-dimensions because there is an accurate approximation to calculate the derivative of a two-dimensional image. The Sobel operator, which is also called as Sobel Filter, is used in image processing and computer vision, particularly within edge detection algorithms. And it creates an image which highlights edges and transitions. This operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives out of which one for horizontal changes and another one for vertical.

Gradient magnitude is calculated by

$$G = \sqrt{G_x^{\,2} + G_y^{\,2}} \qquad (3)$$

The x-coordinate is defined here as increasing in the right direction and the y-coordinate is defined as increasing in the down direction.

B. Thresholding

Once we compute the measurement strength of edges (basically the gradient magnitude), the next step is to apply thresholding. Thresholding is used to decide whether the edges are present or not in an image. Thresholding is the simplest method of image segmentation from a grayscale image, thresholding can be used to create binary images. More edges will be detected if the threshold value is lower, and the result will be increasingly open to the noise and detecting edges of irrelevant features in the image. On the other hand if the threshold is high it can miss minute edges, or result in broken edges. Thresholding with hysteresis is generally used approach for handling the problem of appropriate thresholds for the purpose of thresholding [5]. Multiple thresholds are used in this method to find edges. To find the start of an edge we begin it with upper threshold value. Once we get the start point, we trace the remaining path of the edge through the image pixel by pixel. We keep marking an edge whenever we are above the lower threshold value. We stop marking our edge at the time the value falls below our lower threshold value. This approach makes the assumption that edges are more likely to be in continuous curves and allows us to follow an indistinct section of an edge. As previously seen, every noisy pixel in the image is marked as an edge. Still, we have the problem of choosing appropriate thresholding parameters because of suitable thresholding values can also vary over the image.

C. Multiscale Contrast

Contrast is the most commonly used local feature for attention detection because the contrast operator simulates the human visual receptive fields. Contrast is usually computed at multiple scales Without knowing the size of the salient object.

A Gaussian pyramid is a technique used in an image processing, especially in texture synthesis. This technique involves creating a series of images which are weighted down and scaled down using a Gaussian average which is also known as gaussian blur. Using this technique multiple times, it creates a bunch of successively smaller images with each pixel containing a local average that corresponds to a pixel neighborhood on a lower level of the pyramid.

Multiscale contrast shows the high contrast boundaries by giving low scores to the homogenous regions of the salient object.

III. SALIENCY MAP COMPUTATION

Humans use visual selective attention to try to reduce the computational complexity and save computational resources. Top-down factors are very subjective and difficult to model. But bottom-down factors are based on the visual features like orientation, intensity or color which are easier to estimate. There are many computational models based on the bottom-up mechanisms of the visual attention which are designed to determine the saliency map of the image from visual features of the image [6].

Generating the topographical maps of the visual saliency from the images are called Saliency Maps. Saliency map is the one which integrates the normalized information from the individual feature maps into single global measure of conspicuity. In analogy with the center-surround representations of maps of elementary visual features. We will use bottom-up approach. Bottom-up saliency is determined by how different a stimulus is from its surround in many sub modalities and at many scales [7]. The saliency map was designed for converting selective attention as an input to the control mechanism. Once a topographic map of saliency is established, Winner-Take-All mechanism computes the position of the maximum in this map and the location is obtained. After the selection of the location is made, suppression of activity at the selected location (which may correspond to the psychophysically observed inhibition of return mechanism) leads to selection of the next location at the location of the second-highest value in the saliency map. And repeating this procedure few times, generate a sequential scan of the complete visual scene. This role of the saliency map in the control of which locations in the visual scene are attended is close to that of the master map postulated in the Feature Integration Theory.

- The saliency maps are based in Feature Integration Theory that define the next steps:
  1. An early representation were composed of a set of feature maps which were computed in parallely, permitting separate representations of several stimulus characteristics.
  2. A topographic saliency map is then computed in which each location encodes the combination of properties across all the feature maps as a conspicuity measure.
  3. A selective mapping into a central non-topographic representation is computed through the topographic saliency map of all the properties of a single visual location.
  4. A winner-take-all network implements the selection process based on one major rule: conspicuity of location (minor rules of proximity or similarity preference can also also be suggested).
  5. Hesistancy of this selected location causes an automatic shift to the next visible location. Feature maps codes visibility only within a particular feature dimension.
- Fig.1 shown below is the Itti's computational model for salient object detection. This is the bottom-up computational approach.

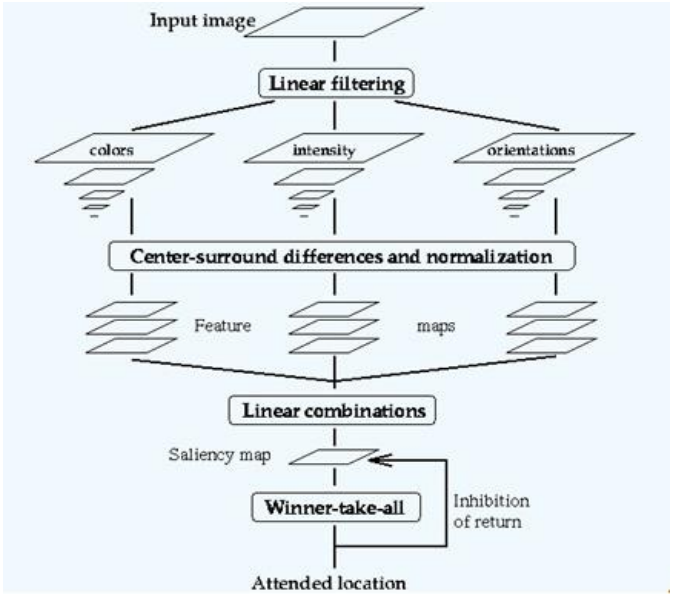

Fig. 1 Bottom-Up Saliency-Based Visual Attention

IV. GROUND TRUTH CONSTRUCTION

The ground-truth based methods measure the deference between the results obtained by segmentation and the human-labeled ground truths together. Ground truths are more intuitive than the factual based measures since they can well represent the human-level interpretation of an image. In this category some measures are aim to count the degree of an overlapping between regions with strategies of being intolerant or tolerant for concentration contain an image. There are also measures matching the boundaries between segmentations in contrast with working on regions of an image [8,9]. These measures are more sensitive to the dissimilarity between the segmentation and the ground truths considering the region boundaries only. There is no standard procedure for segmentation because of the imperfectly defined nature of an image segmentation (i.e. there might be multiple acceptable segmentations which are consistent to the human interpretation of an image). There is also a large difference in the perceptually meaningful segmentations for various images. And these above factors make the evaluation task very complex. Here we focus on evaluating segmentation results with multiple ground truths considerations. The existing methods of this kind of considerations prefer matching the given entire segmentation with ground truths for the evaluation purpose. But the available human-labeled ground truths are having only a small fraction of all the possible interpretations of an image. The available dataset of ground truths might not contain the desired ground truth which is suitable to match the input segmentation. Thus such kind of comparison often leads to a certain partiality on the result or it is far from the aim of objective of the evaluation.

We propose a new framework to solve this problem. Theoretically it is given as below-

Consider a set of ground truths G = {G1,G2,...,GK} of an image

X = {x1,x2,...,xN}

Where Gi = {gi 1,gi 2,...,gi N} denotes a labeling set of X, i = 1,..., K and N is the number of elements in the image(pixels, regions). Let S = {s1, s2,...,sN} be a given segmentation of X, where sj is the label of xj (boundary or non-boundary) j = 1,...,N. To examine the similarity between S and G, we compute the similarity between S and a new ground truth consideration G^*

G^* is computed from G based on S. and denoted as G^* = {g∗ 1,g∗ 2,...,g∗ N}. G^* is constructed by putting together pieces from G, i.e., each piece g∗ j ∈ {g1 j ,g2 j ,..., gK j }. G^* is a geometric ensemble of local pieces from G. We will develop an optimistic strategy to choose the elements of G^* by which S will match G as much as possible. Then G^* can be taken as a new segmentation of X by assigning each pixel label. To construct G^* we introduce a label lgj (l = 1,...,K) to each g∗ j in G^*. Fig. 3 uses an example to illustrate how to construct the new ground truth G^*.

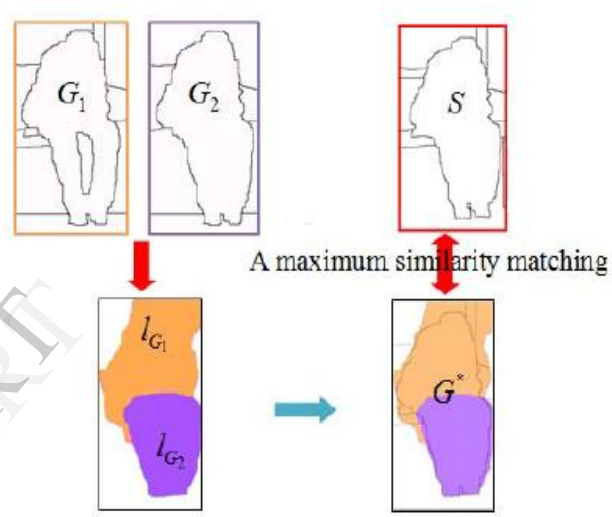

Fig. 2. An example of adaptive ground truth composition for the given segmentation S.

We can see that, given two ground truth images G1 and G2, G^* is found by first computing the optimal labeling set for the ground truths. Then elements of G^* which are labeled as 1 (or 2) will take their values from G1 (or G2). This leads to a maximum similarity matching between S and G^*.

V. EXPERIMENTAL RESULTS

Fig. 3. Result of an image after the application of algorithm.

VI. CONCLUSION

This paper is based entirely on bottom up computational approach. Basic image processing techniques like feature extraction, image extraction, segmentation and object recognition has been used. Using thresholding and boundary determination we have obtained feature maps. Given an input image, this system attempt to predict which location in the image will automatically and unconsciously draw your attention towards them. This algorithm is capable of detecting multiple salient objects provided that they are distinct and prominent in the image and also it detects the most salient object in an image with fair accuracy.

ACKNOWLEDGMENT

This investigation supported by the department of Electrical Engineering of VJTI, Mumbai, University of Mumbai, India.

REFERENCES

[1] A.M. and G. Gelade ,"A feature integration theory of attention ," Cognitive Psychology, Vol.12, No.1, pp. 97-136,1980.

[2] C. Koch and S. Ullman, .Shifts in selective visual attention: towards the underlying neural circuitry,. *Human Neurobiology*, Vol.4,pp.219-227,1985.

[3] L. Itti and C. Koch, .Computational modeling of visual attention,. *Nature Reviews Neuroscience*, Vol. 2, No. 3, pp. 194-203,Mar.2001

[4] E. Argyle. "Techniques for edge detection," Proc. IEEE, vol. 59, pp. 285-286, 1971

[5] L. Itti, C. Koch, and E. Niebur, A model of saliency- based visual attention for rapid scene analysis, IEEE Trans. Pattern Anal. Machine Intell., Vol. 20, No. 11, pp. 1254-1259, Nov. 1998

[6] Hou, X., Zhang, L.: Saliency detection: A spectral residual approach.In:CVPR(2007).

[7] J. Harel. A saliency implementation in mat- lab. [Online] http://www.klab.caltech. edu/~harel/share/gbvs.php, 2010.

[8] Martin, D.: An empirical approach to grouping and segmentation. Ph.D. disser- tation U. C. Berkeley (2002).

[9] Freixenet, J. Munoz X., Raba D. Marti, J. Cuff, X: Yet another survey on image segmentation: Region and boundry information integration. European conference on Computer Vision. Pp.21-25(2002).