

# Fixation-Based Visual Image Segmentation using the concept of Border-ownership

Manjusha Singh<sup>1</sup>, Abhishek Misal<sup>2</sup>

Assistant Professor, Dept of CSE, Chhatrapati Shivaji Institute of Technology, Durg, Chattisgarh, India<sup>1</sup>

Assistant Professor, Dept of ETC, Chhatrapati Shivaji Institute of Technology, Durg, Chattisgarh, India<sup>2</sup>

**Abstract :** The reason behind the choice of segmenting an image based on a fixation point originates from the way humans perceive objects in their surroundings. Even when our eyes continuously fixate at different things in the scene to see/perceive them, most computer vision algorithms are yet to give any significance to fixation. It is believed that fixation (a part of visual attention) is the reason why human visual system works so well. This method proposes an approach where we do not segment the whole image in several regions and then reason on what these regions might be, but segment only the region which contains our fixation point. So in short segmenting an object from the scene implies finding a “fixation point” and finding the contour of the object that encloses our point of interest. Border-ownership is an automated fixation strategy for finding interest points and it means - knowing the boundary of an object. Using the border ownership information, we automatically select fixation points inside all possible objects in the image and segment them. In this paper, some other ways are also briefly discussed which can be used to select fixation points such as Contrast-Saliency model and Symmetry-saliency model but we use Border-ownership method.

**Keywords:** visual, border-ownership, fixation, contrast, symmetry.

## I. INTRODUCTION

With the development of information technology, images have been the major source of information. How to analyze and process these images efficiently has been the focus of research. In image analyzing and processing, people are often only interested in some regions which are called objects or foreground. Therefore, it is needed to separating these objects from the image to do further image processing. Image segmentation is the technology that can be used to divide the image into several regions and separate the objects from the background.

During the research, people find human have the remarkable ability to find the objects of interest in complex scenes quickly which is called “visual attention”. It is very important and necessary to introduce the visual attention mechanism in image processing because it can reduce the computational complexity, save computational resource and improve the efficiency of image processing [1].

The purpose of image segmentation is to partition an image into *meaningful* regions and objects (of related content) with respect to a particular application. Vision is the most advanced of our senses, so images play the single most important role in human observation. The human visual system (HVS) is known to have an attention module that uses the low-level visual cues (such as color, texture etc.) to quickly find the salient locations in the scene. The human eyes are then drawn to these salient points (also called fixations). These fixations points are going to be used as the identification markers for the objects of interest in the scene. The human (primate) visual system observes and makes sense of a dynamic scene (video) or a static scene (image) by making a series of fixations at various salient locations in the scene.

A New Segmentation approach is proposed in [2] that is biologically motivated and connects segmentation with visual attention (Inspired by the Human Visual System). Although fixation is known be an important component of the human visual system, it has largely been ignored by computer vision researchers

The human visual system has two types of attention: overt attention (eye movements) [3] and covert attention (without eye movement) In this work, overt attention is meant whenever the term attention is used, so attention is classified into two categories based on whether its deployment over a visual scene is guided by scene features or by intention: the first is called bottom-up and is driven by low-level processes; the second refers to top-down processes. Most work has happened in bottom-up attention. The feature integration theory [4] suggests that when perceiving a stimuli, features are "registered early, automatically, and in parallel, while objects are identified separately" and at a later stage in processing. Feature integration theory is a theory of attention that has inspired many bottom-up models of visual attention [5, 6]. The most admired is the one proposed in [7] and it has become a standard model of bottom-up visual

attention, in which saliency due to primitive features such as intensity, orientation and color are computed independently. A model like this is not appropriate because it often fails when attention needs to be focused on an object. The attention causes the eye to move and fixate at a new location in the scene. Each fixation will lie on an object, identifying that object (which can be a region in the background too) for the segmentation step. Now, segmenting that fixated region is defined as finding the “optimal” enclosing contour—a connected set of boundary edge fragments—around the fixation. This new formulation of segmenting fixated regions is a well-defined problem. This algorithm takes a fixation point as its input and outputs the region containing the given fixation point in the scene. Advantage-The segmentation becomes a fully automatic process which finds the optimal segmentation of the fixated regions without any user input.

## II. MOTIVATION

Over the years, many different algorithms have been proposed that segment an image into regions, but the definition of what is a correct or “desired” segmentation of an image (or scene) has largely been indefinable to the computer vision community. Most segmentation algorithms depend upon some form of user input, without which the definition of the optimal segmentation of an image is ambiguous.

With respect to a particular object of interest, the correct/ desired segmentation of the scene is the one wherein the object of interest is represented by a single or just a couple of regions. So, the goal of segmenting a scene is intricately linked with the object of interest in the scene and can be well defined only if the object of interest is identified and known to the segmentation algorithm beforehand. But having to know about the object of interest even before segmenting the scene seems to make the problem one of many contradictory problems in computer vision, as we usually need to segment the scene first, to recognize the objects in it. So, how can we identify an object even before segmenting it? So the existing segmentation algorithms cannot give correct/desired results.

## III. FIXATION-BASED SEGMENTATION METHOD

Here, bottom-up image segmentation is considered. That is, we ignore (top down) contributions from object recognition in the segmentation process and we expect to segment images without recognizing objects. For a given fixation point, segmenting the region/object containing that point is a twostep process [2]:

**Cue Processing:** Visual cues such as color, texture, motion and stereo generate a probabilistic boundary edge map wherein the probability of a pixel to be at the boundary of any object in the scene is stored as its intensity.

**Segmentation:** For a given fixation point, the optimal closed contour (connected set of boundary edge pixels) around that point in the probabilistic edge map. This process is carried out in the polar space to make the segmentation process scale invariant.

In order to segment multiple objects, the segmentation process will be repeated for the fixation points from inside each of the objects of interest. However, the edge map contains both types of edges, namely, boundary (or depth) and internal (or texture/intensity) edges so it is important to be able to differentiate between the boundary edges from the non-boundary (e.g. texture and internal) edges.

“Polar space method” traces the closed contour through the probabilistic boundary edge map in the Cartesian (coordinate system) space to polar coordinate system and gives optimal enclosing contour around the fixation show that we obtain segmentation accuracy.

For this segmentation framework, the fixation just needs to be inside the objects in the scene. As long as this is true, the correct segmentation will be obtained. This segmentation method clearly depends on the fixation point, and thus it is important to select the fixations automatically. There can be different ways to automatically select fixations; we use Border-ownership.

## IV. BORDER-OWNERSHIP

The border ownership of a boundary pixel is the information about the side containing the object [8] and it means - knowing the boundary of an object. Besides helping select points inside the objects, the border ownership information also helps differentiate between the closed contours corresponding to the objects (boundary of that object) and the non-object closed contours (a closed contour that does not correspond to the boundary of any object in the image). An overview of this system is as follows: A probabilistic boundary edge map is generated using color, texture and motion cues. For the pixels with significant boundary likelihood, the border ownership is also estimated. Using the border ownership information, a set of points inside different objects is automatically selected. Around each point, [2] finds the closed contour in the probabilistic boundary edge map. A subset of resulting closed contours that uniquely

corresponds to the objects is finally selected as the output of the system. To cleanse any false closed contour still remaining in the output, temporal persistence is also enforced at last.

This method selects only those closed contours that correspond to the objects in the scene while rejects duplicate and non-object closed contours.

**V. METHODS FOR SELECTION OF FIXATION POINTS:**

There are different ways/ algorithms that select fixation points and then segmentation process can be carried out for the selected points:

**A. Contrast-Saliency model:**

This saliency method calculates independently saliency by center-surround contrasts of basic primitive features, that utilizes brightness, color, and orientation contrasts. This model is henceforth referred to as the contrast-saliency model. It is a visual attention system, inspired by the behavior and the neuronal architecture of the early primate visual system. The contrast saliency model often selects fixations close to the border of the object. In this, Multi-scale image features are combined into a single topographical saliency map. A dynamical neural network then selects attended locations in order of decreasing saliency. The system breaks down the complex problem of scene understanding by rapidly selecting, in a computationally efficient manner, conspicuous locations to be analyzed in detail [7].

**B. Symmetry-saliency model:**

When interpreting a scene, humans pay attention to objects, not so much to basic features. Hence, configural features play an important role in human visual attention. A configural feature is a higher-level feature that integrates different parts of a figure. Symmetry, for instance, can be a stronger visual attractor than basic features [9]. Human eye fixations can also be predicted based on symmetry in the image. Contrast focuses near corners and edges in the image, whereas symmetry highlights symmetrical configurations which often coincide with objects in the scene. In [10], symmetry has shown to outperform the contrast saliency model in detecting salient objects in the scene. A segmentation approach is proposed named Automatic Detection And Segmentation (ADAS) based on this model.

For the detection of objects, symmetry is used because the quality of segmentations improves when fixation points lie near the center of the object.

The amount of local symmetry at a given pixel,  $p = (x, y)$ , is calculated by applying a symmetry kernel to the pixel [10]. Pixels pairs in the symmetry kernel contribute to the local symmetry value. A pair consists of two pixels,  $p_i$  and  $p_j$ , so that  $p = (p_i + p_j)/2$ . The contribution of the pixel pair to the local symmetry of  $p$  is calculated by comparing the intensity gradient  $g_i$  at  $p_i$  and gradient  $g_j$  at  $p_j$ :

$$c(i,j) = (1 - \cos(\gamma_i + \gamma_j)) \cdot (1 - \cos(\gamma_i - \gamma_j)) \dots\dots\dots eq. (1)$$

where  $\gamma_i = \theta_i - \alpha$  is the angle between the orientation of the gradient,  $\theta_i$ , and the angle,  $\alpha$ , of the line between  $p_i$  and  $p_j$ . The first term in Eq. (1) has a maximum value when  $\gamma_i + \gamma_j = \pi$ , which is true for gradient orientations that are mirror symmetric in  $p$ . Using only this term would also respond to two pixels that have the same gradient orientation and thus lie on a straight edge. Since we want to find the centers of symmetrical patterns, the second term in the equation demotes pixels pairs with similar gradient orientations.

All experiments show improved performance of symmetry –saliency model over the contrast-saliency model.

**C. Border-Ownership method:**

The border ownership has been reported to be registered by the neurons in a primate’s visual cortex. It is shown that the depth information is most important in determining the border ownership of a pixel at the boundary of an object. They also report that the determination of the border ownership happens as a result of local processing of visual cues. In computer vision literature, a popular term for border ownership is figure/ground assignment. But, compared to static monocular cues, depth and motion information are stronger cues in determining border ownership of a boundary edge pixel [8].

**VI.METHODOLOGY FOR CARRYING OUT SEGMENTATION:**

**A. Probabilistic Boundary Edge Map:**

Edge detection is a very important area in the field of Computer Vision. Edges define the boundaries between regions in an image, which helps with segmentation and object recognition [11]. In the probabilistic boundary edge map, the

intensity of a pixel is set to be the probability to be either depth or contact boundary in the scene. The probability to be at a depth boundary can be determined by checking for a discontinuity in the optical flow map at the corresponding pixel location because depth discontinuity introduces discontinuity in the optical flow map as well. But the exact location of discontinuities in optical flow maps often does not match with the true object boundaries, a well known issue for optical flow algorithms. To account for this, we use static cues such as color and texture to, first, find all possible boundary locations in the image which are the edge pixels with positive color and/or texture gradient. Then, the probability of these edge pixels to be on depth and contact boundary is determined. The maximum of two probabilities is assigned as the probability of an edge pixel to be on object boundary,  $P_B(x, y)$ . Note that color and texture gradient values do not participate in determination of the boundary probability [8].

#### B. Fixation Strategy:

The goal of the fixation strategy is to select points inside objects so that the fixation-based segmentation method [2] takes those points and generates the closed boundaries enclosing those points. To select the points, we first pick the edge pixels in the probabilistic boundary edge map with the boundary probability greater than 0.5, and assume that they lie on the boundary of some object in the image. We represent this subset of boundary edge pixels by  $I_B$  as given below:

$$I_B(x,y) = \begin{cases} 1 & \text{if } P_B(x,y) > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Additionally, we estimate the border ownership (object side) of the edge pixels in  $I_B$ . A boundary edge pixel in  $I_B$  can either be a depth or a contact boundary. For a pixel at depth discontinuity, the object side is the one that is closer to the camera which means will have bigger mean optic flow value.

For a pixel at contact boundary, the object side is the one with the color distribution different than the known color distribution of the surface. We compute the difference between the properties of the two sides of an edge pixel [8].

#### C. Fixation Based Segmentation:

For each selected point found in previous section, the fixation based segmentation approach [2] finds the closed boundary around the given point by combining the edge pixels in the probabilistic boundary edge. The segmentation for each point has two

intermediate steps: first, the probabilistic boundary edge map  $P_B$  is converted from the Cartesian to the Polar space using the given point as the pole for the conversion, in order to achieve scale invariance. Following this, a binary segmentation of the polar edge map generates the optimal path through the polar edge map such that when the curve is mapped back to the original Cartesian space, it encloses the point. The two-step segmentation process is repeated for all fixation points found in previous section using the same  $P_B$  [8].

#### D. Selecting Closed Contours Corresponding To Objects:

We have as many closed contours as the number points selected by the fixation strategy. Since the selection of points depends on the contour fragments in  $I_B$ , the fragments that are part of the same object boundary generate multiple points lying inside the same object; these points give rise to duplicate closed contours. Also, sometimes due to error in the estimation of border ownership of the edge fragment, the corresponding point lies outside of any object in the image which will lead to a closed contour that does not correspond to an object. So, we need a process that sifts through all the closed contours to pick only the ones that uniquely correspond to the objects in the scene. We require a method to differentiate between any two closed contours [8].

#### E. Temporal Persistence:

The regions selected by this solution may still contain false regions arising due to wrongly detected strong edges in the probabilistic boundary edge map. One simple way to get rid of the spurious regions without having to use high level knowledge is based on the observation that the noisy edges in the probabilistic boundary edge map is caused by the noise in the flow map as we have used motion cues to determine the object boundaries in our experiments. Since the effect of noise is not localized but changes its locations in the flow map over time, the wrongly detected boundary edge fragments will change too. This means, the spurious regions formed by these edge fragments would change as well. All of this suggests that if we repeat the entire process for frame 2 and 3 and match the selected regions with the regions obtained for the frames 1 and 2 (as we are considering 3 consecutive frames of a scene), and accept only the regions

that occur in both cases, we end up discarding most of the spurious regions. The regions that persisted in time are more likely to be actual objects [8].

The whole methodology is shown as a flowchart in figure 1:

Figure 1: Modified Flow-chart for the Proposed Methodology:

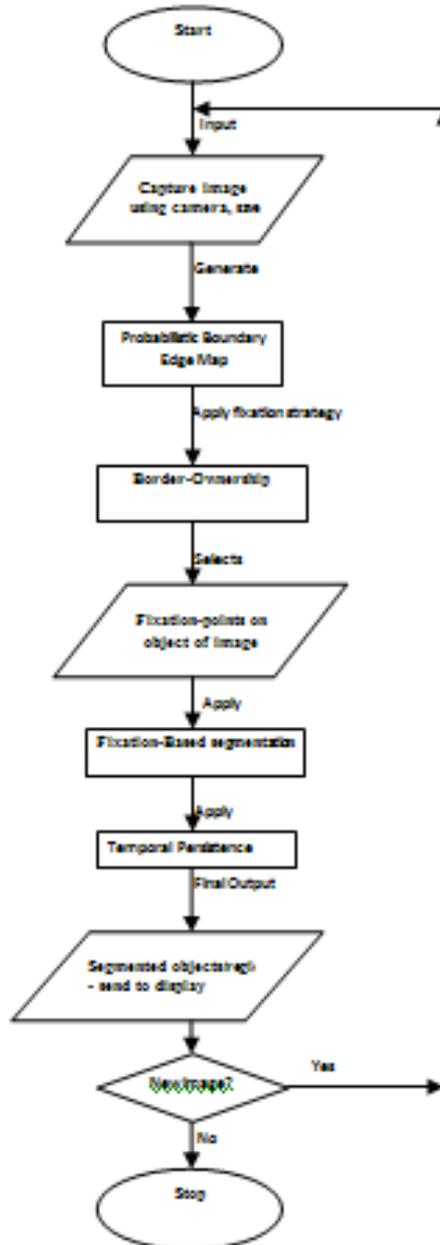


Figure 1: Flow-chart of Methodology

## VII. EXPECTED RESULTS

The performance is measured in terms how many of the objects correspond to one of the segmented contours (recall) and how many of segmented contours correspond to the actual objects (precision).

The temporal persistence helps weed out closed contours arising out of noisy edge pixels in the probabilistic boundary edge map. It improves the precision of the system by almost 25%. Further improvement in precision can be achieved by using a recognition system which will match the contours with its knowledge base to reject any unknown contours.

It is expected that the proposed system will, on average, segment 17 out of 20 objects kept on a table given a three consecutive frames of the scene.

To evaluate the efficiency of the fixation strategy, the percentage of all selected fixation points is calculated that actually lie inside an object and it is expected to be 85%. The final evaluation is about the robustness of predicting the border ownership (or object side) which is computed as the mean of the percentage of the pixels on any object boundary with correct object side prediction. It is expected to be 93%.

## VIII. CONCLUSION

A system is described that segments objects of any size or shape automatically. The system is based on the idea that segmentation should produce a closed contour surrounding the fixation point of attention. Two novel contributions of this work are a fixation strategy (basic attention mechanism) to select the points on the objects, and a strategy to select only the regions corresponding to the objects. The input to the system is a minimum of three images of the scene. Also, the segmentation step can be carried out for all the selected points in parallel as they do not depend on each other. Use of border-ownership appears to be more promising than symmetry-saliency or contrast-saliency model for selection of fixation points

## REFERENCES

- [1] Qiaorong Zhang, Guochang Gu, Huimin Xiao, "Image segmentation based on visual attention mechanism," Journal of Multimedia, Finland, vol. 4(6), pp. 363-370, December, 2009.
- [2] Ajay K. Mishra, Yiannis Aloimonos, Cheong Loong Fah and Ashraf Kassim, "Active Visual Segmentation" IEEE Transaction on Pattern Analysis and Machine Intelligence (TPAMI), vol 34, No. 4, April 2012.
- [3] G. Kootstra and L. R. B. Schomaker. Prediction of human eye fixations using symmetry. In Cognitive Science Conference (CogSci), Amsterdam, The Netherlands, 2009.
- [4] A.M. Treisman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, 12, page no 97-136, 1980.
- [5] C.C. Williams A. Hollingworth and J.M. Henderson. To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. Psychonomic Bulletin and Review, 8:761-768, 2001.
- [6] J. M. Henderson. Human gaze control during real-world scene perception. Trends in Cognitive Sciences, 7:498-504, 2003.
- [7] L. Itti, C. Koch, and E. Niebur. A model of saliency based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11):1254-1259, 1998.
- [8] Mishra, A. and Aloimonos, Y. "Visual Segmentation of "Simple" Objects for Robots" Proceedings: Robotics Science and Systems conference (RSS), Los Angeles, June 27 - July 1, 2011.
- [9] Kootstra, G., Bergström, N., and Kragic, D. (2010) "Fast and Automatic Detection and Segmentation of Unknown Objects" In: *Proceedings of the IEEE-RAS International Conference on Humanoids Robotics (Humanoids 2010)*, December 6-8, 2010, Nashville, TN.
- [10] G. Kootstra, N. Bergstrom, and D. Kragic, "Using symmetry to select fixation points for segmentation," in International Conference on Pattern Recognition, 2010.
- [11] E. Nadernejad, S. Sharifzadeh and H. Hassanpour "Edge Detection Techniques: Evaluations and Comparisons" Applied Mathematical Sciences, Vol. 2, 2008, no. 31, 1507 - 1520
- [12] D. Martin, C. Fowlkes, and J. Malik. "Learning to detect natural image boundaries using local brightness, color and texture cues" T-PAMI, 26(5):530-549, May 2004.

## BIOGRAPHY



**Manjusha Singh** received the BE (CSE) degree from MPC CET, Bhilai (Pt. RSSU, Raipur) in 2008. She is Assistant Professor of CSE dept. Of CSIT, Durg since July 2008. Her research interests are in the field of Digital Image Processing.



**Abhishek Misal** is Assistant Professor of Electronics and Telecommunication Engg. Dept. of CSIT Durg since Feb. 2008. He received the BE (ETC) degree from CSIT, Durg (Pt. RSSU, Raipur) in 2006 and M.Tech. (Instrumentation & control) degree from BIT Durg (CSVTU, Bhilai) in 2011. He has published papers in various national and international conferences and journals and also attended various workshops and seminars. He is a life member of IETE, India and ISTE, India. His research interests are in the field of Digital signal Processing and its applications in field of bio-medical.