International Journal of Advanced Research Methodology in Engineering & Technology, ISSN 2456-6446
Volume 1, Issue 4, July 2017

Image segmentation using hidden markov random field model for saliency detection

Anchitha Sathyan
PG Student
Electronics and Communication Department
LBS Institute of Technology for Women
Thiruvananthapuram, India
anchitha92@gmail.com

Reena M Roy
Assistant Professor
Electronics and Communication Department
LBS Institute of Technology for Women
Thiruvananthapuram, India
reenamroy@gmail.com

***Abstract*— Salient structure detection in an image is required by many applications such as image re-targeting, automatic cropping, object tracking, video encoding, and selective sharpening etc. Salient areas are generally regarded as areas which human eye will typically focus on, and finding these areas is the main step for object identification. That is image segmentation is the key process in saliency detection. In computer vision applications, image segmentation is the process of partioning a digital image into multiple segments, sets of pixels known as super-pixels. The proposed method uses Hidden Markov Random Field (HMRF) model for image segmentation. The main advantage of using this is instead of normal segmentation output, we get a relative depth wise segmentation output. The image is segmented into different clusters by using k-means clustering algorithm and it is modelled by a Gaussian mixture model.**

Keywords*— Salient structure , Computer vision, Hidden markov random field model, EM algorithm

I. INTRODUCTION

Image segmentation is an important technology for image processing. Segmentation partitions an image into distinct regions containing each pixel with similar attributes. To be meaningful and useful for image analysis and interpretation, the regions should strongly relate to depicted objects or features of interest. The first step in low level image processing is transforming a gray scale or color image into one or more other images to high level image description in terms of feature, objects, and scenes. A striking feature is the process of segmentation which is used for the extraction of the main features of the image, like a color input, calculates the number of remaining points of the maps for color, brightness, contrast and orientation at different scales of static image. Segmentation is dividing a digital image into segments (sets of pixels, also known as super-pixel). The purpose of segmentation is to simplify the display of an image into something that's meaningful and easier to analyze change. Image segmentation normally used for objects and boundaries (lines, curves, etc.) to take pictures and visually see. More specifically, image segmentation method for assigning a label for each pixel in the image so the pixels in the same label certain visual properties. The most striking objects have the quality to visual elimination of their environment and are likely to attract attention to the man. An important that makes an object is the striking visual difference in the background.

The main step for saliency detection is the segmentation process. Saliency is a perceptual quality that makes an object, person, or pixel stand out relative to its neighbors and thus capture our attention. Saliency detection is the key to the extraction of image information. Extracting image saliency region is required in most image processing methods. Precisely extracting the salient regions of images effectively facilitates many image applications. How is the saliency detection process achieved in human visual system? It is believed that two stages of visual processing are involved: first, the parallel, fast, but simple pre-attentive process; and then, the serial, slow, but complex attention process. In this stage, certain low level features such as orientation, edges, or intensities can pop up automatically. From a viewpoint of object detection, what pops up in the pre-attentive stage is the candidate object. In order to address a candidate that has been detected but not yet identified as an object. Most of the detection models focus on summarizing the properties of target objects. However, general properties shared by various categories of objects are not likely to exist.

Humans are able to detect visually distinctive (so called salient) scene regions effortlessly and rapidly (pre-attentive stage). These filtered regions are then perceived and processed in finer details for extraction of richer high-level information (attentive stage). This capability has long been studied by cognitive scientists and has recently attracted a lot of interest in the computer vision community mainly because it helps find the objects or regions that efficiently represent a scene and thus harness complex vision problems such as scene understanding. One of the earliest saliency models, which generated the first wave of interest across multiple disciplines including cognitive psychology, neuroscience, and computer vision, is proposed by Itti et.al[1]. This model is an implementation of earlier general computational frameworks and psychological theories of bottom-up attention based on center-surround mechanisms (e.g., Feature Integration Theory

(FIT) by Treisman and Gelade [2], Guided Search Model by Wolfe et.al [3]. Itti et.al show examples where their model is able to detect spatial discontinuities in scenes. Subsequent behavioral and computational studies start to predict fixations with saliency maps to verify saliency models and to understand human visual attention. A second wave of interest appears with works of Liu et.al [4],[5] and Achanta et.al [6] who define saliency detection as a binary segmentation problem. These works themselves are inspired by some earlier models striving for detecting regions. Since then a plethora of saliency models have emerged that have blurred the boundary between these two categories of models. Further, it has been less clear where this new definition stands, as it shares many concepts with other established computer vision areas such as image segmentation algorithms, category independent object proposal generation approaches, fixation prediction models, and object detection methods.

In addition to the fast, bottom-up, involuntary, and stimulus-driven stage of attention which is of main interest in computer vision community, there exists a slower, top down, voluntary, and goal-driven stage of attention which is relatively less explored due to the complexity and variety of daily tasks and behaviors. Further, subjective factors such as age, culture, and experience regulate attention. For example, a detective sees a crime scene differently than a policeman or a pedestrian. Some related topics, closely or remotely, to visual saliency include: object importance, memorability, scene clutter, video interestingness, surprise, image quality assessment, scene typicality, aesthetic, and attributes.

II. LITERATURE SURVEY

The term saliency was first proposed by Tsotsos et.al [7] in the context of visual attention. Since then researchers have shown a great interest towards pre-attentive or bottom-up saliency detection. Early methods have mostly concentrated on human eye fixation prediction and they have introduced the basic principles of saliency detection. Then, the important problem that has been addressed in literature is salient object detection and segmentation. Since, bottom up saliency is stimulus driven and does not look for any particular object in the scene, it can be used for unsupervised segmentation of all the prominent objects in an image. This leads to a solution of the problem of generic object segmentation.

A. Frequency Tuned Salient Region Detection

Achanta et.al[8] describes saliency by using frequency tuned detection. In this method, for salient region detection that outputs full resolution saliency maps. This method exploits the features of color and luminance, is simple to implement, and is computationally efficient. In frequency tuned method of finding the saliency map S for an image I of width W and height H pixels can be formulated as

$$S(x,y) = | I\mu - Iwhc(x,y) | \quad (1)$$

Where Iμ is the arithmetic mean pixel value of the image and Iwhc is the Gaussian blurred version of the original image to eliminate fine texture details as well as noise and coding artifacts. The norm of the difference is used since they are interested only in the magnitude of the differences. This is computationally quite efficient. Also, as they operate on the original image without any downsampling, they obtain a full resolution saliency map

To extend (1) to use features of color and luminance, we rewrite it as:

$$S(x,y) = || I\mu - I_{whc}(x,y) || \quad (2)$$

Where Iμ is the mean image feature vector, Iwhc(x,y) is the corresponding image pixel vector value in the Gaussian blurred version (using a 5*5 separable binomial kernel) of the original image, and || || is the L2 norm. Using the Lab color space, each pixel location is an [L,a, b]T vector, and the L2 norm is the Euclidean distance. This method, summarized in (2) allows to fulfill all of the requirements for salient region detection.

The true usefulness of a saliency map is determined by the application. In this method they consider the use of saliency maps in salient object segmentation. To segment a salient object, we need to binarize the saliency map such that ones (white pixels) correspond to salient object pixels while zeros (black pixels) correspond to the background. This analysis illustrated that the deficiencies of other techniques arise from the use of an inappropriate range of spatial frequencies. Based on this analysis, they presented a frequency tuned approach of computing saliency in images using low level features of color and luminance, which is easy to implement, fast, and provides full resolution saliency maps. The resulting saliency maps are better suited to salient object segmentation, demonstrating both higher precision and better recall.

B. Contex Aware Saliency Detection

Most of the models use contrast as an important cue. These approaches work well for images which have a simple background and high contrast between background and foreground image elements. Stas Goferman et.al [9] proposed a new type of saliency-contex aware saliency. This method aims at detecting the image regions that represent the scene. Normally this definition differs from other saliency definitions whose goal is to either identify fixation points or detect the dominant object. In this method saliency is modelled using both local low-level features and global considerations, as well as visual organization rules and high level features. They have taken overlapping patches at different scales and modeled saliency as distance in color, inversely weighted by distance in position among the patches. In accordance with the saliency definition, they presented a detection algorithm which is based on four principles observed in the psychological literature. The benefits of this approach are evaluated in two applications where the context of the dominant objects is just as essential as the objects themselves. In image retargeting they demonstrate that using saliency prevents distortions in the important regions. In summarization they show that their saliency helps to produce compact, appealing, and informative summaries. The algorithm is mainly based on four principles.

- Local low-level considerations, including factors such as contrast and color.
- Global considerations, which suppress frequently occurring features, while maintaining features that deviate from the norm.
- Visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized.
- High-level factors, such as human faces.

In accordance with the first principle ,areas that have distinctive colors or patterns should obtain high saliency. Conversely, homogeneous or blurred areas should obtain low saliency values. In agreement with second principle frequently-occurring features should be suppressed. According to third principle, the salient pixels should be grouped together, and not spread all over the image.

C. Salient Structure Detection by Contour Guided Visual Search

In this approach Kai-Fu et.al[10] define the task of salient structure detection to unify the saliency related task, such as fixation prediction, salient object detection, and detection of other structures of interest in cluttered environments. To solve such SS detection tasks, a unified framework inspired by the two-pathway-based search strategy of biological vision is proposed in this paper. First, a contour-based spatial prior (CBSP) is extracted based on the layout of edges in the given scene along a fast non-selective pathway, which provides a rough, task-irrelevant, and robust estimation of the locations where the potential salient structures are present. Second, another flow of local feature extraction is executed in parallel along the selective pathway. Finally, Bayesian inference is used to auto-weight and integrate the local cues guided by CBSP and to predict the exact locations of salient structure. However this method fails to detect salient object in cluttered environment.

III. METHODOLOGY

The human pays unequal attention to what is seen in the world. When looking at some images, people are usually attracted by some particular objects within the images. Other subjects appear uninteresting for them. This ability of withdraw from some things in order to deal effectively with others is called attention. The important step for this detection is segmentation. Here I propose a new method for the segmentation of salient object using a statistical approach called Hidden Markov Random Field Model (HMRF). It is a stochastic model for segmentation which uses the spatial information in the input image. Out of the various stochastic models, Hidden Markov random field (HMRF) model provides a better framework for many complex problems. This is due to the fact that, HMRF model is based on the notion of neighborhood structure and therefore helps in understanding global interaction through local spatial interactions. Moreover, the global interaction is governed by Gibbs distribution. In HMRF the observation is a probabilistic function (discrete or continuous) of a state. All observations are dependent on the state that generated them, not on the neighboring observations. HMM is a finite set of states, each of which is associated with a probability distribution. In a particular state an outcome or observation can be generated, according to the associated probability distribution. It is only the outcome, not the state visible to an external observer and therefore states are “hidden” to outside; hence the name Hidden Markov Model. This model is specifically useful where the data is hidden. A special case of HMM is that, the underlying stochastic process is considered as MRF instead of a Markov chain and therefore not restricted to one dimension. This special case is referred to as Hidden Markov Random Field (HMRF) model. The HMRF model is based on the Markov random field theory, in which the spatial information is encoded through a neighborhood system. Hidden Markov random field (HMRF) model is a stochastic process generated by a Markov random field whose state sequence cannot be observed directly but can be observed through observations. The advantage of the HMRF model derives from the way in which the spatial information is encoded through the mutual influences of neighboring pixels. The HMRF model is more flexible for image modeling in the sense that it has the ability to encode both the statistical and spatial properties of an image.

Salient structure detection using Hidden Markov Random Field is a statistical approach for saliency detection which segment the input image depth wise. Here the input image is a color image. The input color image is given to HMRF-EM block which segment the input image into depth wisely. In this one of the plane contain the salient object.

A. Hidden Marko Random Field Model

A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly. A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other.

B. HMRF Algorithm

- Apply Gaussian filter to preprocess the input image

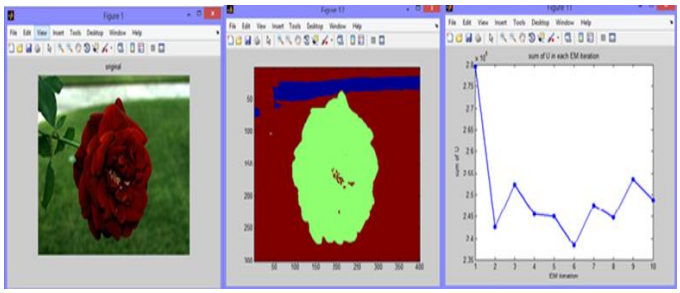
- Apply K means clustering for obtaining the initial labels.
- Calculate the likelihood distribution
- Using current parameter estimate the labels by MAP estimation
- Calculate the posterior (unobserved) distribution for all labels and all pixels.
- Use posterior distribution to update the parameters

The Gaussian blur is used for applying transformation to each pixel in the image. This reduce the image detail and the noise. The initial labels are obtained by performing K-means clustering. The initial labels obtained from K-means are not smooth enough and have morphological holes. The initial segmentation by K-means provides X(0) for MAP algorithm and the initial parameters θ(0) for the Expectation Maximization (EM) algorithm. EM algorithm is used because it is an unsupervised learning problem.

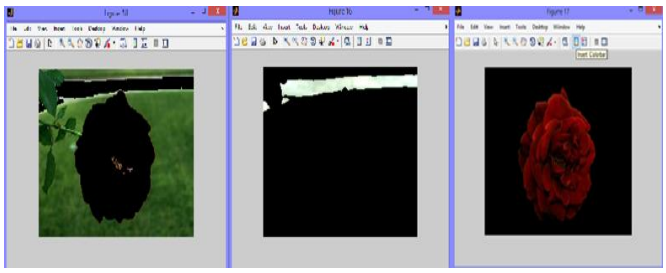
IV. EXPERIMENTAL RESULTS

The experiment was done on two different datasets; MSRA-A and ECSSD. Salient structure is detected by extracting the depth information in an image. It is done by segmenting the input image into foreground and background. For this hidden markov model was used. It segment the input image into different planes. Then extracted different planes and one of the plane contain the salient object.

Gaussian filtering was applied to the input image for reducing the noise and unwanted detail. The initial labels are obtained by K means clustering. Here the number of clusters were set to three. That is the number of planes are three. It is an unsupervised clustering algorithm and classifies the input data points into three classes. It is mainly based on the intensity distribution. The parameters are obtained by EM algorithm. Each of the E and M steps are straight forward assuming the other is solved. There is two assumptions; knowing the label of each pixel we can estimate the parameters, knowing the parameters of the distributions we can assign a label to each pixel. Here the EM iteration is set to 5 which is the optimum value.



(a) (b) (c)



(d) (e) (f)

Fig .1 HMRF Output, (a) Input Image, (b) HMRF Output, (c) EM iteration, (d) Plane 1 of HMRF output, (e) Plane 2 of HMRF output, (f) Plane 3 of HMRF output

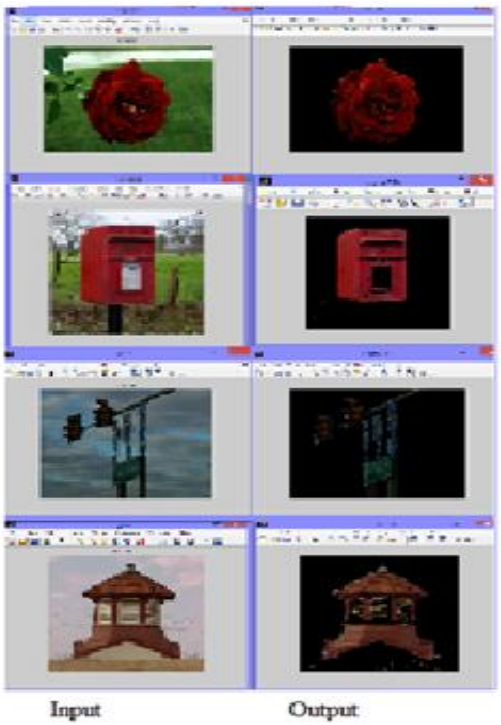


Fig .2 Salient object detection using HMRF model

HMRF segment the input image into depth wise, so the different planes of the output contains objects in different planes. If we separate the planes we get objects in each plane separately. From Fig 1 it is clear that one of the plane contains the salient object. Fig 2 depicts the salient object detection output of different images.

References

[1] L. Itti, C. Koch, and E. Niebur," A model of saliency-based visual attentio for rapid scene analysis", IEEE TPAMI, 1998

[2] M.Treisman and G.Gelade,"A feature-integration theory of attention",Cognitive Psychology, pp. 97–136, 1980

[3] J.M Wolfe, K.R Cave, and S.L Franzel,"Guided search: an alternative to the feature integration model for visual search", J. Exp. Psychol. Human, vol. 15, no 3, p. 419, 1989

[4] T. Liu, J.Sun, N.Zheng, X. Tang, and H.Y Shum,"Learning to detect a salient object", CVPR, pp. 1-8, 2007

[5] T.Liu, J.Sun, N.Zheng, X.Tang, and H.Y Shum,"Learning to detect a salient object", IEEE TPAMI, vol. 33, no. 2, pp. 353-367, 2011

[6] R. Achanta, F.Estrada, P.Wils, and S. Susstrunk,"Salient region detection and segmentation", Comp. Vis. Sys., 2008

[7] Tsotsos, J. K., S. M. Culhane, W. Y. Kei Wai, Y. Lai, N. Davis, and F. Nuo,"Modeling visual attention via selective tuning",Artificial intelligence, 78(1), 507-545, 1995.

[8] R. Achanta, S. Hemami, F. Estrada, and S. S usstrunk, "Frequency-tuned salient region detection", CVPR, p. 1597–1604, 2009

[9] S.Goferman, L.Zelnik-Manor and A.Tal.,"Context-aware saliency detection", CVPR, 2010

[10] Kai-Fu Yang, Hui Li, Chao-Yi Li, and Yong-Jie Li,"A Unified Framework for Salient Structure Detection by Contour-Guided Visual Search", IEEE Transactions On Image Processing, Vol. 25, No. 8, 2016