International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 8 Issue 1 –APRIL 2014.

VISUAL ATTENTION BASED SALIENT OBJECT MOTION DETECTION IN SPATIO TEMPORAL VOLUME.

-A CMPARATIVE STUDY

Bhagyasree D[#1]

[#]*M-tech, Computer vision and image processing, Amrita vishwa vidhyapeetham, Coimbatore, India*

***Abstract*—We present different visual attention based salient object detection methods for effectively detecting object in structured environment. Human brains pay more attention towards some important part of image sequences. Those attentions are extraordinary fast and realistic one. Computation of such salient object detection like intelligence behavior is very difficult task for implementation. We plan to explore the area of computer vision to create a study of salient object detection models based on visual attention. The problems of visual attention based salient object detections are mainly studied by researchers include neural system, physiology; computer vision etc has received much attention over the past few years. There are wide ranges of applications in salient object detections like traffic rule violation detection, object tracking and security surveillances, human action prediction recognition etc. Recent studies demonstrated that it can also be applied in robotic navigation as well in different perspective and object detection in multi dimensional space together with cues for human activity recognition and event understanding.**

***Index Terms*- observability; rarity; scene analysis spatiotemporal saliency detection; visual attention.**

I. INTRODUCTION

Visual saliency gives a specific subjective perceptual quality that makes the items in the environment stand out from their neighborhood and immediately grab our attention even in the presence of complex background. Modern attention theory explains saliency as selective attention. Each attention model gives verity of definition for saliency. Visual attention is unconsciously driven by low level features included in an image sequences like intensity or color, orientation, motion features etc. The existing approaches are mainly deals with bottom-up computational model concepts. These models mainly consist of following three stages. The first stage is feature extraction from the current frame, in which contrast, texture, orientation, motion like visual features are extracted at different scales. Next stage includes saliency computation based on either self information, centre surround approach, or graph based operation with various visual features. Saliency in a frame can be represented by saliency map, which is computed by normalization of previous stage with proper linear or non linear combination. Final saliency of each image pixel is given by master map or saliency map.

In visual saliency approach, that generally searches for the identification of rare features. Temporal information must be considered along with spatial features for each and every frame in a video. Hence the different methods available were studied and an overall view of the techniques of salient object detection based visual saliency with their merits and demerits of each method were explored.

II. COMPARATIVE STUDY

A. Object motion detection using information theoretic spatio temporal saliency.

The paper of [chang liu *et al.* (2009)] proposes object motion detection based on saliency in a video. Salient object from a video frame has been obtained using the concept of both information theory and information saliency map(ISM). Selective attention and saliency is closely related to each other. Information theory will give the inverse relation about the probability of occurrence of an event with respect to information content in each frame. Both information content and occurrence of an event have equivalent importance. Saliencies in each pixel are represented by using the saliency map. Spatio temporal information saliency map concepts have been used. In the paper, the author has done separate saliency for both static and temporal volume. Finally, dynamic fusion approach has to be adopted for tie upping both saliencies together. Some basic concepts of dimensionality reduction and kernel density estimations have been also used for constructing better information saliency map. The ISM used for getting salient object from its background.

Occlusion effect, illumination problems, similarities in colors are the challenges included in this work. Reliable output is obtained without background subtraction.

B. A Linear Dynamical System Framework for Salient Motion Detection

This paper [Viswanath Gopalakrishnan et al. (2012)] proposed a method for salient object from a video sequences has been detected by using the idea of observability while the frames are arranged as a linear dynamical system. According to psychological concept, the abrupt stimuli may leads to a sudden and involuntary creation of attention. In this work Salient motion under complex moving background has been considered. The key concept of the paper is that, a frame work for volumetric saliency with predictable motion and observability. Saliency map were constructed in two aspects, one from pixel level and another from region level. Final saliency map is obtained by fusing both pixel and region level saliency together. Multi input multi output based representation of frames has to be taken. Most salient region from an image sequence is modeled by computing spatio temporal patch between the highest groups of output pixels with in a linear dynamic system frame work. They introduce a distance measure for comparing the most relevant spatio temporal patch with other existing approaches.

The main advantages of this methodology includes, it does not require any kind of background estimation and knowledge of motion in a system. The linear dynamic system is computationally effective. In this work the author consider only the dynamic part of the video instead of static information.

C. A Model of Saliency Based Visual Attention on Rapid Scene Analysis

The paper [Laurent Itti *et al*, (2011)] proposes a practical method for salient object detection on rapid scenes. Selective visual attention is responsible for ordering visual information for recognizing a particular moving object even in the presence of complex background that are continuously changing. The author created a topographical master map by combining frame features in different scales. A dynamical neural network is chosen for getting most visual attention area thereby reducing saliency. This paper comes out with mainly three steps. In First stage, filter the most visible features in a scene throughout by using linear filter. In second stage, taking center surround difference with normalization procedure has been done. Finally, results are joined together by using accurate linear combination method. Each of the stages is carried out in multiple scales. The problem of salient object detection has been done in computationally efficient manner by analyzing the most salient region in each scene.

The main strength and limitations of the paper are, the performance is well in natural scenes, whereas for unimplemented feature type these methods may fail. The method quickly retrieves the salient object with verity of colors, textures, and shapes.

D. Regularized Feature Reconstruction for Spatio-temporal Saliency Detection.

In the paper [Zhixiang Ren *et al*, (2013)] has proposed a method for salient object detection frame work on the basis of regularized feature reconstruction on spatio temporal volume. Most of the existing approaches follow bottom-up visual information, whereas in this work, the author trying to give the best salient region by discarding the rest of the image regions. In videos, temporal information should be considered in addition to spatial information. For temporal saliency, the author modeled the movement of the best salient patch as reconstruction image sequence by considering the neighborhoods patches and coherent motion trajectories model has been found by Laplacian smoothing. For spatial saliency, most centre surround contrast region has been extracted by using same sparse reconstruction model. Finally, the spatial and temporal saliency are summed up and normalized to obtain the redundant visual information.

This work performs better by comparing all existing unsupervised methods. Regularized feature reconstruction method successfully worked on datasets of 1003 images.

E. Spatio-Temporal Saliency Based On Rare Model.

In the paper [Marc Décombas *et al*, (2008)] has proposed to model a system for salient object detection based on rare model. This work has been build up on the basis of RARE concept. Color, orientations, motion amplitude and directions have been taken as the feature for getting targeted salient object. There by adding dynamic features into existing static features. All features have been computed in multi scales, hence called multi scale rarity mechanism. Spatio temporal saliency map has been calculated separately in order to track the salient area. Temporal tracking has been also done here for better salient area detection. Feature extraction step is the first stages followed by multi scale rarity mechanism. Finally fusion is done by using linear combination on both spatial and temporal saliency. Some weight ages have been also added along with this fusion stage. After that, for obtaining correct visual information temporal tracking has been done.

This model has been efficiently computing in verity classes of videos, containing any type of random motion.

III. CONCLUSION

In this study we have found that there are various methods available currently that help in the detection of salient object based on visual attention. It has been noticed that while dealing with video sequences with different challenges and issues, but we may have to compromise on the salient object detection from a video model. Also it was observed that while dealing with object detection, background subtraction approach is need not be necessary. Bottom- up approach was the primary approach used for the salient object detection. Illumination conditions, speed of the

moving objects, background cluttering, color similarities etc are the main challenging issues come under the salient object detection in majority of the works. The major goal of the salient object detection is to extract the most redundant attention area.

Based on the study we are proposing a new method to detecting salient object from a sequence of image that would help in security surveillance, traffic control monitoring, museum monitoring, human action prediction, robotic applications, night time indoor/outdoor monitoring, smart room applications etc. In our method, a sequence of images collected from SLR/normal camera is taken as input. Temporal saliency will give the key idea for highest salient motion patches by considering neighboring frames, whereas the static saliency check for whether the target patch is necessary for the current frame or not. From the above study, local centre surround contrast has been successfully giving the accurate salient detection region. That can be easily measure by reconstruction error.

.

REFERENCES

[1] L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11) (1998) 1254–1259.

[2] J. Han, K.N. Ngan, M. Li, H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images", *IEEE Transactions on Circuits and Systems for Video Technology* 16 (1) (2006) 141–145.

[3] Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. "Learning realistic human actions from movies". In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[4] Navalpakkam and L. Itti. "An integrated model of top-down and bottom-up attention for optimizing detection speed". In *CVPR*, pages 2049–2056, 2006.

[5] Bugeau and P. Perez, "Detection and segmentation of moving objects in complex scenes," *Comput. Vision Image Understand.*, vol. 113, no.4, pp. 459–476, Apr. 2009.

[6] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 22, no.8, pp. 747–757, Aug. 2000.

[7] L. Yuan and H. Y. Shum, "Synthesizing dynamic texture with closedloop linear dynamic system," in *Proc. Eur. Conf. Comput. Vision*, 2004, pp. 603–616.

[8] H. Cheng, A. Tourapis, J. Llanch, and J. Boyce, "Adaptive spatiotemporal filtering for video denoising", *IEEE International Conference on Image Processing*, 2004.

[9] L. Guo, O. Au, M. Ma, and Z. Liang, "Temporal video denoising based on multihypothesis motion compensation," *IEEE Trans. on Circuits Syst. Video Technol.*, vol. 17, no. 10, pp. 1423-1429, Oct. 2007.

[10] D. G. Lowe. Distinctive image features from Scale- Invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

BHAGYASREE D: Recived The Bachlers Degree In Computer Science And engineering From Calicut University ,Kozhikode India, In 2012. She Is Currently Doing Masters Degree In Computer Vision And Image Processing From Amrita Vishwa Vidhyapeetham ,Coimbatore, Tamil Nadu, India.