International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August, 2019

Saliency Object Detection from Video Streams using Salient-Graph Model with High-Level Background Prior

Ruchi Kshatri, Kavita

Abstract*: This study proposes a novel salient graph model with high-level background prior. As usual, the collected data is pre-processed and then used for segmentation analysis. Object detection is still a daunting task due to increased complexity of false positive rate. Thus, a salient graph model is constructed using high-level background prior. Initially, the contrast of an image enhanced for superpixels and used for finding the shortest path of visible region. Then, saliency map is formed by spatial analysis of those visible superpixels. In salient post-processing, the salient graph is constructed by labelling background nodes with minimized cost. Based on formed salient region, each adjacent superpixel with background nodes are used for queries. Atlast, the estimated saliency and objectness measures detects the objects with minimal constraints. The proposed framework is analyzed on SegTrack and SegTrack 2, video segmentation dataset. The results states that the proposed method achieves better results than state of the art models by improved precision, recall, F-measure and computational time.

***Keywords* : Salient object detection, Background prior, Superpixels, Graph construction, Saliency measure and the salient graphs.**

I. INTRODUCTION

Recent developments in image processing technologies have greater impact over variant real-time applications. Human visual system plays a vital role in processing the relevant details, discarding the rest of images. The human eyes compute the most attractive regions and then passes on to visual cortex [1]. Detecting the objects from a scene is a critical task of image processing systems. To analyze and arrange the required details from different set of image analysis is a trivial task. The content of the image determines the importance of those images. The task of salient object detection is to find the appropriate information between foreground and background scene. This analysis differs from segmentation tasks like semantic segmentation models, thus, salient object detection [2] is very interesting and attractive. It serves as an important step to different sorts of object detections on video and image compression, object recognition, supervised segmentation, visual tracking, rendering services, information discovery and the image retrieval.

Saliency object detection is a process of achieving précised visual information from an image. It is done in two ways, namely, a) recognizing most salient objects and b) segmenting the most accurate regions. Hence, the models are designed on those kinds of applications. Mostly, precision-recall values [3] are estimated for area based models. Single (or) multi-objects detection models are designed for object detection systems. Most of the existing models segments the salient objects via prediction maps and the find the objects in a scene. Usually, the following are the metrics studied for better saliency detection models, a) Good detection: probability of higher false positive rate [4] i.e false identification of salient background; b) High resolution: formation of saliency maps should possess high resolution for locating the salient objects; c) Computational efficiency: from first step to last step, the models should detect the salient regions robustly.

The novelty of this study is as follows:

- Presented the significance and scope of the salient object detection systems from perspective of real-time challenges.
- Analyzed high-level background priors that resolved the gap between objectness measures and the salient object detection issues at some possible extent.

The three background priors, namely, contrast, associativity and the spatial distribution priors that leverage the detection results with respect to image boundary.

By formed saliency maps, the salient-graph construction process has better ability to deal with flexible size objects. The extraction of high-level features achieves better objectness measures.

The proposed study is arranged as follows: Prior techniques explored by other researchers explores in Section II. The proposed methodology presents in Section III. It's followed by experimental results in Section IV and concludes in Section V.

II. RELATED WORK

This section presents the existing algorithms suggested by other researchers. In [5], the authors studied deep construction module using background priors.

Revised Manuscript Received on October 30, 2019.
* Correspondence Author
Ruchi Kshatri *, Research Scholar, Jayoti Vidyapeeth Women's University, Jaipur, India. Email: ruchi.kshatri@gmail.com
Dr.Kavita, Associate Professor, Jayoti Vidyapeeth Women's University, Jaipur, India

At first, the background priors are modelled and then the objects are detected. Stacked denoising encoders using deep learning models ensured the better latent patterns of background priors. Finally, the deep autoencoders helped to resolve the separation of salient objects from background.

The training model of background prior consumes higher time. The author in [6] discussed the convex hull priors using graph regularized model. Convex hull is used for detecting the image center which makes better saliency of objects. Pairwise saliency energy function with graph regularization helps to separate the salient objects from background. The precision-recall value depicts higher computational complexity of the saliency mapping.

In [7], they discussed superpixel and background connectivity priors via object segmentation process. Linear iterative clustering algorithm was used for extracting the superpixels of an image. Each superpixel defined by spatial layout using image boundary. The saliency and the background values are connected using grabcut method. The specified grab region is further segmented. Runtime taken for saliency computation is higher. The author in [8] presented a sparse gradient based structure decomposition which utilized laplacian regularization for background priors. Over penalized issue of ranking function not resolved. Motion cues of optical field were studied by [9] that analyzed the motion history via objectiveness measure for background and the object regions. Saliency models with deep features are not detected properly ie. High false positive rate. The author in [10] studied background defocus in single portrait images. Acquiring depth information is a challenging task. Foreground details enhanced by gradient domain guided image filter and thus degraded background pixels. Convex posture of an image reduces the accuracy. In [11], they discussed about learning affinity via multi-scale graph of stating graph affinities model for accurate foreground extraction. Energy function on graph priors limits the error functionalities.

In [12], the author discussed Markov absorption probabilities [13] for region detection by weighted graph via partial image borders. Most of the markov models are enabled for building the hidden patterns of data. A reasonable amount of visual saliency models were designed using cognitive psychology [14]. Prior methods depend on local contrast priors and thus object detected with high contrast values. Then, a centred surrounded contrast is estimated using multi-scale differencing of Gaussian [15]. Different methods were suggested by using variant contrast information. Objects interior is finding by degrading the object contours and thus contrast based models were used. Thus, a global based contrast models were used for computing the frequency of the image. The object attenuation exists due to some restriction in local methods [17]. Though, it's a segmentation issue, the salient object detection without background priors are not possible. The computation of saliency maps using hierarchical representation in tree structure [18-21]. The inference of saliency can easily be achieved by belief propagation. Resolving three layers of hierarchical models using weight distribution leverages the single layer maps. Optimization of saliency values [22] of all superpixels that solves all salient criteria like visual rarity, center-bias and the mutual correlation were estimated. Study is further extended using quadratic programming models that influences the superpixels. Common salient objects can share by multiple salient information like similar visual appearances stated by [23]. Thus, analysis on such models needs a single input image.

III. PROPOSED METHODOLOGY

This section presents the proposed model of the research study. The review analysis states the significance of background priors to find the salient objects. The main aim of the study is to detect the salient objects from video streams input without compromising the accuracy of background priors. The framed objectives based on aim are as follows,

a) To enhance the accuracy of saliency objects from video streams
b) To maintain the integrity of salient information of video streams during background prior analysis.

The proposed, salient graph model with high-level background prior is explained as follows:

1.1 Data collection: It is the first step in the proposed model. The SegTrack and SegTrack V2 datasets are collected from [24]. The acquired dataset is prone to irrelevant noises and its being eliminated.

1.2 Image preprocessing: The acquired data is being pre-processed by median filtering that resolves the noise by non-linear method. The edge of a frame is maintained while removing the noises. Thus, it's employed by replacing the pixel value with median value of neighboring pixels. The filtered image is then converted grayscale image from RGB value 0-255 to grayscale of binary value 0 and 1. Then, by applying bounding box, the background of a scene is segmented

1.3 Image Segmentation: It is the most core part of the proposed model that composes of detecting the salient objects via saliency computation and the saliency post-processing.

1.3.1 Saliency Computation: Here, the high-level information of background analysis is done. In order to appropriately map the salient objects, the following processes are done.

a) Background contrast: It, simply, enhances the color quality of the background scene. Further, it helps to get the high-level priors. Average color and centroid are estimated for denoting the appearance and the location of superpixels. The arbitrate background priors (S_k) are estimated using mahalanobis distance of background cluster C_k, and the saliency map is given as:

$$S_{bcp} = \sum_{k=1}^{n} m_k S_k \qquad (3.1)$$

Where, m_k= Error Rate

$S_k = \sqrt{(C_k^{ij} - C`_k)O_k^{-1}(C_k^{ij} - C_k)}$ are then normalized into [0,1]. (3.2)

Retrieval Number F9023088619/2019©BEIESP
DOI: 10.35940/ijeat.F9023.088619
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

From this, the superpixels with same appearance are denoted as saliency.

b) Background associativity priors: It estimates the visibility region by shortest path cost to its image border. Mostly, geodesic and minimum barrier distances are used for associativity.

Here, a novel Robust MBD++ is designed for developing associativity priors. It derives better cost estimation and the background. Let μ(x) be the path between image boundary and the superpixels. The distance cost β(x) of robust MBD is measured as:

$$\beta(x) = \min(\beta(x), \theta(\mu(x))) \quad (3.3)$$

In general, the MBD takes the barrier strength of a path between two confined regions and makes the process very simple. In addition to this, we include regions based associativity which eradicates higher computational costs. The advantages of regions based robust MBD++ are: It takes only similar appearance superpixels that reduces the iteration complexity and it assists for background and foreground consistent saliency values. The result of β(x) is further normalized for forming saliency map.

c) Spatial Distribution Prior: It is performed on segmented image that exhibits a spatially distributed in an image. Initially, the associated superpixels are formed into m clusters. It measures the spatial variance between clusters that divides the Image I into n clusters, $I = \{m_1, m_2, ...m_n\}$. Here, pairwise mahalanobis distance between superpixels measured in color space is given as input to clusters. The spatial variance H_x of m_n is given as:

$$H_x(m_n) = \frac{1}{|m_n|}\sum_{p \in m_n} |p_x - R_x(m_n)| \quad (3.4)$$

Where,

$R_x(m_n) = \frac{1}{|m_n|}\sum_{p \in m_n} p_x$, px is the x-coordinate of superpixels p.

The spatial distribution saliency map is given as:

$$S_{sdp} = 1 - (H_x(m_n).V_y(m_n)) \quad (3.5)$$

Atlast, the high level features such as background contrast, background connectivity and spatial distribution are analyzed and it's further used for salient post-processing.

1.3.2 Saliency post- processing: It's the most core part of detecting the salient objects. It composes of three processes, explained as follows:

- Graph construction:
    - It defines the spatial adjacency between each pair of superpixels. Based on above information, each superpixel is treated as a node. Hence, it is represented as G= (V, E) where V defines the set of nodes with possible edge set E. Depends on spatial distribution priors, the weight between neighboring nodes p_i and p_j under affinity matrix is given as:

- $A_{ij}^f = \exp(-\frac{1}{\delta^2} d^f_{man}(p_i \cdot p_j))$ (3.6)
- Using the above eqn., the distance between two neighboring nodes are estimated. Likewise, the affinity matrix is constructed, the diagonal matrix that holds the aggregate of weights associated with each node and it's calculated by:

$$D_f = \mathrm{diag}\{d_{11}, d_{22}, ..d_{mm}) \quad (3.7)$$

- Developing boundary superpixels queries: Let G is the graph with N nodes. Here, background priors considered as non-salient, and thus, background node as query. Each node is formulated based on query. The ranking function is learnt from queries relevance of unlabelled and background nodes. The background nodes are then computed as superpixels based on queries.
- Saliency optimization: It refines the salient graph by its background nodes for detecting the objects. The basic idea of our proposed framework lies in integrating saliency maps generated by using different features in an ensemble manner. We thus extract color features and objectness features by considering saliency and objectness, respectively.

Furthermore, the saliency maps for different features are measured and thus obtain the final results as, object detected with fine-grain background priors. The below fig.3.1 explain the workflow of proposed system.

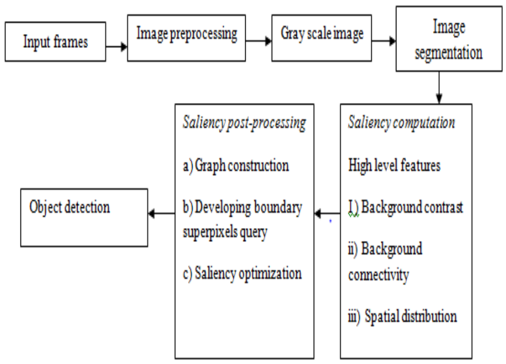

Fig 3.1. Proposed workflow

IV. EXPERIMENTAL RESULTS

This section explains experimental analysis of the proposed study. In our study, we analyze the simple video dataset collected from SegTrack and SegTrack2 [24]. It composes of six videos with pixel-wise ground truth, namely, birdfall, cheetah, girl and the monkeydog were taken from stationery camera. The table 4.1 &f fig.4.1 presents the sample information about SegTrack and SegTrack2.

Retrieval Number F9023088619/2019©BEIESP
DOI: 10.35940/ijeat.F9023.088619
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

TABLE 4.1 SAMPLE INFORMATION ABOUT DATASET

Images	No. Of frames	No. Of objects	Motion blur	Appearance change	Objects interaction	Complex deformation
Girl	21	1	0	0	0	0
Birdfall	30	1	0	0	0	0
Parachute	51	1	0	0	0	0
Monkeydog	71	2	0	0	0	0
Frog	279	2	0	0	0	0

Fig.4.1 Sample images of bird of paradise, birdfall, parachute, frog, girl, humming bird, monkey and monkeydog

The following are the metrics studied for analyzing the performance of the detection system:

- Precision: It is defined as the accurate analysis of an input image. It is observed from two cases, true positive and false positive. Thus, the precision is given as:

$$\text{Precision}=\frac{TP}{TP+FP} \quad (3.8)$$

Where, TP (True positive) correctly identified as foreground.

FP (False Positive) correctly identified as not foreground.

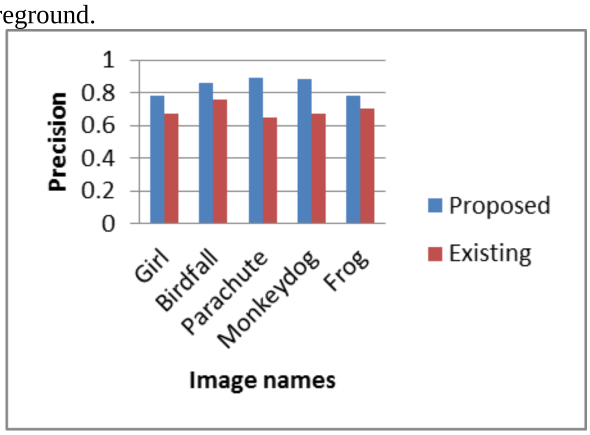

Fig.4.1. Precision analysis

- Recall: It is defined as the accurate salient analysis of an input image. In general, a restricted area bounds to be salient and thus its detection is of important. Thus, the recall is given as:

$$\text{Recall}=\frac{TP}{TP+FN} \quad (3.9)$$

Where, TP (True Positive) correctly identified as background.

FN (False Negative) correctly not identified as background.

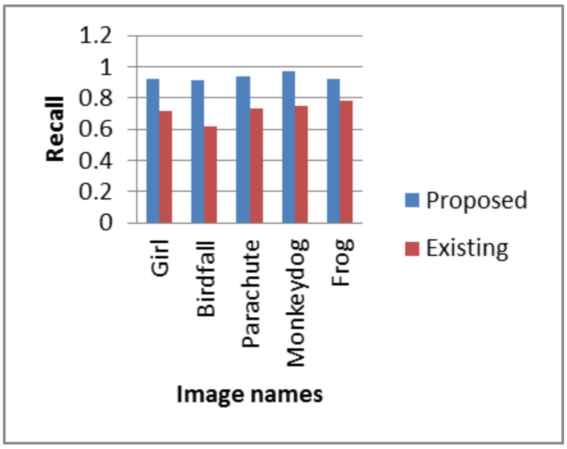

Fig.4.2 Recall analysis

- F-Measure: It measures from analysis of precision and recall. The salient post-processing analysis depicts how far the graph queries resolved for detecting the objects using saliency maps. It is measured as:

$$F(\alpha)=\frac{(1+\alpha^2)\text{PrecisionXRecall}}{\alpha^2(\text{Precision}+\text{Recall})} \quad (3.10)$$

Where, α dictates the weight rate and set to 1.

TABLE 4.2 PERFORMANCE ANALYSIS BETWEEN EXISTING AND PROPOSED SYSTEM

Image Name	Precision		Recall		F-measure	
	Proposed	Existing [25]	Proposed	Existing	Proposed	Existing
Girl	0.78	0.67	0.92	0.72	0.86	0.65
Birdfall	0.86	0.76	0.91	0.62	0.89	0.63
Parachute	0.89	0.65	0.94	0.73	0.88	0.69
Monkeydog	0.88	0.67	0.97	0.75	0.79	0.72
Frog	0.78	0.70	0.92	0.78	0.84	0.78

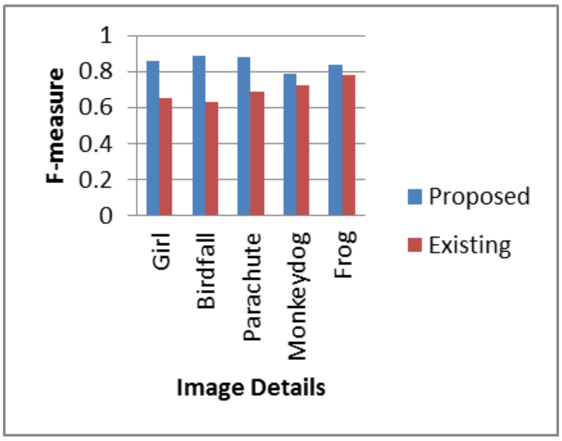

Fig.4.3 F-measure analysis

The table 4.2 presents the performance analysis between existing and proposed system. The proposed technique achieves optimized results due to high-level feature background analysis. Prior techniques on background priors treated the image boundary regions as background that leads to higher false positive rate.

Here, the superpixel processing on contrast, associativity and the spatial distribution helps to reduce the false positive rate with reduced computational costs. Likewise, the average time taken of each step in proposed method analysis defines the significance of the high-level background priors and given in table 4.3.

TABLE 4.3 AVERAGE TIME COST OF EACH STEP IN THE PROPOSED METHOD

Step	Pre-process	Robust MBD++	Graph model	Query analysis	Object detection
Time (s)	0.145	0.245	0.994	1.023	1.563

V. CONCLUSION

Presently, the advancements made in image analysis and interpretation techniques have attracted the researchers, to provide fast solutions to complex processes. Salient object detection is one of the fields that require robust and simple solutions for better visual information analysis. This paper suggests a better salient object detection models by analyzing high-level features of background priors. Firstly, the video segmentation dataset, SegTrack and SegTrack2 are collected which composes of natural and simple scenes objects like girl, waterfall etc. Median filtering is applied to remove the irrelevant noises of an image. Image boundary restricts the proper identification of the background priors and thus, color appearance information of superpixels is examined. The saliency computation and the saliency post-processing are the two vital steps of the proposed framework. The high level features like contrast, associativity and spatial distribution of the background priors are analyzed. By enhancing the appearance of image, the visibility region with respect to image boundary is computed and then location of each superpixel helps to form saliency maps. Based on formed saliency region, the adjacencies between neighboring superpixels are used for developing object related queries and thus objects are detected from salient information. This helps to achieve better detection rate in terms of precision, recall and F-measure. Compared to state of the art method, the high level features of background priors improve the speed and reduced computational cost.

REFERENCES

1. Q. Hou, M.-M. Cheng, X. Hu, A. Borji, Z. Tu, and P. Torr, "Deeply supervised salient object detection with short connections," in IEEE Conf. Comput. Vis. Pattern Recog., 2017.
2. C. Guo and L. Zhang, "A novel multiresolution spatiotemporalsaliency detection model and its applications in image and video compression," IEEE Trans. Image Process., vol. 19, no. 1, pp. 185–198, 2010.
3. J. Guo, T. Ren, L. Huang, X. Liu, M.-M. Cheng, and G. Wu, "Video salient object detection via cross-frame cellular automata," in Int.Conf. Multimedia and Expo, 2017, pp. 325–330.
4. M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in Int. Conf. Comput. Vis.,2009, pp. 817–824.
5. M.-M. Cheng, F.-L. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Repfinder: finding approximately repeated scene elements for image editing," in ACM Trans. Graph., vol. 29, no. 4, 2010, p. 83.
6. G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, "A shapepreserving approach to image resizing," Comput. Graph. Forum,vol. 28, no. 7, pp. 1897–1906, 2009.
7. U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottomup attention useful for object recognition?" in IEEE Conf. Comput.Vis. Pattern Recog., 2004.
8. Y. Wei, X. Liang, Y. Chen, X. Shen, M.-M. Cheng, Y. Zhao,and S. Yan, "Stc: A simple to complex framework for weaklysupervised semantic segmentation," IEEE Trans. Pattern Anal.Mach. Intell., vol. 39, no. 11, pp. 2314–2320, 2017.
9. Q. Hou, P. K. Dokania, D. Massiceti, Y. Wei, M.-M. Cheng, and P. H. S. Torr, "Mining pixels: Weakly supervised semantic segmentation using image labels," in EMMCVPR, 2017.
10. Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan, "Object region mining with adversarial erasing: A simple classification to semantic segmentation approach," in IEEE Conf. Comput.Vis. Pattern Recog., 2017.
11. Y. Wei, X. Liang, Y. Chen, Z. Jie, Y. Xiao, Y. Zhao, and S. Yan, "Learning to segment with image-level annotations," Pattern Recognition, vol. 59, pp. 234–244, 2016.
12. A. Borji, S. Frintrop, D. N. Sihite, and L. Itti, "Adaptive object tracking by learning background context," in IEEE Conf. Comput.Vis. Pattern Recog. Worksh., 2012.
13. P. L. Rosin and Y.-K. Lai, "Artistic minimal rendering with lines and blocks," Graphical Models, vol. 75, no. 4, pp. 208–229, 2013.
14. J. Han, E. J. Pauwels, and P. De Zeeuw, "Fast saliency-aware multi-modality image fusion," Neurocomputing, pp. 70–80, 2013.
15. T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: Internet image montage," ACM Trans. Graph., vol. 28, no. 5, pp. 124:1–10, 2009.
16. S.-M. Hu, T. Chen, K. Xu, M.-M. Cheng, and R. R. Martin,"Internet visual media processing: a survey with graphics and vision applications," The Vis. Comput., vol. 29, no. 5, pp. 393–405, 2013.
17. H. Liu, L. Zhang, and H. Huang, "Web-image driven best views of 3d shapes," The Vis. Comput., pp. 1–9, 2012.
18. J.-Y. Zhu, J. Wu, Y. Wei, E. Chang, and Z. Tu, "Unsupervised object class discovery via saliency-guided multiple class learning," in IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 3218–3225.
19. Y. Gao, M. Wang, D. Tao, R. Ji, and Q. Dai, "3-D object retrieval and recognition with hypergraph analysis," IEEE Trans. Image Process., vol. 21, no. 9, pp. 4290–4303, 2012.
20. M.-M. Cheng, Q.-B. Hou, S.-H. Zhang, and P. L. Rosin, "Intelligent visual media processing: When graphics meets vision," J. Comput. Sci. Tech., 2017.
21. A. Abdulmunem, Y.-K. Lai, and X. Sun, "Saliency guided local and global descriptors for effective action recognition," Computational Visual Media, vol. 2, no. 1, pp. 97–106, 2016.
22. L. Itti and C. Koch, "Computational modeling of visual attention," Nature reviews neuroscience, vol. 2, no. 3, pp. 194–203, 2001.
23. A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," IEEE Trans. Image Process., vol. 24, no. 12, pp. 5706–5722, 2015.
24. Dataset downloaded from: https://web.engr.oregonstate.edu/~lif/SegTrack2/dataset.html
25. Tao Xi. Wei Zhao., Han Wang & Weisi Lin, "Salient object detection with spatiotemporal background priors for video", IEEE transaction on Image Processing, 2016.