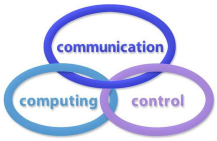
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 17, Issue: 2, Month: April, Year: 2022
Article Number: 4299, https://doi.org/10.15837/ijccc.2022.2.4299



CCC Publications

Video Saliency Detection by using an Enhance Methodology Involving a Combination of 3DCNN with Histograms

Suresh Kumar R., Mahalakshmi P., Jothilakshmi R., Kavitha M.S., Balamuralitharan S.

Suresh Kumar R
Centre for System Design,Department of ECE
Chennai Institute of Technology
Chennai, India
Corresponding author:sureshkumarrphd@yahoo.com

Mahalakshmi P
Department of CSE
SRM Institute of Science and Technology
Chennai, India

Jothilakshmi R
Department of IT, R.M.D. Engineering College
Chennai, India

Kavitha M S
Department of EEE
R.M.K. Engineering College
Chennai, India

Balamuralitharan S
Department of Mathematics
SRM Institute of Science and Technology
Chennai, India

Abstract

When watching pictures or videos, the Human Visual System has the potential to concentrate on important locations. Saliency detection is a tool for detecting the abnormality and randomness of images or videos by replicating the human visual system. Video saliency detection has received a lot of attention in recent decades, but due to challenging temporal abstraction and fusion for spatial saliency, computational modelling of spatial perception for video sequences is still limited.Unlike methods for detection of salient objects in still images, one of the most difficult aspects of video saliency detection is figuring out how to isolate and integrate spatial and temporal features.Saliency detection, which is basically a tool to recognize areas in images and videos that catch the attention of the human visual system, may benefit multimedia applications such as video or image retrieval, copy detection, and so on. As the two crucial steps in trajectory-based video classification methods are feature point identification and local feature extraction. We suggest a new spatio-temporal saliency detection using an enhanced 3D Conventional neural network with an inclusion of histogram for optical and orient gradient in this paper.

Keywords: Histogram of optical flow (HoF), Histogram of oriented gradient (HoG), Human Visual System (HVS), Saliency detection, salient object detection, salient region detection.

1 Introduction

In these images, human vision has the extraordinary ability to distinguish visually variable objects and regions. Human vision studies can help solve computer vision application issues. And scientists are intrigued by its ability to detect artefacts or regions that reflect a scene, a process known as saliency detection. The principle of saliency detection is depicted in Figure 1. The first column includes the original pictures. As shown in Figure, significant objects or sectors are identified and separated from the context in saliency detection. The human visual system acts as a filter, directing further focus to the most appealing and interesting regions or objects for further analysis. Humans may show visual fixation, which is the practice of keeping one's gaze fixed on a single point. Some visual saliency models are based on this visual perception phenomenon and aim to predict human fixations [1]. Furthermore, some visual saliency models are powered by computer vision applications and seek to classify the salient regions in a picture or video [2].

The automated measurement of significant regions and objects of images with no preceding inference or information is referred to as saliency detection. Saliency is typically defined as the difference between a pixel and its immediate surroundings. Visual models, strictly combinations of computational models are all used in saliency models. Purely computational models do not take into account the visual features of the human eye, while others do. Visual models, which can be loosely divided into two categories: attention models and salient region detection models, receive more attention. The former employs a selective visual attention system to dynamically sample the scene's most relevant visual content. These models develop a set of visual focus points that pinpoint the important items. The latter is responsible for detecting and segmenting the entire object or area.

The identification of saliency is a hot topic in computer vision. Intensive study has been undertaken out for salient object prediction to classify the most significant or visible object regions with the advancement of object-based computing vision applications [3][4][5]. These methods vary from early saliency estimation algorithms, which concentrated on positioning human eye fixations [6], in that they aim to highlight the entire salient object regions. In general, salient object tracking models produce large, well-connected salient areas. A video salient object detection system for locating primary salient objects in complex scenes in this paper. For each video frame, it generates a gray-scale saliency diagram, with brighter pixels indicating higher saliency values. Bottom-up mechanisms are used in most conventional salient object detection for images. Low-level features along with heuristics are used in these models, [7],[8]. While these bottom-up saliency models yield remarkable results, they are not without flaws. They don't produce clear saliency values for the entire salient object and the entire context, particularly when the object has multiple components or the context is cluttered. Bottom-up methods use pixels or super pixels as core elements to infer saliency, which causes this phenomenon. Working on the full object level will be more normal from the standpoint of human experience. Since pixel and superpixel level methodology lack object-level functionality, they are unable to fully achieve the goal of retrieving salient object regions. This problem, however, has yet to be tackled in traditional approaches. Visual focus is an important process in the Human Visual System (HVS), and has been

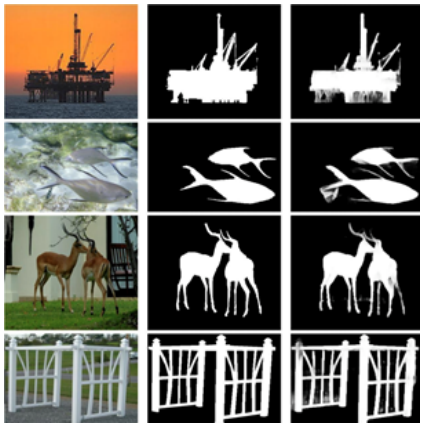


Figure 1: Salient object or region

extensively studied in visual perception studies and neuroscience. While observing at a visual scene, our visual focus system filters out the bulk of the irrelevant details and concentrates on the important bits. In general, there are two approaches to visual attention: top-down and bottom up, a task-driven mechanism also termed as Top-down attention, is dictated by specific prior experience, such as assignments, aspirations, and current objectives, while bottom-up attention, also known as a stimulus-driven methodology, determines salient regions spontaneously based on the feature distinction from low-level aspects such as, colour, and texture in visual scenes and luminance.

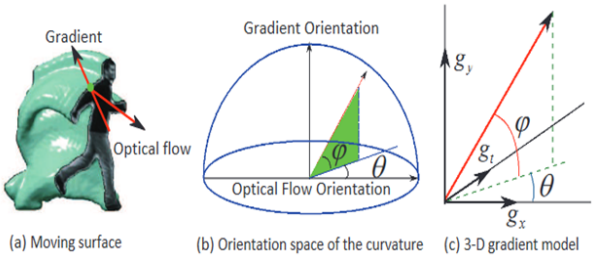


Figure 2: Motion Surface Capture

Saliency regions can be derived from visual sections using visual attention modelling. In general, an image pixel's saliency is well-defined as the likelihood of that pixel being beheld at. There are two types of saliency detection models currently available: eye fixation prediction and salient object detection models. Salient object predicting models focus in locating the entire salient objects in visual scenes, while human fascinationestimate models much focusfor locating fixation regions where human eyes search during scene viewing. Due to their broad applications, various types of saliency detection models have been investigated over the last decade [9][10][11][12]. Itti et al. suggested a multi-scale low-level function model of visual saliency based on colour, luminance, and orientation[13]. In the study [14], a biologically inspired framework was suggested using graph theory as a result of its work. The HVS seems to deal with goals from different scales, according to some psychological research [15,16]. Yan et al. [17], inspired by these studies developed a classified saliency estimate model to solve the issue of small-scale salient object discovery.

2 Existing Methods

2.1 Predictive video saliency method (PVSM):

Saliency, which is described as irregularity, unpredictability, rarity, or surprise, is a perceptual quality that distinguishes an item, individual, or pixel from its surroundings and thus attracts our attention.Saliency methods such as [18][19][20] are important for computer vision. There are two types of video detection methods. One considers optical flow only as a part of features before taking into account video saliency continuity between consecutive frames, whereas the other handles a video frame by frame without considering motion detail. The pitfall of this model lies at detection of still images.

2.2 Deep learning for video saliency detection (DVSD):

The proposed deep video saliency model is made up of two modules: a dynamic saliency network and a static saliency network, both of which are used to capture the spatial - spectral statistics of dynamic scenes. The static saliency network's saliency estimates are integrated into the self-motivated saliency network, allowing our method to learn how to fuse static and dynamic saliency detection automatically and generate final saliency results have spatiotemporal featureare much less computational load. Using video sequences or scene comprehension [21]inferring the important case. The concept of saliency has been broadened to include apprehending popular saliency among related videos and images [22]. However, the above approaches and conventional saliency detection have

significant variations. Saliency maps generated with high quality for still images on two data sets as DAVIS and FBMS. The failure of this approach lies at Optimal results need to be achieved.

2.3 Video Saliency Detection using object proposals (OVSD):

Identification of salient object regions in videos with an effective usage of object proposals. A proposal ranking with object candidates, and voting scheme is used to screen out non-salient pieces, pick object regions with salient features, and conclude an initial saliency approximationfounded on various saliency cues object-level.Saliency features related to visual tracking [23], center-surround contrast prior and both background prior are utilized for effective detection of video saliency [24][25]. Relatively than individually measuring saliency values are obtained at different levels, our approach used a more instinctual graphical saliency analysis to specifically locate candidate salient object proposals. Even in the presence of a cluttered backdrop, the proposed system is having maximum salient parts with complex motion patterns, compared to state-of-the-art methods. The drawback of the proposed method lies in not determining tasks relevant to stereo saliency.

2.4 Video saliency detection by gestalt theory (GVSD):

Based on Gestalt theory and cues with spatiotemporal nature, a novel framework for video saliency detection has been created. To calculate the feature contrast for spatial saliency estimation, extracted spatial features such as luminance, colour, and texture. Extracted motion features are determined by optical flow for temporal saliency prediction. The spatiotemporal saliency mapsare created by combining temporal and spatial saliency, as well as uncertainty weighting transversely modalities, as determined by Gestalt theory's two laws common fate of similarity. Yuan et al. developed a denser and sparse labelling system for saliency detection based on deep neural networks in [26]. Using a convolutional LSTM (Long Short-Term Memory) architecture, Wang et al. [27] proposed a video saliency prediction model. On three public databases, experimental findings resembles that the proposed video saliency detection model outclasses other existing models. The failure of this approach lies at exclusion of High-level spatial features as the approach was limited to low level spatial features.

Author	Contribution	Methodology	Advantages	Disadvantages
PVSM Qian Li et al. [28]	The saliency map within each frame can be determined by the outcome of the previous frame by creating correspondences using motion knowledge.	Optical flow features, frame by frame video processing	Frame-by-frame video processing is performed without taking into account motion information.	Does not work for still image saliency detection
DVSD Wenguan wang et al. [29]	deep video saliency training using pixel wise annotated video, saliency detection and training fast videos,	Conventional Neural Networks	Saliency maps generated with high quality for still images on two data sets as DAVIS and FBMS	Optimal results need to be achieved
OVSD Fang Guo et al. [30]	Various discriminative saliency cues and conventional saliency assumptions.	Cues with various saliency levels along with proposal ranking for object candidate and voting scheme	Even in the occurrence of a cluttered background, our system accurately located full salient parts with complex motion patterns.	Unable to detect tasks relevant to stereo saliency
(GVSD) Yuming Fang etal. [31]	Extraction of spatial features which include texture, color and luminance features in order to carry out computing contrast features.	Gestalt theory and cues with Spatiotemporal	Successful calculation of saliency map with high values compared to other existing methods	High level spatial features are not included as the approach was limited to low level spatial features

Table 1: Showing the comparison of existing methods

3 Architecture of the proposed model

Spatiotemporal and combining spatiotemporal features can be learnt using Conv3DNet and Deconv3DNet. Three successive video frames ((I_(t-1),I_t,I_(t+1)) loss propagation in forward direction can be measured in Conv3DNet using a classification model with a video frame (It) and point cloud map Gt. The pooling 3D layer, convolutional 3D layer and pooling 3D layer are denoted as k*k*d, which are considered in 3D convolutional layer with kernel/stride size. The output form of 3D convolutional and 3D de convolutional layers are denoted as f*h*w*c, where h represents the height, f as input video frames number, w and c as width and channels for the considered video frames.

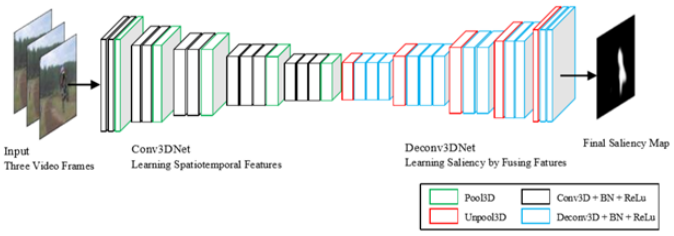


Figure 3: Architecture of the proposed video saliency detection model

Architecture of the proposed video saliency detection model
Input: Let ν,τ,λ,β, γ be the video, temporal window and parameters For the given video at each patch do
1: Temporal surroundings of the window are identified.
For
t$\varepsilon\Omega$
2: Temporal dictionary is generated by using all temporal patches
End for
3: temporal saliency is calculated
4: except the overlapped patch all other patches are used for building dictionary temporal
5: special saliency is calculated
End for
6: combine temporal and spatial saliency Sal =(sal)_T + (Sal)_S
Output: saliency maps for all frames in the video
Where temporal saliency is given as And spatial saliency is given as The weight inversely proportional

$$Sal_T(y_n) = \sum_{t\in\Omega} || y_n - D_{t,n}\hat{x}_{t,n} | \left| x^2 + \lambda \sum_{t\in\Omega} \left|\left|\hat{x}_{t,n}\right|\right|_p + \beta \sum_{j=1}^{N} ||\hat{X}_n - \hat{X}_j \right| |2W_{nj}$$

$$Sal_s\,(y_n) = || y_n - D_n||2 + \gamma\, ||\hat{X}_n||$$

to uncertainty since more uncertainty should be assigned a lower ultimate saliency prediction weight are set. To summarize, Ut denotes spatial saliency whereas Us denotes temporal saliency. Each video frame's spatiotemporal saliency map may be estimated with a combination of spatial and temporal saliency with uncertain grading as follows:

$$S = \frac{U^t S_s \ + U^s t}{U^t + U^s}$$

3.1 Conv3DNet

Five blocks of Conv3DNet, each with three 3D convolution layers and three 3D pooling layers, with Reluoperations andbatch normalization. each block in order to create feature maps with a 7 * 7 size, we construct five team blocks for the suggested model. The 3D convolutional procedure is shown in Fig. 3. The output extracted features will be bottom and sparse as a result of the convolutional and pooling processes' rapid pace. This is why we utilize Deconv3DNet to learn elevated spatial and temporal characteristics.Down sampling and sparsification occur as a result of convolutional and pooling processes. As a result, we utilize Deconv3DNet to teach ourselves about high-level temporal and spatial characteristics. the video frames (It-1, It, It+1)are sent into the Conv3DNet, which learns the coherence and motion between them. This information is crucial for video saliency recognition. After performing a convolutional operation, the resulting feature map Y is designated as follows:

$$Y = f(\Sigma w * X + b)$$

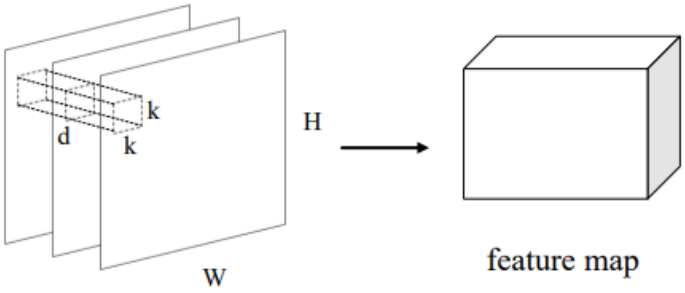


Figure 4: 3D convolutional operations

For a 3-D convolutional layer, the kernel is a cube of size (in depth, in spatial filter size) d * k * k. Feature maps are measured in terms of their width (W) and height, (H) respectively.

3.2 Deconv3DNet

The proposed Deconv3DNet uses 3D deconvolutional and unpooling operations to combine spatiotemporal data in video sequences and up-sample feature maps' resolutions. The Deconv3DNet we created has five group operations in it: three 3D deconvolutional layers, three 3D unpooling levels, and a set of four batch normalization layers. For the first and second groups, we offer a 3D deconvolutional and 3D unpooling layers with Relu operation and batch normalization. In the last three groups, we use a 3D convolution layers, two 3D unpooling and deconvolutional layers strides are set to 122 in order to increase the 3-D size of extracted features, whereas the temporal measurement is resolved to 1 because we want to determine the saliency map of a time with unitary frame.
To maintain as many high-level saliency signals as feasible, we use an additional 3D convolutional layer with a size of 224 224 1 when generating the final saliency map. Here, we utiliseRelu as the input signal of deconvolutional and convolutional layers. The squared Euclidean error is utilised as the gradient descent. With (I, G), we have a set of training samples consisting of three frames with the form h * w * 3 and the video frame's ground truth map Gt, which is denoted by (I, G),It.

4 Experimental setup

A Graphic User Interface is designed dedicatedly using Visual Studio IDE to gather the performance of existing method. Visual C++ is used to script the proposed method. Existing work implementations are fetched from GitHub code repository as Common Language Runtime (CLR) [32][33][34][35]Library. Implementation of the proposed method is done with 32-bit application compatibility to maintained uniformity of the existing methods. All methods are blended as x86 assembly CLR Runt time Library. DAVIS and SegTrackV2 datasets are used to evaluate the existing and proposed methods. Precision and F-Measure values are measured for existing and proposed methods with the above-mentioned datasets to make an upright comparison. Precision is measured for the corresponding recall values from 0 to 1 in 0.1 steps. Similarly, F-Measure values are calculated for the threshold values from 25 to 250 in step 25.

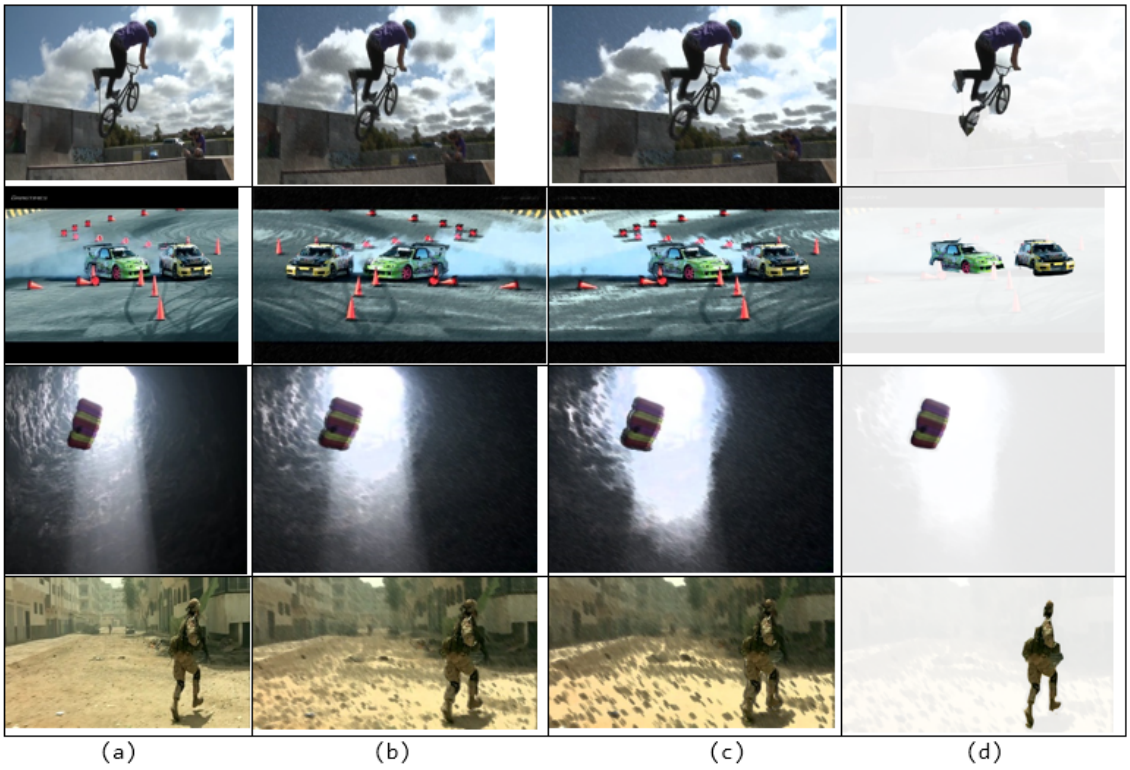


Figure 5: saliency map samples (a): input file (b): special saliency map (c): temporal saliency map (d): output fie

4.1 EXPERIMENTAL SETUP

4.1.1DAVIS (Densely Annotated VIdeo Segmentation): [36] consists of fifty full HD video sequences in high quality that span multiple instances of prevalent video object segmentation difficulties like occlusions, motion blur, and appearance changes. Each video is accompanied by ground truth segmentation that is densely annotated, pixel-accurate, and per-frame.

4.1.2 SegtrackV2: The SegTrackv2 dataset [37] is made up of 14 videos that have frame-by-frame ground-truth annotation. There are single and multiple objects, as well as slow and fast motion, bisecting and communicating objects.

5 Results

The results are obtained with the help of two parameters such as F-Measure and Precision which are obtained from two different considered dataset as DAVIS and SegtrackV2. Precisionis a metric for how accurate a classifier is. A higher precision indicates less false positives, whereas a lower precision

indicates a higher number of false positives. The correctness, or sensitivity, of a classifier is measured

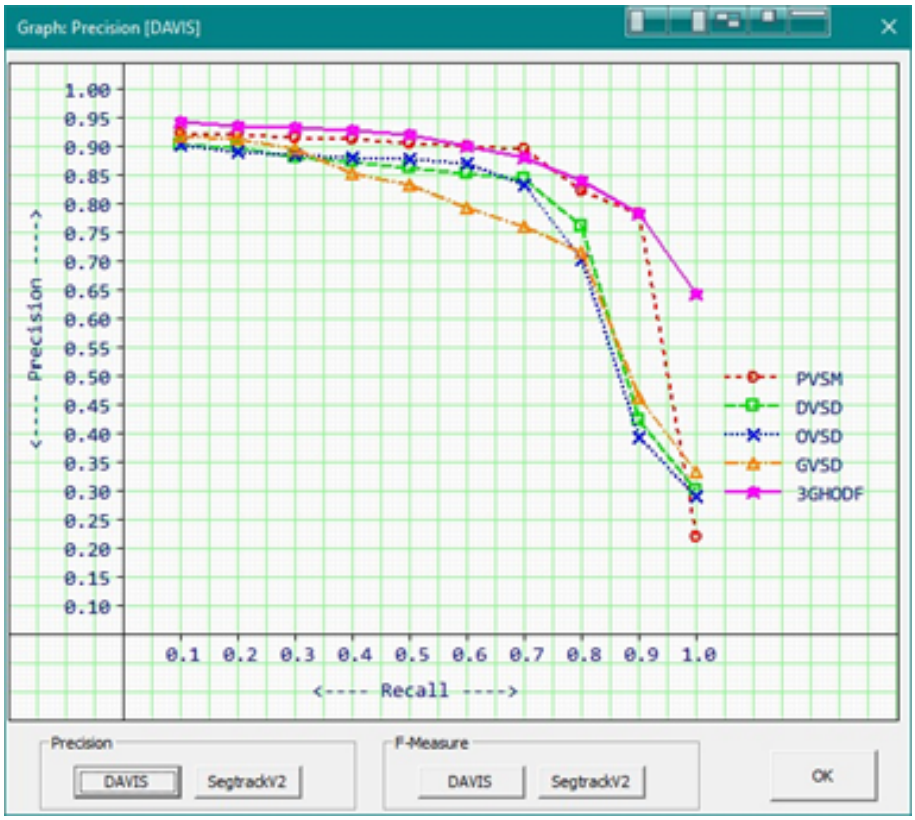


Figure 6: Obtained Precision values graph of proposed and existing methods From DAVIS Dataset

Precision [DAVIS]					
Recall	**PVSM**	**DVSD**	**OVSD**	**GVSD**	**3DHOGF**
0.1	0.9233	0.9036	0.9029	0.9193	0.9446
0.2	0.9201	0.8991	0.892	0.9138	0.9368
0.3	0.9161	0.8807	0.8851	0.8967	0.9332
0.4	0.9145	0.874	0.8803	0.8549	0.9294
0.5	0.9065	0.863	0.8777	0.8341	0.9206
0.6	0.9018	0.8545	0.8703	0.793	0.9007
0.7	0.8972	0.8435	0.8347	0.761	0.8822
0.8	0.823	0.7603	0.7026	0.7168	0.8421
0.9	0.7834	0.4228	0.3939	0.4633	0.7844
1	0.2208	0.301	0.2912	0.3332	0.6434

Table 2: Obtained Precision values of proposed and existing methods From DAVIS Dataset

by recall. The higher the recall, the less the false negatives, because when lower the recall, the falser negatives. High precision values refer to high reliability of the proposed system.

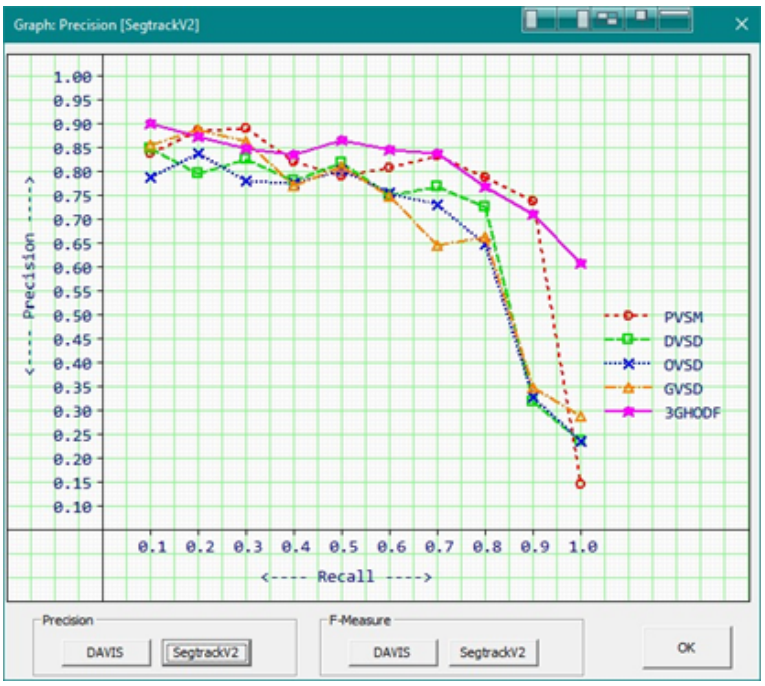


Figure 7: Obtained Precision values graph of proposed and existing methods from segtrackV2 Dataset

Precision [SegtrackV2]					
Recall	**PVSM**	**DVSD**	**OVSD**	**GVSD**	**3DHOGF**
0.1	0.8393	0.8496	0.7889	0.8553	0.9006
0.2	0.8861	0.7951	0.838	0.8898	0.8728
0.3	0.8921	0.8267	0.7811	0.8627	0.8492
0.4	0.8205	0.78	0.7763	0.7709	0.8354
0.5	0.7925	0.819	0.8037	0.8101	0.8666
0.6	0.8078	0.7505	0.7563	0.749	0.8467
0.7	0.8332	0.7695	0.7307	0.647	0.8382
0.8	0.789	0.7263	0.6486	0.6628	0.7681
0.9	0.7394	0.3188	0.3299	0.3493	0.7104
1	0.1468	0.237	0.2372	0.2892	0.6094

Table 3: Obtained Precision values of proposed and existing methods from SegtrackV2 Dataset

As it is observed from the precision values for both the data sets when our proposed system is compared with other existing state of art models for the considered two different datasets it is observed that higher precision values are obtained which indicates the reliability of the system.

F-Measure: The F measure is a metric for determining how accurate a test is. It's measured using the test's precision and recall.

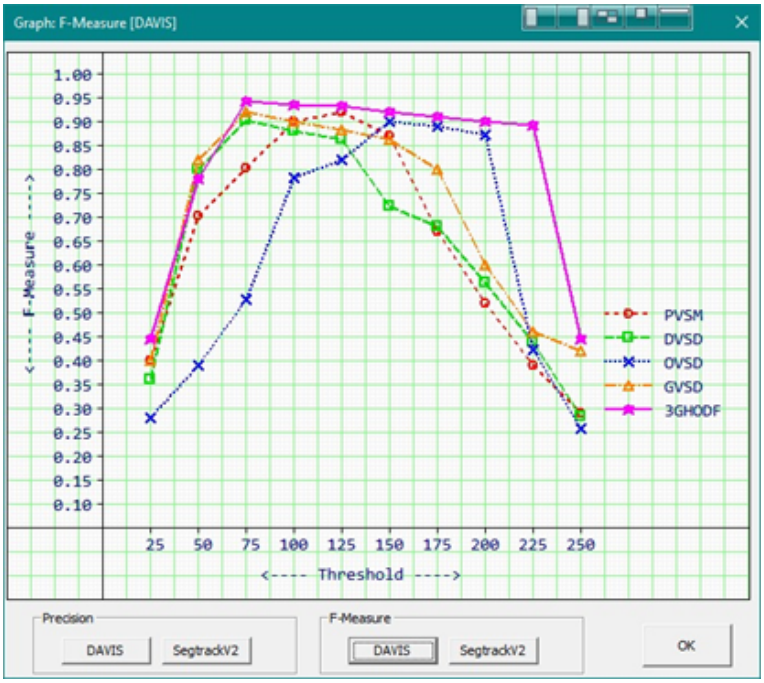


Figure 8: Obtained F-Measure values Graph for proposed and existing methods from DAVIS Dataset

F-Measure [DAVIS]					
Threshold	**PVSM**	**DVSD**	**OVSD**	**GVSD**	**3DHOGF**
25	0.4014	0.3616	0.2818	0.402	0.4466
50	0.7028	0.8022	0.3916	0.821	0.7804
75	0.8028	0.9049	0.5297	0.9208	0.9429
100	0.9022	0.8801	0.7847	0.9025	0.9372
125	0.92	0.8642	0.8216	0.884	0.9332
150	0.8705	0.7225	0.9014	0.8634	0.9222
175	0.67	0.6802	0.8902	0.8002	0.9102
200	0.521	0.5636	0.8743	0.6018	0.9025
225	0.3919	0.44	0.4231	0.4612	0.8942
250	0.2917	0.2842	0.2598	0.4201	0.4456

Table 4: Obtained F-Measures values of proposed and existing methods from DAVIS Dataset

with precision equaling the numbers of successfully predicted positive outcomes divided by the total number of positive results, including those that weren't, and sensitivity equaling the numbers of successfully predicted positive results divided by the total number of samples that should have been positive.

From both the datasets when our proposed system results are compared with the existing state of art methods it is found that our obtained results are more benefited and higher than the existing system there by indicating that the probability of getting high accurate reliable values are possible by our proposed system. Which indicates that in video saliency the classification of both temporal and special saliency can be achieve with correct values from the proposed system.

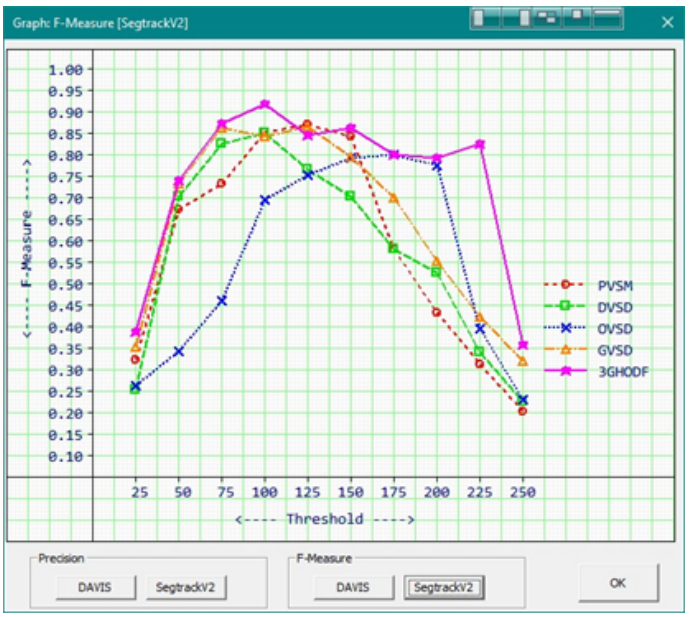


Figure 9: Obtained F-measured values graph of proposed and existing methods fromSegtrckV2 Dataset

F-Measure [SegtraclV2]					
Threshold	**PVSM**	**DVSD**	**OVSD**	**GVSD**	**3DHOGF**
25	0.3234	0.2536	0.2638	0.354	0.3886
50	0.6748	0.7042	0.3436	0.733	0.7424
75	0.7348	0.8269	0.4617	0.8628	0.8749
100	0.8542	0.8521	0.6967	0.8445	0.9192
125	0.872	0.7662	0.7536	0.866	0.8452
150	0.8425	0.7045	0.7934	0.7954	0.8642
175	0.582	0.5822	0.8022	0.7022	0.8022
200	0.433	0.5256	0.7763	0.5538	0.7945
225	0.3139	0.342	0.3951	0.4232	0.8262
250	0.2037	0.2262	0.2318	0.3221	0.3576

Table 5: Dataset character mentioned

6 Conclusion and future extension

For effective saliency detection in video, the results obtained from our proposed method when compared with the existing state of art method it is clearly noticed that the precision and F measure values which are obtained are better and fine when compared with the existing methods. According to the findings of the experiments, building a video saliency detection algorithm using 3D convolutional operations for learning spatio - temporal features is far more effective than using time-consuming hand-crafted characteristics to learn these features. An efficient method for obtaining spatiotemporal information between successive frames is the Conv3DNet (It-1, It, It+1)which consists of a collection of 3D convolutional layers. In addition to learning the ultimate spatiotemporal saliency map for video, the Deconv3DNet integrates the spatial and temporal characteristics from the Conv3DNet. This indicates that the classification for both the special and temporal features can be achieved with the proposed reliable system. In future we will extend the same patters on still image special and temporal saliency detection which is merely challenge task.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Conflict of interest

The authors declare no conflict of interest.

References

[1] A. Borji and L. Itti, State-of-the-art in visual attention mod-eling, *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no.1,pp.185–207,2013.*

[2] 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.*

[3] X. Shen and Y. Wu, A unified approach to salient object detection via low rank matrix recovery, *Proc. IEEE CVPR, Providence, RI, USA, 2012, pp. 853–860.*

[4] H. Kim, Y. Kim, J.-Y. Sim, and C.-S. Kim, Spatiotemporal saliency detection for video sequences based on random walk with restart, *IEEE Trans. Image Process., vol. 24, no. 8, pp. 2552–2564, Aug. 2015.*

[5] W. Wang, J. Shen, and L. Shao, Video salient object detection via fully convolutional networks, *IEEE Trans. Image Process., to be published, doi: 10.1109/TIP.2017.2754941.*

[6] J. Peng, J. Shen, and X. Li, High-order energies for stereo segmentation, *IEEE Trans. Cybern., vol. 46, no. 7, pp. 1616–1627, Jul. 2016.*

[7] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, Saliency filters: Contrast based filtering for salient region detection, *in Proc. IEEE CVPR, Providence, RI, USA, 2012, pp. 733–740.*

[8] .M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, Global contrast based salient region detection, *IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 569–582, Mar. 2015.*

[9] W. Wang, Q. Lai, H. Fu, J. Shen, H. Ling, Salient object detection in the deep learning era: an in-depth survey, *CoRR abs/1904.09146 (2019).*

[10] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Anal. Mach. Intell. 20 (11) (1998) 1254–1259.*

[11] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: *International Conference on Neural Information Processing Systems, 2006, pp. 545–552.*

[12] P. Zhang, T. Zhuo, W. Huang, K. Chen, M. Kankanhalli, Online object tracking based on CNN with spatial-temporal saliency guided sampling, *Neurocomputing 257 (2017) 115–127.*

[13] J. Zhang, K.A. Ehinger, H. Wei, K. Zhang, J. Yang, A novel graph-based optimization framework for salient object detection, *PatternRecognit. 64 (1) (2017) 39–50.*

[14] H. Chen, Y. Li, D. Su, Multi-modal fusion network with multi-scale multi– path and cross-modal interactions for RGB-D salient object detection, *Pattern Recognit. 1 (1) (2018).1–1.*

[15] E. Macaluso, C.D. Frith, J. Driver, Directing attention to locations and to sensory modalities: multiple levels of selective processing revealed with PET, *Cerebral Cortex 12 (4) (2002) 357–368.*

[16] T.S. Lee, D. Mumford, Hierarchical bayesian inference in the visual cortex, *JOSAA 20 (7) (2003) 1434–1448.*

[17] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, *in: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1155–1162.*

[18] . Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. *In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 1597–1604. IEEE (2009)*

[19] Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. *In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 409–416. IEEE (2011)*

[20] Cui, X., Liu, Q., Metaxas, D.: Temporal spectral residual: fast motion saliency detection. *In: Proceedings of the ACM International Conference on Multimedia (2009).*

[21] B. X. Nie, P. Wei, and S.-C. Zhu, Monocular 3D human pose estimation by predicting depth on joints. *In IEEE International Conference on Computer Vision, 2017*

[22] D. Zhang, J. Han, C. Li, J. Wang, and X. Li, Detection of co-salient objects by looking deep and wide, *International Journal of Computer Vision, vol. 120, no. 2, pp. 215–232, 2016.*

[23] X. Dong et al., Occlusion-aware real-time object tracking, *IEEE Trans. Multimedia, vol. 19, no. 4, pp. 763–771, Apr. 2017.*

[24] X. Dong, J. Shen, L. Shao, and L. Van Gool, Sub-Markov random walk for image segmentation, *IEEE Trans. Image Process., vol. 25, no. 2, pp. 516–527, Feb. 2016.*

[25] J. Shen et al., Real-time superpixel segmentation by DBSCAN clustering algorithm, *IEEE Trans. Image Process., vol. 25, no. 12, pp. 5933–5942, Dec. 2016.*

[26] Y. Yuan, C. Li, J. Kim, W. Cai, D.D. Feng, Dense and sparse labeling with multidimensional features for saliency detection, *IEEE Trans. Circuits Syst. Video Technol. 28 (5) (2018) 1130–1143.*

[27] W. Wang, J. Shen, F. Guo, M.-M. Cheng, A. Borji, Revisiting video saliency: a large-scale benchmark and a new model, *In IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4894–4903.*

[28] Li Q., Chen S., Zhang B. (2012) Predictive Video Saliency Detection. *In: Liu CL., Zhang C., Wang L. (eds) Pattern Recognition. CCPR 2012. Communications in Computer and Information Science, vol 321. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-33506-8_23.*

[29] Wang, Wenguan et al. Deep Learning For Video Saliency Detection. *ArXiv abs/1702. 00871 (2017)*

[30] F. Guo et al., Video Saliency Detection Using Object Proposals, *In IEEE Transactions on Cybernetics, vol. 48, no. 11, pp. 3159-3170, Nov. 2018, doi: 10.1109/TCYB.2017.2761361.*

[31] Karthik, A., MazherIqbal, J.L. Efficient Speech Enhancement Using Recurrent Convolution Encoder and Decoder. *Wireless Pers Commun 119, 1959–1973 (2021).*

[32] Yuming Fang, Xiaoqiang Zhang, Feiniu Yuan, NevrezImamoglu, Haiwen Liu, Video saliency detection by gestalt theory, *Pattern Recognition, Volume 96,2019,106987, ISSN 0031-3203.*

[33] https://docs.microsoft.com/en-us/cpp/build/reference/clr common language runtime compilation? View = msvc-160

[34] https://docs.microsoft.com/en-us/cpp/dotnet/walkthrough-compiling-a-cpp-program-that-targets-the-clr-in-visual-studio?view=msvc-160

[35] https://en.wikipedia.org/wiki/Common_Language_Runtime

[36] https://www.red-gate.com/simple-talk/dotnet/net-development/creating-ccli-wrapper/

[37] Wang, Bofei et al., Object-based Spatial Similarity for Semi-supervised Video Object Segmentation. (2019).

[38] Li F., Kim T., Humayun A., Tsai D., Rehg J. M., Video Segmentation byTracking Many Figure-Ground Segments, *In:IEEE International Conference onComputer Vision (ICCV), 2013.*

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Suresh Kumar R.; Mahalakshmi P.; Jothilakshmi R.; Kavitha M.S.; Balamuralitharan S.(2022). Video Saliency Detection by using an Enhance Methodology Involving a Combination of 3DCNN with Histograms, *International Journal of Computers Communications & Control*, 17(2), 4299, 2022.
https://doi.org/10.15837/ijccc.2022.2.4299