NOVATEUR PUBLICATIONS
INTERNATIONAL JOURNAL OF INNOVATIONS IN ENGINEERING RESEARCH AND TECHNOLOGY [IJIERT]
ISSN: 2394-3696
VOLUME 3, ISSUE 8, Aug.-2016

# DETECTING SALIENT OBJECTS IN A VIDEO'S BY USING SPATIO-TEMPORAL SALIENCY & COLOUR MAP

KHADAKE SUHAS .B
*Department of Electronics Engineering,*
*Bharatratna Indira Gandhi College of Engineering,*
*Kegaon, Solapur, India * suhaskhadake@gmail.com*

## ABSTRACT

When we watch a video, a detection of salient object is not a simple task, because in background many other processes are performed. So for detection of salient object we design a novel method, which uses the particle filters. In this new approach, spatial temporal saliency map & colour feature are used this will quickly recover from false detection. Actually the proposed method is based on the local feature comparing with dominant feature present in the frame.
Above method is a marked object of salient, where there is a large difference between local and dominant features. Here we have to use two types of saliency. First saliency method is a spatio saliency which uses hue & saturation feature and second method called as motion saliency which uses optical flow vector feature.
So in the above proposed saliency technique the updation of motion Saliency & spatial saliency will be considered in an every iteration, so complexity of detection of salient object is reduced.

**INDEX TERMS:** HVS, pixel-level, phase spectrum, resolution, Gaussian, Particle filter, Spatio-Temporal saliency.

## INTRODUCTION

Human 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.* [3] (see Fig. 1). This model is an implementation of earlier general computational frameworks and psychological theories of bottom-up attention based on center-surround mechanisms. The distinguishing aspect between image and video saliency is the temporal motion information in the latter that introduces the notion of the first step of a salient motion detection algorithm is the foreground-background segmentation and subtraction step.

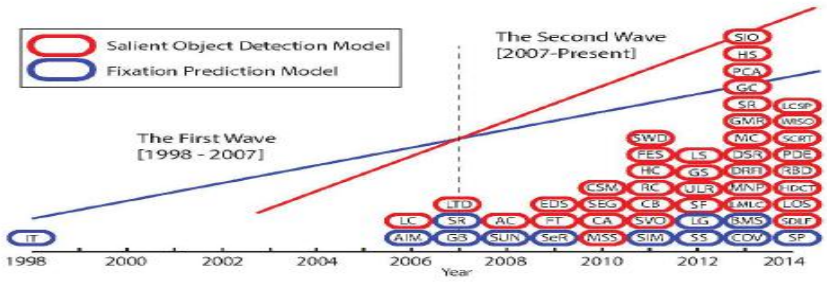


Fig. 1. A simplified chronicle of saliency modeling. Models in the first wave (1998-2007) are mainly dealing with fixation prediction while models in the seconds wave (2007-now) mainly addressed the detection and segmentation of the most salient objects. While both trends are still active research areas in computer vision and cognitive science, salient object detection has attracted more interest recently.

To segment out the foreground from background there are lot of techniques available. The pixels intensity values are added with the colour illumination into the algorithm, so it is easy to improve the image subtraction techniques. The background subtraction model in [4] models each pixel as a mixture of Gaussians. Mixture of several Gaussians is preferred to model the background. The model is updated regularly and evaluated to determine which of the Gaussians are sampled from the background process. When changes in the background are too fast then the variance of the Gaussians becomes too large and non parametric approaches are more suited. The concept of observably from the output pixels is provides a good clue to the salient motion detection in natural videos.

## LITERATURE REVIEW

Salient object detection in videos is challenging because of the competing motion in the background, resulting from camera tracking an object of interest, or motion of objects in the foreground. The authors present a fast method to detect salient video objects using particle filters, which are guided by spatio-temporal saliency maps and colour feature with the ability to quickly recover from false detections. The proposed method for generating spatial and motion saliency maps is based on comparing local features with dominant features present in the frame[1]. A region is marked salient if there is a large difference between local and dominant features. For spatial saliency, hue and saturation features are used, while for motion saliency, optical flow vectors are used as features. Experimental results on standard datasets for video segmentation and for saliency detection show superior performance over state-of-the-art methods.

A new method for automatic Salient object segmentation is an important research area in the field of object recognition, image retrieval, image editing, scene reconstruction, and 2D/3D conversion. In this work, salient object segmentation is performed using saliency map and colour segmentation. Edge, colour and intensity feature are extracted from mean shift segmentation (MSS) image, and saliency map is created using these features [2]. First average saliency per segment image is calculated using the colour information from MSS image and generated saliency map. Then, second average saliency per segment image is calculated by applying same procedure for the first image to the thresholding, labelling, and hole-filling applied image. Above are applied to the mean image of the generated two images to get the final salient object segmentation. The effectiveness of proposed method is proved by showing

80%, 89% and 80% of precision, recall and F-measure values from the generated salient object segmentation image and ground truth image.

A dynamical neural network then selects attended locations in order of decreasing saliency. The system breaks down the complex problem of scene understanding by rapidly selecting, in a computationally efficient manner, conspicuous locations to be analyzed in detail [3]. We propose a principled approach to summarization of visual data (images or video) based on optimization of a well-defined similarity measure. The problem we consider is re-targeting (or summarization) of image/video data into smaller sizes. A good "visual summary" should satisfy two properties: (1) it should contain as much as possible visual information from the input data; (2) it should introduce as few as possible new visual artifacts that were not in the input data (i.e., preserve visual coherence). We propose a bi-directional similarity measure which quantitatively captures these two requirements: Two signals S and T are considered visually similar if all patches of S (at multiple scales) are contained in T, and vice versa. The problem of summarization/re-targeting is posed as an optimization problem of this bi-directional similarity measure. We show summarization results for image and video data. We further show that the same approach can be used to address a variety of other problems, including automatic cropping, completion and synthesis of visual data, image collage, object removal, photo reshuffling and more[4] However, computational modeling of this basic intelligent behavior still remains a challenge. This paper presents a simple method for the visual saliency detection. Our model is independent of features, categories, or other forms of prior knowledge of the objects. By analyzing the log-spectrum of an input image, we extract the spectral residual of an image in spectral domain, and propose a fast method to construct the corresponding saliency map in spatial domain. We test this model on both natural pictures and artificial images such as psychological patterns. The results indicate fast and robust saliency detection of our method [5].a Salient image regions permit non-uniform allocation of computational resources. The selection of a commensurate set of salient regions is often a step taken in the initial stages of many computer vision algorithms, thereby facilitating object recognition, visual search and image matching. In this study, the authors survey the role and advancement of saliency algorithms over the past decade. The authors first offer a concise introduction to saliency. Next, the authors present a summary of saliency literature cast into their respective categories then further differentiated by their domains, computational methods, features, context and use of scale. The authors then discuss the achievements and limitations of the current state of the art. This information is augmented by an outline of the datasets and performance measures utilized as well as the computational techniques pervasive in the literature [6]. A Primate demonstrates unparalleled ability at rapidly orienting towards important events in complex dynamic environments. During rapid guidance of attention and gaze towards potential objects of interest or threats, often there is no time for detailed visual analysis. Thus, heuristic computations are necessary to locate the most interesting events in quasi real-time. We present a new theory of sensory surprise, which provides a principled and computable shortcut to important information. We develop a model that computes instantaneous low-level surprise at every location in video streams. The algorithm significantly correlates with eye movements of two humans watching complex video clips, including television programs (17,936 frames, 2,152 saccadic gaze shifts). The system allows more sophisticated and time-consuming image analysis to be efficiently focused onto the most surprising subsets of the incoming data [7]. A spatiotemporal saliency algorithm based on a centre-surround framework is proposed. The algorithm is inspired by biological mechanisms of motion-based perceptual grouping and extends a discriminant formulation of center-surround saliency previously proposed for static imagery. Under this formulation, the saliency of a location is equated to the power of a predefined set of features to discriminate

between the visual stimuli in a centre and a surround window, centred at that location. The features are spatiotemporal video patches and are modelled as dynamic textures, to achieve a principled joint characterization of the spatial and temporal components of saliency. The combination of discriminant centre-surround saliency with the modelling power of dynamic textures yields a robust, versatile, and fully unsupervised spatio temporal saliency algorithm, applicable to scenes with highly dynamic backgrounds and moving cameras. The related problem of background subtraction is treated as the complement of saliency detection, by classifying non salient (with respect to appearance and motion dynamics) points in the visual field as background. The algorithm is tested for background subtraction on challenging sequences, and shown to substantially outperform various state-of-the-art techniques. Quantitatively, its average error rate is almost half that of the closest competitor [8].

## METHODOLOGY

We employ a particle filter for its ability to approximate the posterior distribution of a system based on a finite set of weighted samples. Particle filters are also highly robust to partial occlusion and computationally inexpensive. The weight of each particle is initialized using a uniform distribution. The first set of particles is initialized around the centre of the first frame of the video. Weights are subsequently calculated as the weighted sum of distance measures of the candidate regions to the reference distribution. The spatio-temporal saliency and the colour maps are used to calculate the weight of the samples, which allows subsequent iterations to move the particles closer to the most salient object. In the proposed framework, we detect only one object of interest. Fig. 2 illustrates a block diagram of how the particle filter framework is used to detect the salient object. Colour versions of all the Figures used in this are available online.

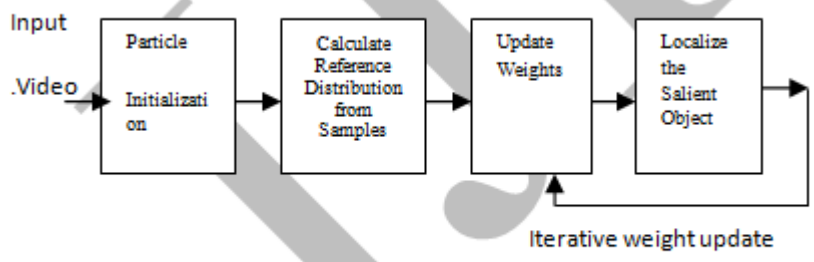


Fig.2. Proposed System Block Diagram

### 3.1.1 PARTICLE FILTER INITIALIZATION

Unlike a tracking framework, where the filter is initialized with a bounding box containing the object to be tracked, our method initializes the first set of particles at the centre of the frame, with each particle carrying the same weight. The initial samples are obtained within an ellipse whose two axes are set to half the width and height of the frame, respectively.Fig.3 shows the particle initialization and the final result of salient object detection. Fig. 3a shows the frame in which the particles are initialized. Once the particles are initialized, the subsequent steps iteratively.

Fig 3(a) Particle initialization

Fig3 (b) Salient object detection

### 3.1.2 REFERENCE DISTRIBUTION

In this section, the proposed approach utilizes colour & spatio temporal map as feature to detect salient map. we initialize the weight of each particle using uniform transition. the first set of particles is initialized around the centre of first frame of the video & provide a brief review to particle filters. A particle filter is a sequential Monte Carlo method that recursively approximates the posterior distribution from a finite set of weighted samples:

$$\{x_t^i, w_t^i\}_{i=1,\ldots,N},$$

where each sample xit represents a hypothetical state of the system with a corresponding weight wit . Considering that we have the observations Yt = y0… yt of the system up to time t, the goal is to estimate the state of the system xt from the posterior distribution p(xt|Yt).

### 3.1.3 UPDATION OF WEIGHT

The state-space model of the system can be represented as

$$xt = Axt{-}1 + ht{-}1yt = Cxt + et \quad (1)$$

Where A and C are the state transition and measurement matrices and ηt–1 and εt are the system and measurement noises. Similar to any linear Bayesian technique, a particle filter employs predict and update approach. During the prediction step, the posterior probability density at time t is calculated using the state transition model.

$$p(xt\,|Yt{-}1) = p(xt\,|xt{-}1)p(xt{-}1\,|Yt{-}1)\,dxt{-}1 \quad (2)$$

$$xt = Axt{-}1 + ht{-}1$$

$$yt = Cxt + et$$

Where A and C are the state transition and measurement matrices and ηt–1 and εt are the system and measurement noises. Similar to any linear Bayesian technique, a particle filter employs predict and update approach. A particle filter is driven by the state vector and the dynamic model of the system., we sample an ellipse whose state is defined as x = [x, y, ˙x, ˙y, Hx, Hy, ˙a], where x, y provides the location of the ellipse, ˙x, ˙y are the velocity components, Hx, Hy are the length of the half axes of the ellipse and ˙a indicates the scale change. The state propagation is done using first-order auto-regressive (AR) process as xt = Axt–1 + ηt–1.

The state transition matrix used in the dynamic model is a constant velocity and scale model. The observation likelihood for each sample is updated using (8). Once the weights of the samples are updated, the mean state or the location of the salient object is calculated as

$$E[w_t] = \sum_{n=1}^{N} w_t^n x_t^n$$

The location of the ellipse in frame t is updated using the state information calculated from [9]. The top row of Fig. 4 shows a tracking shot of a lynx running on a snow covered region. Even though the background consists of large motion, the spatio-temporal saliency map (to be described in the next section) is able to clearly demarcate the salient object allowing the particles to converge and track the lynx reliably. The bottom row is also an example of a tracking shot of a skier in action. The shot demonstrates the ability of the framework to detect different salient objects at different times. Initially, fig.3 a, 3b,3c, are the spatio saliency map of different objects & fig 3d,3e,3f are the fig. of object detection. The weights of the particles are reinitialized every t frames in order to avoid fixating on a particular object for the entire duration of the video sequence. In our experiments, we set t = F/2, where F is the frame rate of the video. We provide another example in Section 5 describing how the proposed method might erroneously detect the wrong object as salient, but with particle weight reinitialization, there is quick recovery from the error. Different colour models that are obtained from each cluster and whose weights are calculated based on the colour composition of the segment. The salient object detection approach presented assumed the availability of a spatio-temporal saliency map. In the next section, we describe how such a map can be obtained at the pixel level with computational efficiency. It can be noted that the proposed model can be extended to detect multiple objects of interest by clustering and segmenting the video frame based on the spatio-temporal saliency values. This would allow pixels with similar saliency measures and those that are spatially close to be clustered together. The cluster, which represents a region object, is associated with a rank based on the average saliency value of the cluster. Multiple particle trackers can be initialized around each of the cluster centroids as see points, thus, building a framework that would consider multiple hypothesis based on different colour models that are obtained from each cluster and whose weights are calculated based on the colour composition of the segment. The salient object detection approach presented assumed the availability of a spatio-temporal saliency map. In the next section, we describe how such a map can be obtained at the pixel level with computational efficiency. Spatio-temporal saliency is an important method for detecting salient objects in a video.

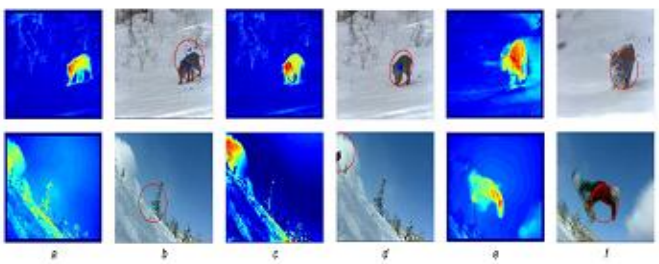

Fig 4. Results of salient object detection (dots indicate the location of the particles, the ellipse indicates the salient object)

a, c, e Spatio-temporal saliency mapsb, d, f Object detection

### 3.1.4 LOCALIZING SALIENCY OBJECTS

Salient region detection approaches in [3–5, 17] measure saliency at the patch-level resulting in saliency maps at reduced resolutions. 3 types of saliency map

1 Motion saliency map

2 Spatial saliency maps

3 Spatio-temporal saliency maps

Fig. 5 shows frames from three different videos and their corresponding spatial, motion and spatio-temporal saliency maps. The top row shows a frame from a video of a sauntering rhinoceros, shot with a stationary camera while the middle and bottom rows are videos of a wolf roaming in the forest and a skier performing a stunt, respectively. The last two videos are of tracking shots. Similar to [7], we adopt the Kullback–Leibler (KL) divergence and area under receiver operating characteristic (ROC) curve to evaluate the performance of the spatio-temporal saliency maps. KL divergence measures the correspondence of the salient regions to human saccade positions; larger the measure, the closer is the proposed saliency model to the human attention mechanism. We adopt the method followed in [8] wherein the saliency value that corresponds to a human saccade position is computed as the maximum of the human saccade positions over a circle of diameter 128 pixels, centered at the human saccade position. The saliency values collected over the entire database are discretized into 10 bins which are subsequently normalized to obtain the probability. Distribution P.The distribution for Q is calculated in a similar manner from spatio-temporal saliency maps. As the positions are sampled randomly, we repeat the experiment 100 times to obtain a fair evaluation of the performance. Particle filters implement the prediction-updating transitions of the filtering equation directly by using a genetic type mutation-selection particle algorithm. The samples from the distribution are represented by a set of particles; each particle has a likelihood weight assigned to it that represents the probability of that particle being sampled from the probability density function. Weight disparity leading to weight collapse is a common issue encountered in these filtering algorithms; however it can be mitigated by including a resampling step before the weights become too uneven. Several adaptive resampling criteria can be used, including the variance of the weights and the relative entropy w.r.t. the uniform distribution.

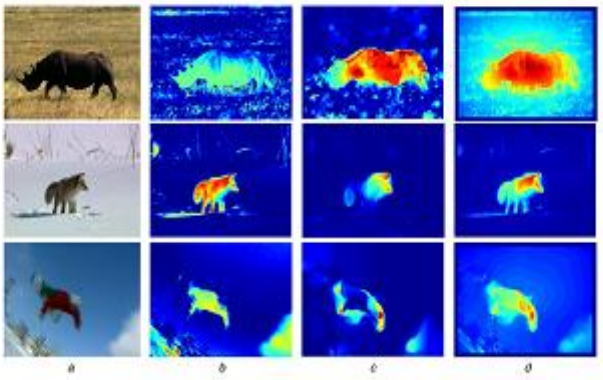

Fig.5 Saliency maps

a Original frame
b Spatial saliency map
c Motion saliency map
d Spatio-temporal saliency map

## CONCLUSION

This method followed wherein the saliency value that corresponds to a human saccade position is computed as the maximum of the human saccade positions over a circle of diameter 128 pixels, centered at the human saccade position. The saliency values collected over the entire database are discredited into 10 bins which is subsequently normalized to obtain the probability distribution P. The distribution for Q is calculated in a similar manner from spatio-temporal saliency maps. As the positions are sampled randomly, we repeat the experiment.

We compare the speed of the proposed spatio-temporal saliency framework with that of the saliency frameworks compared above, based on the number of frames that are processed per second. Although the proposed method is not as fast as PS, the quality of the saliency map is much higher. Detecting salient objects need generation of saliency map is very important because use of saliency map gives correct identification of the saliencies in a video.

## REFERENCES

1) Karthik Muthuswamy, Deepu Rajan," *Particle filter framework for salient object detection in videos". Centre for Multimedia and Network Technology, School of Computer Engineering,* Nanyang Technological University, 50 Nanyang Avenue,N4-02C-92 639798, Singapore

2) Han, S.-H., Jung, G.-D., Lee, S.-Y., Hong, Y.-P., Lee, S.-H.: '*Automatic salient object segmentation using saliency map and color segmentation', J. Central South Univ., 2013, 20, (9), pp. 2407–2413.*

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

4) Simakov, D., Caspi, Y., Shechtman, E., Irani, and M.: '*Summarizing visual data using bidirectional similarity'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008.*

5) Hou, X., Zhang, L.: *'Saliency detection: a spectral residual approach'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 2007*

6) Duncan, K., Sarkar, S.: *'Saliency in images and video: a brief survey', IET Comput. Vis., 2012, 6, (6), pp. 514–523*

7) Itti, L., Baldi, P.: '*A principled approach to detecting surprising events in video'.Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005, pp. 631–637*

8) Mahadevan, V., Vasconcelos, N.: '*Spatiotemporal saliency in dynamic scenes',IEEE Trans. Pattern Anal. Mach. Intell., 32, (1), pp. 171–177*

9) Gopalakrishnan, V., Hu, Y., Rajan, and D. *'Sustained observability for salient motion detection'.Proc. 10th Asian Conf. on Computer Vision, Queenstown, New Zealand, 2010, pp. 732–743*

10) Muthuswamy, K., Rajan, D.:' *Salient motion detection through state controllability' .Proc. IEEE Int. Conf on Acoustics, Speech and Signal Processing, Kyoto, Japan, 2012, pp. 1465–1468*

11) Xia, Y., Hu, R., Wang, Z.: *'Salient map extraction based on motion history map'. Proc. 4th Int, Congress on Image and Signal Processing, Shanghai, China, 2011, pp. 427–430*

12) Luo, Y., Tian, Q.: '*Spatio-temporal enhanced sparse feature selection for video saliency estimation'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, 2012, pp. 33–38*

13) Ren, Z., Gao, S., Chia, L., Rajan, D.: '*Regularized feature reconstruction for spatio-temporal saliency detection', IEEE T. Image Process., 2013, 22, (8), pp. 3120–3132*

14) Li, Y., Zhou, Y., Xu, L., Yang, X., Yang, J.: *'Incremental sparse saliency detection'. Proc. 16th IEEE Int. Conf. on Image Processing, Cairo, Egypt, 2009, pp. 3093–3096*

15) Guo, C., Ma, Q., Zhang, L.: '*Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 2008*

16) Zhai, Y., Shah, M.: '*Visual attention detection in video sequences using spatiotemporal cues'. Proc. 14th Annual Int. Conf. of Multimedia, Santa Barbara, CA, USA, 2006, pp. 815–824*

17) Seo, H.J., Milanfar, P.: '*Static and space-time visual saliency detection by self-resemblance', J. Vis., 2009, 9, (12), pp. 1–27*

18) Li, Q., Chen, S., Zhang, B.: *'Predictive video saliency detection'. Proc. Chinese Conf. on Pattern Recognition, Beijng, China, 2012, pp. 178–185*

19) Marat, S., Phuoc, T.H., Granjon, L., Guyader, N., Pellerin, D., Guérin-Dugué, *A.: 'Modelling spatio-temporal saliency to predict gaze direction for short videos', Int. J. Comput. Vis., 2009, 82, (3), pp. 231–243.*