Multimed Tools Appl (2018) 77:8511–8529
DOI 10.1007/s11042-017-4748-0

Performance enhancement of salient object detection using superpixel based Gaussian mixture model

Navjot Singh[1] · **Rinki Arya**[2] · **R. K. Agrawal**[2]

Received: 19 September 2016 / Revised: 4 April 2017 / Accepted: 24 April 2017 /
Published online: 20 May 2017

Abstract Humans possess an intelligent system which effortlessly detect salient objects with high accuracy in real-time. It is a challenge to develop a computational model which can mimic human behavior such that the model achieves better detection accuracy and takes less computation time. So far the research community have suggested models which achieve better detection accuracy but at the cost of computation time and vice versa. In this paper, we attempted to realize a model that takes less computational time and simultaneously achieves higher detection accuracy. In the proposed model the original image is divided into m superpixels using SLIC superpixels algorithm and then these superpixels are clustered into k regions using k-means algorithm. Thereafter the result of the k-means clustering is used to build Gaussian mixture model whose parameters are refined using Expectation-Maximization algorithm. Finally the spatial variance of the clusters is computed and a center-weighted saliency map is computed. The performance of the proposed model and seventeen related models is evaluated both qualitatively and quantitatively on seven publicly available datasets. Experimental results show that the proposed model outperforms the existing models in terms of precision, recall and F -measure on all the seven datasets and in terms of area under curve on four datasets. Also, the proposed model takes less computation time in comparison to many methods.

Keywords Salient object detection · Superpixels · Gaussian mixture model · Expectation maximization · Spatial variance · Saliency map

✉ Navjot Singh
navjot.singh.09@gmail.com

[1] National Institute of Technology Uttarakhand, Srinagar, Uttarakhand 246174, India

[2] School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi 110067, India

1 Introduction

Salient object detection [5] refers to the extraction of dominant objects (salient objects) in an image which automatically attracts visual attention. It is a challenging problem in the field of computer vision and has many real-time applications in surveillance systems, remote sensing and image retrieval. It is helpful in automatic target detection, robotics, image and video compression, automatic cropping/ centering to display objects on small portable screens, medical imaging, advertising a design, image enhancement and many more.

Salient object detection involves the transformation of the original image to a saliency map [14] such that the salient objects are highlighted while the background is suppressed. Saliency map generally take the values between [0, 1]. Higher the value of a pixel, higher is its chances to become a salient pixel. The approaches for salient object detection can be broadly classified into two main categories [7]: bottom-up and top-down. Bottom-up approaches involves the extraction of low-level features from the image and then combining them into a saliency map. They are fast, stimulus driven and task independent. While in the top-down approaches, human observation behavior is exploited to accomplish certain goals and is task dependent. Usually top-down approaches are combined with the bottom-up approaches to detect salient objects.

Most of research works mostly focussed on the bottom-up aspect of visual attention. With the advancement of these bottom-up approaches, researchers started distinguishing the two very similar terms: fixation prediction and salient object detection. The fixation prediction models try to mimic the human vision with an objective that the human eyes mainly focus on some of the points in a given scene if shown for a few seconds. These points are helpful in eye movement prediction. The second category of models which are salient object detection models detects the most salient object in an image by segmenting the image into two regions, a salient object and background, by drawing accurate silhouettes of the salient object. Both categories of models construct saliency maps which are useful for different purposes. In literature, the research community has suggested different combination schemes in order to yield a saliency map from a set of low level features. The research work of Itti et al. [14] is motivated by the neuronal activity of the receptive fields in the human visual system. The three features such as intensity, color and orientation were considered equally important and were linearly combined to obtain a saliency map. While Liu et al. [19] proposed a supervised approach to learn a weight vector in order to combine the multi-scale contrast, center-surround histogram and the color spatial distribution features into a saliency map. We also investigated some of the other most popular related models like the one given by Bruce and Tsotsos [6] who modeled visual saliency by utilizing the concept of information maximization. Han et al. [10] applied region growing techniques over the saliency map obtained using the research work of Itti et al. [14] and extracted salient regions. Meur et al. [22] used the subband decomposition based energy for the chromatic as well as the achromatic channels to compute the saliency. Harel et al. [11] extended the work of Itti et al. [14] and gave a graph based visual saliency model. Hou and Zhang [12] gave a simple and fast method for visual saliency detection by extracting the spectral residual of the image. Yu and Wong [29] extracted the salient objects at the grid level instead at the pixel level. Zhang et al. [30] used Bayesian framework to compute the probability of a target at every location in the image. Achanta et al. [2] used an image subtraction technique to generate a frequency tuned saliency model. Achanta and Susstrunk [1] gave the visual saliency model by utilizing the maximum symmetric surround difference for every pixel in the image. Zhang et al. [31] combined position,

area and intensity saliency based on the outcome of scalable subtractive clustering, and employed Bayesian framework to classify a pixel into an attention pixel or a background pixel. Goferman et al. [9] proposed a context-aware saliency detection algorithm to detect salient objects. Liu et al. [20] used kernel density estimation method and two-phase graph cut approach to detect salient objects. Shen and Wu [24] incorporated the low rank matrix and a sparse noise in some feature space to detect the salient object. Vikram et al. [27] randomly sampled the image into a number of rectangular regions and computed local saliency over these regions. İmamoğlu et al. [13] proposed a saliency detection model by extracting low-level features based on wavelet transform. Singh and Agrawal [25] modified the Liu et al. [19] model at the feature level and employed a combination of Kullback-Leibler divergence and Manhattan distance to detect salient objects. Liu et al. [21] proposed a novel saliency tree approach to extract salient objects from the image. Zhu et al. [34] used a multisize superpixel approach based on multivariate normal distribution estimation for salient object detection. Peng et al. [23] suggested a saliency-aware image-to-class distances for image classification. Jiang et al. [15] proposed multi-level image segmentation technique which utilizes the supervised learning approach to map the regional feature vector to a saliency score.

Few researchers have extended saliency detection to co-saliency detection, like the one suggested by Fu et al. [8]. They used two layer clustering, where one layer focusses on groups the pixels on each image (single image), and the other layer associates the pixels on all images (multi-image).

Recently researchers have also suggested few models based on deep learning. Zhao et al. [33] proposed a multi-context deep learning framework using deep convolutional neural networks for salient object detection. Lin et al. [18] suggested a model which uses midlevel features on the basis of low-level k-means filters within a unified deep framework in a convolutional manner for saliency detection. Zhang et al. [32] proposed a co-saliency detection method based on intrasaliency prior transfer and deep intersaliency mining. Li and Yu [16, 17] suggested a deep contrast learning method for salient object detection using deep convolutional neural networks.

The common thing that is witnessed from related models is that they explored multiple low-level features of the image and then combined those using different strategies. The features involved were either of the same size of the image or of reduced size. The evaluation of the models is done on publicly available datasets to find their detection accuracy and its computation time. Experimental results demonstrated that most of the models [1, 2, 12, 14, 22, 27, 29–31] take less computation time but provide degraded detection accuracy because of either reduced size of image or simpler combination strategies. On the other hand, the models such as [9, 13, 19, 20, 24, 25] achieve better detection accuracy at the cost of higher computation time because they involved either full resolution image or some kind of learning technique is involved in combining the low-level features. However, there is need to develop a model which takes less computation time and simultaneously achieves high detection accuracy. One possible way to realize this objective is to utilize a single dominant feature in a model that is sufficient to describe an image instead of multiple features as commonly used in most of the state-of-the-art methods. In most of the state-of-the-art models dealing with multiple features, we have observed experimentally that color feature is most commonly used and dominates the remaining features. Snowden [26] also suggested that a purely chromatic signal is sufficient to capture visual attention. Color feature can be extracted either at the local level or the global level. Since colors are widely spread in an image, so color as a global feature may be more appropriate.

In this paper we propose an approach which utilizes color feature at the global level to detect the salient object. The motivation of the model came from the fact that image is constructed from several signals (say k), assumed to be Gaussians. Here a signal can be formed from various shades of a color present in the image. Then a mixture of Gaussians needs to build over these signals using a parametric estimation technique. Generally the images present in the datasets consist of thousands of pixels. Estimating the parameters of k Gaussians (strength, mean and covariance) using these large number of pixels will require huge computation time. Instead of this, if these large numbers of pixels are reduced to a smaller number of regions of similar pixels, then the estimation of parameters of k signals will take less computation time.

In the proposed model, the original RGB image of size $W \times H$, where of W and H represent the width and height of the image respectively, is first divided into m superpixels using SLIC superpixels algorithm [3] as it is fast and efficient. Since the superpixel comprises of pixels which are similar in color, hence each superpixel is represented by the mean value of its pixels, thereby reducing the image pixels to only m pixels. The colors of these m superpixels are further clustered into k color components using k-means algorithm. The result of the clustering procedure is used to build Gaussian mixture model, whose parameters are further refined using Expectation-Maximization algorithm. Thereafter, spatial variance of these color components is computed and a center-weighted saliency map is formulated.

It is found that the researchers have adopted superpixels for computing saliency at the local level [28] (i.e. in a specific neighborhood of a superpixel) and not at the global level (i.e. considering the complete image as a whole). The problem that arises here is that only smaller objects are captured and gets higher saliency value, while the larger objects are discarded and gets lower saliency value. To capture the details of the larger objects as well, we used superpixels at the global level. So the use of superpixels and GMM to capture saliency at the global level in a computationally efficient manner is the innovation in the proposed method.

In order to check the efficacy of the proposed model, experiments are carried out on seven publicly available image datasets. The performance is evaluated in terms of precision, recall, F -measure, area under curve and computation time and compared with existing seventeen other popular models.

The paper is organized as follows. Section 2 describes the proposed model. The experimental setup and results are included in section 3. Conclusion and future work are presented in Section 4.

2 Proposed model

In general, humans can effortlessly detect salient objects with high accuracy in real-time. It is a challenge to develop a model which can mimic human behavior such that the model achieves high detection accuracy and takes less computation time. One way of accomplishing this task is to utilize a single dominant feature in the model that best characterizes an image. We have investigated a number of features that are used in different state-of-the-art models and have found that the feature computed in terms of color is most commonly used. Also, Snowden [26] has very well suggested that a purely chromatic signal is sufficient to capture visual attention. There are two different ways of extracting a feature in salient object detection, at local or the

global level. In the local level a certain region is picked within an image and saliency is computed over it, while in the global level the complete image is considered while computing the saliency. As far as color is concerned, it is widely spread in an image, so using color as a global feature may be more appropriate.

The proposed model employs the concept of SuperPixels and Gaussian Mixture Model (SP-GMM) which is discussed in detail underneath.

2.1 Gaussian mixture model construction

In the color space, clustering of the RGB image $\mathbf{I}$, i.e. $\mathbf{I}(p) = [\mathbf{R}(p)\ \mathbf{G}(p)\ \mathbf{B}(p)]^T$ of size $W \times H$ into k regions and then constructing Gaussian mixture model is a time consuming process. But if the number of pixels is decreased to m such that $m \ll W \times H$, then the computation time can be considerably reduced. So the input RGB image is first divided into m superpixels using SLIC superpixels algorithm [3]. Let $\mathbf{SP}$ be the set containing the RGB values of m superpixels given by

$$\mathbf{SP} = \{\boldsymbol{SP}_i\}_{i=1}^{m}; \boldsymbol{SP}_i = \frac{1}{|\boldsymbol{S}_i|} \sum_{p \in \boldsymbol{S}_i} [\boldsymbol{R}(p)\ \boldsymbol{G}(p)\ \boldsymbol{B}(p)]^T \tag{1}$$

where $\boldsymbol{SP}_i$ is the RGB value of the i -th superpixel, $\boldsymbol{S}_i$ is the set of pixels in the i -th superpixel and $|\boldsymbol{S}_i|$ represents its size such that $\sum_{i=1}^{m} |\boldsymbol{S}_i| = W \times H$. Now the set $\mathbf{SP}$ is partitioned into k clusters using k-means algorithm. The result of the clustering algorithm is used as samples to build Gaussian mixture model (GMM).

The parameters of the GMM include the weights, means, and co-varainces of the Gaussians. The initial weight w_i^0 of the i -th cluster is given as

$$w_i^0 = \frac{n_i}{m}\ i = 1, 2, \ldots, k \tag{2}$$

where n_i is the number of superpixels belonging to the i -th cluster. Assuming that the j –th superpixel belongs to the i -th cluster, the initial mean of the i-th cluster $\boldsymbol{\mu}_i^0$ is given as

$$\boldsymbol{\mu}_i^0 = \frac{1}{n_i} \sum_{j \in \boldsymbol{P}_i} \boldsymbol{SP}_j\ i = 1, 2, \ldots, k \tag{3}$$

where $\boldsymbol{P}_i$ is the set of superpixels belonging to the i -th cluster. The initial co-variances $\boldsymbol{\Sigma}_i^0$ are defined as

$$\boldsymbol{\Sigma}_i^0 = \frac{1}{n_i - 1} \sum_{j \in \boldsymbol{P}_i} \left(\boldsymbol{SP}_j - \boldsymbol{\mu}_i^0\right)\left(\boldsymbol{SP}_j - \boldsymbol{\mu}_i^0\right)^{\boldsymbol{T}}; i = 1, 2, \ldots, k \tag{4}$$

Thereafter, the expectation maximization (EM) algorithm is applied to update the parameters of the GMM until convergence is achieved. Using the current parameters of the l -th iteration the probability of a superpixel j to belong to the i -th cluster is calculated as

$$Pr^l\left(i|\boldsymbol{SP}_j\right) = \frac{w_i^l \mathcal{N}\left(\boldsymbol{SP}_j|\boldsymbol{\mu}_i^l, \boldsymbol{\Sigma}_i^l\right)}{\sum_{t=1}^{k} w_t^l \mathcal{N}\left(\boldsymbol{SP}_j|\boldsymbol{\mu}_t^l, \boldsymbol{\Sigma}_t^l\right)} \tag{5}$$

Then weight, mean and co-variance of the Gaussians are updated as

$$\begin{aligned} w_i^{l+1} &= \frac{1}{m}\sum_{j=1}^{m}\Pr^{l}\left(i|\mathbf{SP}_j\right) \\ \boldsymbol{\mu}_i^{l+1} &= \frac{\sum_{j=1}^{m} Pr^{l}\left(i|\mathbf{SP}_j\right).\mathbf{SP}_j}{\sum_{j=1}^{m} Pr^{l}\left(i|\mathbf{SP}_j\right)} \\ \sum\nolimits_i^{l+1} &= \frac{\sum_{j=1}^{m}\Pr^{l}\left(i|\mathbf{SP}_j\right).\left(\mathbf{SP}_j-\boldsymbol{\mu}_i^{l}\right).\left(\mathbf{SP}_j-\boldsymbol{\mu}_i^{l}\right)^{T}}{\sum_{j=1}^{m}\Pr^{l}\left(i|\mathbf{SP}_j\right)} \end{aligned} \tag{6}$$

The log-likelihood for $l+1$ iteration is computed as

$$\text{loglik}^{l+1} = \sum_{j=1}^{m}\left(log\left(\sum_{i=1}^{k} w_i^{l+1}.\mathcal{N}\left(\boldsymbol{SP}_j|\boldsymbol{\mu}_i^{l+1},\boldsymbol{\Sigma}_i^{l+1}\right)\right)\right) \tag{7}$$

Eqs. (5–7) are repeated until convergence is achieved. The inequality for the convergence condition is given as

$$\text{abs}\left(\text{loglik}^{l+1}-\text{loglik}^{l}\right) < 1.0e{-}3 \tag{8}$$

Using the final parameter values of the GMM, each and every pixel p of the original RGB image $\mathbf{I}$ of size $W \times H$ is assigned to the i -th cluster with a probability given as

$$Pr^{\text{final}}(i|\boldsymbol{I}(p)) = \frac{w_i\mathcal{N}(\boldsymbol{I}(p)|\boldsymbol{\mu}_i,\boldsymbol{\Sigma}_i)}{\sum_{j=1}^{k} w_j\mathcal{N}\left(\boldsymbol{I}(p)|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right)} \tag{9}$$

where w_i, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the weight, mean and covariance matrix of the i -th cluster respectively.

2.2 Spatial variance and saliency map computation

The spatial variance measures the distribution of a color component in an image. Lower the spatial variance of a color component better is its chances to be salient and vice-versa. In the spatial domain, variance of the i -th cluster is computed both in the horizontal as well as the vertical direction. The horizontal variance V_i^h of the i -th cluster is given as

$$V_i^h = \frac{\sum_{p\in\mathbf{P}} Pr^{\text{final}}(i|\boldsymbol{I}(p)).\left(x_p-M_i^h\right)^2}{\sum_{p\in\mathbf{P}} Pr^{\text{final}}(i|\boldsymbol{I}(p))} \tag{10}$$

where $M_i^h = \frac{\sum_{p\in\mathbf{P}} Pr^{\text{final}}(i|\boldsymbol{I}(p)).x_p}{\sum_{p\in\mathbf{P}} Pr^{\text{final}}(i|\boldsymbol{I}(p))}$, x_p is the x-coordinate of the p -th pixel and $\mathbf{P}$ is the set of all the pixels present in the image. Similarly the vertical variance V_i^v is computed. The total spatial variance is given by

$$V_i = V_i^h + V_i^v \tag{11}$$

V_i is normalized between [0,1] computed as

$$V_i = \frac{V_i - min(V_i)}{max(V_i) - min(V_i)} \tag{12}$$

Thereafter, a center-weighted scheme is applied to give more weightage to the clusters present near the center of the image. The position weight D_i of the i -th cluster is given by

$$D_i = \sum_{p \in \boldsymbol{P}} Pr^{\text{final}}(i|\boldsymbol{I}(p)).d_p \tag{13}$$

where d_p is the distance between the pixel p and the image center using the L2 norm. D_i is also normalized between $[0, 1]$ computed as

$$D_i = \frac{D_i - min(D_i)}{max(D_i) - min(D_i)} \tag{14}$$

Finally the pixel-wise saliency map **SM** is given as

$$\mathbf{SM}(p) = \sum_{i=1}^{k} Pr^{\text{final}}(i|\boldsymbol{I}(p)).(1-\boldsymbol{V_i}).(1-\boldsymbol{D_i}) \tag{15}$$

The values of the saliency map **SM** are normalized between $[0, 1]$ computed as

$$\mathbf{SM} = \frac{\boldsymbol{SM} - min(\boldsymbol{SM})}{max(\boldsymbol{SM}) - min(\boldsymbol{SM})} \tag{16}$$

A threshold is applied on the saliency map to generate an attention mask. Fig. 1 depicts the working of the model on certain images.

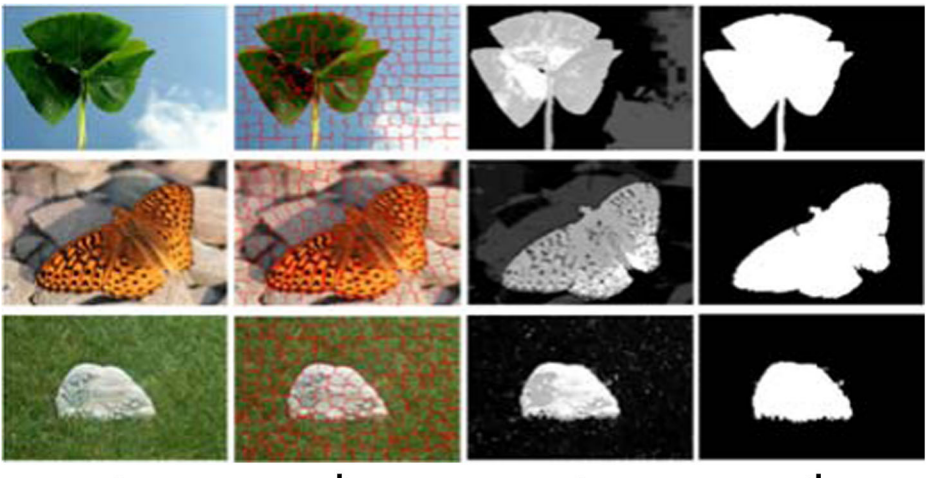

Fig. 1 **a** Original image **b** SLIC Superpixels **c** Saliency map **d** Ground Truth

Table 1 Parameter values of various models

Models	Parameters
Itti [14]	Center scale $c = \{2, 3, 4\}$, difference between surround and center scale $\delta = \{3, 4\}$
AIM [6]	Standard deviation of Gaussian filter $\sigma = 20$
GBVS [11]	Standard deviation $\sigma = 2.5$ in computing weight of the edge
SR [12]	Standard deviation of Gaussian filter $\sigma = 8$ for smoothing the saliency map
Liu [19]	Linear weight vector $\overrightarrow{\lambda} = \{0.24, 0.54, 0.22\}$ for combining features into the saliency map
SUN [30]	Standard deviation of Difference of Gaussian filter $\sigma = 4$
FT [2]	Ratio of standard deviations of two Gaussian filters for computing Difference of Gaussians $\rho = 1.6$
ASS [1]	Standard deviation of Gaussian filter $\sigma = 10$
Gof [9]	Most similar patches $K = 64$
Shen [24]	Step size $\alpha = 0.02$
Vikram [27]	Standard deviation of Gaussian filter $\sigma = 0.5$
WT [13]	Gaussian filter $k \times k$ where $k = 3$
SA [25]	No. of clusters for construction of Gaussian Mixture Model $C = 6$
COSAL [8]	Standard deviation of Gaussian filter $\sigma = 8$
DRFI [15]	Different levels of segmentations $M = 15$
DCL [16]	Conditional Random Field Parameters $w_1 = 3$, $w_2 = 5$, $\sigma_\alpha = 3$, $\sigma_\beta = 50$, $\sigma_\gamma = 3$
MDF [17]	Learning rate of 3-layer perceptron network $\eta = 0.2$, Conditional Random Field Parameters $w_1 = 3$, $w_2 = 5$, $\sigma_\alpha = 3$, $\sigma_\beta = 50$, $\sigma_\gamma = 3$
SP-GMM	No. of superpixels $m = 200$, no. of Gaussian signals $k = 5$

3 Experimental setup and results

Intensive care has been taken while evaluating the related models. The parameters as suggested in the related papers have been set accordingly and saliency maps are computed. Table 1 list the parameter values of various models. A qualitative as well as a quantitative evaluation is done in order to measure the performance of the proposed model, and is compared with the existing approaches. All the experiments are carried out using Windows 7 environment over Intel (R) Xeon (R) processor with a speed of 2.27 GHz and 4GB RAM.

3.1 Salient object database

The performance of the proposed model and seventeen other related models is examined using the following seven publicly available datasets (Table 2):

The test dataset comprises of all these 12,500 images and is used for performance evaluation.

3.2 Qualitative evaluation

The qualitative evaluation of the proposed model and seventeen other related models can be seen in Fig. 2. We have chosen some of the images from the test data set that contain objects differing in shape, size, position, type etc. It can be clearly seen from Fig. 2 that the proposed model yields better saliency maps in comparison to the related methods.

3.3 Quantitative evaluation

The quantitative evaluation of the proposed model and seventeen other models is done in terms of precision, recall, F measure, area under curve (AUC), and computation

time. Using the ground truth **G** and the detection result **R**, precision, recall, F - measure are calculated as

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ \text{Recall} &= \frac{Tp}{Tp+FN} \\ F_\beta &= \frac{\left(1+\beta^2\right)\times \text{Precision}\times \text{Recall}}{\beta^2\times \text{Precision}+\text{Recall}} \\ TP &= \sum_{G(x,y)=1}\mathbf{R}(x,y); \text{FP} = \sum_{G(x,y)=0}\mathbf{R}(x,y) \\ FN &= \sum_{\mathbf{R}(x,y)=0}\mathbf{G}(x,y); \text{TN} = \sum_{G(x,y)=0}\mathbf{R}(x,y) \end{aligned} \tag{17}$$

where $\beta = 1$ as we are giving equal weightage to both precision and recall, and TP (true positives) is the number of salient pixels that are detected as salient pixels. FP (false positives) is the number of background pixels that are detected as salient pixels. FN (false negatives) is the number of salient pixels that are detected as background pixels.

AUC is computed by drawing a receiver operator characteristic (ROC) curve. ROC curve is plotted between the true positive rate (TPR) and the false positive rate (FPR). TPR and FPR are given by

$$\begin{aligned} TPR &= \frac{TP}{\sum_{(x,y)}\mathbf{G}(x,y)} \\ FPR &= \frac{FP}{W\times H - \sum_{(x,y)}\mathbf{G}(x,y)} \end{aligned} \tag{18}$$

where W and H represents the width and height of the image respectively. The saliency maps corresponding to the proposed model as well as state-of-the-art models are first normalized between [0,255]. Then 256 thresholds are chosen one by one and the values of TPR and FPR are computed and the ROC curve is plotted and finally area under the curve (AUC) is calculated. Table 3 shows the quantitative performance

Table 2 Datasets used for salient object detection

SNO	Dataset	# Images	# Objects
1	MSRA-B[a]	5000	~1
2	ASD[b]	1000	~1
3	SAA_GT [4]	5000	~1
4	SOD[c]	300	~3
5	SED1[d]	100	1
6	SED2[e]	100	2
7	ECSSD[f]	1000	~1

[a] http://www.research.microsoft.com/enus/um/people/jiansun/salientobject/salient_object.htm

[b] http://ivrgwww.epfl.ch/supplementary_material/RK_CVPR09/GroundTruth/binarymasks.zip

[c] http://elderlab.yorku.ca/~vida/SOD/index.html

[d] http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB

[e] http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB

[f] http://www.cse.cuhk.edu.hk/leojia/projects/hsalie ncy/

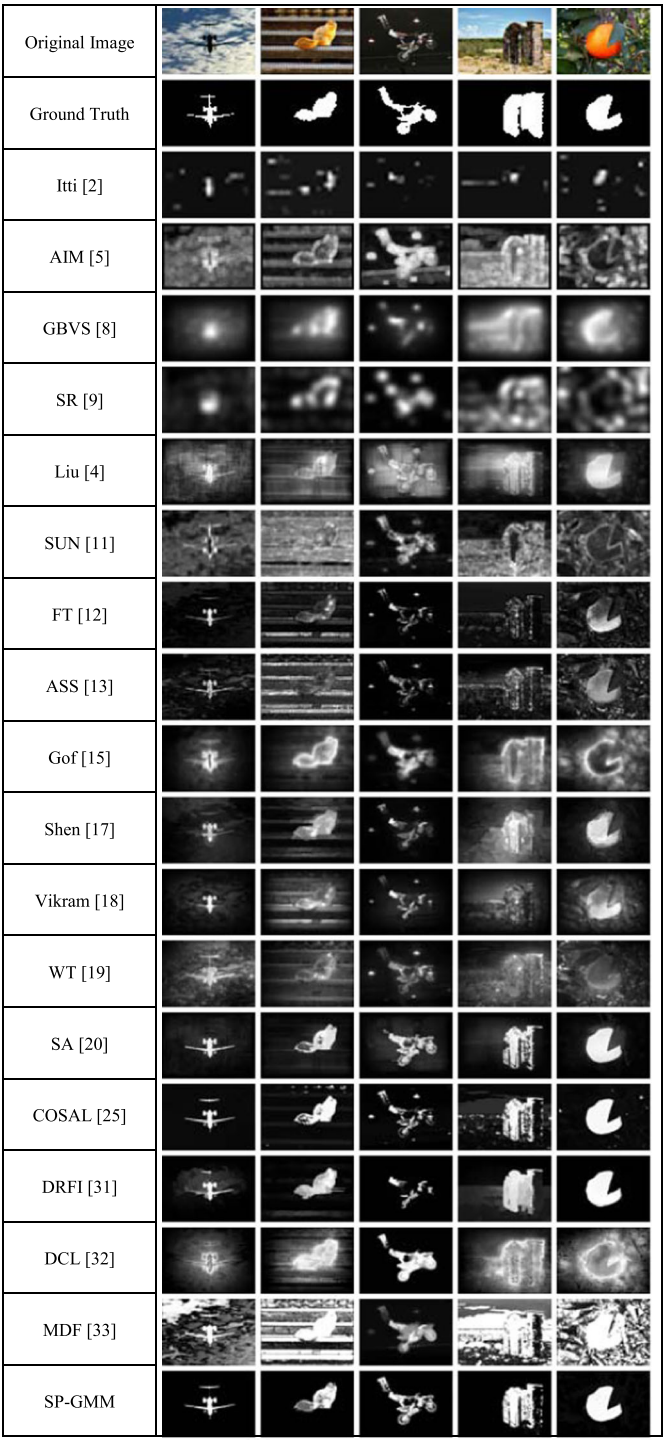

Fig. 2 Saliency maps for different state-of-the-art models and the proposed model

Table 3 Quantitative comparison on seven datasets and their computation time

	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	ECSSD	Time (in sec) per image
Itti [14]	$P = 0.672$	$P = 0.550$	$P = 0.545$	$P = 0.685$	$P = 0.720$	$P = 0.676$	$P = 0.311$	1.70
	$R = 0.614$	$R = 0.695$	$R = 0.609$	$R = 0.154$	$R = 0.150$	$R = 0.297$	$R = 0.462$	
	F = 0.642	F = 0.614	F = 0.575	F = 0.252	F = 0.248	F = 0.413	F = 0.372	
	A = 0.663	A = 0.529	A = 0.590	A = 0.550	A = 0.623	A = 0.601	A = 0.414	
AIM [6]	$P = 0.728$	$P = 0.535$	$P = 0.527$	$P = 0.412$	$P = 0.562$	$P = 0.470$	$P = 0.429$	50.8
	$R = 0.762$	$R = 0.859$	$R = 0.777$	$R = 0.625$	$R = 0.790$	$R = 0.816$	$R = 0.659$	
	F = 0.745	F = 0.659	F = 0.628	F = 0.500	F = 0.657	F = 0.597	F = 0.5197	
	A = 0.705	A = 0.631	A = 0.673	A = 0.796	A = 0.880	A = 0.861	A = 0.573	
GBVS [11]	$P = 0.800$	$P = 0.666$	$P = 0.658$	$P = 0.520$	$P = 0.695$	$P = 0.542$	$P = 0.447$	59.8
	$R = 0.692$	$R = 0.634$	$R = 0.612$	$R = 0.584$	$R = 0.597$	$R = 0.600$	$R = 0.584$	
	F = 0.742	F = 0.650	F = 0.634	F = 0.550	F = 0.642	F = 0.570	F = 0.506	
	A = 0.698	A = 0.579	A = 0.636	A = 0.813	A = 0.868	A = 0.821	A = 0.537	
SR [12]	$P = 0.761$	$P = 0.502$	$P = 0.588$	$P = 0.479$	$P = 0.614$	$P = 0.504$	$P = 0.460$	**0.02**
	$R = 0.526$	$R = 0.440$	$R = 0.372$	$R = 0.336$	$R = 0.360$	$R = 0.450$	$R = 0.383$	
	F = 0.622	F = 0.469	F = 0.456	F = 0.395	F = 0.454	F = 0.476	F = 0.418	
	A = 0.658	A = 0.505	A = 0.581	A = 0.732	A = 0.780	A = 0.796	A = 0.622	
Liu [19]	$P = 0.674$	$P = 0.700$	$P = 0.763$	$P = 0.423$	$P = 0.589$	$P = 0.431$	$P = 0.526$	25.7
	$R = 0.889$	$R = 0.921$	$R = 0.895$	$R = 0.737$	$R = 0.806$	$R = 0.803$	$R = 0.812$	
	F = 0.767	F = 0.795	F = 0.824	F = 0.538	F = 0.681	F = 0.561	F = 0.639	
	A = 0.802	A = 0.733	A = 0.767	A = 0.796	A = 0.868	A = 0.812	A = 0.662	
SUN [30]	$P = 0.598$	$P = 0.542$	$P = 0.668$	$P = 0.379$	$P = 0.561$	$P = 0.417$	$P = 0.363$	3.64
	$R = 0.857$	$R = 0.848$	$R = 0.764$	$R = 0.431$	$R = 0.611$	$R = 0.659$	$R = 0.469$	
	F = 0.704	F = 0.661	F = 0.713	F = 0.403	F = 0.585	F = 0.511	F = 0.409	
	A = 0.681	A = 0.602	A = 0.641	A = 0.716	A = 0.851	A = 0.776	A = 0.577	
FT [2]	$P = 0.717$	$P = 0.599$	$P = 0.800$	$P = 0.608$	$P = 0.735$	$P = 0.830$	$P = 0.571$	0.17
	$R = 0.575$	$R = 0.606$	$R = 0.517$	$R = 0.300$	$R = 0.347$	$R = 0.533$	$R = 0.361$	
	F = 0.638	F = 0.603	F = 0.628	F = 0.402	F = 0.471	F = 0.649	F = 0.443	
	A = 0.669	A = 0.625	A = 0.648	A = 0.595	A = 0.650	A = 0.676	A = 0.549	
ASS [1]	$P = 0.786$	$P = 0.635$	$P = 0.801$	$P = 0.655$	$P = 0.817$	$P = 0.757$	$P = 0.664$	0.31
	$R = 0.704$	$R = 0.670$	$R = 0.524$	$R = 0.366$	$R = 0.452$	$R = 0.589$	$R = 0.433$	
	F = 0.743	F = 0.652	F = 0.634	F = 0.470	F = 0.580	F = 0.663	F = 0.524	
	A = 0.698	A = 0.630	A = 0.664	A = 0.790	A = 0.840	A = 0.797	A = 0.630	
Gof [9]	$P = 0.712$	$P = 0.697$	$P = 0.679$	$P = 0.492$	$P = 0.659$	$P = 0.551$	$P = 0.500$	124.0
	$R = 0.763$	$R = 0.782$	$R = 0.726$	$R = 0.518$	$R = 0.496$	$R = 0.559$	$R = 0.545$	
	F = 0.737	F = 0.737	F = 0.702	F = 0.505	F = 0.566	F = 0.555	F = 0.522	
	A = 0.776	A = 0.705	A = 0.741	A = 0.791	A = 0.833	A = 0.813	A = 0.533	
Shen [24]	$P = 0.703$	$P = 0.716$	$P = 0.680$	$P = 0.476$	$P = 0.658$	$P = 0.590$	$P = 0.548$	71.9
	$R = 0.907$	$R = 0.903$	$R = 0.841$	$R = 0.693$	$R = 0.771$	$R = 0.790$	$R = 0.770$	
	F = 0.792	F = 0.799	F = 0.752	F = 0.564	F = 0.710	F = 0.676	F = 0.640	
	A = 0.783	A = 0.713	A = 0.753	A = 0.794	A = 0.860	A = 0.814	A = 0.683	
Vikram [27]	$P = 0.716$	$P = 0.605$	$P = 0.648$	$P = 0.508$	$P = 0.673$	$P = 0.684$	$P = 0.580$	1.47
	$R = 0.801$	$R = 0.738$	$R = 0.677$	$R = 0.585$	$R = 0.593$	$R = 0.619$	$R = 0.640$	
	F = 0.756	F = 0.665	F = 0.662	F = 0.544	F = 0.631	F = 0.650	F = 0.608	
	A = 0.769	A = 0.613	A = 0.690	A = 0.795	A = 0.839	A = 0.769	A = 0.579	
WT [13]	$P = 0.662$	$P = 0.606$	$P = 0.612$	$P = 0.451$	$P = 0.622$	$P = 0.575$	$P = 0.484$	6.55
	$R = 0.840$	$R = 0.801$	$R = 0.702$	$R = 0.564$	$R = 0.608$	$R = 0.720$	$R = 0.612$	
	F = 0.741	F = 0.690	F = 0.654	F = 0.501	F = 0.615	F = 0.639	F = 0.540	
	A = 0.743	A = 0.693	A = 0.718	A = 0.785	A = 0.824	A = 0.817	A = 0.588	
SA [25]	$P = 0.806$	$P = 0.818$	$P = 0.826$	$P = 0.698$	$P = 0.801$	$P = 0.780$	$P = 0.641$	21.6
	$R = 0.874$	$R = 0.903$	$R = 0.869$	$R = 0.645$	$R = 0.737$	$R = 0.793$	$R = 0.574$	
	F = 0.817	F = 0.858	F = 0.847	F = 0.671	F = 0.768	F = 0.786	F = 0.606	
	A = 0.840	A = 0.778	A = 0.796	A = 0.796	A = 0.880	A = 0.872	A = 0.795	
COSAL [8]	$P = 0.797$	$P = 0.813$	$P = 0.795$	$P = 0.608$	$P = 0.820$	$P = 0.752$	$P = 0.634$	1.14
	$R = 0.793$	$R = 0.847$	$R = 0.698$	$R = 0.457$	$R = 0.631$	$R = 0.782$	$R = 0.550$	

Table 3 (continued)

	MSRA-B	ASD	SAA_GT	SOD	SED1	SED2	ECSSD	Time (in sec) per image
	F = 0.795	F = 0.830	F = 0.743	F = 0.522	F = 0.713	F = 0.767	F = 0.589	
	A = 0.716	A = 0.644	A = 0.694	**A = 0.844**	**A = 0.897**	A = 0.763	A = 0.617	
DRFI [15]	*P* = 0.726	*P* = 0.695	*P* = 0.711	*P* = 0.634	*P* = 0.746	*P* = 0.665	*P* = 0.684	3.52
	R = 0.697	*R* = 0.724	*R* = 0.683	*R* = 0.666	*R* = 0.691	*R* = 0.786	*R* = 0.555	
	F = 0.711	F = 0.709	F = 0.697	F = 0.650	F = 0.717	F = 0.721	F = 0.613	
	A = 0.699	A = 0.664	A = 0.660	A = 0.803	A = 0.866	A = 0.806	A = 0.666	
DCL [16]	*P* = 0.801	*P* = 0.768	*P* = 0.732	*P* = 0.689	*P* = 0.774	*P* = 0.683	*P* = 0.624	106.4
	R = 0.628	*R* = 0.755	*R* = 0.618	*R* = 0.703	*R* = 0.784	*R* = 0.777	*R* = 0.616	
	F = 0.704	F = 0.761	F = 0.670	F = 0.696	F = 0.779	F = 0.727	F = 0.620	
	A = 0.804	A = 0.672	A = 0.692	A = 0.755	A = 0.855	A = 0.795	A = 0.652	
MDF [17]	*P* = 0.799	*P* = 0.741	*P* = 0.648	*P* = 0.629	*P* = 0.711	*P* = 0.684	*P* = 0.591	117.5
	R = 0.812	*R* = 0.806	*R* = 0.782	*R* = 0.689	*R* = 0.735	*R* = 0.729	*R* = 0.634	
	F = 0.805	F = 0.772	F = 0.709	F = 0.658	F = 0.723	F = 0.706	F = 0.612	
	A = 0.783	A = 0.609	A = 0.784	A = 0.805	A = 0.868	A = 0.817	A = 0.571	
SP-GMM	***P* = 0.824**	***P* = 0.819**	***P* = 0.832**	***P* = 0.730**	***P* = 0.821**	***P* = 0.799**	***P* = 0.742**	1.87
	***R* = 0.931**	***R* = 0.930**	***R* = 0.908**	***R* = 0.727**	***R* = 0.827**	***R* = 0.817**	***R* = 0.755**	
	F = 0.874	**F = 0.871**	**F = 0.868**	**F = 0.729**	**F = 0.824**	**F = 0.808**	**F = 0.748**	
	A = 0.821	**A = 0.865**	**A = 0.829**	A = 0.836	A = 0.885	**A = 0.887**	**A = 0.843**	

P Precision, *R* Recall, *F* F-measure, *A* Area under Curve

evaluation of the proposed method in comparison to the other state-of-the-art methods on all the seven datasets including their average computation time per image. Their ROC curves are shown in Fig. 3.

The number of superpixels (m) and clusters (k) required to build the Gaussian mixture model play vital role. The number of superpixels were varied from 50 to 500 and it was found that with the increase in the number of superpixels the performance increases till $m = 200$ and remains constant thereafter. It can be observed from Fig. 4 and Fig. 5 that the best value of performance measures can be obtained at $m = 200$ and $k = 5$.

Table 3 shows the quantitative evaluation of the proposed model in comparison with seventeen related models. The best results are shown in bold.

MSRA-B

- The proposed model gives fine shape information that fetches it the highest precision, recall and F-measure.
- The proposed model has the best AUC value except SA [25] model.

ASD

- The proposed model gives the highest precision, recall, F-measure and AUC values.

SAA GT

- The proposed model gives the highest precision, recall, F-measure and AUC values.

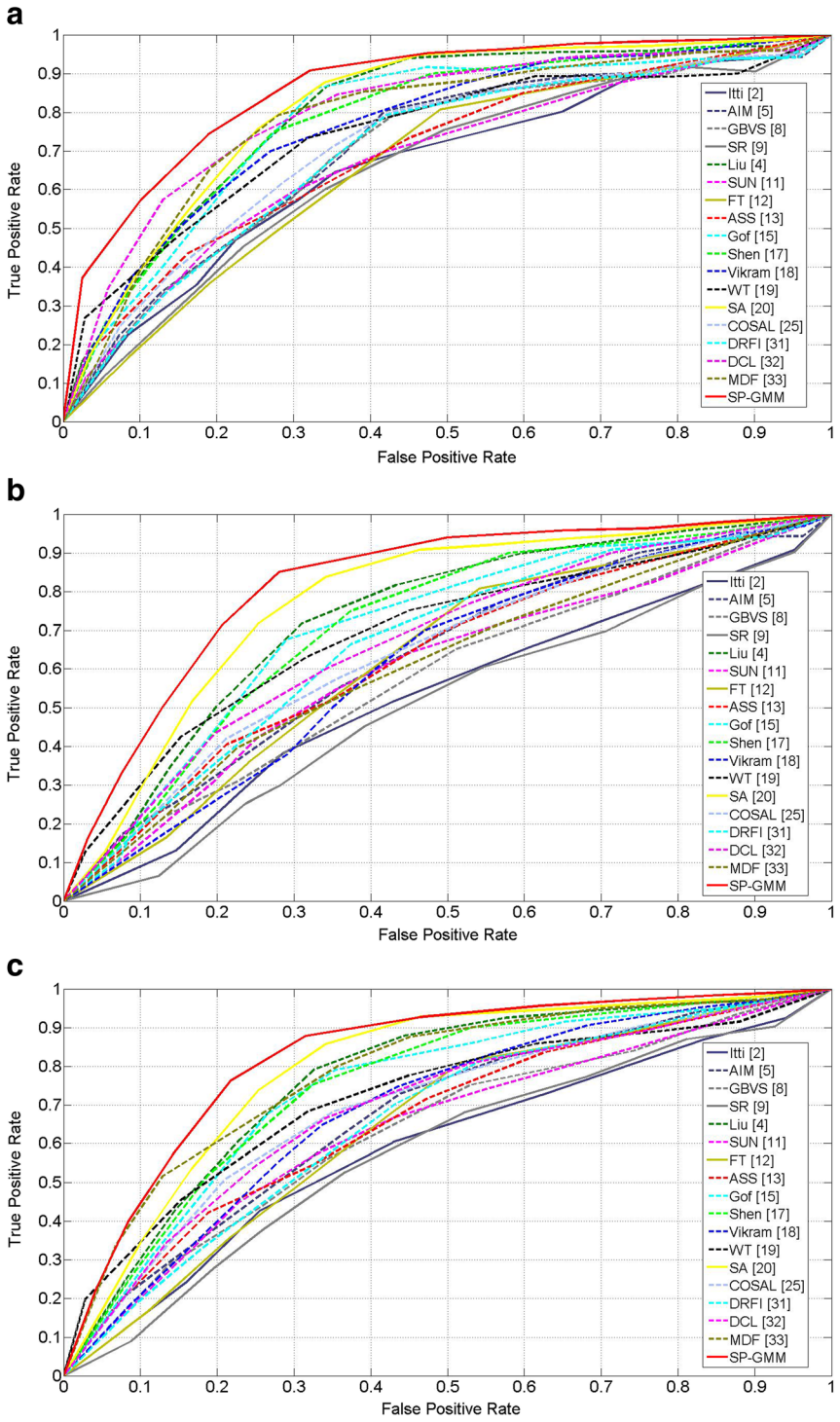

Fig. 3 ROC for the seven datasets (**a**) MSRA-B (**b**) ASD (**c**) SAA_GT (**d**) SOD (**e**) SED1 (**f**) SED2 (**g**) ECSSD

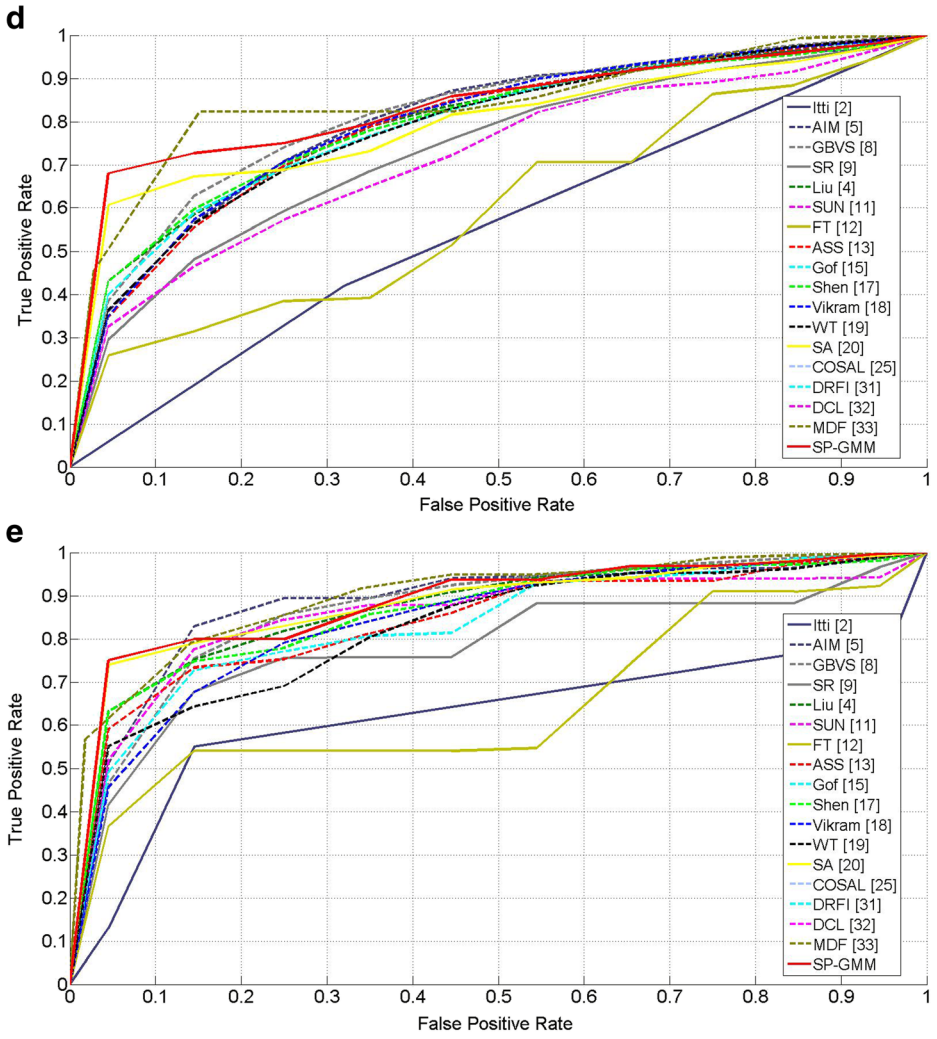

Fig. 3 (continued)

SOD

- The proposed model fetches it the highest precision, recall and F-measure.
- The proposed model has the best AUC value except COSAL [8] model.

SED1

- The proposed model has the highest precision, recall and F-measure.
- The proposed model has the best AUC value except COSAL [8] model.

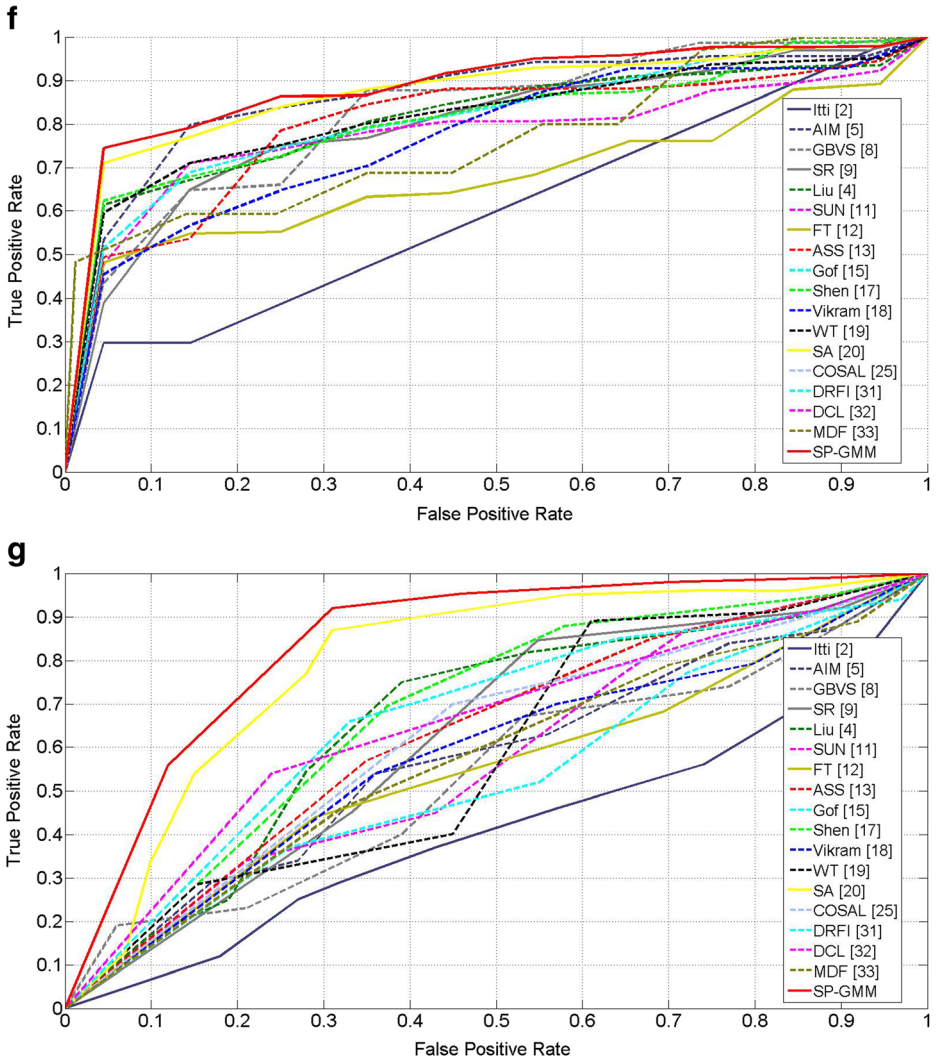

Fig. 3 (continued)

SED2

- The proposed model gives the highest precision, recall, F-measure and AUC values.

ECSSD

- The proposed model fetches the highest precision, recall, F-measure and AUC values.

Computation Time

- The SR [12] model takes the least computational time.

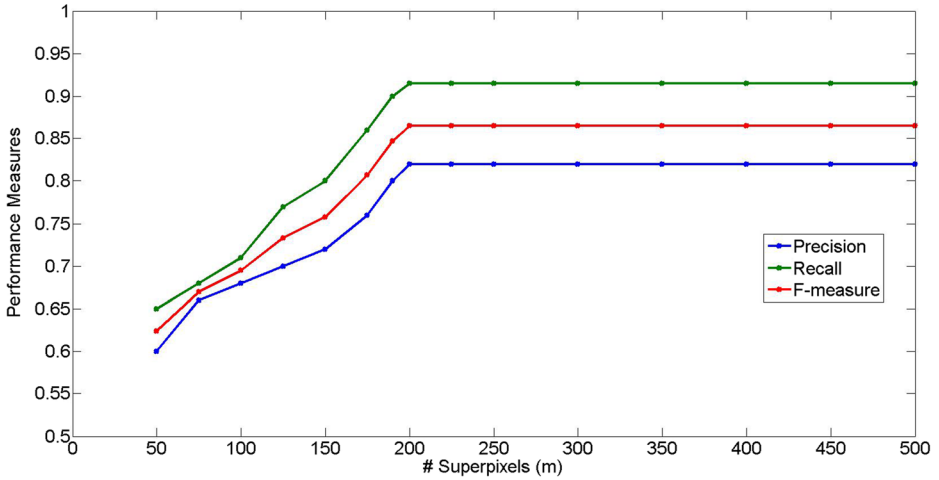

Fig. 4 Parameter analysis of the no. of Superpixels (m)

- As compared to the models like Liu [19], AIM [6], GBVS [11], SUN [30], Gof [9], Shen [24], WT [13], SA [25], DRFI [15], DCL [16], MDF [17], the proposed model achieves better detection accuracy and requires very less time.

4 Conclusion and future work

Salient object detection can be achieved by either exploring the bottom-up components or its integration with the top-down components. The research community is mostly fascinated by the bottom-up components as these methods are fast and task independent. Researchers have tried to improve the detection accuracy at the cost of complexity of model which is

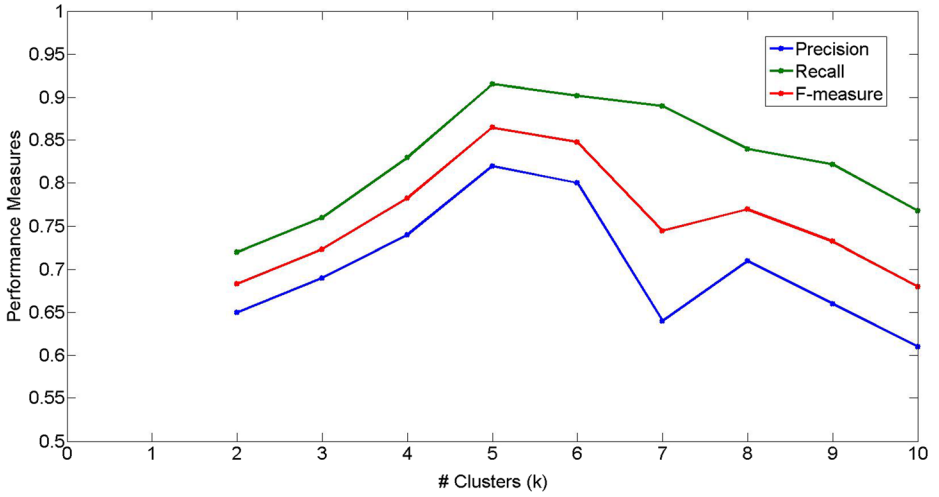

Fig. 5 Parameter analysis of the no. of Clusters (k)

computationally expensive. Some research efforts are made to reduce the computation time but degraded the detection accuracy. In the proposed model, we attempted to improve the salient object detection accuracy with less computation time. The model employed the use of SLIC superpixels, Gaussian mixture model and Expectation-Maximization algorithm to detect a salient object. Generally the images present in the datasets are of size 300 × 400, i.e. around 0.12 million pixels. Estimating the parameters of Gaussians (strength, mean and covariance) using 0.12 million samples is time consuming. The manuscript attempted to reduce the pixels, say to 200, where there is not much of a loss in the estimated values of the parameters, and then the computation time can be reduced to a considerable extent.

Experimental results demonstrate better performance of the proposed model in comparison to the existing methods in terms of precision, recall and F-measure on all the seven datasets and AUC on four datasets. In comparison to many state-of-the-art models, the proposed model requires less computation time.

There are certain more challenges in detecting salient objects. These include partial occlusion, background clutter, articulation, etc. The datasets used in our experiments contain images with only one salient object. Research work may be extended to detect any number of salient objects or no salient object at all.

References

1. Achanta R, Susstrunk S (2010) Saliency Detection using Maximum Symmetric Surround. Proc. International Conference on Image Processing, pp. 2653–2656
2. Achanta R, Hemamiz S, Estraday F and Susstrunk S (2009) Frequency-tuned Salient Region Detection. Proc IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1597–1604
3. Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Susstrunk S (2010) SLIC Superpixels. EPFL Technical Report 2010:149–300
4. Arya R, Singh N, Agrawal RK A novel hybrid approach for salient object detection using local and global saliency in frequency domain. Multimedia Tools and Applications. doi:10.1007/s11042-015-2750-y
5. Borji A, Itti L (2013) State-of-the-art in visual attention modeling. Proc IEEE Trans Pattern Anal Mach Intell 35(1):185–207
6. Bruce NDB, Tsotsos JK (2006) Saliency based on information maximization. Proc Adv Neural Inf Proces Syst 18:155–162
7. Frintrop S, Rome E and Christensen HI (2010) Computational Visual Attention Systems and their Cognitive Foundation: A Survey. Proc. ACM Transactions on Applied Perception, Vol. 7, no. 1
8. Fu H, Cao X, and Tu Z (2013) Cluster-based co-saliency detection. IEEE Transactions on Image Processing, 2013.
9. Goferman S, Zelnik-Manor L, Tal A (2012) Context-aware saliency detection. Proc IEEE Trans Pattern Anal Mach Intell 34(2012):1915–1926
10. Han J, Ngan KN, Li MJ, Zhang HJ (2006) Unsupervised extraction of visual attention objects in color images. Proc IEEE Trans Circuits Syst Video Technol 16:141–145
11. Harel J, Koch C and Perona P (2007) Graph Based Visual Saliency. Proc. Advances in Neural Information and Processing Systems, pp. 545–552
12. Hou X and Zhang L (2007) Saliency Detection: A Spectral Residual Approach, Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8
13. İmamoğlu N, Lin W, Fang Y (2013) A saliency detection model using low-level features based on wavelet transform. Proc IEEE Trans Multimedia 15:96–105
14. Itti L, Koch C, Niebur E (1998) A model of saliency based visual attention for rapid scene analysis. Proc IEEE Trans Pattern Anal Mach Intell 20:1254–1259
15. Jiang, H., Wang, J., Yuan, Z., Wu, Y., Zheng, N. and Li, S., (2013). Salient object detection: A discriminative regional feature integration approach. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2083–2090).
16. Li, G. and Yu, Y., (2016). Deep contrast learning for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 478–487).

17. Li G, Yu Y (2016) Visual saliency detection based on multiscale deep CNN features. IEEE Trans Image Process 25(11):5012–5024
18. Lin, Y., Kong, S., Wang, D. and Zhuang, Y., (2014). Saliency detection within a deep convolutional architecture. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.
19. Liu T, Yuan Z, Sun-Wang J, Zheng N, Tang X, Shum HY (2011) Learning to detect a salient object. Proc IEEE Trans Pattern Anal Mach Intell 33:353–366
20. Liu Z, Shi R, Shen L, Xue Y, Ngan KN, Zhang Z (2012) Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. Proc IEEE Trans Multimedia 14:1275–1289
21. Liu Z, Zou W, Meur OL (2014) Saliency tree: a novel saliency detection framework. IEEE Trans Image Process 23:1937–1952
22. Meur OL, Callet PL, Barba D, Thoreau D (2006) A coherent computational approach to model bottom up visual attention. Proc IEEE Trans Pattern Anal Mach Intell 28:802–817
23. Peng P, Shao L, Han J, Han J (2015) Saliency-aware image-to-class distances for image classification. Neurocomputing 166:337–345
24. Shen X, Wu Y (2012) A unified approach to salient object detection via low rank matrix recovery. Proc IEEE Conf Comput Vis Pattern Recognit 2012:853–860
25. Singh N, Agrawal RK (2013) Combination of Kullback–Leibler divergence and Manhattan distance measures to detect salient objects. SIViP 9(2015):427–435
26. Snowden RJ (2012) Visual attention to color: parvocellular guidance of attentional resources? Proc Psychol Sci 13:180–184
27. Vikram TN, Tscherepanow M, Wrede B (2012) A saliency map based on sampling an image into random rectangular regions of interest. Proc Pattern Recognit 45(9):3114–3124
28. Yang, C., Zhang, L., Lu, H., Ruan, X. and Yang, M.H., 2013. Saliency detection via graph-based manifold ranking. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3166–3173).
29. Yu Z, Wong HS (2007) A rule based technique for extraction of visual attention regions based on real time clustering. Proc IEEE Trans Multimedia 9:766–784
30. Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW (2008) SUN: a Bayesian framework for saliency using natural statistics. Proc J Vis 8:1–20
31. Zhang W, Wu QMJ, Wang G, Yin H (2010) An adaptive computational model for salient object detection. IEEE Trans Multimedia 12:300–315
32. Zhang D, Han J, Han J, Shao L (2016) Cosaliency detection based on intrasaliency prior transfer and deep intersaliency mining. IEEE transactions on neural networks and learning systems 27(6):1163–1176
33. Zhao, R, Ouyang W, Li H, and Wang X (2015) "Saliency detection by multi-context deep learning." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1265–1274
34. Zhu L, Klein DA, Frintrop S, Cao Z, Cremers AB (2014) A Multisize Superpixel approach for salient object detection based on multivariate normal distribution estimation. IEEE Trans Image Process 23:5094–5107

Navjot Singh is working as an Assistant Professor in National Institute of Technology, Uttarakand, India. He obtained M.Tech (Computer Science and Technology) and Ph.D. from Jawaharlal Nehru University, New Delhi, India. His current research areas are: Computer Vision, image processing, object detection, pattern recognition, feature extraction, and classification.

Rinki obtained M.Tech (Computer Science and Technology) from Jawaharlal Nehru University, New Delhi. Presently she is pursuing Ph.D. (Computer Science) from Jawaharlal Nehru University, New Delhi. Her current research areas are: Computer Vision, object detection, pattern recognition, and feature extraction.

R.K. Agrawal obtained M.Tech (Computer Application) from Indian Institute of Technology Delhi, New Delhi and Ph.D. (Computational Physics) from University of Delhi, Delhi. Presently, he is working as a Professor at the School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi. His current research areas are: Classification, feature extraction and selection for pattern recognition problems in domains of image processing, security, and bioinformatics.