

# Combination of Kullback–Leibler divergence and Manhattan distance measures to detect salient objects

Navjot Singh · R. K. Agrawal

Received: 3 July 2012 / Revised: 7 March 2013 / Accepted: 8 March 2013 / Published online: 2 April 2013  
© Springer-Verlag London 2013

**Abstract** Salient object detection is a computer vision technique that filters out redundant visual information and considers potentially relevant parts of our visual field. In this paper, we modify the Liu et al. model for salient object detection, which combines multi-scale contrast, center–surround histogram and color spatial distribution with conditional random fields. A combination of Symmetric Kullback–Leibler divergence and Manhattan distance instead of chi-square measure is employed to determine center–surround histogram difference. The modified Liu et al. model also uses a less computational intensive color spatial distribution map. To check the efficacy of the modified Liu et al. model, the performance is evaluated in terms of precision, recall,  $F$ -measure, area under curve and computation time. Experiment is carried out on a publicly available image datasets, and performance is compared with Liu et al. model and six other popular state-of-the-art models. Experimental results demonstrate that the modified Liu et al. model outperforms Liu et al. model and other existing state-of-the-art methods in terms of precision,  $F$ -measure, area under curve and has comparable performance in terms of recall with Liu et al. model.

**Keywords** Salient object detection · Kullback–Leibler divergence · Manhattan distance · Gaussian mixture model

## 1 Introduction

Salient object detection [2,3,13,15] is an active area of research which has drawn the attention of research community of computer vision and image processing in recent past. Visual saliency assists our ability to swiftly locate the vital information in an image. Such information can be a pixel, person or an object that stands out relative to its neighbors and grabs greater attention. The outcome of the salient object detection process is a map that contains the saliency value of each and every pixel in the image. An attention filter is generated from the saliency map that segregates the interesting information into a foreground object and ignores the irrelevant information that corresponds to the background object. Salient object detection is used for automatic target detection such as finding traffic signs [13,14] along the road or military vehicles in a savanna [14], in robotics to find salient objects in the environment as navigation landmarks, in image and video compression [14] by giving higher quality to salient objects at the expense of degrading background clutter, automatic cropping/centering [21] of images for display on small portable screens [4], finding tumors in mammograms [16], advertising a design [14], image collection browsing [20], image enhancement [7] and many more.

Various methods for obtaining visual attention have been suggested in the literature, which are broadly classified into two major categories [5,23]: bottom-up and top-down. A bottom-up visual attention is driven by low-level features (like intensity, color, orientation, etc.) in the scene. It extracts visual features from the image at multiple scales and combines them into a saliency map. Salient locations are identified using winner-take-all [13] and inhibition-of-return [13] operations. On the other hand, the top-down models [23] are task-dependent and exploit human observation behavior to achieve specific goals. They are always integrated with the

---

N. Singh (✉) · R. K. Agrawal  
School of Computer and Systems Sciences,  
Jawaharlal Nehru University, New Delhi, India  
e-mail: navjot.singh.09@gmail.com

R. K. Agrawal  
e-mail: rkajnu@gmail.com

bottom-up models and use a priori knowledge of the visual system to generate saliency maps for localizing objects of interest.

Our work is motivated by the model proposed by Liu et al. [18] which combined multi-scale contrast, center-surround histogram and color spatial distribution with conditional random fields. In the proposed model, we have extended the work of Liu et al. [18]. In this work, we have used a combination of Manhattan distance and symmetric Kullback–Leibler divergence (KLD), an information theoretic approach, instead of chi-square measure to determine center-surround histogram difference. Additionally, a less computational intensive color spatial distribution map is also used. To check the efficacy of the extended Liu et al. model, the performance is evaluated in terms of precision, recall,  $F$  measure, area under the curve and computation time. The experiment is carried out on a publicly available image datasets, and performance is compared with Liu et al. model [18] and six other popular state-of-the-art models.

The paper is organized as follows. Section 2 includes the state-of-the-art methods to obtain visually salient objects. The Liu et al. model and its modifications are discussed in Sect. 3. Experimental setup and results are included in Sect. 4. Conclusion and future work are presented in Sect. 5.

## 2 Related work

Itti et al. [13] proposed a neurobiological model that incorporated the center-surround difference of intensity, color and orientation features at multiple scales. It works on local features and not on the image as whole. Harel et al. [9] modeled the bottom-up visual attention by employing graph theoretic ideas to determine activation maps from the raw features. The model gives high saliency values to the nodes which are at the center of the image. However, the use of a fully connected directed graph for every feature map makes the model more complex. Han et al. [8] integrated the Itti et al. model with Markov random field and region growing techniques to extract attention objects. Le Meur et al. [19] employed visibility, perception and perceptual grouping in their model. However, the model considered the achromatic structure in general and gave unclear boundaries. Yu and Wong [22] used a real-time clustering algorithm for image segmentation at the grid cell level. The model gives good initial centers of the clusters in lesser number of iterations but highly depends on the image segmentation accuracy. In Liu et al. [18] model, a generic salient object is separated from the image background. The features are extracted at the local, regional and global level. The local feature is obtained by computing the contrast information of a pixel in a given neighborhood at different levels of details. The regional feature is made up of a center-surround histogram map. The global feature is

represented in terms of a color spatial distribution map using Gaussian mixture models (GMM). To combine these features into a saliency map, linear weights are determined using conditional random field (CRF) learning under maximum likelihood (ML) criteria. Gao and Vasconcelos [6] used the discriminant saliency concept which is based on center-surround mechanism to detect salient object. Klein and Frin-trop [17] computed the KLD distance between the center and surround regions of an image for different feature channels and then fused them into the saliency map. Hou and Zhang [11] extracted the spectral residual of an image by analyzing its log-spectrum. Achanta et al. [1] computes the saliency map by subtracting the Gaussian blurred version of the image from the mean pixel value of the image. Hou and Zhang [12] used the incremental coding length to measure the perspective entropy gain of each feature channel to detect salient object. Itti and Baldi [15] gave the notion of Bayesian surprise theory and computed the difference between the posterior and prior beliefs to locate the surprising items. Zhang et al. [23] combined the position, area and intensity saliencies using Bayesian framework and Gaussian mixture models. The model can detect objects even if the image segmentation results are not accurate. The model fails when the intensity difference between the object and background is less. Bruce and Tsotsos [3] proposed a visual saliency model based on information maximization.

## 3 Modified Liu et al. model

Liu et al. [18] proposed a model which involves image segmentation or binary labeling where salient object is to be separated from the image background. The primary objective of their work was to detect a generic salient object instead of a specific object category. To achieve this, a set of features at the local, regional and global levels were incorporated in the model. Dyadic Gaussian image pyramids [13] were formed which gives the contrast information at different levels of details ranging from the finer to the coarser details. These details were linearly combined into a multi-scale contrast feature map [18] that depicts the object at the local level and is defined as

$$f_c(x, I) = \sum_{l=1}^L \sum_{x' \in N(x)} \left\| I^l(x) - I^l(x') \right\|^2 \quad (1)$$

where  $x$  is the pixel in the image  $I$ ,  $N(x)$  is a  $9 \times 9$  neighborhood,  $I^l$  is the  $l$ -th level image in the dyadic Gaussian pyramid and  $L$  is set to 6.

Regional feature includes the computation of a center-surround histogram [18] map. In the center-surround histogram each and every pixel is enclosed by a center rectangle  $R_c$ . Then, a surrounding rectangle  $R_s$  having the same area

is constructed around  $R_c$ . The size range of these rectangles [18] is set to  $[0.1, 0.7] \times \min(W, H)$ , where  $W$  is the width and  $H$  is the height of the image. A set of aspect ratios  $\{0.5, 0.75, 1.0, 1.5, 2.0\}$  is used to select the most distinct rectangle at each pixel. This can be achieved by employing chi-square distance between the histograms of a center rectangle  $R_c$  and surrounding rectangle  $R_s$  given by

$$\chi^2(R_c, R_s) = \frac{1}{2} \sum_{i=1}^b \frac{(R_c(i) - R_s(i))^2}{R_c(i) + R_s(i)} \tag{2}$$

where  $R_c(i)$  and  $R_s(i)$  are the histograms of RGB color for center and surround rectangles respectively,  $b = 1000$  is the number of bins. Now, the difference between the histograms of the two rectangles will give the amount of dissimilarity. The higher is the difference, the higher is the dissimilarity, and the more is that pixel salient with respect to its surroundings. So for every pixel  $x$  in the image, the most distinct rectangle  $R_c^*$  is computed as

$$R_c^*(x) = \operatorname{argmax}_{R_c(x)} \chi^2(R_c(x), R_s(x)) \tag{3}$$

Finally, the center-surround histogram feature [18] is given by

$$f_h(x, I) \propto \sum_{\{x'|x \in R_c^*(x')\}} w_{xx'} \cdot \chi^2(R_c^*(x'), R_s^*(x')) \tag{4}$$

where  $w_{xx'} = \exp(-0.5\sigma_{x'}^{-2} \|x - x'\|^2)$  is the Gaussian falloff weight with variance  $\sigma_{x'}^2$  set to one-third the size of  $R_c^*(x)$ .

Color spatial distribution [18] is used as a global feature. For this, an image is initially clustered into  $C$  colors (clusters) using k-means algorithm. These colors are then represented by Gaussian Mixture Models (GMMs)  $\{\omega_c, \mu_c, \Sigma_c\}_{c=1}^C$  whose weight, mean and covariance matrix are estimated by the Expectation Maximization algorithm. Experimentally,  $C$  is set to 6. The probability of a pixel  $x$  to be assigned to the  $c$ -th color component is given by

$$p(c|I_x) = \frac{\omega_c \mathfrak{N}(I_x|\mu_c, \Sigma_c)}{\sum_c \omega_c \mathfrak{N}(I_x|\mu_c, \Sigma_c)} \tag{5}$$

The color spatial distribution feature map [18] is calculated as

$$f_{csd}(x, I) \propto \sum_c p(c|I_x) \cdot (1 - V(c)) \cdot (1 - D(c)) \tag{6}$$

where  $V(c)$  is the spatial variance [18] in both horizontal as well as vertical direction,  $D(c) = \sum_x p(c|I_x) d_x$  is the weight and  $d_x$  is the distance between the pixel  $x$  and the central pixel.

These three features were normalized to  $[0,1]$  and linearly combined using conditional random field (CRF). The block diagram of Liu et al. model is shown in Fig. 1. A few basic

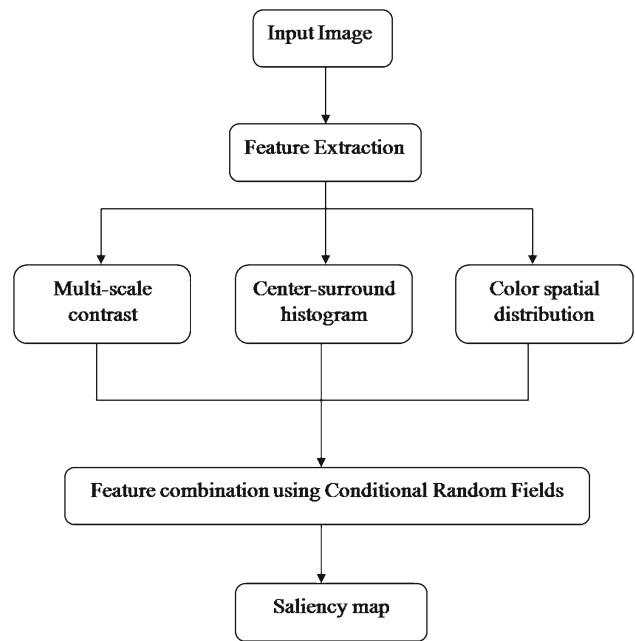


Fig. 1 Block diagram of the Liu et al. model

modifications have been done in the model suggested by Liu et al. [18]. These modifications are discussed underneath.

### 3.1 Center-surround histogram feature based on combination of Manhattan distance and Kullback-Leibler divergence

Liu et al. [18] employed chi-square test to compute the difference between center and surround histograms. It is well known that the chi-square test is sensitive to sample size of pixels. It is also sensitive to small expected frequencies in one or more of the intervals in the histogram. Hence, the chi-square test does not give much information about the strength of the dissimilarity. In the literature, many distance measures or metric have been suggested to capture the dissimilarity between the two tuples or data distributions. Among them, Manhattan distance (or L1 norm) and L2 norm are most commonly used. Both Manhattan distance [10] and L2 norm [10] satisfy triangular inequality. However, Manhattan distance is more robust to outliers than L2 norm. Also, Manhattan distance is effective when the data are highly sparse. The limitation of Manhattan distance and L2 norm is that a high value in a single histogram bin dominates the distance between two histograms. In the literature, KLD [10], an information theoretic approach, is successfully used in many applications as a dissimilarity measure. It is an information divergence and measures the difference between two probability distributions. KLD does not satisfy symmetric and triangular inequality properties. However, KLD has the following unique properties [10]: (i) invariant to the permutation of the order in which the components of the vector are arranged, (ii)

amplitude scaling, (iii) monotonic nonlinear transformation, (iv) less sensitive to noise, (v) generalizes easily to the multivariate case where equivalence on more than one parameter is required and (vi) applicable over a wide range of distributions of the response variable. Both Manhattan distance and KLD have their own advantages and disadvantages. The combination of Manhattan distance and KLD may bring advantages of both and thus can be a suitable choice to measure dissimilarity between the center and surround histograms.

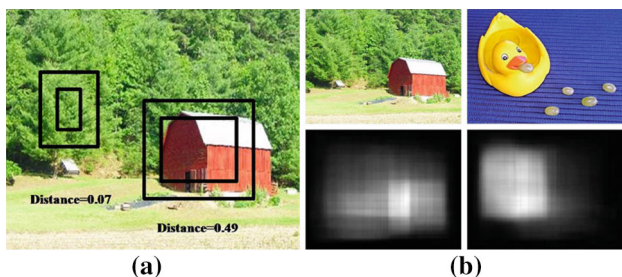
In order to compute the dissimilarity between the histograms of the two rectangles  $R_c$  and  $R_s$ , the proposed distance measure (CSKLMD) which is a combination of symmetric KLD and Manhattan Distance can be used which is given by

$$D_{\text{CSKLMD}}(R_c, R_s) = \frac{1}{2} \sum_{i=1}^b (R_c(i) - R_s(i)) \log \left( \frac{R_c(i)}{R_s(i)} \right) + \sum_{i=1}^b |R_c(i) - R_s(i)| \quad (7)$$

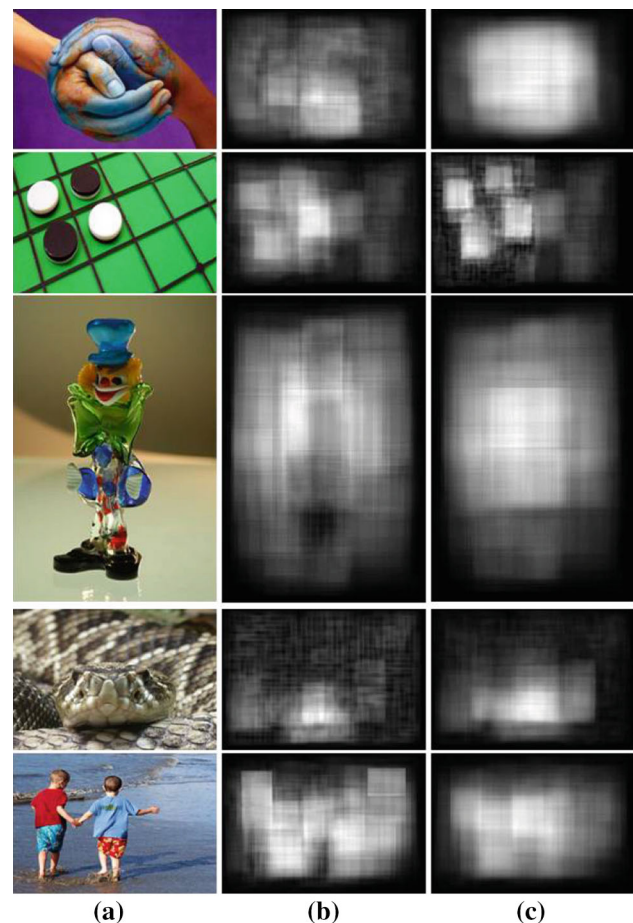
where  $b = 1000$  is the number of bins and as per convention  $0 \log(0) = 0$  and  $0 \log(0/0) = 0$ . Figure 2 shows the center and surround rectangles for some images with their corresponding center–surround histogram feature maps. Figure 3 shows the comparison of center–surround histogram feature map for some images using chi-square and CSKLMD.

### 3.2 Area concept integrated color spatial distribution feature

Liu et al. [18] model involves computation of the color spatial distribution map by taking the weighted sum of the all colormaps generated by the Gaussian mixture models (GMM). However, it does not consider the fact that in general the salient object or region of interest in the image with the maximum area is sufficient to describe the color spatial distribution map. Rest of the connected components has lower color spatial distribution saliency values and is just an overhead to the computational cost.



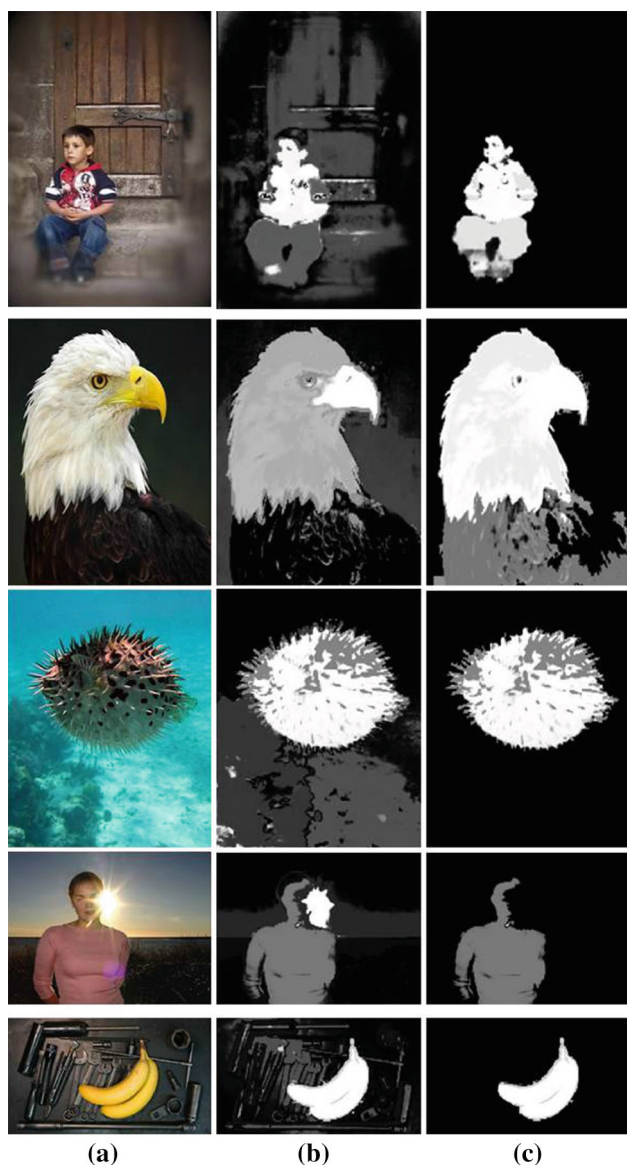
**Fig. 2** **a** Center and surround rectangles for different objects with their CSKLMD distance. **b** Original images with their corresponding center–surround histogram maps



**Fig. 3** **a** Original image. **b** Center–surround histogram map by chi-square measure. **c** Center–surround histogram map by CSKLMD

From Eq. (5), we get  $C$  colormaps. These colormaps undergo a two-stage post-processing. In first stage, the holes are filled by using morphological operation. A hole is a set of background pixels that lie within the object and can be of any size. Then, a number of connected components are determined using 8-pixel connectivity and are labeled, with label 0 as the background, label 1 for object 1, label 2 for object 2 and so on. We have discarded the component labeled 0 as it is the background. Now, we are left with object 1, object 2, etc. In the second stage, we chose only that colormap which contains the connected component having the maximum area. Here, area is the number of pixels used to constitute an object. In most of the images, background covers approximately half of the image area. After the background is removed, we are left with many small objects. We pay more heed to object that has the maximum area among them. This is done to avoid those connected components that have very low saliency values. So instead of combining all the  $C$  colormaps, we only chose that colormap which contains the connected component with the maximum area to represent as color spatial distribution map. The results of color spatial distribution map for the modified





**Fig. 4** **a** Original image. **b** Color spatial distribution feature map of Liu et al. method. **c** Color spatial distribution feature map of the proposed method (color figure online)

Liu et al. model and the Liu et al. [18] method can be seen in Fig. 4.

#### 4 Experimental setup and results

In order to check the efficacy of the modified model of Liu et al., we compared its performance with Liu et al. [18] method and six other state-of-the-art methods. We modified the Liu et al. model in two ways. The first model (Modified\_model\_1) employs CSKLM instead of chi-square distance measure to compute center-surround histogram feature map. While in the second model (Modified\_model\_2), the modifications are done both in terms of distance measure (CSKLM instead of chi-square) and color spatial distribution map.

##### 4.1 Salient object database

The MSRA SOD (Microsoft Research Asia Salient Object Database) database<sup>1</sup> has been used to evaluate these eight models, both subjectively and objectively. The database contains high-quality images of various object categories and scene types. The MSRA salient object database includes two datasets: MSRA SOD image set A which contains 20843 color images in 71 subfolders with their ground truth manually labeled by three users and MSRA SOD image set B which contains 5000 color images in 10 subfolders with their ground truth manually labeled by nine users. All the images are of size  $400 \times 300$  or  $300 \times 400$  having intensity values  $[0,255]$ . Three thousand images are randomly chosen from image set A for training and 5000 images of image set B for testing. The test dataset is used for performance evaluation. All the experiments are carried out using Windows 7 environment over Intel(R) Xeon(R) processor with a speed of 2.27 GHz and 4 GB RAM.

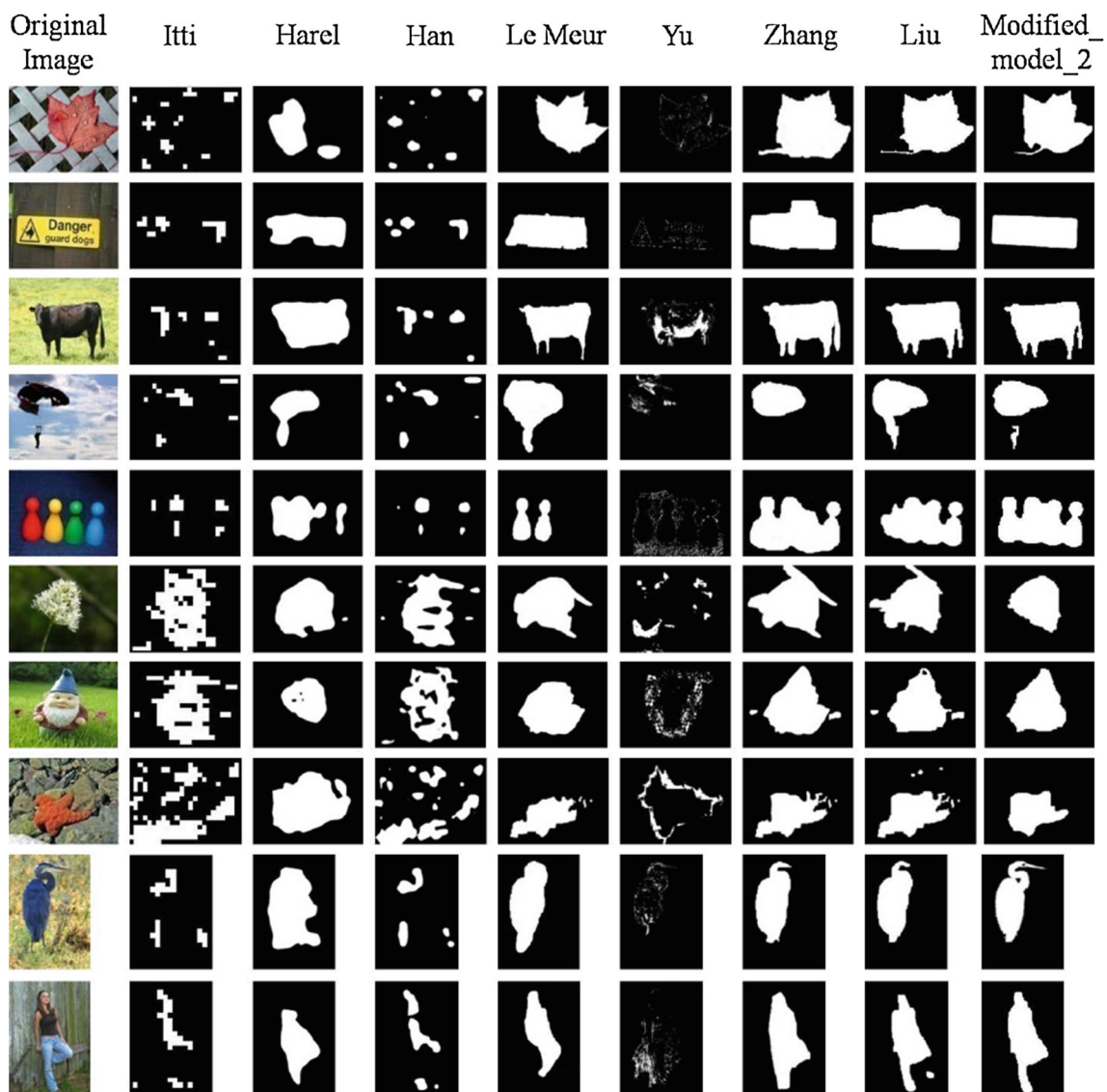
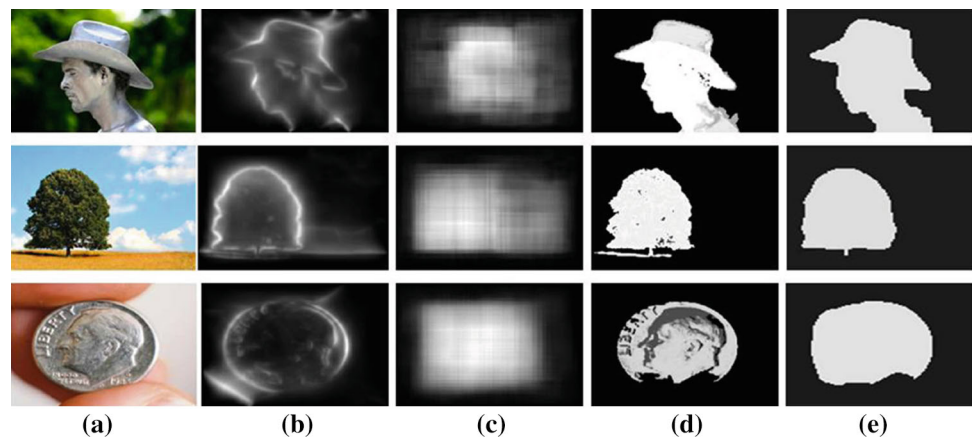
The saliency maps are generated after combining the multi-scale contrast, center-surround histogram and color spatial distribution feature maps according to our proposed modifications. Then, a threshold is used to obtain attention masks from the saliency maps as shown in Fig. 5. The comparison of the attention masks of the Modified\_model\_2 with other state-of-the-art models is shown in Fig. 6. The objects chosen differ in size, shape, color, texture, etc.

The following observations regarding the attention masks can be drawn from Fig. 6:

- Itti et al. [13] method gave disappointing results as it considered the local features and neglected the global one.
- Harel et al. [9] method which is an extended work of Itti et al. gave better saliency maps but the shape information was poor.
- Han et al. [8] used the concept of region growing, and the seed determination may not be done correctly, so it gave unsatisfactory results.
- Le Meur et al. [19] method was far better than the above three bottom-up models, but it missed the finer shape details. Their model provides smooth contours and lacked detailed information of object silhouettes.
- Yu and Wong [22] model depends on the accuracy of image segmentation. A bad segmentation furnished inadequate results.
- Zhang et al. [23] saliency results were very close to the Liu et al. but it fails when the intensity difference between the object and the background is low.

<sup>1</sup> [http://www.research.microsoft.com/en-us/um/people/jiansun/salientobject/salient\\_object.htm](http://www.research.microsoft.com/en-us/um/people/jiansun/salientobject/salient_object.htm).

**Fig. 5** **a** Original image. **b** Multi-scale contrast map. **c** Center-surround histogram map. **d** Color spatial distribution feature map. **e** The resultant attention mask (color figure online)



**Fig. 6** Comparison of the proposed method with state-of-the-art methods based on their attention masks generated after thresholding the saliency maps

- Liu et al. [18] method gave better saliency results but it bestowed superfluous information of the object that was not required.
- The Modified\_model\_2 gave the best saliency results. The shape information is clear and fine.

#### 4.2 Ground truth construction and quantitative evaluation

In order to get an objective evaluation, the performance of the proposed method is also evaluated quantitatively. The qualitative measures used are precision, recall, *F*-measure, area under curve (AUC) and computational time. Since the ground truth is represented as rectangular region in the database, so the processed salient object must also be enclosed within a rectangle. Using the ground truth rectangle *G* and the detection results rectangle *R*, precision, recall and *F*-measure are calculated as

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} & \text{Recall} &= \frac{TP}{TP + FN} \\ F_\alpha &= \frac{(1 + \alpha) \times \text{Precision} \times \text{Recall}}{\alpha \times \text{Precision} + \text{Recall}} \end{aligned} \tag{8}$$

where  $\alpha = 1$  and

$$\begin{aligned} TP &= \sum_{G(x,y)=1} R(x,y) & FP &= \sum_{G(x,y)=0} R(x,y) \\ FN &= \sum_{R(x,y)=0} G(x,y) \end{aligned} \tag{9}$$

where

TP (true positives) is the number of salient pixels that are detected as salient object.

FP (false positives) is the number of background pixels that are detected as salient object.

FN (false negatives) is the number of salient pixels that are detected as background objects.

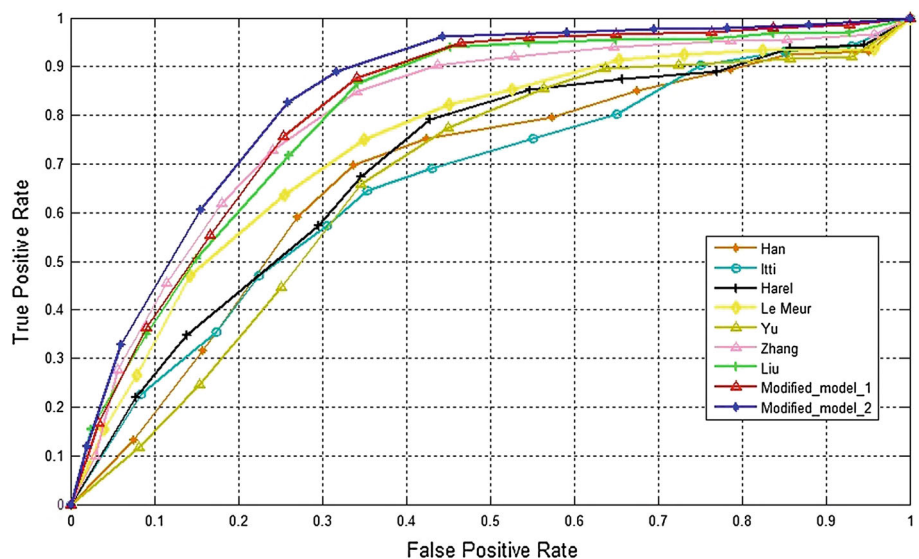
The results of the nine users are averaged. A receiver operator characteristic (ROC) curve is drawn to compute AUC. Figure 7 shows the ROC curve between the true positive rate (TPR) and the false positive rate (FPR). TPR and FPR are given by

$$\begin{aligned} TPR &= \frac{TP}{\sum_{(x,y)} G(x,y)} \\ FPR &= \frac{FP}{W \times L - \sum_{(x,y)} G(x,y)} \end{aligned} \tag{10}$$

In methods [9,13,18,19] and [23] and the proposed method, a threshold is used on the saliency map to obtain an attention mask. This threshold is adjusted to depict the ROC plots. For method [8] the parameter  $\delta$  and for [22]  $\epsilon$  is tuned to plot the ROC. AUC is calculated to measure the efficiency of the different models. Table 1 shows the qualitative performance evaluation of the proposed method in comparison with the other state-of-the-art methods. The best results are shown in bold. We observed the following from Table 1 and Fig. 7:

- The performance of Modified\_model\_1 is improved in terms of precision, *F*-measure and AUC and deteriorated in terms of recall in comparison with Liu et al. model.
- The performance of Modified\_model\_2 is slightly improved in terms of precision, recall, *F*-measure and AUC in comparison with Modified\_model\_1.
- The Modified\_model\_2 gave the highest precision because the shape information was fine and did not consider additional information of the object like shadow or some neighboring portion of the object’s background.

**Fig. 7** Comparison of ROC curve of proposed method with existing state-of-the-art methods



**Table 1** Quantitative performance of the proposed method and the state-of-the-art methods

	Itti [13]	Harel [9]	Han [8]	Le Meur [19]	Yu [22]	Zhang [23]	Liu [18]	Modified_model_1	Modified_model_2
Precision	0.672	0.800	0.688	0.686	0.745	0.785	0.674	0.792	<b>0.806</b>
Recall	0.614	0.692	0.751	0.860	0.721	0.733	<b>0.889</b>	0.829	0.874
<i>F</i> -measure	0.642	0.742	0.718	0.763	0.733	0.758	0.767	0.810	<b>0.817</b>
AUC	0.6627	0.6976	0.6754	0.7413	0.6686	0.7983	0.8019	0.8145	<b>0.8403</b>
Time (in seconds) per image	1.70	59.8	12.4	1.90	<b>1.60</b>	4.31	25.7	25.4	21.6

- With the perfect shape information, the detection rectangle for some images becomes smaller than the ground truth rectangle. This resulted into a slightly lower recall of the Modified\_model\_2 in comparison with Liu et al. [18] method.
- Maximum value of *F*-measure is obtained for the Modified\_model\_2 which is the weighted harmonic mean of precision and recall.
- It is well known that the model that covers the maximum area under the ROC curve is better in terms of performance. The Modified\_model\_2 gave the highest AUC value.
- The use of information theoretic approach makes the Modified\_model\_2 computationally more expensive than the other models but is better than Liu et al. [18] method and Harel et al. [9] method.

## 5 Conclusion and future work

In this paper, we proposed two modifications of Liu et al. [18] model for salient object detection, which combined multi-scale contrast, center-surround histogram and color spatial distribution with conditional random field. The Modified\_model\_1 uses the combination of symmetric KLD and Manhattan distance instead of chi-square measure to determine the center-surround histogram difference. The Modified\_model\_2 also uses a less computational intensive color spatial distribution map. To check the efficacy of the Modified\_model\_2, the performance is evaluated in terms of precision, recall, *F*-measure, AUC and computation time. The experiment is carried out on a publicly available image datasets, and performance is compared with Liu et al. model and six other popular state-of-the-art models. Experimental results demonstrate that the proposed model (Modified\_model\_2) outperforms Liu et al. model and other existing state-of-the-art methods in terms of precision, *F*-measure, AUC and gives comparative performance in terms of recall with Liu et al. model. However, the Modified\_model\_2 is found computationally more expensive than the other models but is better than Liu et al. [18] method and Harel et al. [9] method.

One of the most challenging tasks is to examine the interaction between attention and object recognition. A system can profit from the combination of an attentional front-end and a recognition back-end. Foreground items can be portrayed by means of geometrical models that take into consideration the chromatic and structural characteristics of the items. In future, nonlinear combination of the features can be investigated to evaluate the performance. The model with non-rectangular regions can also be investigated to improve the detection. The work needs to be extended to detect any number of salient objects or no salient object at all. Multi-level observation models need to build for complex scenes incorporating the face detection and human skin detection framework.

**Acknowledgments** The authors are indebted to the reviewers for their constructive suggestions which significantly helped in improving the quality of this paper. In addition, the first author expresses his gratitude to the University Grant Commission (UGC), India, for the obtained financial support in performing this research work.

## References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: Proceedings of International Conference on Computer Vision and, Pattern Recognition, pp. 1597–1604 (2009)
2. Borji, A., Sihite, D.N., Itti, L.: Salient object detection: a benchmark. In: Proceedings of the European Conference on Computer Vision, pp. 414–429 (2012)
3. Bruce, N., Tsotsos, J.: Saliency, attention, and visual search: an information theoretic approach. *Proc. J. Vis.* **9**(3), 1–24 (2009)
4. Chen, L., Xie, X., Fan, X., Ma, W., Shang, H., Zhou, H.: A visual attention model for adapting images on small displays. Technical report, Microsoft Research Redmond(2002)
5. Frintrop, S., Rome, E., Christensen, H.I.: Computational visual attention systems and their cognitive foundations: a survey. *Proc. ACM Trans. Appl. Percept.* **7**, 1–39 (2010)
6. Gao, D., Vasconcelos, N.: Bottom-up saliency is a discriminant process. In: Proceedings 11th International Conference on Computer Vision, pp. 1–6 (2007)
7. Gasparini, F., Corchs, S., Schettini, R.: Low quality image enhancement using visual attention. *Proc. Opt. Eng.* **46**(4), 040502-1–040502-3 (2007)
8. Han, J., Ngan, K.N., Li, M.J., Zhang, H.J.: Unsupervised extraction of visual attention objects in color images. *Proc. IEEE Trans. Circuits Syst. Video Technol.* **16**(1), 141–145 (2006)



9. Harel, J., Koch, C., Perona, P.: Graph based visual saliency. *Proc. Adv. Neural Inf. Process. Syst.* **15**, 545–552 (2007)
10. Haykin, S.: *Neural Networks: A Comprehensive Foundation*. Pearson Prentice Hall, Englewood Cliffs, NJ (2005)
11. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8 (2007)
12. Hou, X., Zhang, L.: Dynamic visual attention: searching for coding length increments. *Proc. Adv. Neural Inf. Process. Syst.* **21**, 681–688 (2008)
13. Itti, L., Koch, C., Niebur, E.: A model of saliency based visual attention for rapid scene analysis. *Proc. IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998)
14. Itti, L.: *Models of bottom up and top down visual attention*. Ph.D. dissertation, California Institute of Technology, Pasadena (2000)
15. Itti, L., Baldi, P.: Bayesian surprise attracts human attention. *Proc. Visi. Res.* **49**(10), 1295–1306 (2009)
16. Karssemeijer, N.: Detection of stellate distortions in mammograms. *Proc. IEEE Trans. Med. Imaging* **15**, 611–619 (1996)
17. Klein, D.A., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: *Proceedings of the International Conference on Computer Vision*, pp. 2214–2219 (2011)
18. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. *Proc. IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 353–366 (2011)
19. Meur, O.L., Callet, P.L., Barba, D., Thoreau, D.: A coherent computational approach to model bottom up visual attention. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**, 802–817 (2006)
20. Rother, C., Bordeaux, L., Hamadi, Y., Blake, A.: Autocollage. In: *Proceedings of ACM SIGGRAPH*, pp. 847–852 (2006)
21. Santella, A., Agrawala, M., Decarlo, D., Salesin, D., Cohen, M.: Gaze based interaction for semi automatic photo cropping. In: *Proceedings of the Conference on Human Factors in Computing Systems*, pp. 771–780 (2006)
22. Yu, Z., Wong, H.S.: A rule based technique for extraction of visual attention regions based on real time clustering. *Proc. IEEE Trans. Multimed.* **9**, 766–784 (2007)
23. Zhang, W., Wu, Q.M.J., Wang, G., Yin, H.: An adaptive computational model for salient object detection. *Proc. IEEE Trans. Multimed.* **12**, 300–315 (2010)