





Saliency Map Improvement Using Edge-Aware Filtering

Diptiben Patel^(✉)  and Shanmuganathan Raman 

Electrical Engineering, Indian Institute of Technology Gandhinagar,
Gandhinagar 382355, Gujarat, India
{diptiben.patel, shanmuga}@iitgn.ac.in

Abstract. Content-aware applications in computational photography define the relative importance of objects or actions present in an image using a saliency map. Most saliency detection algorithms learn from the human visual system and try to find relatively important content as a salient region(s). This paper attempts to improve the saliency map defined by these algorithms using an iterative process. The saliency map of an image generated by an existing saliency detection algorithm is modified by filtering the image after segmenting into foreground and background. In order to enhance the saliency map values present in the salient region, the background is filtered using an edge-aware guided filter and the foreground is enhanced using a local Laplacian filter. The number of iterations required varies according to the image content. We show that the proposed framework enhances the saliency maps generated using the state-of-the-art saliency detection algorithms both qualitatively and quantitatively.

1 Introduction

The human visual system and the brain are remarkably fast in processing visual information in a fraction of time. The human visual system makes this process faster by focusing on a “distinctive and attentive” object or action and processing it first before the other regions. The human eye is fixated to a distinctive region with higher priority and spends much of the processing time on it as compared to the other non-distinctive regions. The distinctive region(s) is/are named as salient region(s) and the map describing its distinctiveness is called saliency map in computer vision. The goal of saliency detection algorithms is to estimate the fixation of a human eye according to the distinctiveness. The estimated saliency map is used in different applications to mimic the human visual system such that the modification of the scene induces as less visible artifacts as possible. The estimated saliency map is used in many computer vision applications such as image segmentation [1], object detection [2], image compression [3] and image enhancement [4], to name a few. The efficiency of these applications depends on the accuracy of the underlying saliency detection algorithm.

Rather than improving the existing saliency detection algorithms, can we improve the saliency maps generated by the state-of-the-art saliency detection

algorithms using some iterative process? In this work, the values in terms of the saliency map generated by a saliency detection algorithm are improved in the salient region of an image by modifying the original image in an iterative manner. Most of the saliency detection algorithms look for local contrast or edge details at first stage to estimate the saliency map of an image. Therefore, if the local contrast of an image is modified such that the edge details in the non-salient region are suppressed and those present in the salient region are enhanced, it can enable us to modify and improve the saliency map of an image. This motivation is behind the proposed framework to enhance the saliency map of an image generated using the existing saliency detection algorithms. The number of iterations required for the improvement in a given saliency map depends on the image content making it an adaptive algorithm.

The paper is organized as follows. Section 2 surveys various saliency detection algorithms and different criteria considered for defining saliency. The framework for improvement in the saliency map generated by existing saliency detection algorithms is proposed in Sect. 3. The results and the comparisons are discussed in Sect. 4. The paper is concluded with some pointers to future work in Sect. 5.

2 Related Work

Saliency detection algorithms use different features of an image to estimate the human eye fixations. In late 90's, Itti *et al.* have defined the saliency map using early visual features comprising of image intensity contrast, color contrast, and local orientation contrast computed at different Gaussian scales [5]. Harel *et al.* have proposed a graph-based visual saliency model in feature space using bottom-up approach [6]. The above-mentioned saliency detection algorithms do not take care of exact boundaries of salient regions. Achanta *et al.* have proposed frequency tuned saliency detection algorithm providing exact boundaries of salient regions imparting more frequency content across the boundaries using color and luminance features [7]. The saliency map of an image is estimated using local feature color contrast, sparse sampling, kernel density estimation and Bayesian theory model in [8]. Liu *et al.* have proposed learning based salient object detection algorithm which trains Conditional Random Field (CRF) algorithm using features: multi-scale contrast (local), center-surround histogram (regional), and color spatial distribution (global) [9]. While detecting eye fixations or salient regions, the context of the salient region is lost. Using local contrast, global features, visual organization rules, and some of the high-level features, the context of an image alongside salient region is preserved in [10]. Cheng *et al.* have proposed a saliency detection algorithm by combining histogram based contrast and region based contrast [11]. Murray *et al.* have proposed a saliency model based on low-level vision system at multi-scale decomposition using color and luminance channels [12]. Li *et al.* have proposed a learning based saliency detection algorithm which combines eye fixation and segmentation models in order to perform segmentation of the salient objects [13]. In [14], salient objects and distractors are separated by learning the distribution of projected features using principal component analysis. Borji *et al.* have proposed a

patch based saliency detection algorithm which defines saliency of patches based on how they are different from surrounding patches and how often they occur in an RGB and Lab color space of an image [15]. Another patch based saliency detection algorithm defines the saliency of a patch using the distance from an average of all patches in the color space and the pattern space along the principal component directions [16]. The weighted dissimilarity between patches using multiple parameters is used to form a saliency map in [17].

Improving the saliency map generated by the saliency detection algorithms is a novel research area. Lei *et al.* have proposed a framework using Bayesian decision theory after finding rough saliency map from different saliency detection algorithms [18]. They enhanced the saliency map of an image using a conditional probability of pixels having similar color values with that of pixels with higher saliency value in rough saliency map. This framework will fail if the salient object contains different colors and is not captured in rough saliency map. Alternatively, we have proposed to improve the saliency map by modifying the image in order to enhance the saliency value present in the salient region in an iterative manner. From the saliency map generated using an existing saliency detection algorithm, foreground and background regions are found using image segmentation. The image is modified differently for foreground and background regions iteratively. This enables us to enhance the saliency values in the salient (foreground) region and to suppress the saliency values in the non-salient (background) region.

3 Proposed Approach

The purpose of saliency detection algorithm is to look for distinctive regions in an image where human eye fixates. The more the distinctiveness of a pixel, the more the saliency value assigned to a pixel. Most saliency detection algorithms fail to properly distinguish distinctive regions and sometimes assign the same value to a region in an image. Sometimes, they may not be able to assign consistent saliency value to the same object in an image. Our goal is to improve and make the saliency values concentrated in distinctive regions by enhancing the energy present in the distinctive regions iteratively. This can be achieved by smoothing out the non-distinctive region (background) and coarsening the distinctive region (foreground) while generating the saliency map in an iterative manner.

3.1 Methodology

The proposed framework for an improvement in a given saliency map S_0 of image I_0 is described below. Let the image I_0 be of size $M \times N$.

1. The saliency map of an image I_i generated using an existing saliency detection algorithm is S_i . The energy E_i of a saliency map S_i is defined as a squared sum of gray level co-occurrence matrices (GLCM [19]) C_{θ_j} in 4 directions 0° , 45° , 90° , and 135° with the distance between given two pixels being 1 and is shown in Eq. (1).

$$E_i = \sum_{a=1}^P \sum_{b=1}^P \left(\sum_{j=1}^4 C_{\theta_j}(a, b) \right)^2 \quad (1)$$

$$C_{\theta_j}(a, b) = \sum_{x=1}^M \sum_{y=1}^N \begin{cases} 1, & S_i(x, y) = a \ \& \\ & S_i(x + x_{\theta_j}, y + y_{\theta_j}) = b \\ 0, & \text{Otherwise} \end{cases}$$

here, $i = 0, 1, \dots, K$, $x = 1, 2, \dots, M$, $y = 1, 2, \dots, N$, $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, $x_{\theta_j} \in \{-1, 0, 1\}$, $y_{\theta_j} \in \{-1, 0, 1\}$, K is the number of iterations, P is number of intensity levels in an image and E_i is the energy of the saliency map after the i^{th} iteration.

2. The image I_i is segmented into foreground (FG_i) and background (BG_i) using kernel k-means algorithm described in [20]. The segmentation method proposed in [20] requires rectangular box R_i as a seed that includes foreground region (FG_i) which is a salient region in our case. We find this rectangular box R_i from binary map P_i using the saliency map S_i as shown in Eq. (2).

$$P_i(x, y) = \begin{cases} 1, & \text{if } S_i(x, y) \geq \text{average}(S_i) \\ 0, & \text{Otherwise.} \end{cases} \quad (2)$$

R_i of size $M \times N$ is a minimum area rectangle which contains all the 1's in P_i . The segmentation method outputs an image G_i with foreground region intact and background region with values 1. We generate the binary mask B_i (indicating foreground region) using G_i as shown in Eq. (3).

$$B_i(x, y) = \begin{cases} 1, & \text{if } G_i(x, y) \neq 1 \\ 0, & \text{Otherwise.} \end{cases} \quad (3)$$

3. As our goal is to propagate energy towards the salient region (foreground), we exaggerate the foreground region and apply smoothing on the background region of the image. We use local Laplacian filter T_l described in [21] to exaggerate the details in the foreground with $\sigma_r = 0.1$. The low pass filtering of the background region is performed using the guided filter T_g with filter size $w = 5$ and the guide image to be the same as the input image to be filtered. The modified image $I_{(i+1)}$ is generated by combining differently filtered foreground and background regions and it is as described in Eq. (4).

$$I_{(i+1)} = FG_{(i+1)} + BG_{(i+1)} \quad (4)$$

$$= (B_i \times T_l(I_i, \sigma_r)) + ((1 - B_i) \times T_g(I_i, I_i, w))$$

4. We repeat the steps 1–3 iteratively with the modified input image $I_{(i+1)}$. The summarized block diagram of iterative process is shown in Fig. 1. The number of iterations to be performed is dependent on the image content and is estimated as described in Sect. 3.2.

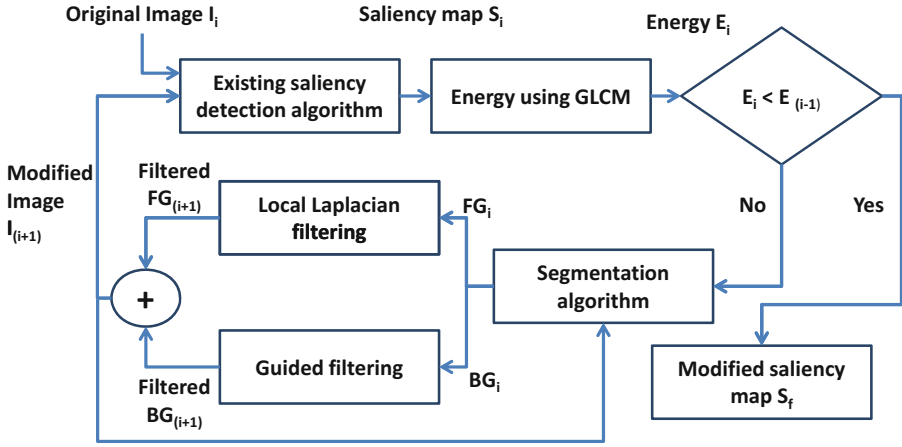


Fig. 1. The proposed saliency improvement framework.

3.2 Optimal Number of Iterations

The energy variation of a saliency map of modified images over the number of iterations is shown in Fig. 2. Observing the energy variations over the number of iterations, the energy E_i starts decreasing after some iteration. During initial iterations, as the smoothing of a background region tends to gain a constant intensity value, energy value E_i increases to achieve a value of 1. (Energy of a gray co-occurrence matrix of a constant intensity image is 1). After some iteration, exaggeration of foreground region starts dominating the effect of background region on energy and detail enhancement in foreground region reduces the energy value. As more enhancement of details in the foreground leads to saturation of intensity values, we stop the iterations when energy values start decreasing. The

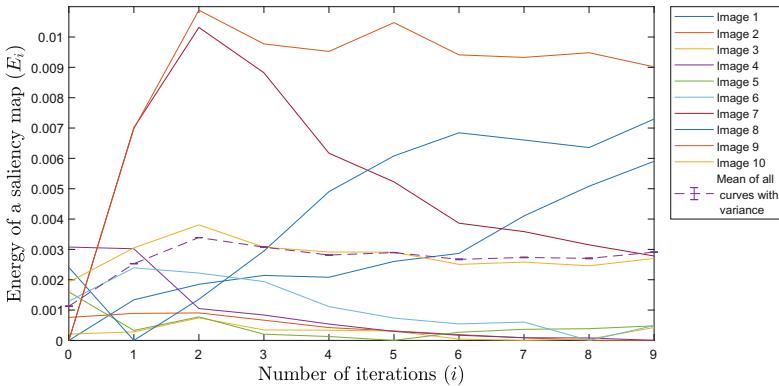


Fig. 2. Energy of the improved saliency map as a function of number of iterations. Each curve shows energy variation of a saliency map of modified images over the iterations.

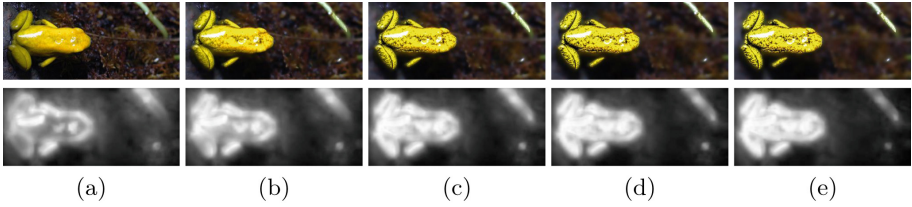


Fig. 3. Effect of iterative process on saliency map: Top row: modified images: (a) Original image, (b)–(e) modified images after $i = 1, 2, 3$ and 4 iterations, Bottom row: saliency maps of the corresponding modified images using existing saliency detection algorithm by [10].

same can be observed from Fig. 3. Figure 3 shows the modified images and their saliency maps after each iteration. From Fig. 3, it is observed that after a certain number of iterations, the intrusion of constant values in the background region and the saturation of exaggerated edges in the foreground region lead to decrease in the energy of a saliency map. Hence, the iterative process has to be stopped as soon as the energy value starts decreasing. The saliency map (S_f) of the modified image after the last iteration provides the improved version of the saliency map (S_0) of the original image. In this way, the number of iterations required for an improvement in the saliency map is adaptive depending on the given image.

4 Results and Discussions

We have tested the proposed framework on the MSRA salient object dataset [9] and compared the improvement in the saliency maps with their corresponding original saliency maps. The comparison is performed using several existing state-of-the-art saliency detection algorithms such as graph-based visual saliency (GBVS) approach [6], spatially weighted dissimilarity (SWD) based approach [17], non-parametric low level vision (NPL) based approach [12], context-aware saliency (CS) based approach [10], distinct patch (Patch) based approach [16], discriminative subspaces (DSRC) based approach [14], and kernel density estimation (KDE) based approach [8]. The maximum number of iterations is kept to 10 which is way higher than 4, the average number of iterations required to enhance the saliency map.

Visual comparison of the proposed approach for a number of images is shown in Fig. 4. Figure 4 shows the original images (Image), their ground truth binary saliency maps (GT), the saliency maps using the existing saliency detection algorithms (X), and their corresponding improved version (X_{SI}) obtained using the proposed framework. One can observe that the saliency map generated using an existing saliency detection algorithm is improved after the execution of the proposed approach for every state-of-the-art saliency detection algorithm.

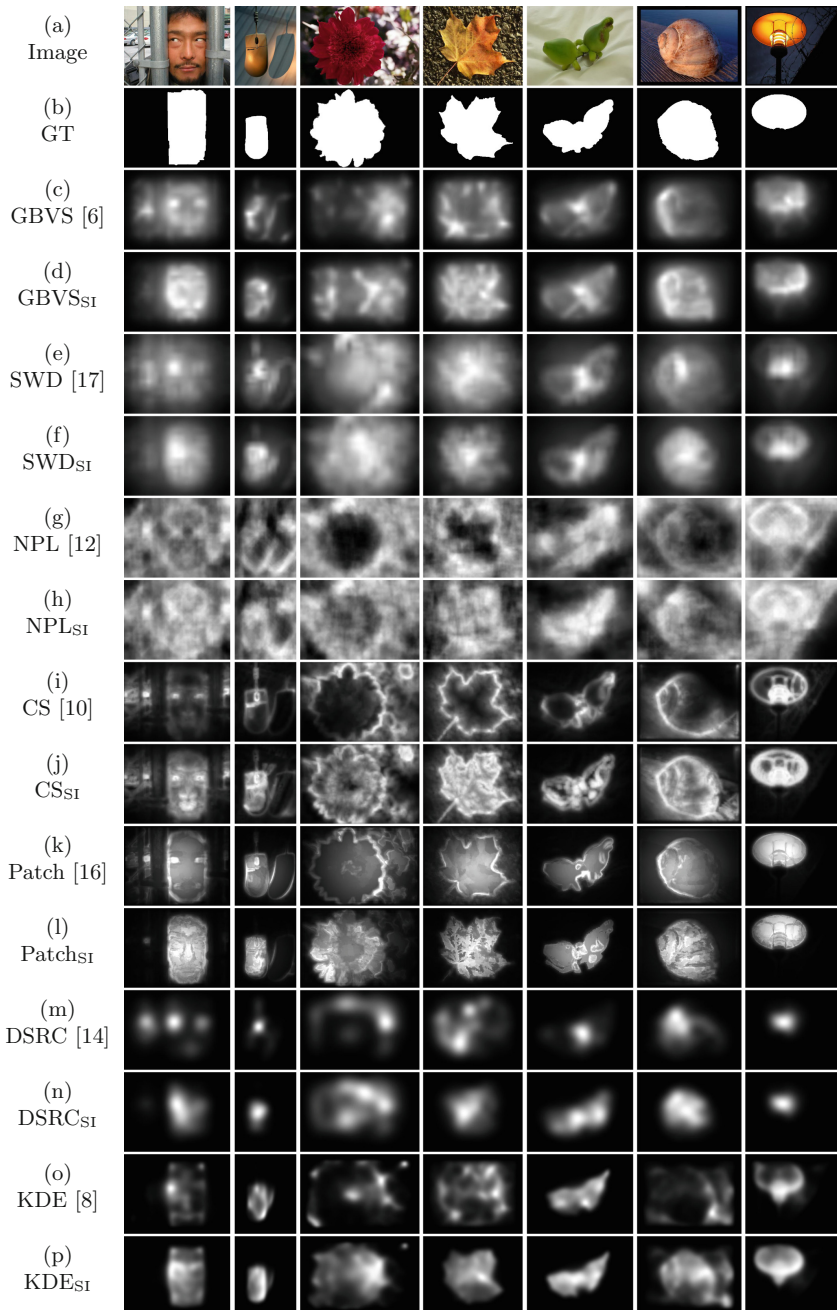


Fig. 4. Visual comparison of improvement in saliency map: (a) Input original images and (b) their ground truths, (c, e, g, i, k, m, o) Saliency maps using existing saliency detection algorithms and (d, f, h, j, l, n, p) modified saliency maps using proposed approach.

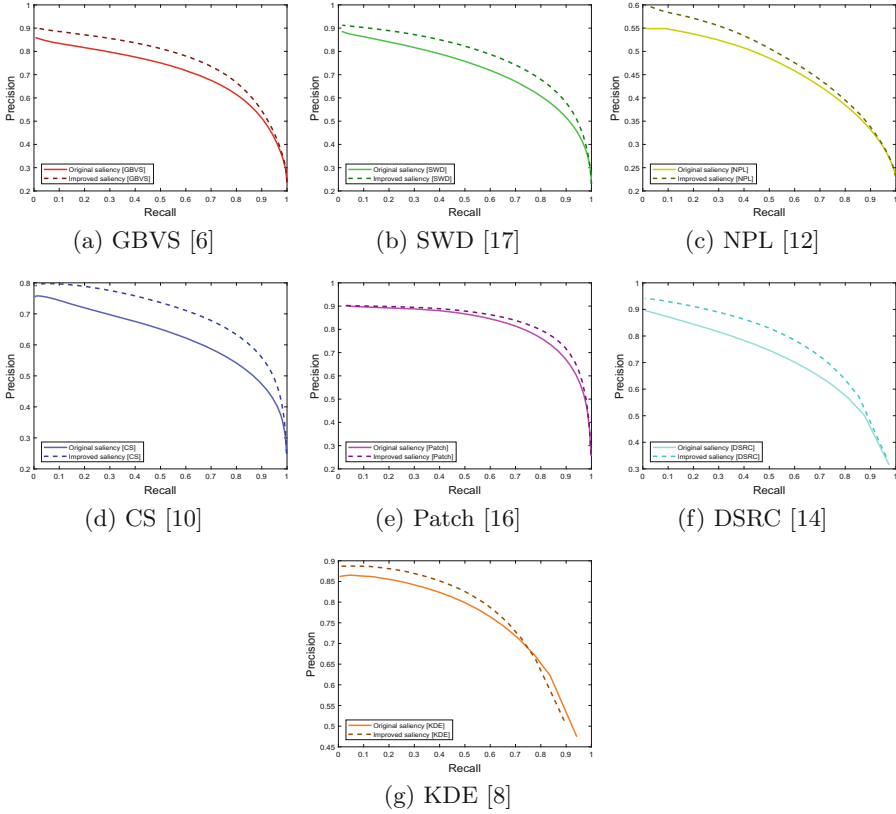


Fig. 5. Average precision - recall curves for saliency maps generated using existing saliency detection algorithms and modified saliency maps using proposed approach.

The objective evaluation of the results obtained using proposed framework is carried out using two measures: precision-recall measure and recently proposed structure measure [22]. Precision (Pr) and recall (Re) values with respect to a fixed threshold and the ground truth binary saliency map are calculated as shown in Eq. (5). S_T is thresholded binary saliency map and GT is ground truth map available with the dataset. For a fixed threshold value, better performance is identified by higher precision value and higher recall value. For each threshold, precision and recall values are averaged across the number of images. Figure 5 shows the graph of the average precision-recall values for different saliency detection algorithms considered in this study. It can be observed from Fig. 5 that the implementation of the proposed framework improves the saliency map generated using all the state-of-the-art saliency detection algorithms. We evaluate the quality of the proposed framework using F-measure F which is defined in Eq. (6).

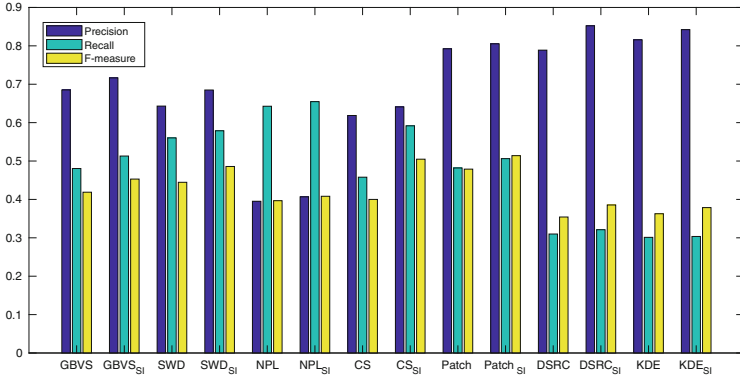


Fig. 6. Precision, recall and F-measure for saliency maps using existing saliency detection algorithms and their corresponding improvement using proposed approach.

$$Pr = \frac{\sum_{x=1}^M \sum_{y=1}^N (S_T \cap GT)}{\sum_{x=1}^M \sum_{y=1}^N S_T}, Re = \frac{\sum_{x=1}^M \sum_{y=1}^N (S_T \cap GT)}{\sum_{x=1}^M \sum_{y=1}^N GT} \quad (5)$$

$$F = 2 \left(\frac{Pr \times Re}{Pr + Re} \right) \quad (6)$$

Using Eqs. (5) and (6), average Pr , Re and F values for the saliency maps obtained using different algorithms and their corresponding improved saliency maps obtained using the proposed framework are shown in Fig. 6. Higher precision and higher recall values indicate that the proposed framework improves the saliency map generated by all the saliency detection methods considered. We can observe that we tend towards salient object segmentation with a better suppression of non-salient regions using the proposed approach. Recently Fan *et al.* have proposed a structural similarity measure to evaluate region-aware and object-aware similarities between non-binary saliency map and ground truth (GT) [22]. Region-aware and object-aware structure similarities try to capture global structure and global distributions of foreground objects. The structure measure overcomes the pixel-wise comparison algorithms for better overall global structure comparison. Table 1 shows the structure measure values: the first row shows the

Table 1. Structure measure for saliency maps generated using existing saliency detection algorithms and their corresponding improvement using proposed approach.

	GBVS [6]	SWD [17]	NPL [12]	CS [10]	Patch [16]	DSRC [14]	KDE [8]
S_0	0.6677	0.6604	0.5032	0.6246	0.7096	0.5984	0.5900
S_f	0.6965	0.6989	0.5114	0.6934	0.7320	0.6170	0.5983

average values of structure measures of saliency maps generated using different existing saliency detection algorithms and the second row shows that of using proposed improvement framework. It is noted that S-values are increased by applying proposed improvement framework.

5 Conclusion

The proposed method introduces an iterative process to improve the saliency map obtained by an existing saliency detection algorithm. The saliency values are forced to be more concentrated in distinctive regions and suppressed in non-distinctive regions. This is achieved by smoothing non-distinctive or background regions and enhancing the details of distinctive or foreground regions using edge-preserving filters. The performance of the saliency improvement framework is shown to be effective using precision, recall, F-measure and recently proposed structure measure. The proposed saliency improvement technique can be used for various applications of computer vision which require salient object detection and segmentation. In future, we would like to extend the work for improving the salient object detection in videos.

References

1. Qin, C., Zhang, G., Zhou, Y., Tao, W., Cao, Z.: Integration of the saliency-based seed extraction and random walks for image segmentation. *Neurocomputing* **129**, 378–391 (2014)
2. Peng, H., Li, B., Ling, H., Hu, W., Xiong, W., Maybank, S.J.: Salient object detection via structured matrix decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(4), 818–832 (2017)
3. Ouerhani, N., Bracamonte, J., Hugli, H., Ansoerge, M., Pellandini, F.: Adaptive color image compression based on visual attention. In: *Proceedings of 11th International Conference on Image Analysis and Processing*, pp. 416–421. IEEE (2001)
4. Zhao, J., Chen, Y., Feng, H., Xu, Z., Li, Q.: Infrared image enhancement through saliency feature analysis based on multi-scale decomposition. *Infrared Phys. Technol.* **62**, 86–93 (2014)
5. Itti, L., Koch, C., Niebur, E., et al.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(11), 1254–1259 (1998)
6. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems*, pp. 545–552 (2006)
7. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, pp. 1597–1604. IEEE (2009)
8. Rezaadegan Tavakoli, H., Rahtu, E., Heikkilä, J.: Fast and efficient saliency detection using sparse sampling and kernel density estimation. In: Heyden, A., Kahl, F. (eds.) *SCIA 2011. LNCS*, vol. 6688, pp. 666–675. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21227-7_62
9. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.-Y.: Learning to detect a salient object. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(2), 353–367 (2011)

10. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(10), 1915–1926 (2012)
11. Cheng, M.-M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.-M.: Global contrast based salient region detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 569–582 (2015)
12. Murray, N., Vanrell, M., Otazu, X., Parraga, C.A.: Saliency estimation using a non-parametric low-level vision model. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 433–440. IEEE (2011)
13. Li, Y., Hou, X., Koch, C., Rehg, J.M., Yuille, A.L.: The secrets of salient object segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 280–287 (2014)
14. Fang, S., Li, J., Tian, Y., Huang, T., Chen, X.: Learning discriminative subspaces on random contrasts for image saliency analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 1095–1108 (2016)
15. Borji, A., Itti, L.: Exploiting local and global patch rarities for saliency detection. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 478–485. IEEE (2012)
16. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1139–1146 (2013)
17. Duan, L., Wu, C., Miao, J., Qing, L., Fu, Y.: Visual saliency detection by spatially weighted dissimilarity. In: *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 473–480. IEEE (2011)
18. Lei, J., Wang, B., Fang, Y., Lin, W., Le Callet, P., Ling, N., Hou, C.: A universal framework for salient object detection. *IEEE Trans. Multimed.* **18**(9), 1783–1795 (2016)
19. Haralick, R.M., Shanmugam, K., et al.: Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **3**(6), 610–621 (1973)
20. Tang, M., Ben Ayed, I., Marin, D., Boykov, Y.: Secrets of GrabCut and kernel k-means. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1555–1563 (2015)
21. Paris, S., Hasinoff, S.W., Kautz, J.: Local Laplacian filters: edge-aware image processing with a Laplacian pyramid. *ACM Trans. Graph.* **30**(4), 68 (2011)
22. Fan, D.-P., Cheng, M.-M., Liu, Y., Li, T., Borji, A.: Structure-measure: a new way to evaluate foreground maps. *arXiv preprint [arXiv:1708.00786](https://arxiv.org/abs/1708.00786)* (2017)