International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 4 (2015), pp. 125-134
http://dx.doi.org/10.14257/ijsip.2015.8.4.11

Salient Object Detection Based on Context and Location Prior

Duzhen Zhang[1,2] and Chuancai Liu[1]

[1] *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, PR China*
[2] *School of Computer Science and Technology, Jiangsu Normal University, Xuzhou 221116, PR China*
zhduzhen@aliyun.com

Abstract

A novel automatic salient object detection algorithm, which integrates context-based saliency with location computation based on the boundary priors, is proposed. Input image is expressed as a close-loop graph with superpixels as nodes and salient object of image has a well-defined graph-based manifold ranking location. The saliency of the image elements is defined based on their relevances to the given seeds or queries. Saliency object location is carried out in a two-stage scheme to extract background regions and foreground salient objects efficiently. We introduce a location weight to measure the relationship of superpixels and the centroid of the detected salient regions to eliminate the background. Saliency map is computed through context analysis and location computing based on multi-scale superpixels. Experimental results on three public benchmark datasets demonstrate that our approach performs well compared to existing state-of-the-art methods.

Keywords: *Saliency Map, Manifold Ranking, Context, Bottom-up, Boundary Prior*

1. Introduction

Visual saliency plays important roles in natural vision in that saliency can direct eye movements, deploy attention, and facilitate tasks like object detection and scene understanding. Many models have been built to compute saliency map. There are two major categories of factors that drive attention: bottom-up factors and top-down factors [1]. Bottom-up factors are derived solely from the visual scene. Regions of interest that attract our attention are in a bottom-up way and the responsible feature for this reaction must be sufficiently discriminative with respect to surrounding features. Inspired by the feature-integration theory [2], Itti, *et al.*, [3] proposed one of the earliest bottom-up selective attention models by utilizing color, intensity and orientation of images. Most computational models [4-17] are data-driven and focused on bottom-up attention. Bottom-up attention can be biased toward targets of interest by top-down cues such as object features, priors, scene context and task demands. Top-down methods [20, 21] are task-driven that entails supervised learning with class labels. Bottom-up and top-down factors should be combined to direct attentional behavior. Attention models were reviewed recently by Borji and Itti [22], Borji, *et al.*, [23]. Saliency models have been developed for eye fixation prediction and salient object detection. The former focuses on identifying a few fixation locations on natural images, which is important for understanding human attention. The latter also called salient object segmentation is to accurately detect where the salient object should be, which is useful for many high-level vision tasks [18, 23]. Since saliency models, whether they address salient object segmentation or fixation predictions, both generate saliency maps, they are interchangeably applicable. In this paper, we focus on the salient object detection task in which integrates with context and location prior.

Yang, *et al.,* [18] represent the image as a close-loop graph with superpixels as nodes. Based on the boundary priors or foreground cues, which show that people tend to gaze at the center of images, they propose a bottom-up algorithm to detect salient regions in images through manifold ranking on it. In order to further eliminate the background, we consider the detected salient regions as location prior and introduce a location weight to measure the relationship of all superpixels and the centroid of the detected salient regions.

Jiang, *et al.,* [19] propose an automatic salient object segmentation method which integrates context analysis based on multi-scale superpixels with object-level shape prior. A region (superpixel) is salient if it is distinguished from its immediate context (spatial neighbors). They assume that the salient object in an image is most probably placed near the center of the image. This assumption, known as *Rule of Thirds* [25], as a location prior is a subjective aesthetic viewpoint. In our work, we incorporate the fore mentioned location prior, computing weights between superpixels and centroid of the object, with local context analysis to implement salient object segmentation.

We compare our method with six state-of-the-art models of saliency detectors. Experimental results show that our method performs well for visual saliency detection task, having highest recall and F-measure values. The rest of this paper is organized as follows. Section 2 introduces and analyzes graph-based manifold ranking object location model, a novel location weight is proposed. Section 3 is context and location based saliency computation scheme. Experimental results and comparisons with state-of-the-art models are presented in Section 4, and conclusions are given in Section 5.

2. Graph-based Manifold Ranking Object Location

Since background often presents local or global appearance connectivity with each of four image boundaries and foreground presents appearance coherence and consisetncy, salient object rarely occupies three or all sides of an image; these cues are called the boundary priors. In this work, we exploit these cues to compute pixels saliency and object location based on the ranking of superpixels. For each image, we construct a close-loop graph where each node is a superpixel. We model salient object detection as a manifold ranking problem and propose a two-stage scheme for graph labelling, using ranking with background and foreground queries, respectively. The object location is identified by its centroid of the final map.

2.1. Graph Construction and Manifold Ranking

A ranking method exploits the intrinsic manifold structure of data for graph labelling. Given a dataset $X = \{x_1, \ldots x_l, x_{l+1}, \ldots, x_n\} \in \mathbb{R}^n$, some data points are labelled queries and the rest need to be ranked according to their relevances to the queries. Let $f : X \to R^n$ denote a ranking function which assigns a ranking value f_i to each point x_i , and f can be viewed as a vector $f = [f_1, \ldots, f_n]^T$. Let $y = [y_1, \ldots, y_n]^T$ denote an indication vector, in which $y_i = 1$ if x_i is a query, and $y_i = 0$ otherwise. We define a graph $G = (V, E)$ on the dataset, where the nodes V are the dataset X and the edges E are weighted by an affinity matrix $W = [w_{ij}]_{n \times n}$. Given G, the degree matrix is $D = \mathrm{diag}\{d_{11}, \ldots, d_{nn}\}$, where $d_{ii} = \sum_j w_{ij}$. In our work, we represent the image as a graph, where V is a set of nodes and E is a set of undirected edge, each node is a superpixel generated by the SLIC algorithm [24]. In order to improve the performance of the proposed method, we enforce that the nodes on the four sides of image are connected, *i.e.,* any pair of boundary nodes are considered to be adjacent. Thus, we denote the graph as a close-loop graph.

The weight between two nodes is defined by

$$w_{ij} = e^{-\frac{\|c_i - c_j\|}{\gamma^2}} \qquad (1)$$

where c_i and c_j denote the mean of the superpixels corresponding to two nodes in the CIE LAB color space, γ is a constant that controls the strength of the weight.

The best ranking of queries is computed by solving the following ranking function [18]:

$$f^* = (D - \alpha W)^{-1} y \quad (2)$$

where α is a constant. Given an input image represented as a graph and some salient query nodes, the saliency of each node is defined as its ranking score computed by Eq. (2). It is noted that we measure the saliency of nodes using the normalized ranking score $\bar{f}^*$ when salient queries are given, and using $1 - \bar{f}^*$ when background queries are given.

2.2. Salient Object Detection and Location

We employ a two-stage scheme for bottom-up saliency detection using ranking with background and foreground queries. In the first stage, we exploit the boundary priors by using the nodes on each side of image as labelled background queries (*i.e.*, background seeds or query samples). From each labelled result, we compute the saliency of nodes based on their relevances (*i.e.*, rankings) to those queries as background labels. The four labelled maps are then integrated to generate a saliency map. In the second stage, we apply binary segmentation on the resulted saliency map from the first stage and take the labelled foreground nodes as salient queries. The saliency of each node is computed based on its relevance to foreground queries for the final map. The object location is identified by its centroid of the final map.

The weight between superpixel *i* and centroid of the object is defined by

$$w_i = e^{-\frac{\|c_i - c_{loc}\|}{\sigma^2}} \quad (3)$$

where c_i and c_{loc} denote the centroid coordinates of the superpixel and the object location, σ is a constant that controls the strength of the weight.

3. Context and Location based Saliency Computation

Jiang, *et al.*, [19] propose an automatic salient object segmentation method which integrates context analysis based on multi-scale superpixels with object-level shape prior. They introduce three characteristics to define a salient object: ① the salient object is always different from its surrounding context; ② the salient object in an image is most probably placed near the center of the image; ③ a salient object has a well-defined closed boundary. We agree that a region (superpixel) is salient if it is distinguished from its immediate context (spatial neighbors) and a salient object has a well-defined closed boundary. They assume that the salient object in an image is most probably placed near the center of the image. The salient object should emerge from anywhere in the image. In our work, we incorporate our fore mentioned location prior with local context analysis to implement salient object segmentation.

Firstly, we adopt a method to detect a salient object on multiple superpixel scales, which is obtained by fragmenting the image with *N* groups of different parameters. A region (superpixel) is salient if it is distinguished from its immediate context, defined as a set of spatial neighbors in our scheme. Specifically, on superpixel scale *n* , we fragment input image *I* into regions $\{r_i^{(n)}\}_{i=1}^{R(n)}$. Given region $r_i^{(n)}$ and its spatial neighbors $\{r_k^{(n)}\}_{k=1}^{K(n)}$, the saliency of $r_i^{(n)}$ is defined as:

$$S(r_i^{(n)}) = -w_i^{(n)} \log(1 - \sum_{k=1}^{K(n)} \alpha_{ik}^{(n)} d_{color}(r_i^{(n)}, r_k^{(n)})) \quad (4)$$

where $\alpha_{ik}^{(n)}$ is the ratio between the area of $r_k^{(n)}$ and total area of the neighbors of $r_i^{(n)}$; $d_{color}(r_i^{(n)}, r_k^{(n)})$ is the color distance between regions $r_i^{(n)}$ and $r_k^{(n)}$, computed as χ^2 distance between the CIE Lab and hue histograms of two regions; $w_i^{(n)}$ is the weight between region $r_i^{(n)}$ and centroid of the object (Section 2.2).

Finally, we propagate saliency value from multiple regions to pixels. Saliency of pixel *p* is defined as:

$$S_m(p) = \frac{\sum_{n=1}^{N}\sum_{i=1}^{R(n)} S(r_i^{(n)})(\| I_p - c_i^{(n)} \| + \varepsilon)^{-1} \delta(p \in r_i^{(n)})}{\sum_{n=1}^{N}\sum_{i=1}^{R(n)} (\| I_p - c_i^{(n)} \| + \varepsilon)^{-1} \delta(p \in r_i^{(n)})} \qquad (5)$$

where *i* is the index of region, *n* is the index of superpixel scale, ε is a small constant (0.1 in our implementation), $c_i^{(n)}$ is the color center of region $r_i^{(n)}$, $\| I_p - c_i^{(n)} \|$ is the color distance from the pixel p to the color center of $r_i^{(n)}$ and $\delta(\cdot)$ is the indicator function.

The main steps of the proposed salient object detection algorithm are summarized below:

Algorithm 1 Salient Object Detection Based on Context and Location Prior

```
Input: An image and required paramerers
1. Segment the input image into superpixels, construct a graph $G$ with superpixels as nodes, compute its degree matrix $D$ and weight matrix $W$ by Eq. 1.
2. Compute $(D - \alpha W)^{-1}$ and set its diagonal elements to 0.
3. Two-stage approach with the background and foreground queries for manifold ranking to generate the final map.
4. The salient object location is identified by the centroid of the final map.
5. The weight between superpixel and centroid of the object is computed by Eq. 3.
6. Calculate our saliency map $S_m$ according to Eq. 4 and Eq. 5.
Output: A saliency map $S_m$ of input image.
```

4. Experimental Results

We evaluate the proposed method on three datasets. The first one is the MSRA-B dataset [20] which contains 5,000 images with the ground truth of salient region marked by bounding boxes. We use the pixel-wise annotation of MSRA-B salient object dataset provided by Jiang, *et al.*, [26]. The second one is the MSRA-1000 dataset, a subset of the MSRA-B dataset, which contains 1,000 images provided by Achanta, *et al.*, [7] with accurate human-labelled masks for salient objects. Each image in the database contains a salient object or a distinctive foreground object. The third one is the ECSSD dataset [26] with 1,000 images, which includes many semantically meaningful but structurally complex images for evaluation. The images are acquired from the internet and 5 helpers were asked to produce the ground truth masks.

We compare our saliency method with six state-of-the-art saliency detection algorithms. The six saliency detectors are Jiang, *et al.*, [19], Yang, *et al.*, [18], Zhang, *et al.*, [27], Achanta, *et al.*, [7, 6], and Goferman, *et al.*, [9], hereby referred to as CBsal, GBMR, SDSP, FT, AC, and Goferman. The CBsal and Goferman are two detectors of the top-performance methods for saliency detection in the survey [23]. SDSP [27] is recently developed method combining three priors.

There are several paramerers in the proposed algorithm: the edge weight γ in Eq. (1), controls the strength of weight between a pair of nodes; the balance weight α in Eq. (2), balances the smooth and fitting constraints in the regularization function of manifold ranking algorithm; the weight σ in Eq. (3), controls the strength of the weight. As in Yang, *et al.*, [18], $\gamma^2 = 0.1$ and $\alpha = 0.99$ for all the experiments. The weight σ is empirically chosen, set to 10 in our implementation.

The SLIC superpixel software is used to generate superpixels. The superpixel number is set to 200 in experiments considering the trade-off between performance and speed.

4.1. MSRA-1000

4.1.1. Quantitative Evaluation

To evaluate a model that outputs a saliency map, there are several metrics. We evaluate all methods by ROC curve and precision, recall, and F-measure bars.

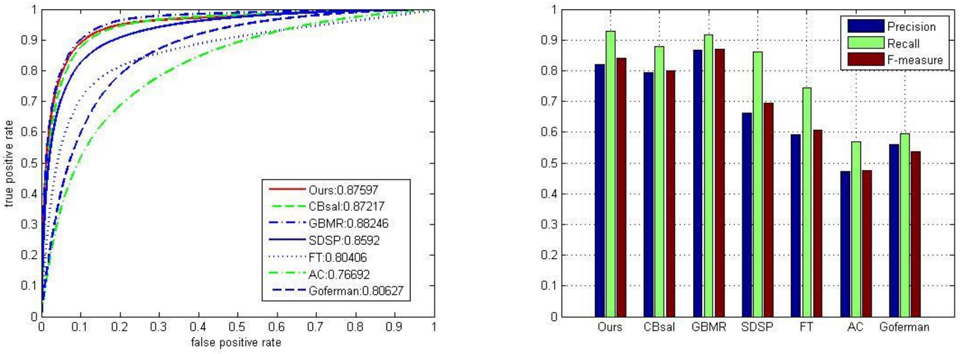
As the most popular measure, ROC (Receiver Operating Characteristic) is used for the evaluation of a binary system with a variable threshold. Using this measure, the model's estimated saliency map is treated as a binary classifier on every pixel in the image; pixels with larger saliency values than a threshold are classified as salient object while the rest of the pixels are classified as non-salient object. Human-labelled masks are then used as ground truth. By varying the threshold, the ROC curve is drawn as the false positive rate vs. true positive rate, and the area under curve (AUC) indicates how well the saliency map predicts actual human-labelled masks.

We also compute the precision, recall and F-measure with an adaptive threshold proposed in Achanta, *et al.*, [7], defined as twice the mean saliency of the image. The F-measure is the overall performance measurement computed by the weighted harmonic of precision and recall:

$$F_\beta = \frac{(1+\beta^2)Precision \times Recall}{\beta^2 Precision + Recall} \qquad (6)$$

where we set $\beta^2 = 0.3$ to quantitatively evaluate the performance.

To obtain a quantitative evaluation we compare ROC curves and AUC on the database MSRA-1000. Figure 1 is the result of our method and other six methods. It shows our method performs well, and outperforms the CBsal [19], SDSP [27], FT [7], AC [6], and Goferman [9]; achieving the highest recall values. The performance almost approaches to GBMR [18].



(a) (b)

Figure 1. (a) ROC Curves and AUC Scores of Different Methods (b) Precision, Recall and F-measure, Every Group of the Bars from Left to Right Represents Precision, Recall and F-measure Respectively

4.1.2. Comparison of Saliency Maps

We choose some images from the MSRA-1000 dataset, the salient object appearing mainly in the corner or near the image boundaries, to visually verify the effect of our location weight. Figure 2 is the output of the six state-of-the-art methods and our method for comparison. Compared to other saliency detectors, our method can eliminate the background ‘noise’ almost completely, the salient objects are perfectly highlighted.

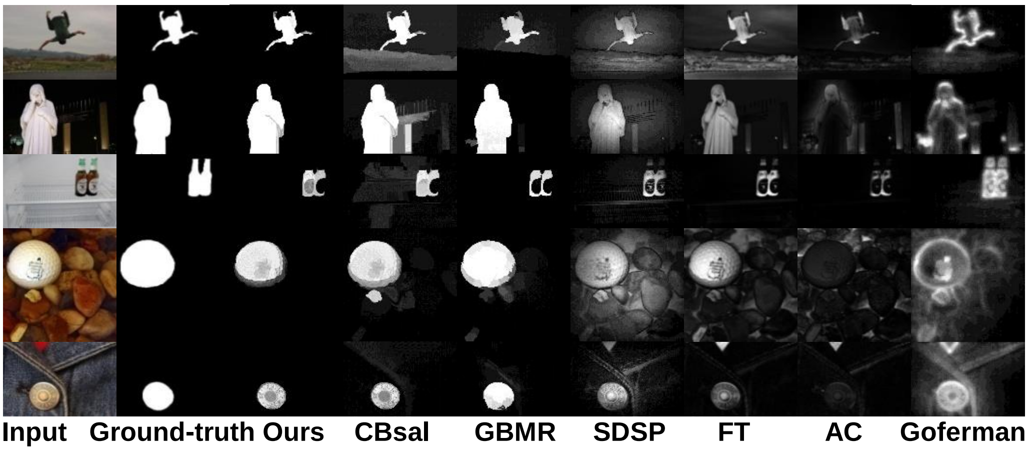


Figure 2. Visual Comparison of Saliency Maps

4.2. MSRA-B

Since Figures 1 and 2 indicate that the performance of CBsal, GBMR, SDSP, and ours are better than FT, AC, and Goferman, we compare our method with CBsal [19], GBMR [18], SDSP [27] on the dataset MSRA-B. Figure 3 is the ROC curves of our method and other three methods. It shows our method performs better on this larger dataset which contains 5,000 images.

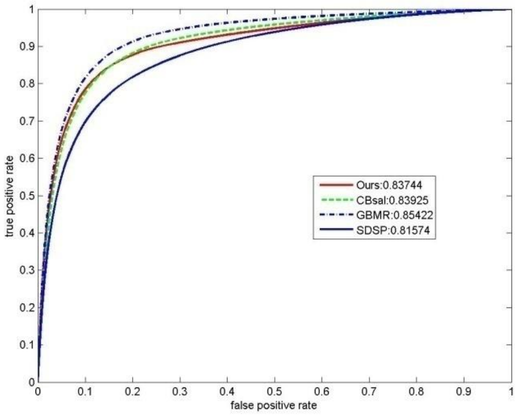
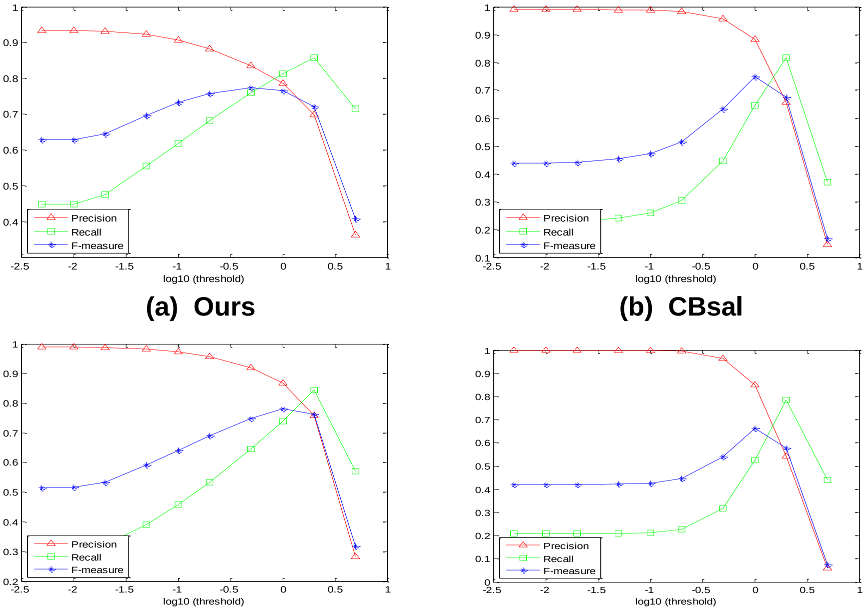


Figure 3. ROC Curves and AUC Scores of our Method and CBsal, GBMR, SDSP

It also shows that the performance of our method is lower than CBsal’s by a finger's breadth. We think it is because of center-bias and CBsal [19] add it as a prior (Section 3). There is a strong center-bias in the single-object datasets, including MSRA-B and MSRA-1000 datasets, most probably due to the tendency of photographers to frame interesting objects at the image center [23]. Therefore, we should build a more complicated dataset which has no center-bias to evaluate our method.

In Section 4.1.1, we compute the precision, recall and F-measure bars with an adaptive threshold defined as twice the mean saliency of the image. Since our method can eliminate the background almost completely, we will select an optimized value of the adaptive threhold.

Figure 4 is the precision, recall and F-measure over denary logarithm of threshold. As the threshold increases, the value of recall and F-measure increase to the maxima, then decrease, while the value of precision decreases persistently.



(a) Ours (b) CBsal

(c) GBMR (d) SDSP

Figure 4. Precision, Recall and F-measure Over Denary Logarithm of Threshold

We set adaptive threshold 0.2 in experiment. Figure 5 illustrates the precision, recall and F-measure results of our method and other three methods. It shows precision value of our method is lower than CBsal [19], GBMR [18], and SDSP [27], but overall, our method outperforms the other three methods on this large dataset, having highest recall and F-measure values.

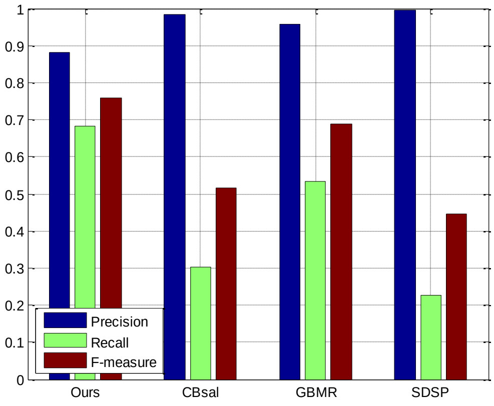


Figure 5. Precision, Recall and F-measure of our Method and CBsal, GBMR, SDSP Every Group of the Bars from Left to Right represents Precision, Recall and F-measure Respectively

4.3 ECSSD

We test the proposed model on the ECSSD dataset in which images contain diversified patterns in both foreground and background. Similar to the experiments on the MSRA-B dataset, the ROC curves and Precision, recall and F-measure (threshold=0.2) of our method and CBsal, GBMR, SDSP are computed. Figure 6 and Figure 7 are the results. Though the performance of all models descends, we can come to the same conclusions as Section 4.2 that our method performs better and has highest recall and F-measure values on this dataset.

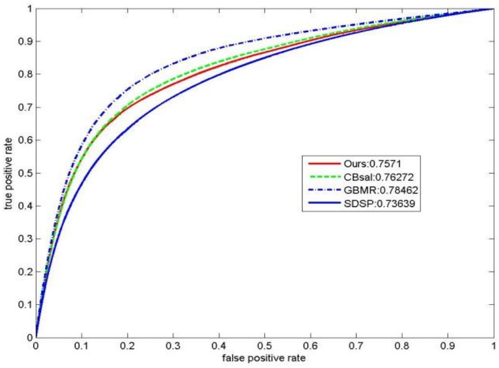


Figure 6. ROC Curves and AUC Scores of our Method and CBsal, GBMR, SDSP

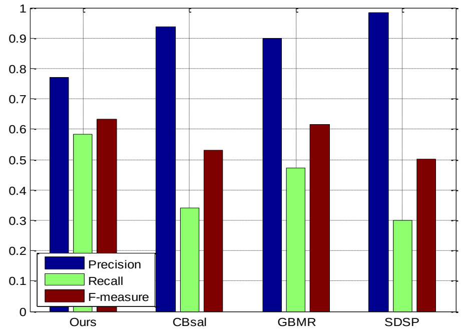


Figure 7. Precision, Recall and F-measure of our Method and CBsal, GBMR, SDSP Every Group of the Bars from Left to Right Represents Precision, Recall and F-measure Respectively

5. Conclusions

In this paper, we propose a salient object detection algorithm which integrates context-based saliency with location computation based on the boundary priors. We adopt a two-stage approach with the background and foreground queries for manifold ranking to generate the salient regions. A location weight to measure the relationship of superpixels and the centroid of the detected salient regions is introduced to eliminate the background. Our saliency map is computed through context analysis and location computing based on multi-scale superpixels. We evaluate the proposed algorithm on three popular benchmark datasets (MSRA-1000, MSRA-B, ECSSD) and demonstrate promising results with comparisons to six state-of-the-art methods. Our future work will focus on cluttered scenes with multiple objects, ameliorate the performance of our method, scale up current model in the spatio-temporal domain.

Acknowledgments

This work is supported by the National Natural Science Fund of China (Grant Nos. 61373063, 61373062) and by the project of Ministry of Industry and Information Technology of China (Grant No. E0310/1112/02-1).

References

[1] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention", Annual Reviews of Neuroscience, vol. 18, no. 193, **(1995)**.

[2] A. Treisman and G. Gelade, "A feature-integration theory of attention", Cognitive Psychology, vol. 12, no. 1, **(1980)**.

[3] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, **(1998)**.

[4] N. Bruce and J. Tsotsos, "Saliency based on information maximization", Advances in neural information processing systems, vol. 18, no. 155, **(2006)**.

[5] L. Zhang, M. H. Tong, T. K. Marks, H. Shan and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics", Journal of Vision, vol. 8, no. 1, **(2008)**.

[6] R. Achanta, F. Estrada, P. Wils and S. Süsstrunk, "Salient region detection and segmentation. Computer Vision Systems", 6th International Conference on Computer Vision Systems, **(2008)** May 12-15, Santorini, Greece.

[7] R. Achanta, S. Hemami, F. Estrada and S. Süsstrunk, "Frequency-tuned salient region detection", 2009 IEEE Conference on Computer Vision and Pattern Recognition, **(2009)** June 20-25, Florida, USA.

[8] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency", Advances in Neural Information Processing Systems, vol. 19, no. 545, **(2007)**.

[9] S. Goferman, L. Zelnik-Manor and A. Tal, "Context-aware saliency detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 1915, **(2012)**.

[10] K.-Y. Chang, T.-L. Liu, H.-T. Chen and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection", IEEE International Conference on Computer Vision (ICCV), vol. 914, **(2011)**.

[11] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang and S. M. Hu, "Global contrast based salient region detection", 2011 IEEE Conference on Computer Vision and Pattern Recognition, **(2011)** June 20-25, Colorado Springs, CO, USA.

[12] V. Gopalakrishnan, Y. Hu and D. Rajan, "Random walks on graphs for salient object detection in images", IEEE Transactions on Image Processing, vol. 19, no. 3232, **(2010)**.

[13] D. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection", International Conference on Computer Vision, vol. 2214, **(2011)**.

[14] Y. Lu, W. Zhang, H. Lu and X. Y. Xue, "Salient object detection using concavity context", International Conference on Computer Vision, vol. 233, **(2011)**.

[15] F. Perazzi, P. Krahenbuhl, Y. Pritch and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection", Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, vol. 733, **(2012)**.

[16] L. Wang, J. Xue, N. Zheng and G. Hua, "Automatic salient object extraction with contextual cue", International Conference on Computer Vision, **(2011)**.

[17] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues", Proceedings of the 14th ACM International Conference on Multimedia, **(2006)** October 23-27, Santa Barbara, CA, USA.

[18] C. Yang, L. Zhang, H. Lu and X. Ruan, "Saliency detection via graph-based manifold ranking", Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, vol. 3166, **(2013)**.

[19] H. Jiang, J. Wang, Z. Yuan and T. Liu, "Automatic salient object segmentation based on context and shape prior", In BMVC, **(2011)**.

[20] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang and H. Y. Shum, "Learning to detect a salient object", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 353, **(2011)**.

[21] J. Yang and M. Yang, "Top-down visual saliency via joint crf and dictionary learning", Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, vol. 2296, **(2012)**.

[22] A. Borji and L. Itti, "State-of-the-art in Visual Attention Modeling", IEEE Transactions on Pattern Analysis and Machine Intelligence, **(2013)**.

[23] A. Borji, Dicky N. Sihite and Laurent Itti, "Salient object detection: a benchmark", In Computer Vision–ECCV 2012, Springer Berlin Heidelberg, **(2012)**, pp. 414-429.

[24] R. Achanta, K. Smith, A. Lucchi, P. Fua and S. Susstrunk, "Slic superpixels compared to state-of-the-art superpixel methods", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 2274, **(2010)**.

[25] S. Bhattacharya, R. Sukthankar and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics", ACM Multimedia, vol. 271, **(2010)**.

[26] Q. Yan, L. Xu, J. Shi and J. Jia, "Hierarchical saliency detection", IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, **(2013)**.

[27] L. Zhang, Z. Gu and H. Li, "SDSP: A novel saliency detection method by combining simple priors", IEEE International Conference on Image Processing (ICIP), vol. 171, **(2013)**.

Authors

Duzhen Zhang, is currently pursuing his Ph.D. degree in Nanjing University of Science and Technology, China. His research interests include visual attention, saliency and machine learning.

Chuancai Liu, is a full professor at the school of computer science and engineering, Nanjing University of Science and Technology, China. He obtained his Ph.D. degree from the China Ship Research and Development Academy in 1997. His research interests include AI, pattern recognition and computer vision.