SALIENCY AND CO-SALIENCY DETECTION BY LOW-RANK MULTISCALE FUSION

Rui Huang[1,2], *Wei Feng*[1,2,*], *Jizhou Sun*[1,2]

[1]School of Computer Science and Technology, Tianjin University, Tianjin, China
[2]Tianjin Key Laboratory of Cognitive Computing and Application, Tianjin University, Tianjin, China
{ruihuang,wfeng,jzsun}@tju.edu.cn

ABSTRACT

To facilitate efficiency, most recent successful saliency detection methods are built on superpixel level. However, saliency detection with single-scale superpixel segmentation may fail in capturing the intrinsic salient objects in complex natural scenes with small-scale high-contrast backgrounds. To tackle this problem and realize more reliable saliency detection, we present a simple strategy using multiscale superpixels to jointly detect salient object via low-rank analysis. Specifically, we construct a multiscale superpixel pyramid and derive the corresponding saliency map using multiple saliency features and priors for each single scale at first. Then, we show that by joint low-rank analysis of multiscale saliency maps, we can obtain a more reliable adaptively fused saliency map that takes all scales saliency results into account. We further propose a GMM-based co-saliency prior to enable the above approach to detecting co-salient objects from multiple images. Extensive experiments on benchmark datasets validate the effectiveness and superiority of the proposed approach over state-of-the-art methods.

Index Terms— Saliency, co-saliency, low-rank analysis, GMM-based co-saliency prior

1. INTRODUCTION

Saliency detection aims to highlight the conspicuous objects in an image. Due to its effectiveness and importance, both single image saliency and multi-image co-saliency detection become popular research topics in recent years and have been widely applied in a number of computer vision applications.

Generally, recent saliency detection methods can be categorized into two categories: i.e., bottom-up model and top-down model. Bottom-up models are often based on low-level features, like color, intensity and texture, etc. Histogram-based contrast (HC) is a typical bottom-up method that computes the saliency values based on color histogram [1]. Region-based contrast (RC) [1] improves HC by further measuring regional contrast and spatial distances to other regions

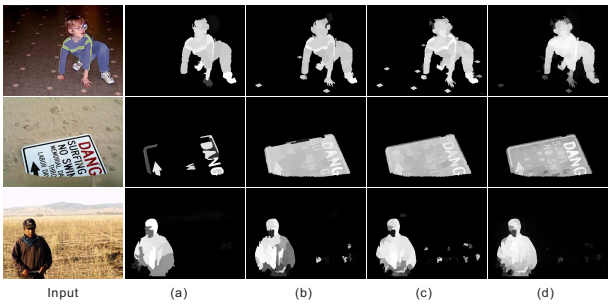


Fig. 1. The comparison of saliency detection of different over-segmentation scales. The number of the superpixels used in (a)-(c) are 100, 250 and 400, respectively. The last column shows our multiscale saliency detection results.

to compute the saliency values. Recently, some top-down models have been used to further boost the performance of saliency detection. Representative methods include various saliency priors, such as center prior, object prior and background prior, which enlighten the possible position of salient objects within an image. Bayesian framework is also used to model multiple low- and mid-level cues to detect salient object [2]. Graph-Based Manifold Ranking (GBMR) [3] weighs the similarity of the image regions with foreground and background cues via graph-based manifold ranking to get the saliency objects. In [4], the integrate uniqueness, focusness and objectness are proposed to detect saliency object. The objectness acts as a high-level prior that ensures the completeness of the detected salient object. Besides, CRF and discriminative dictionary learning are also explored in saliency detection [5, 6]. Compared to bottom-up methods, top-down methods are often task driven and more robust, but usually with higher complexity. Recently, by reasonably assuming that the salient object is sparse and the backgrounds are homogenous, compressive sensing techniques are used to detect salient object and achieve the state-of-the-art performance [7, 8].

Since in most computer vision applications, e.g., image segmentation [9], saliency detection is usually used as a preprocessing step, thus being fast is very important for the success of saliency detection. To facilitate efficiency, most recent methods build their saliency detection models on superpixel

* is corresponding author. Email: wfeng@tju.edu.cn. This work is supported by NSFC (61100121, 61272266), and the National Science and Technology Support Project (2013BAK01B01 and 2013BAH62F02).

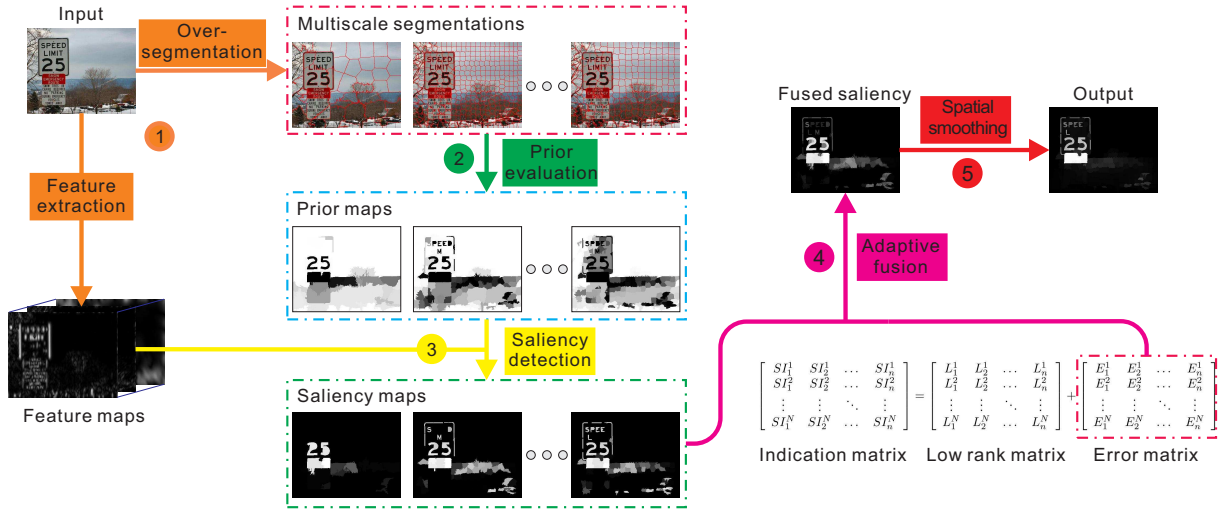


Fig. 2. The flow chart of our single-image multiscale saliency fusion method. The circles and arrows in the same color denote one step of the process. The numbers in circles denote the processing sequence.

level. However, as shown in Fig. 1, saliency maps detected on different superpixel scales may be quite different and inconsistent. Specifically, from Fig. 1, we have two observations: 1) for some images, there may exist a single best superpixel scale, on which we can extract high-quality salient object regions, e.g., the coarsest scale of 1st and 3rd rows in Fig. 1; 2) in other cases, such single best superpixel scale may not necessarily exist, e.g., the 2nd row of Fig. 1, thus we cannot achieve reliable saliency detection at any single scale. Therefore, generally speaking, saliency detection with single-scale superpixel segmentation may fail in capturing the intrinsic salient objects in complex natural scenes with small-scale high-contrast backgrounds. Even for the images with some particular "best" single scales, it is difficult to automatically determine the best scales. Moreover, different superpixel scales have respective pros and cons (see Fig. 1 and Fig. 2). For instance, for coarser scales with a small number of superpixels, the saliency maps are prone to be larger regions and miss the small-scale structures. In contrast, for finer scales, the saliency results tend to highlight all small regions with strong local contrast, thus may inevitably lead to higher false alarm.

In [10], the authors get multiscale saliency by fixed patch size and changed image size. Then the multiple saliency maps are fused with fixed weights. In this paper, to tackle the scale dilemma and realize more reliable saliency detection, we present a simple strategy using multiscale superpixels to jointly detect salient object via low-rank analysis. Specifically, we first construct a multiscale superpixel pyramid. Based on it, we concurrently derive the saliency maps using multiple saliency features and priors at all scales. Then, we present joint low-rank analysis of the multiscale saliency maps, which results in a more reliable adaptively fused saliency map that properly takes all scales saliency results into account. Furthermore, we extend the single image multiscale saliency detection to co-saliency detection from multiple images via the GMM-based co-saliency prior. Extensive experiments on benchmark datasets validate the effectiveness and superiority of our approach over state-of-the-art methods. The main contributions of this paper are twofold:

- We propose a multiscale low-rank saliency fusion method that achieves more reliable saliency detection than single-scale detection.
- We present a GMM-based co-saliency prior to extend the proposed multiscale low-rank saliency detection to co-saliency detection from multiple images, which is able to highlight the co-salient object and suppress the background.

2. LOW-RANK BASED SALIENCY DETECTION

Recently, low-rank analysis has shown great potentials in saliency detection. For instance, [7] decomposes an image into redundancy and saliency parts. Redundancy denotes the information with high regularities, while saliency represents the novelty part. Besides, [8] further models this decomposition as a low-rank matrix recovery problem:

$$(\widehat{B}, \widehat{S}) = \arg\min_{B,S} \operatorname{rank}(B) + \lambda \|S\|_0 \quad s.t. \quad F = B + S, \tag{1}$$

where $F = [f_1, f_2, \ldots, f_N]$ is a feature matrix by stacked N feature vectors, B is recovered low-rank matrix which models background and S is sparse matrix which models saliency.

Description	Dim
The position features	**2**
The average normalized x coordinates The average normalized y coordinates	2
The color features	**40**
The average RGB, CIELAB, HSV values	9
The color histogram of RGB, CIELAB, hue and saturation	4
The global, local contrast and element distribution of the color features	27
The texture features	**80**
The steerable pyramids	12
The Gabor filters	36
The HOG features	31
The singular value (SVF)	1

Table 1. Description and dimension of saliency features used in our method.

Since the above problem is NP, the alternative convex surrogate to solve the above problem is

$$(\widehat{B}, \widehat{S}) = \arg\min_{B,S} \|B\|_* + \lambda\|S\|_1 \quad s.t. \quad F = B + S, \tag{2}$$

where $\|B\|_*$ denotes nuclear norm, and $\|S\|_1$ is the ℓ_1 norm.

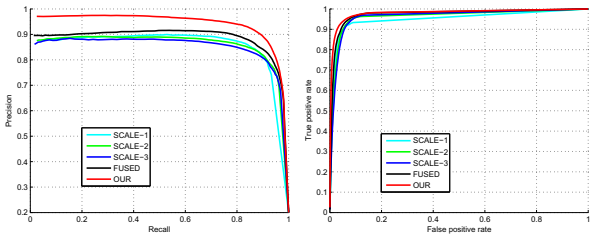
3. THE METHOD

As shown in Fig. 2, multiscale saliency fusion is divided into several steps. Firstly, we introduce how to compute the single-scale saliency map in Section 3.1 (corresponding to the steps 1, 2 and 3). Then we explain how to fuse the multiscale saliency maps in Section 3.2 (corresponding to the step 4). Finally, the fused saliency map is smoothed by considering the relation of the neighborhoods in Section 3.3 (corresponding to the step 5).

3.1. Single-Scale Saliency Detection

Over-segmentation and feature extraction. We use the SLIC [11] to segment the image into superpixels, and extract 122 dimensional features including: location, color and texture. The 40 dimensional color features are extracted as [12]. We extract steerable pyramids [13] features with 4 directions and 3 different scales filter response. The Gabor filters [14] responses with 12 orientations and 3 scales are extracted as Gabor features. We also utilize 31 dimensional HOG implemented by Felzenszwalb *et al.*. The description and the dimensions of the features are stated in Table 1.

Saliency prior. In [15], the authors propose a robust background probability w_i^{bg} with intuitive geometrical interpretation, where i represents the index of the superpixel.



(a) PR cureves (b) ROC curves

Fig. 3. The comparison of PR and ROC curves of multiscale saliency detection results. The scales which we used are 100, 250, 400 for SCALE-1, SCALE-2, SCALE-3 respectively.

In this paper, we incorporate the robust background in our saliency detection method.

Saliency computation. Unfortunately, decomposing F in original feature space always shows inferior saliency object detection results. In [8], to achieve better saliency object results, a transformation matrix T is learnt. By multiplying the transformation matrix T with feature matrix F, we obtain a transformed feature matrix TF. In this space, the transformed features of the image background lies in a low dimensional sub-space such that they can be represented as a low-rank matrix. The readers should refer to [8] for more details about how to get T. The prior P can be easily incorporated by multiplying TF with P. So the final saliency model is

$$(\widehat{B}, \widehat{S}) = \arg\min_{B,S} \|B\|_* + \lambda\|S\|_1 \quad s.t. \quad TFP = B + S. \tag{3}$$

Let $\widehat{S}$ be the optimal solution of Eq. (3). The saliency score $SP(i)$ for the i-th superpixel is

$$SP(i) = \frac{\|\widehat{S}(:,i)\|_2}{\sum_i \|\widehat{S}(:,i)\|_2} = \frac{\sqrt{\sum_j (\widehat{S}(j,i))^2}}{\sum_i \sqrt{\sum_j (\widehat{S}(j,i))^2}}. \tag{4}$$

3.2. Multiscale Saliency Fusion

As mentioned in Section 1, detecting saliency in a single scale always fail in capturing the intrinsic salient object. To obtain a reliable solution for saliency detection, we propose a multiscale fusion method. At first, we construct a multiscale superpixel pyramid with different segmentation scales. Then the single scale saliency map is computed as mentioned in Section 3.1. Finally, we compute the fused saliency map by the adaptive weights that takes all scales saliency results into account.

To compute the adaptive fusion weights, we need to reassign the saliency value that obtained at each scale to the finest over-segmentation scale. The saliency value of a superpixel is the mean saliency value of the pixels contained in it. By placing the saliency value of the superpixels of a single scale

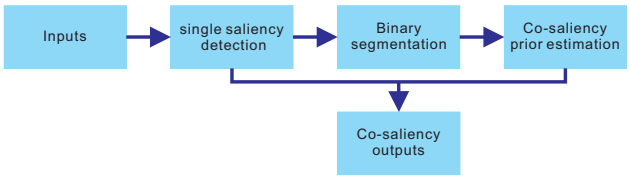


Fig. 4. The flowchart of co-saliency detection.

into a row vector. Then stacking all vectors together forming a saliency indication matrix SI. The ideal situation is that the saliency detection results are consistent in all scale, so the rank is one. We can model this problem as low-rank matrix recovery:

$$(\widehat{L}, \widehat{E}) = \arg\min_{L,E} \|L\|_* + \lambda\|E\|_1 \quad s.t. \quad SI = L + E. \tag{5}$$

The optimal solution $\widehat{E}$ represents the disparity of the multi-scale salient object detection results. Summing up the E's absolute value by rows, then we get a vector $[\widehat{E}_1, \widehat{E}_2, \ldots, \widehat{E}_n]^T$, where n is the number of scales. Large $\widehat{E}_i$ represents the i-th saliency map are not consistent well with others. So the corresponding saliency map should be endowed with a small weight. The adaptive weights are

$$\omega_i = \frac{1}{Z}\exp(-\widehat{E}_i^2), \tag{6}$$

where $Z = \sum \exp(-\widehat{E}_i^2)$ is partition function. Then the fused saliency maps can be obtained by

$$S_{\text{map}}^{\text{fuse}} = \sum \omega_i \cdot S_{\text{map}}^i. \tag{7}$$

As shown in the last column of Fig. 1, the multiscale results are better than the results in each single scale. We also make a quantitative comparison of the results with different segmentation scale in Fig. 3 on MSRA dataset [5].

3.3. Saliency Refinement

We do not consider the smoothness between the neighboring superpixels in aforementioned processing. To do this when finishing the multiscale saliency fusion. Inspired by [15], we optimize the fused saliency map by considering an energy function:

$$E = \sum_i \omega_i^{\text{bg}} * s_i^2 + \sum_i \omega_i^{\text{fg}}(s_i - 1)^2 + \sum_{i,j\in \text{Nei}(i)} \omega_{ij}(s_i - s_j), \tag{8}$$

where s_i denotes the saliency value of the superpixel i, w_i^{bg} is the background probability, w_i^{fg} is the foreground probability (in our method, $w_i^{\text{fg}} = S_i^{\text{fuse}}$), $\text{Nei}(i)$ denotes the neighbors of superpixel i. The weight ω_{ij} is defined as

$$\omega_{ij} = \exp(-\frac{d_{\text{app}}^2(p_i, p_j)}{2\sigma^2}), \tag{9}$$

Fig. 5. Co-saliency detection example of our method. The images of the first to last row are input images, single saliency maps, binary segmentations, co-salient prior maps and co-saliency results.

where $d_{\text{app}}^2(p_i, p_j)$ denotes the L2 distance of their average colors in the CIE-LAB color space, σ is 10 in our experiment.

4. CO-SALIENCY DETECTION

Co-saliency detection is discovering the similar object in a group of images. The co-saliency objects should be saliency in their own image and have high similarity with certain features (e.g., intensity, color, texture or shape) in a set of images [16]. We show our co-saliency detection flowchart in Fig. 4.

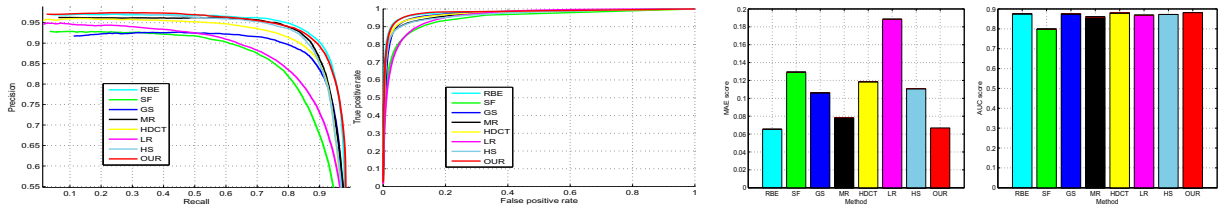
Single saliency detection. Given a set of images $I_{\text{set}} = \{I_1, \ldots, I_i, \ldots, I_n\}$, we compute the single saliency map S_i for i-th image as we proposed in Section 3.

Binary segmentation. Then we segment the saliency map S_i into binary mask M_i with adaptive threshold defined as:

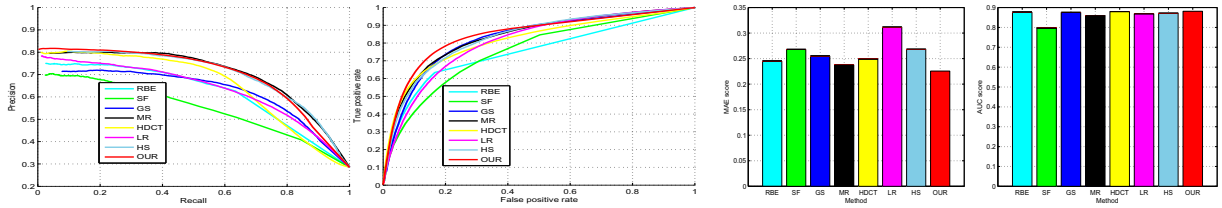
$$T_i = \alpha \cdot \text{mean}(S_i), \tag{10}$$

where $\alpha = 2$ in our experiment. The pixel (or superpixel) with the saliency value which is bigger than the adaptive threshold is identified as foreground, otherwise it is identified as background.

Co-saliency prior estimation. The key problem is how to find the co-salient pattern for a given image set. To tackle this problem, we employ GMM to build the co-saliency prior. The GMM with 5 Gaussian distributions is used to build foreground color models G_i for the foreground pixels for i-th image. We use G_i to estimate the foreground probability of the pixels in a mask M_j of the j-th image. For each image we get n estimations. By averaging them, we get the co-saliency prior for each image.



(a) The quantitative comparison on MSRA dataset [5].



(b) The quantitative comparison on ESCCD dataset [17].

Fig. 6. From left to right column are the PR curves, ROC curves, MAE bar and shuffled AUC bar. The scene of the images in the ESCCD dataset [17] are more complex than that in MSRA dataset [5]. The MAE of our method is lower than others in ESCCD dataset [17].

Co-saliency computation. Finally, we incorporate the co-saliency prior into the single saliency detection model and get the final co-saliency maps. An example of co-saliency detection is shown in Fig. 5.

5. EXPERIMENTAL RESULTS

We evaluate the performances of our method and compare with the state-of-the-art methods on benchmark datasets.

Evaluation methods. Although there exist numerous saliency detection method, the performance of the selected 7 methods are better than others, including the geodesic saliency (GS) [18], high dimensional color transformation (HDCT) [12], low-rank (LR) [8], manifold rank (MR) [3], hierarchical saliency (HS) [17], saliency filter (SF) [19] and robust background estimation (RBE) [15].

Evaluation datasets. For single image saliency detection we use MSRA [5] and ESCCD [17] datasets, both of which have 1000 images. The images in MSRA [5] have complex backgrounds and low contrast object. The ESCCD [17] is an extension of Complex Scene Saliency Dataset [17], which includes many semantically meaningful but structurally complex images for evaluation. To co-saliency detection, we use image pair dataset [20] which contains 105 image pairs, totalling 210 images.

Evaluation criterions. We use standard precision, recall curves (PR curves), ROC curves, AUC, MAE and F-measure to evaluate the performance. PR curves are obtained by average all curves on the same dataset. The single PR curve is obtained by normalizing saliency maps to $[0, 255]$, generating binary mask by a threshold from 0 to 255. ROC curve reflects the relation FPR with TPR at different threshold. To get the robust AUC measure, we use shuffled AUC proposed in [21].

Experimental analysis. As shown in Fig. 7, our saliency detection method can highlight the whole salient object. The saliency results have a clearer background, especially around the salient object. In Fig. 6, we make quantitative comparisons with other single saliency object detection methods on MSRA [5] and ESCCD [17] dataset.

On MSRA dataset [5], our PR curve and ROC curve (shown in red) are higher than other methods. We have highest shuffled AUC measure and lowest MAE measure. On ESCCD dataset [17], the MAE measure is lower than others which represents our multiscale low-rank based saliency object detection is more robust in comparing with the state-of-the-art.

Co-saliency comparison. For demonstrating the effectiveness of our proposed co-saliency object detection method, we compare the proposed method with Liu [22], Fu [23], Li [20] on image pair datasets [20]. We show a quantitative comparison of MAE and Precision, Recall and F-measure results in Fig. 9. Note our precision, F-measure and MAE are better than others. We also show some typical co-saliency detection results compared with the state-of-the-art methods in Fig. 8.

6. CONCLUSION

In this paper, we have proposed a multiscale saliency fusion approach that uses low-rank analysis to adaptively weighs the saliency maps derived from multiple coarse-to-fine scales and realize more reliable saliency detection. We show the effectiveness of joint low-rank analysis in deriving the most consistent parts of saliency maps on different scales. We also propose an effective GMM-based co-saliency prior to extend the proposed single-image saliency detection approach to detect co-salient objects from multiple images. Note, the proposed multiscale saliency detection can be easily implemented in parallel. Hence, in practice, the complexity of our approach

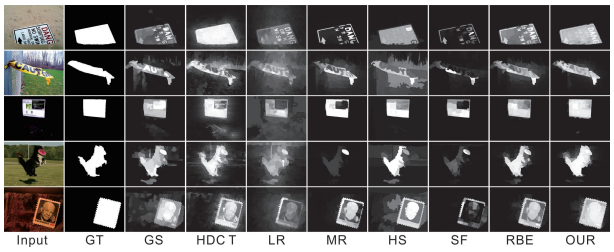


Fig. 7. The comparison of the state-of-the-art saliency detection methods on MSRA dataset [5]. From left to right columns are input image, ground truth, the results of GS [18], HDCT [12], LR [8], MR [3], HS[17], SF [19], RBE [15] and our method.

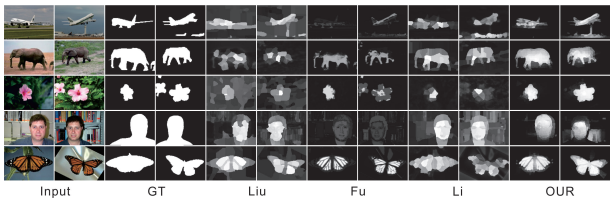


Fig. 8. The comparison of the state-of-the-art co-saliency detection methods on image pair dataset [20]. From left to right columns are input images, ground truth, the results of Liu [22], Fu [23], Li [20] and our method.

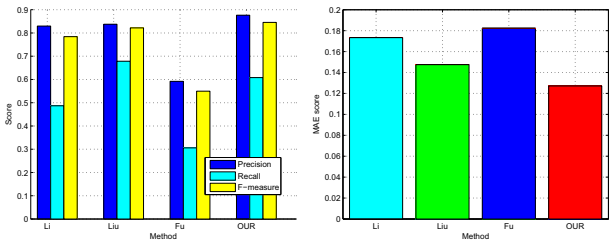


Fig. 9. The quantitative comparison of co-saliency detection on image pair dataset [20].

is comparable to single-scale detection. Experimental results on benchmark datasets validate the effectiveness and superiority of the proposed approach over state-of-the-art methods.

References

[1] M. Cheng, G. Zhang, N.-J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," *CVPR*, 2011.

[2] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE TIP*, vol. 22, no. 5, pp. 1689–1698, 2013.

[3] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," *CVPR*, 2013.

[4] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: Uniqueness, focusness and objectness," *ICCV*, 2013.

[5] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE TPAMI*, vol. 32, no. 2, pp. 253–367, 2011.

[6] J. Yang and M. Yang, "Top-down visual saliency via joint CRF and dictionary learning," *CVPR*, 2012.

[7] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE TIP*, vol. 21, no. 3, pp. 1327–1338, 2012.

[8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," *CVPR*, 2012.

[9] W. Feng, J. Jia, and Z.Q. Liu, "Self-validated labeling of Markov random fields for image segmentation," *IEEE TPAMI*, vol. 32, no. 10, pp. 1871–1887, 2010.

[10] J. Kim, J. Sim, and C. Kim, "Multiscale saliency detection using random walk with restart," *IEEE TCSVT*, vol. 24, no. 2, pp. 198–210, 2014.

[11] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. S usstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE PAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.

[12] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," *ECCV*, 2014.

[13] E.-P. Simoncelli and W.-T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," *ICIP*, 1995.

[14] H.-G. Feichtinger and T. Strohmer, "Gabor analysis and algorithms: theory and applications," 1998.

[15] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," *CVPR*, 2014.

[16] H. Li, F. Meng, and K. Ngan, "Co-salient object detection from multiple images," *IEEE TMM*, vol. 15, no. 8, pp. 1896–1909, 2013.

[17] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," *CVPR*, 2013.

[18] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," *ECCV*, 2012.

[19] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," *CVPR*, 2012.

[20] H. Li and K Ngan, "A co-saliency model of image pairs," *IEEE TIP*, vol. 20, no. 12, pp. 3365–3375, 2011.

[21] L. Zhang, M. Tong, T. Marks, H. Shan, and G.-W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *Journal of Vision*, vol. 8(7), 2008.

[22] Z. Liu, W. Zou, and L. et al. Li, "Co-saliency detection based on hierarchical segmentation," *IEEE SPL*, vol. 21(1), 2014.

[23] H. Fu, X. Cao, and Z. Tu, "Cluster-based co-saliency detection," *IEEE TIP*, vol. 22, no. 10, pp. 3766–3778, 2013.