An image saliency detection method by constructing graph model

Zhenghao Hu, Qiuping Jiang, Zhutuan Li, Feng Shao*
Faculty of Information Science and Engineering
Ningbo University
Ningbo 315211
CHINA
shaofeng@nbu.edu.cn

Abstract: - In this paper, we present an image saliency detection method by constructing graph model. We extract color, texture and compactness features and segment superpixels from an input image to construct a graph model. Then, saliency for each is measured by calculating random walker probability on the node. Extensive results on MSRA dataset containing 1000 test images with ground truths demonstrate that the proposed saliency model outperforms the state-of-the-art saliency models with higher precision and recall performances.

Key-Words: - **S**aliency detection, graph model, superpixel segmentation, random walk

1 Introduction

Visual attention plays an important role in various visual applications by allocating the limited computational resources to those perceptually significant regions. Through detecting saliency map, we can assign relatively high saliency values to those perceptually significant regions while low saliency values to the other regions. The existing saliency methods can be classified into two categories: eye-fixation prediction and salient object detection. Currently, salient object detection models are widely applied in various computer vision and image processing tasks, such as object tracking [1], image segmentation [2,3], quality assessment [4], content-aware image compression [5,6], and adaptive image retargeting [7,8]. In this paper, we mainly focus on designing efficient salient object detection model.

In the past decades, numerous saliency detection models have been proposed to highlight the salient objects based on various low-level, middle-level and high-level properties. The most representative work is the biologically plausible saliency detection model proposed by Itti *et al*., in which various simple low-level feature (e.g., color, intensity and orientation) are integrated using center-surround mechanism [9]. Inspired by Itti's model, various bottom-up saliency detection models were devised. Hou *et al*. proposed a Fourier spectrum residual analysis to detect salient regions based on the log magnitude spectrum representation of image [10]. Cheng *et al*. segment the input image into sub-regions, and calculate the saliency for each region by comparing global region contrast to all other regions in the image [11]. Achanta *et al*. computed the conspicuity likelihood of each pixel based on its color contrast over the whole image [12]. Erdem *et al*. constructed covariance matrix for saliency estimation by incorporating low-level meta-features [13]. Margolin *et al*. integrated both pattern and color distinctiveness for salient object detection [14]. Liu *et al*. proposed an effective superpixel-based saliency model, in which global contrast and spatial sparsity are computed to derive superpixel-level saliency map [15]. Perazzi *et al*. decomposed a given image into compact, perceptually homogeneous superpixels, and computed two contrast measures (uniqueness rate and spatial distribution) at superpixel-level [16]. Many other models can be found in [17-18].

However, most above-mentioned methods detect saliency by measuring local center-surround difference and rarity of features over the entire image. In contrast, graph-based techniques provide another way for saliency detection. Harel *et al*. proposed Graph-based algorithm by redefining the edge weights based on the feature dissimilarity and the spatial proximity between the two connected nodes [19]. Yang *et al*. presented a novel saliency detection method by ranking the similarity of the superpixels with foreground and background cues via graph-based manifold ranking [20]. Lu *et al*. developed a hierarchical graph model and utilize concavity context to compute weights between nodes [21]. Gopalakrishnan *et al*. employed a connected graph and a *k*-regular graph to consider both global and local features in saliency detection

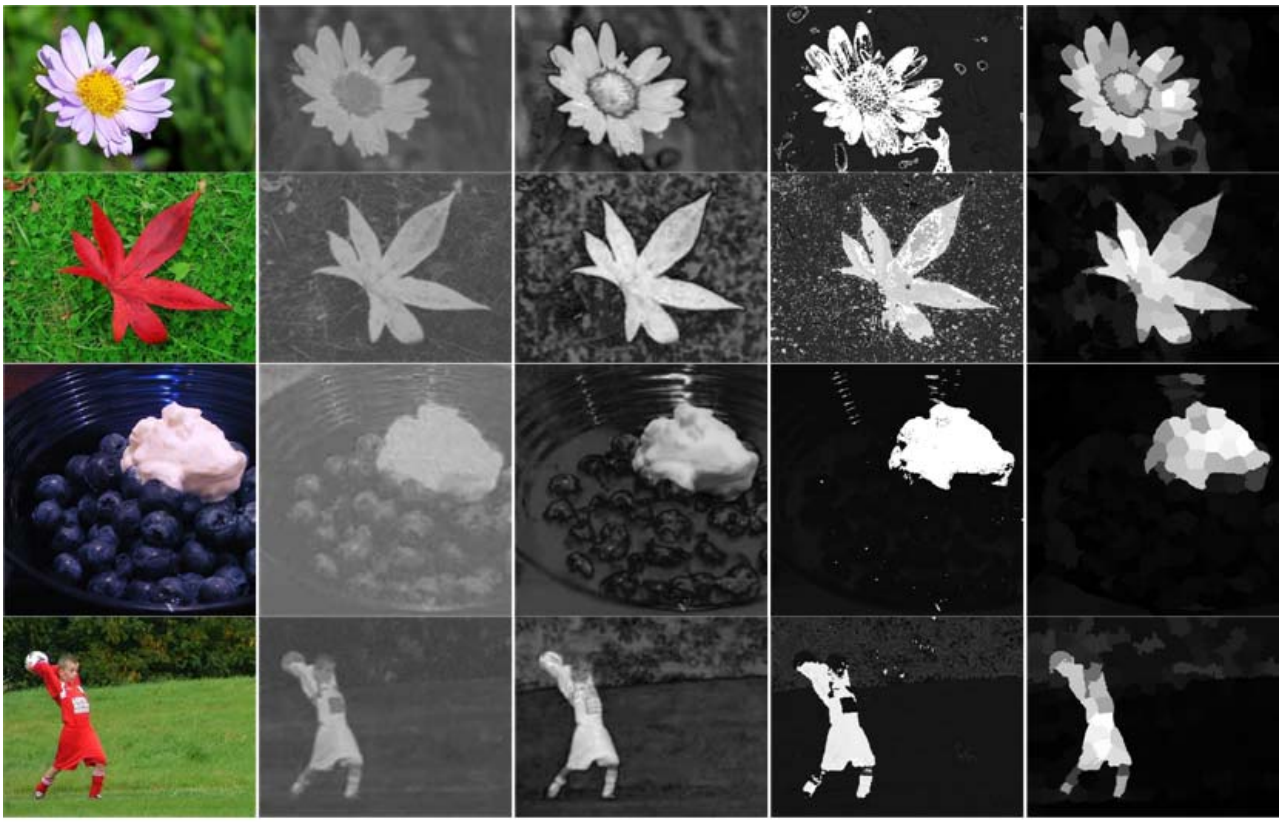

Fig.1. Examples of saliency detection: (a) Input image, (b) color feature based saliency map, (c) texture feature based saliency map, (d) compactness feature based saliency map, (e) the final combined saliency map.

[22]. Kim *et al*. developed a coarse-to-fine multiscale saliency detection method based on the random walk with restart [23].

In this paper, we proposed a salient object detection method by constructing graph model. In the method, color, texture and compactness features are extracted and superpixels are segmented to construct a graph model. Saliency for each node is measured by calculating random walker probability on the node. Experimental results on MSRA dataset demonstrate that the proposed saliency model can achieve a higher saliency detection performance than the state-of-the-art saliency models. The rest of this paper is organized as follows. Section 2 describes the proposed salient object detection model in detail. Experimental results are presented and analyzed in Section 3, and the conclusions are drawn in Section 4.

2 Proposed image saliency detection method

In this paper, we propose an image saliency detection method by constructing graph model. Specially, we extract color, texture and compactness low-level visual features from images. Then, we use SLIC algorithm to extract superpixels and construct a graph model by establishing node and edge information, and finally calculate the saliency for each node by measuring the probability that the random walker stays on the node.

2.1 Feature extraction

Given an image, we extract different types of visual features for each pixel, including

Color. Three Lab color values are extracted for each pixel, producing 3 features, denoted by $\mathbf{c}_{x,y}$.

Texture. Log-Gabor filter responses with 3 scales and 12 orientations are extracted, denoted by $\mathbf{t}_{x,y}$.

Compactness. Since the above features are distributed independently, we present a new feature (i.e., compactness) to quantify feature compactness based on the nearest and farthest neighbours. A feature is useful for clustering if it satisfies the feature compactness for the neighbours. Given these neighbours for each features are different, we evaluate the compactness for each feature. Assume

that we consider $\{\mathbf{c}_{x,y}\}$ to be partitioned into multiple clusters using a simple K-mean clustering algorithm. Then, similar features $\{\mathbf{c}^{\Theta}_{(x,y)\to k}\}_{k=1}^{K}$ would be grouped together, where K is the number of clusters, and Θ denotes the neighbour of the k-th cluster.

After the clustering, we measure the compactness of each cluster. Since each cluster may contain nearest and farthest neighbours, we use spatial variance of all pixels in the cluster to characterize the feature compactness, which is defined as

$$\tau_k = \exp\left(-\alpha \frac{\sigma_{x,k} + \sigma_{y,k}}{\sqrt{W^2 + H^2}}\right) \quad (1)$$

where $\sigma_{x,k}$ and $\sigma_{y,k}$ are the standard deviations of the x and y coordinates of pixels in the k-th cluster. Also, we consider $\{\mathbf{t}_{x,y}\}$ repeating the above process to obtain the texture feature compactness. Finally, we define the compactness of each pixel as a 2 dimensional feature vector that inherits the compactness of the cluster.

The above three feature vectors are normalized, and all these features are formed a 41 dimensional feature vector for each pixel, which captures color, texture and compactness low-level visual features. Fig.1 shows the examples of the proposed saliency detection with different features. It is obvious that the regions belonging to different objects are highlighted with the above features, but the background is still not well suppressed, while the background is suppressed more effectively in the final combined saliency, and the complete salient object is highlighted with well-defined boundaries.

2.2 Graph construction

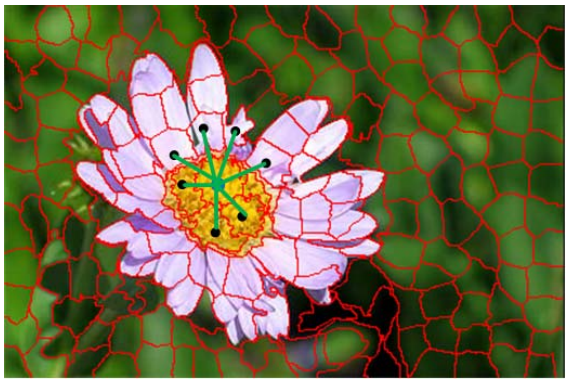

Fig.2. Illustration of the constructed graph model.

Let us assume that an image is represented by a weighted and undirected graph $G = (V, E)$, where V is a set of nodes and E is a set of edges. In our case, a node $v \in V$ corresponds to a superpixel of the image, and an edge $v \in V$ represents the connection between the nodes. An edge connecting two node v_i and v_j is denoted by ω_{ij}. The illustration of the constructed graph model is shown in Fig.2. In this work, each node is a superpixel generated by the SLIC algorithm [24]. Since neighboring nodes are likely to share similar appearance and saliency values, we represent the weight ω_{ij} by enforcing constraints as follows

$$\omega_{ij} = \left\|\mathbf{f}_i - \mathbf{f}_j\right\| \times \exp\left(-\frac{\beta d_{ij}}{\sqrt{W^2 + H^2}}\right) \quad (2)$$

where d_{ij} denotes the Euclidean distance between v_i and v_j, $\mathbf{f}_i$ and $\mathbf{f}_j$ denote the feature vector means of the corresponding superpixels with the three features (color, texture and compactness).

Note that, if only the first term is used as constraint, there may exist zero relevance between any pair of nodes on the graph (this naturally is not always true). By enforcing the spatial constraint, the relevance between nodes is increased when their spatial distance is decreased, which is an important cue for saliency detection. Then, the degree matrix $\mathbf{D}$ is defined as a diagonal matrix that records the sum of the weights connected to each node, whose i-th diagonal element in the degree of node v_i is written as

$$d_i = \sum_j \omega_{ij} \quad (3)$$

We define the affinity matrix $\mathbf{A}$ that represents the reverence of nodes as

$$a_{ij} = \begin{cases} d_i & \text{if } i = j \\ \omega_{ij} & \text{if } v_i \text{ and } v_j \text{ are adjacent nodes} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The use of affinity matrix allows us to formulate the transition matrix $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A} \quad (5)$$

Based on the connection between the random walks, random walker probabilities is equivalent to solve the transition matrix with Markov chain assumption. It satisfies the following equation with the equilibrium distribution $\boldsymbol{\pi}$

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \quad (6)$$

Since the weight ω_{ij} imposed on an edge describes the cost affecting a random walker's choice, the i-the element in the equilibrium distribution $\boldsymbol{\pi}$ represents the expected probability staying in the node v_i. Thus, the saliency for the i-th superpixel can be derived by

$$s_i = \frac{\boldsymbol{\pi}_i - \pi_{\min}}{\pi_{\max} - \pi_{\min}} \quad (7)$$

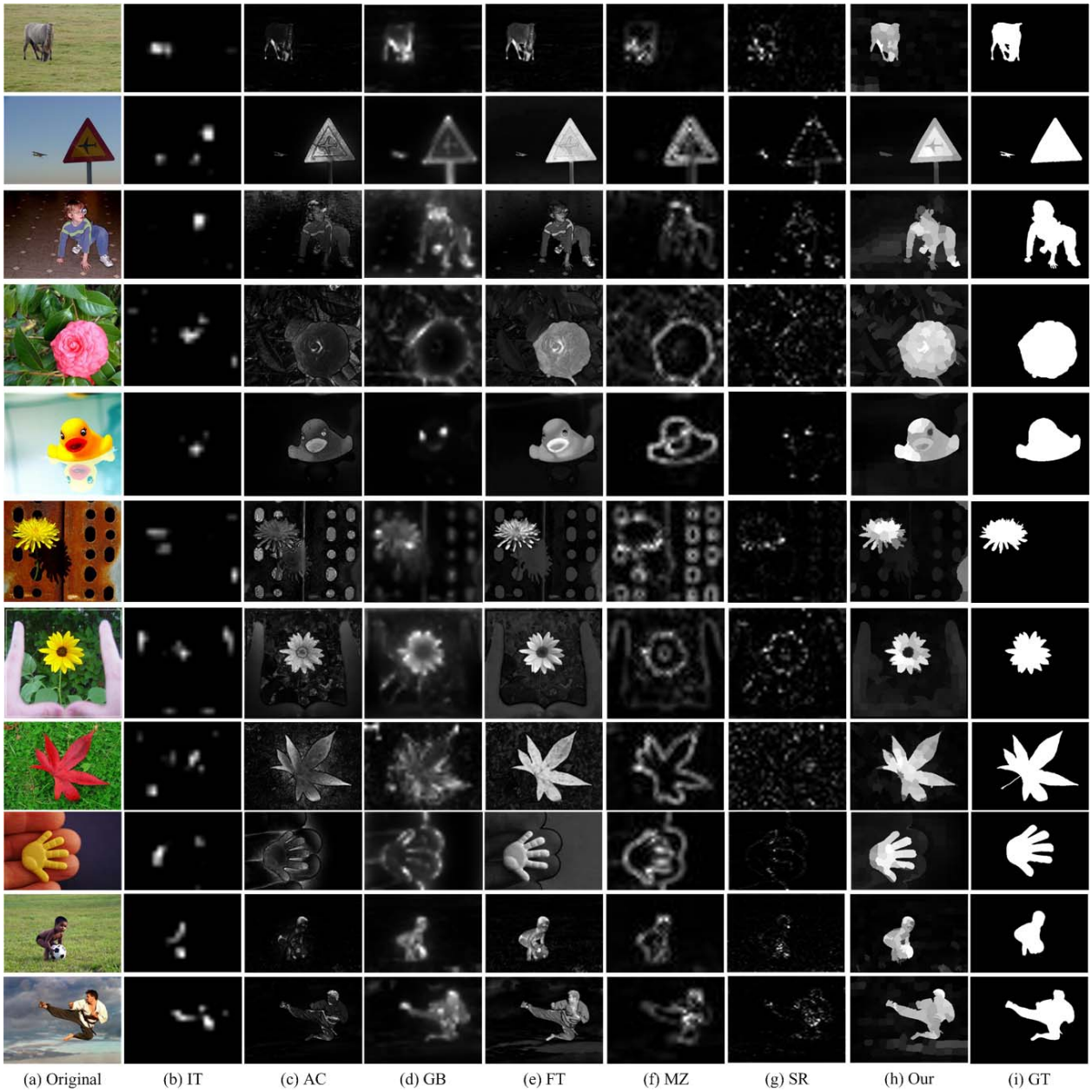

Fig.3. Examples of saliency maps achieved by different methods: (a) original images (a), (b)-(h) saliency maps achieved by the methods IT [9], AC[25], GB[19], FT[12], MZ[26], SR[10] and the proposed method; (i) ground truth masks.

where $\pi_{\max}$ and $\pi_{\min}$ denote the maximum and minimum values in the equilibrium distribution $\boldsymbol{\pi}$.

3 Experimental results and analyses

In order to evaluate the performance of the proposed saliency detection model, we conduct experiments on commonly used MSRA dataset [10]: it includes 1000 images as well as the ground truth binary salient object masks. We compare our method with six state-of-the-art methods including: Itti et al.'s method (IT) [9], Achanta et al.'s method (AC) [25], graph-based saliency method (GB) [19], frequency-tuned method (FT) [12], Ma and Zhang et al.'s method (MZ) [26], spectral residual method (SR) [10]. These methods involve a variety of saliency models. For the existing methods, we use the source codes or executable codes provided by the authors.

In Fig.3, we compare the performance of the proposed method with those six conventional methods. The original images are shown in Fig.3(a),

while the results of the six state-of-the-art methods are presented in Fig.3(b)-(g). The saliency maps obtained by the proposed method are given in Fig.3(h), and the ground truth masks are illustrated in Fig.3(i). The comparison results shows that the proposed method can lead to the improved performance for salient objects extraction from images. Some of the methods, such as GB and MZ, tend to detect only object boundaries as salient, and some methods, such as AC and FT, also declare undesired background as salient regions. On the contrary, the proposed method can highlight the salient regions uniformly and effectively suppress the cluttered backgrounds.

We also compare the results quantitatively, by computing the precision, recall, and F-measure. The precision and recall rates are defined as

$$\text{Precison} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{TP}{FN + TP} \tag{9}$$

where *TP*, *FP*, and *FN* denote the numbers of true positive, false positive, and false negative pixels, respectively. The F-measure rate is defined as

$$\text{F-measure} = \frac{(1+\beta^2)\cdot \text{Precision} \cdot \text{Recall}}{\beta^2 \cdot \text{Precision} + \text{Recall}} \tag{10}$$

where β^2 is set to 0.3 as in [23] to weight precision more than recall. The precision versus recall curves in Fig.4 clearly demonstrate that our method outperforms the other methods, which proves that the saliency maps obtained by the proposed method have the best performance in highlighting salient regions. The comparison results are shown in Fig.5.

It can be clearly observed from Fig. 5 that our method demonstrates the highest precision, recall and F-measure on the MSRA dataset compared with the other six saliency models.

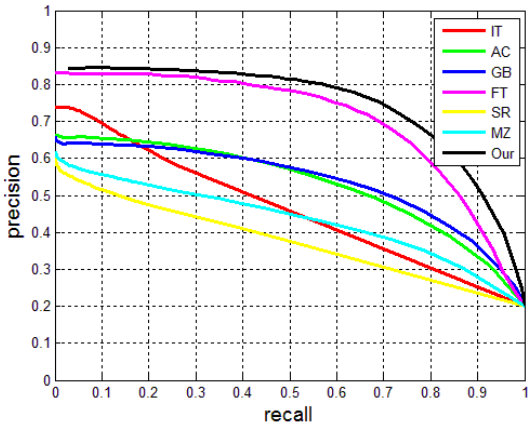

Fig.4. Precision-recall curves of different saliency models

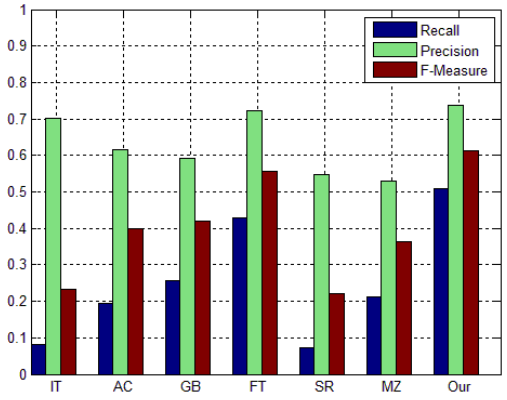

Fig.5. Average precision, recall and F-measure of different saliency models

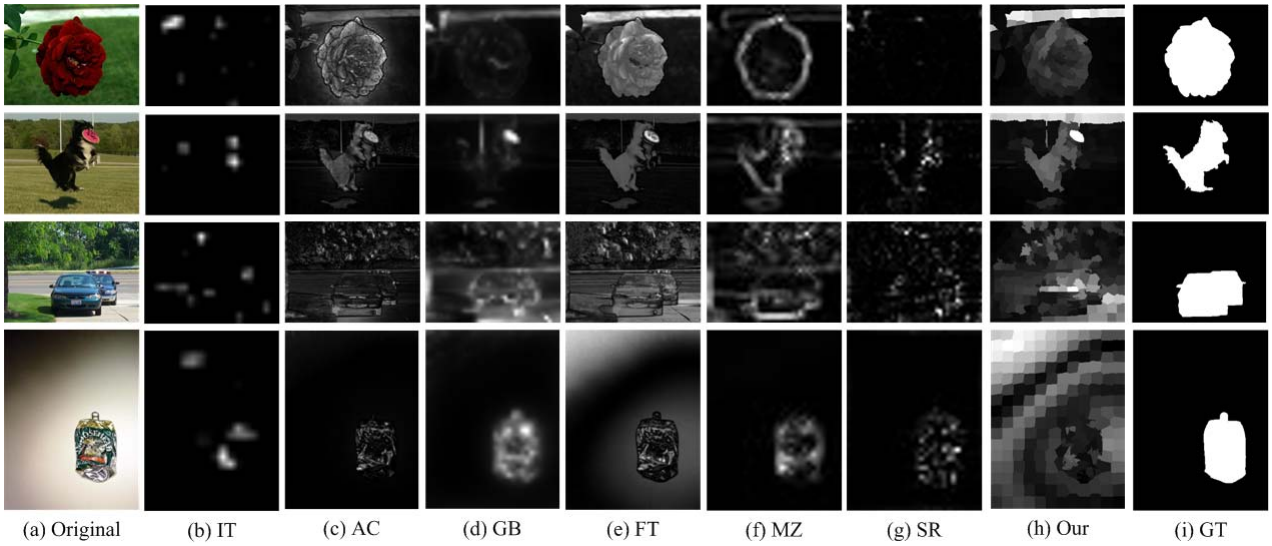

Fig.6. Failure case: (a) original images (a), (b)-(h) saliency maps achieved by the methods IT [9], AC[25], GB[19], FT[12], MZ[26], SR[10] and the proposed method; (i) ground truth masks.

However, although the proposed method performs well in the experiments described above, it may fail to locate the salient objects in some cases. As shown in Fig.6, if the appearance of the object and background regions is similar (i.e., color contrast between the object and background regions are small), our saliency model can not effectively highlight the saliency regions, because color contrast is an important factor but not including in our saliency model. In fact, this issue is challenging for the existing bottom-up saliency detection models without object prior and task information.

4 Conclusion

In this paper, we present an image saliency detection method by constructing graph model. The main contributions of the proposed method are that: (1) color, texture and compactness features are utilized and superpixels are extracted to construct a graph model; (2) saliency for each is measured by calculating random walker probability on the node. Experimental results show that the proposed method outperforms the state-of-the-art methods. In the future work, we will extend our saliency detection framework by incorporating more low-level and high-level feature cues.

Acknowledgement:
This work was supported by the Natural Science Foundation of China (grant 61271021), and the Natural Science Foundation of Ningbo (grant 2012A610039).

References:

[1] N. J. Butko, L. Y. Zhang, G. W. Cottrell, and J. R. Movellan, "Visual saliency model for robot cameras," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2398-2403, 2008.

[2] H. Li and K. N. Ngan, "Saliency model-based face segmentation and tracking in head-and-shoulder video sequences," *Journal of Visual Communication and Image Represent*, vol. 19, no. 5, pp. 320-333, 2008.

[3] Q. Zhang and K. N. Ngan, "Multi-view video based multiple objects segmentation using graph cut and spatiotemporal projections," *Journal of Visual Communication and Image Represent*, vol. 21, no. 5, pp. 453-461, 2010.

[4] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185-1198, May 2011.

[5] Z. Li, S. Qin, and L. Itti, "Visual attention guided bit allocation in video compression," *Image and Vision Computing*, vol. 29, no. 1, pp. 1-14, Jan 2011.

[6] Z. Chen, J. Han, and K. N. Ngan, "Dynamic bit allocation for multiple video object coding," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1117-1124, Jun 2006.

[7] V. Setlur, T. Lechner, M. Nienhaus, and B. Gooch, "Retargeting images and video for preserving information saliency," *IEEE Computer Graphics and Applications*, vol. 27, no. 5, pp. 80–88, 2007.

[8] T. D. Basha, Y. Moses, and S. Avidan, "Stereo seam carving: a geometrically consistent approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2513-2525, 2013.

[9] 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, pp. 1254-1259, 1998.

[10] X. D. Hou and L. Q. Zhang, "Saliency detection: a spectral residual approach," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.

[11] M. M. Cheng, G. X. Zhang, N. J. Mitra, X. Huang, and S. M. Hu, "Global contrast based salient region detection," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 409-416, 2011.

[12] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency tuned salient region detection," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1597-1604, 2009.

[13] E. Erdem and A. Erdem, "Visual saliency estimation by nonlinearly integrating features using region covariance," *Journal of Vision*, vol.13, no. 4, article 11, 2013.

[14] R. Margolin, A. Tal, and L. Manor, "What makes a patch distinct?" in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013.

[15] Z. Liu, L. Meur, and S. H. Luo, "Superpixel-based saliency detection," in *Proc. of International Workshop on Image Analysis for Multimedia Interactive Services(WIAMIS)*, July 2013.

[16] F. Perazzi, P. Krahenbul, Y. Pritch, and A. Hornung, "Saliency filters: contrast based

filtering for salient region detection," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.
[17] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185-207, 2013.
[18] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol.11, no. 3, article 9, 2011.
[19] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," *Advances in Neural Information Processing Systems*, pp. 545-552, 2007.
[20] C. Yang, L. H. Zhang, H. C. Lu, and M. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3166-3173, 2013.
[21] Y. Lu, W. Zhang, H. Lu, X. Xue, "Salient object detection using concavity context," in *Proc. of IEEE International Conference on Computer Vision (ICCV)*. Barcelona, Spain. 2011.
[22] V. Gopalakrishnan, Y. Q. Hu, D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Transactions on Image Processing*, vol. 19, no. 12, pp. 3232-3242, 2010.
[23] J.-S. Kim, J.-Y Sim, C.-S Kim, "Multiscale saliency detection using random walk with restart," IEEE Transactions on Circuits and Systems for Video Technology, vol. 24, no. 2, pp. 198-210, 2014.
[24] R. Achanta, A. Shaji, 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. 11, pp. 2274-2282, 2012.
[25] R. Achanta, F. Estrada, P. Wils, S. Süsstrunk, "Salient region detection and segmentation," in *Proc. of IEEE International Conference on Computer Vision Systems*, pp. 66-75, 2008.
[26] Y. F. Ma and H. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. of ACM Multimedia*, pp. 374-381, 2003.