Probabilistic Hypergraph Optimization for Salient Object Detection

Jinxia Zhang[1,2], Shixiong Fang[1], Krista A. Ehinger[3], Weili Guo[1], Wankou Yang[1], and Haikun Wei[1(✉)]

[1] Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, 2 Sipailou Street, Nanjing 210096, Jiangsu, China
hkwei@seu.edu.cn

[2] Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education, Nanjing University of Science and Technology, Nanjing 210094, China

[3] Centre for Vision Research, York University, Toronto, Canada

Abstract. In recent years, many graph-based methods have been introduced to detect saliency. These methods represent image regions and their similarity as vertices and edges in a graph. However, since they only represent pairwise relations between vertices, they give an incomplete representation of the relationships between regions. In this work, we propose a hypergraph based optimization framework for salient object detection to include not only the pairwise but also the higher-order relations among two or more vertices. In this framework, besides the relations among vertices, both the foreground and the background queries are explicitly exploited to uniformly highlight the salient objects and suppress the background. Furthermore, a probabilistic hypergraph is constructed based on local spatial correlation, global spatial correlation, and color correlation to represent the relations among vertices from different views. Extensive experiments demonstrate the effectiveness of the proposed method.

Keywords: Hypergraph · Optimization framework · Saliency detection

1 Introduction

Since Itti and Koch et al. [1] introduced the first computational model of visual saliency detection, a large amount of approaches have been proposed to detect salient objects in images [2–12]. Visual saliency detection can be applied to many computer vision related applications, including image segmentation, image compression, image quality assessment, object detection, and object recognition.

One series of salient object detection approaches are unsupervised and graph-based [6–10], which model each image as a graph and propagate saliency information via weighted edges connecting image parts. These approaches do not need manually annotated samples and overcome a major drawback of classical,

Y. Sun et al. (Eds.): IScIDE 2017, LNCS 10559, pp. 368–378, 2017.
DOI: 10.1007/978-3-319-67777-4_32

pixel-level bottom-up saliency methods: namely, they tend to highlight object edges more than homogeneous object interiors. Graph-based approaches [6–10] use a pairwise simple graph to model the relations between two image regions. In a simple graph, image regions are vertices; two vertices are connected by an edge and the edge weight represents the similarity between regions.

However, representing the relations among image regions in a pairwise simple graph gives an incomplete representation of the image. It is important to consider not only the pairwise relation between two vertices but also local grouping information among two or more vertices [13]. These higher-order relations can be represented by a hypergraph, a generalization of a simple graph [14]. In a hypergraph, hyperedges are used to describe complex relations among any number of vertices. Hypergraphs have been proved to be useful in many applications, such as image retrieval [13] and image segmentation [15]. In the salient object detection field, very few works have made use of the hypergraph [16]. In the work of Li et al. [16], a binary hypergraph is constructed using a clustering algorithm at multiple image scales, which describes the binary relations between vertices and hyperedges, i.e. whether a vertex belongs to a hyperedge. A hypergraph modeling approach is then proposed to capture the contextual saliency information on image superpixels and the final hypergraph-based saliency map is essentially the average of saliency detection results from different scales.

In this work, we introduce a new hypergraph-based method for salient object detection, which has two main contributions. Firstly, a novel hypergraph based optimization framework is proposed. Salient object detection aims to separate the salient foreground from the non-salient background. Thus, we explicitly exploit both the foreground queries and the background queries in this optimization framework to uniformly highlight all the salient regions and suppress the background. These two kinds of queries are generated based on the boundary prior of visual saliency and the local grouping information among different vertices.

Our second contribution is that a probabilistic hypergraph is constructed to better represent the local grouping information, which captures not only the binary relations but also the probability that a vertex belongs to a hyperedge. We fuse local spatial correlation, global spatial correlation, and color correlation among multiple vertices into a probabilistic hypergraph by adding different types of hyperedges. Each vertex in the hypergraph serves as the centroid for a local spatial hyperedge which connects a centroid vertex to its local spatial neighbors, a global spatial hyperedge which connects a centroid vertex to the global border vertices of the image, and a color hyperedge which connects a centroid vertex to its neighbors in the color space. We further define the probability that a vertex belongs to a hyperedge to be the similarity between this vertex and the corresponding centroid vertex of the hyperedge. We then define the weight of each hyperedge as the quadratic sum of the probabilities that different vertices belong to this hyperedge.

We test our probabilistic hypergraph based optimization framework on three databases and demonstrate that our method outperforms eleven state-of-the-art models. The remainder of this paper is organized as follows: The hypergraph based optimization framework is introduced in Sect. 2; Sect. 3 describes how we construct a probabilistic hypergraph; Our testing procedure and experimental comparison are described in Sect. 4; Sect. 5 concludes the paper.

2 Hypergraph Based Optimization Framework

Let V denote a finite set of vertices and E denote a series of hyperedges, which is a family of subsets of V such that $\bigcup_{e \in E} = V$. A hypergraph can be represented by a $|V| \times |E|$ incident matrix $H(v_i, e_j)$ where i is an index of a vertex and j is an index of a hyperedge.

$$H(v_i, e_j) = \begin{cases} p(v_i|e_j) & \text{if } v_i \in e_j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

If a vertex v_i is contained in a hyperedge e_j, the corresponding value $H(v_i, e_j)$ equals to the probability that v_i belongs to e_j; otherwise, the value equals to 0. Thus, each column of the incident matrix H represents the composition information of a hyperedge in the hypergraph. Let a diagonal matrix W record the hyperedge weights, where each hyperedge e is assigned a positive weight $W(e)$. Based on the incident matrix H and the weight matrix W, a diagonal matrix D_v which records the degree of each vertex $D_v(v_i)$ can be defined as $D_v(v_i) = \sum_{e_j \in E} W(e_j)H(v_i, e_j)$, and a diagonal matrix D_e which records the degree of each hyperedge $D_e(e_j)$ can be defined as $D_e(e_j) = \sum_{v_i \in V} H(v_i, e_j)$.

Based on hypergraph, a ranking algorithm has been proposed for classification [14]. In this algorithm, a smoothness term is used to include the relations among different vertices and a fitting term is used to constrain the final result to be close to the original queries. The hypergraph based ranking algorithm is suitable for the situation where one kind of specific queries has been given, e.g. image retrieval [13]. The goal of salient object detection is to separate the salient foreground from the non-salient background. Thus, both the foreground queries and the background queries are important to detect saliency in images. In this work, we propose a novel hypergraph-based optimization framework which includes both the foreground queries and the background queries for saliency detection. The proposed hypergraph-based optimization framework aims to compute a vector $S : V \Rightarrow \mathbb{R}$, which assigns a saliency value $S(v_i)$ to each vertex v_i.

$$\begin{aligned} S^* &= \arg\min_S(\frac{1}{2}\Omega + \lambda_f \Psi + \lambda_b \Phi), where \\ \Omega &= \sum_{e_k \in E} \sum_{v_i, v_j \in e_k} \frac{W(e_k)}{D_e(e_k)} H(v_i, e_k)H(v_j, e_k)(S(v_i) - S(v_j))^2 \\ \Psi &= \sum_{i=1}^{n} Q_f(v_i) \left\| S(v_i) - L_f \right\|^2 \\ \Phi &= \sum_{i=1}^{n} Q_b(v_i) \left\| S(v_i) - L_b \right\|^2 \end{aligned} \tag{2}$$

In the term Ω, $S(v_i)$ and $S(v_j)$ are the saliency values of the vertices v_i and v_j separately. $H(v_i, e_k)$ and $H(v_j, e_k)$ indicate the probabilities that v_i and v_j belong to the hyperedge e_k. $W(e_k)/D_e(e_k)$ can be understood as the normalized weight of the hyperedge e_k. From the formula, we can see that Ω indicates the smoothness constraint defined in a hypergraph: If two vertices v_i and v_j have high probabilities to belong to the same hyperedge frequently and these hyperedges have high weights, the saliency values of these vertices should be close.

Among the term Ψ, L_f is the label of the salient foreground and is set as 1 in our paper. $Q_f(v_i)$ indicates whether the vertex v_i is a foreground query: If v_i is a foreground query, $Q_f(v_i) = 1$; otherwise, $Q_f(v_i) = 0$. Thus, Ψ can be understood as a foreground fitting term, which encourages the saliency value of a vertex v_i to be close to the label of the salient foreground if this vertex is a foreground query.

Similarly, the term Φ can be understood as a background fitting term. L_b is the label of the non-salient background and is set as -1. $Q_b(v_i)$ indicates whether the vertex v_i is a background query: If v_i is a background query, $Q_b(v_i) = 1$; otherwise, $Q_b(v_i) = 0$. This term Φ encourages a vertex v_i which is a background query to take a saliency value close to the background label.

In Formula 2, λ_f and λ_b are the weighting parameters, which specify the relative balance of three terms. By setting the derivative of the function in Formula 2 to be zero, the saliency values of different vertices can be computed as:

$$S^* = (D_v - HWD_e^{-1}H^T + \lambda_f Q_f + \lambda_b Q_b)^{-1}(\lambda_f Q_f - \lambda_b Q_b) \qquad (3)$$

In this work, we first choose initial foreground queries Q_f and initial background queries Q_b based on the boundary prior of visual saliency. According to the boundary prior, the vertices on the four sides of the image are more likely to be background. We then use the vertices which are not on the image borders as the initial foreground queries and set each of the four image borders as the initial background queries separately to get four maps based on Formula 3. These maps are then normalized to be in [0, 1] and multiplied to get a temporal result T. The final foreground and background queries are acquired based on this result T using different thresholds th_f and th_b: If a vertex v_i has a value $T(v_i) > th_f$, then this vertex is a foreground query; If a vertex v_i has a value $T(v_i) < th_b$, this vertex is a background query.

According to Formula 3, besides the foreground and background queries, the hypergraph represented by H and W is also critical to compute the saliency value of each vertex. In Sect. 3, we describe how we construct the hypergraph for salient object detection.

3 Probabilistic Hypergraph Construction

We first employ the SLIC algorithm [17] to segment an input image I into n homogeneous superpixels and then define these superpixels as the vertices of our hypergraph. The number of the homogeneous superpixels in each input image is set to 300 in this work. For each superpixel, we compute its mean in the CIELab color space and the image spatial coordinates as the color and spatial features of this superpixel.

3.1 Hyperedges and Weights

To get a probabilistic hypergraph, we need to construct the incident matrix H, i.e. choose different hyperedges and compute the probabilities that different vertices belongs to these hyperedges. In this work, given a set of vertices V, we choose different hyperedges based on three kinds of correlations: local spatial correlation, global spatial correlation and color correlation. Thus, for each vertex v_i, three hyperedges are chosen: a local spatial hyperedge, which contains v_i itself as a centroid vertex and its immediate spatial neighbors that share a boundary with the centroid vertex in the input image; a global spatial hyperedge, which contains v_i as a centroid vertex and the global boundary vertices on the four borders of the image; and a color hyperedge which contains v_i as a centroid vertex and its neighbors in the CIELab color space. We further compute the probability that a contained vertex belongs to a hyperedge as the similarity between this vertex and the centroid vertex of the hyperedge. Our method to compute the similarity between two vertices are introduced in Sect. 3.2.

Besides the incident matrix H, the diagonal hyperedge weight matrix W is also important for salient object detection. In our work, we define the weight of each hyperedge $w(e_j)$ as the sum of the squares of the probability that each contained vertex belongs to the hyperedge e_j.

$$w(e_j) = \sum_{v_i \in e_j} H(v_i, e_j)^2 \tag{4}$$

If all vertices belonging to a hyperedge e_j have a higher probability to be part of the hyperedge, this hyperedge has higher inner group similarity and thus should be assigned a higher weight. Otherwise, this hyperedge should be assigned a lower weight.

3.2 Similarity Matrix

Let a $n * n$ matrix SIM contain the similarity between each pair of the vertices. In this work, we compute the similarity between two vertices using exponential function based on both the color distance and the spatial distance.

$$SIM(v_i, v_j) = exp(-\frac{D_c(v_i, v_j) + D_s(v_i, v_j)}{2\sigma}) \tag{5}$$

A scale parameter σ is used to control the strength of the color and spatial distance. If σ has a small value, only vertices with close color features and spatial positions would make a contribution. If σ has a large value, the vertices with large distances would also have big influence on each other. In this work, σ is set to 0.1 for all experiments. $D_c(v_i, v_j)$ and $D_s(v_i, v_j)$ respectively represents the distance of the color features and the spatial positions between the vertices v_i and

v_j. We use the Euclidean distance to compute the color distance $D_c(v_i, v_j)$ and employ the sine spatial distance proposed in our previous work [18] to compute the spatial distance $D_s(v_i, v_j)$, as shown in Formula 6.

$$D_s(v_i, v_j) = \sqrt{(\sin(\pi \cdot |x_i - x_j|))^2 + (\sin(\pi \cdot |y_i - y_j|))^2} \tag{6}$$

In this formula, x_i and y_i represent the horizontal and vertical coordinates of a vertex i in the image plane, which have been normalized to be in [0, 1]. Experimental comparison in the work [18] has demonstrated that the sine spatial distance helps to uniformly suppress the non-salient background.

4 Experimental Comparison

We compare with eleven state-of-the-art saliency detection methods: GB [2], FT [3], RC [4], CB [5], HM [16], GR [6], BD [7], CL [8], GP [9], PM [10] and MST [11]. Following [3,4], we chose different methods based on various principles: recency (PM, MST, CL and GP) and variety (FT is frequency tuned; RC uses regional contrast; CB is based on shape and context knowledge; GB, GR, CL, GP and PM are graph based; HM is hypergraph based; and MST employs minimum spanning tree).

We test the performances of different methods on three standard salient object databases: the 10000-image MSRA10K database [4], the 643-image iCoseg database [19] and the 1000-image Extended Complex Scene Saliency Database (ECSSD) [20]. The MSRA10K database contains many images with one salient object. The iCoseg database contains a lot of images with multiple salient foreground objects. The ECSSD database contains many natural images with complex foreground and background patterns.

The most common evaluation metric for salient object detection, i.e. precision-recall curves (PR curves) is used in this work for experimental comparison. To get PR curves, the saliency map is binarized at each threshold in the range [0:1:255] and the precision and recall values are computed at each threshold by comparing the binary masks against the ground truth.

There are four main parameters in the proposed probabilistic hypergraph-based optimization framework: the balance weights λ_f and λ_b in Formula 3; the thresholds th_f and th_b to get the final foreground and background queries. These parameters are empirically chosen, $\lambda_f = 0.05$, $\lambda_b = 0.1$, $th_f = 0.05$ and $th_b = 0.3$, for all the experiments.

Quantitative Comparison: The quantitative comparison between our method and eleven state-of-the-art methods based on PR curves is shown in Fig. 1.

Figure 1(a) shows PR curves of different methods in images which have one salient object in the content. Previous graph-based methods, e.g. CL, GP and PM already have competitive performances. Our proposed method outperforms these methods, demonstrating the effectiveness of our method. HM and our method are both hypergraph based models, but the algorithms proposed by

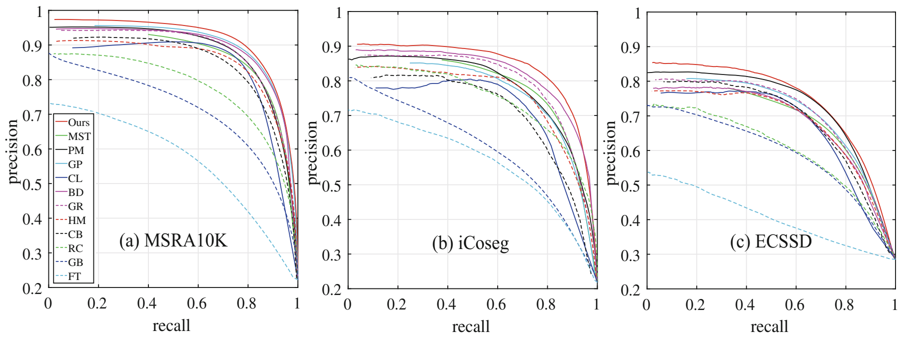

Fig. 1. The quantitative comparison of different methods on MSRA10K, iCoseg and ECSSD databases. Our method gives the best performance on all three databases.

these two models are totally different. Firstly, HM constructs a binary hypergraph based on a clustering algorithm while our method constructs a probabilistic hypergraph based on local spatial correlation, global spatial correlation and color correlation. Secondly, HM detects salient objects based on a hypergraph modeling algorithm which result is essentially the average of saliency detection maps from different scales, while in this work we propose a novel hypergraph-based optimization framework which make use of both the foreground queries and the background queries. The quantitative comparison demonstrates that our method outperforms HM by a large margin.

We further evaluate our method on the iCoseg database which contains a lot of images with multiple foreground objects, shown in Fig. 1(b). We note that the PR curves of different saliency detection methods drop a little in comparison to the PR curves on the MSRA10K database. This phenomenon indicates that it's harder to detect all of the foreground regions when there are multiple salient objects in an image. From the comparison, it can be seen that the proposed method outperforms other saliency detection methods. It demonstrates that the construction of the probabilistic hypergraph and the proposed optimization framework can help to detect multiple salient objects in an image.

The ECSSD database contains 1000 complex natural images. The performance comparison in Fig. 1(c) demonstrates that the proposed probabilistic hypergraph based optimization framework can better detect salient foreground objects in complex images than other state-of-the-art methods.

We further compare the performances of different hypergraphs, which use all or part of local spatial correlation, global spatial correlation and color correlation on the ECSSD database, shown in Fig. 2. The hypergraphs with only one correlation have the worst performances. The hypergraphs with two of the correlations outperforms the hypergraphs considering only one correlation. And our method using all three correlations has the best performance.

Visual Comparison: Different evaluated saliency detection methods can be visually compared in Fig. 3, which displays representative images to highlight

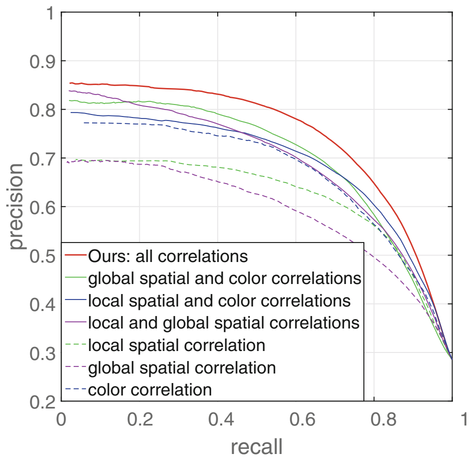

Fig. 2. The PR curves on ECSSD database with different hypergraphs. (Color figure online)

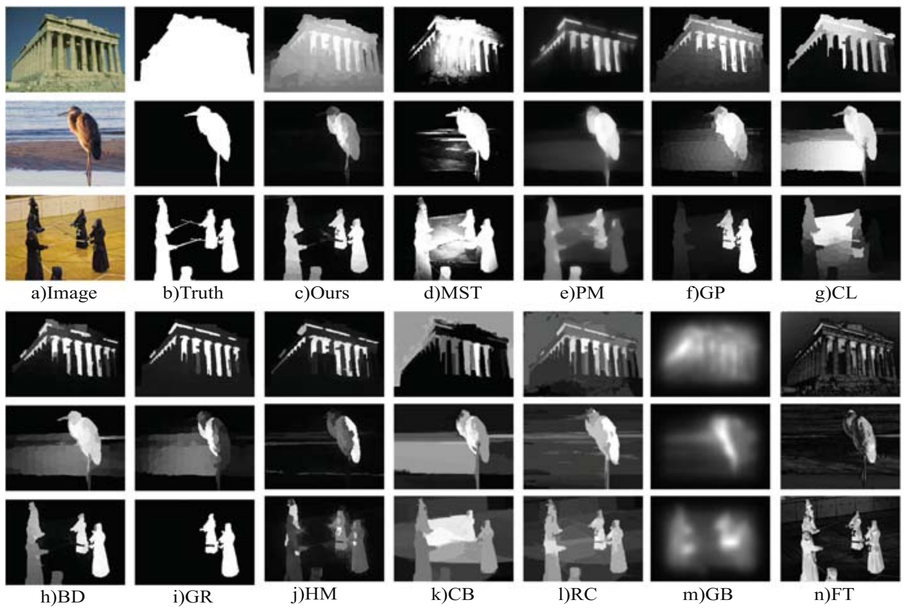

Fig. 3. Comparison of different salient object detection methods. The first column is the original image, the second is the ground truth. The third row is our saliency detection result, and the remaining columns are results of other evaluated methods.

the differences between these methods. An image with a complex salient object is shown in the first row. Our method uniformly highlights the building while most other methods only detect part of it. The second row shows an image with a complex background. Our method can well suppress the background while a lot of other methods wrongly highlight a portion of the background. When there are multiple salient objects in an image (the last row), our method can uniformly highlight all of the salient objects and largely suppress the background while other methods only detect part of the salient objects or wrongly highlight the background. In short, the proposed probabilistic hypergraph based optimization framework outperforms other saliency detection methods in generating results which are more consistent with the ground truth.

5 Conclusion

In this work we have proposed a probabilistic hypergraph based optimization framework to detect salient objects in images. In this framework, both foreground queries and background queries are explicitly exploited to better detect salient objects. To construct a probabilistic hypergraph, we use three different correlations in natural images, including local spatial correlation, global spatial correlation, and color correlation. Each vertex in the hypergraph serves as the centroid for a local spatial hyperedge which connects the centroid vertex to its immediate spatial neighbors, a global spatial hyperedge which connects the centroid vertex to the vertices on the borders of the image, and a color hyperedge which connect the centroid vertex to its neighbors in color space. The probability that a vertex belongs to a hyperedge and the weight of each hyperedge are computed based on a similarity matrix, which encodes the similarities between pairs of vertices and is computed based on both color distance and sine spatial distance. Extensive experimental comparisons demonstrate that the proposed method outperforms state-of-the-art methods on a variety of images for salient object detection.

Acknowledgments. This work was supported by the National Natural Science Fund of China (Grant numbers 61233011, 61374006, 61473086, 61703100); Major Program of National Natural Science Foundation of China (Grant number 11190015); Natural Science Foundation of Jiangsu (Grant number BK20131300, BK20170692); the Innovation Fund of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education (Nanjing University of Science and Technology, Grant number JYB201601); the Innovation Fund of Key Laboratory of Measurement and Control of Complex Systems of Engineering (Southeast University, Grant number MCCSE2017B01); and the Fundamental Research Funds for the Central Universities (2242016k30009).

References

1. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)

2. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Advances in Neural Information Processing Systems, pp. 545–552 (2006)
3. Achanta, R., Hemami, S., Estrada, F., et al.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
4. Cheng, M.-M., Zhang, G.-X., Mitra, N.J., et al.: Global contrast based salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416 (2011)
5. Jiang, H., Wang, J., Yuan, Z., et al.: Automatic salient object segmentation based on context and shape prior. In: British Machine Vision Conference, vol. 6, p. 7 (2011)
6. Yang, C., Zhang, L., Lu, H., et al.: Saliency detection via graph-based manifold ranking. In: Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2013)
7. Zhu, W., Liang, S., Wei, Y., et al.: Saliency optimization from robust background detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2814–2821 (2014)
8. Gong, C., Tao, D., Liu, W., et al.: Saliency propagation from simple to difficult. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2531–2539 (2015)
9. Jiang, P., Vasconcelos, N., Peng, J.: Generic promotion of diffusion-based salient object detection. In: IEEE International Conference on Computer Vision, pp. 217–225 (2015)
10. Kong, Y., Wang, L., Liu, X., Lu, H., Ruan, X.: Pattern mining saliency. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 583–598. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4_35
11. Tu, W.-C., He, S., Yang, Q., et al.: Real-time salient object detection with a minimum spanning tree. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2334–2342 (2016)
12. Zhang, J., Ehinger, K.A., Wei, H., et al.: A novel graph-based optimization framework for salient object detection. Pattern Recogn. **64**, 39–50 (2017)
13. Huang, Y., Liu, Q., Zhang, S., et al.: Image retrieval via probabilistic hypergraph ranking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3376–3383 (2010)
14. Zhou, D., Huang, J., Schölkopf, B.: Learning with hypergraphs: clustering, classification, and embedding. In: Advances in Neural Information Processing Systems, pp. 1601–1608 (2006)
15. Kim, S., Nowozin, S., Kohli, P., et al.: Higher-order correlation clustering for image segmentation. In: Advances in Neural Information Processing Systems, pp. 1530–1538 (2011)
16. Li, X., Li, Y., Shen, C., et al.: Contextual hypergraph modeling for salient object detection. In: IEEE International Conference on Computer Vision, pp. 3328–3335 (2013)
17. Achanta, R., Shaji, A., Smith, K., et al.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
18. Zhang, J., Ehinger, K.A., Ding, J., et al.: A prior-based graph for salient object detection. In: IEEE International Conference on Image Processing, pp. 1175–1178 (2014)

19. Batra, D., Kowdle, A., Parikh, D., et al.: iCoseg: interactive co-segmentation with intelligent scribble guidance. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3169–3176 (2010)
20. Yan, Q., Xu, L., Shi, J., et al.: Hierarchical saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1155–1162 (2013)