Journal of Electronic Imaging

JElectronicImaging.org

Multiple Gestalt principles-based graph for salient region detection

Jinxia Zhang
Shixiong Fang
Haifeng Zhao
Guang-Hai Liu
Haikun Wei
Lihuan Chen
Kanjian Zhang

Jinxia Zhang, Shixiong Fang, Haifeng Zhao, Guang-Hai Liu, Haikun Wei, Lihuan Chen, Kanjian Zhang, "Multiple Gestalt principles-based graph for salient region detection," *J. Electron. Imaging* **27**(5), 051227 (2018), doi: 10.1117/1.JEI.27.5.051227.

Journal of Electronic Imaging 27(5), 051227 (Sep/Oct 2018)

Multiple Gestalt principles-based graph for salient region detection

Jinxia Zhang,[a] Shixiong Fang,[a] Haifeng Zhao,[b] Guang-Hai Liu,[c] Haikun Wei,[a,*] Lihuan Chen,[a] and Kanjian Zhang[a]
[a]Southeast University, School of Automation, Ministry of Education, Key Laboratory of Measurement and Control of CSE, Nanjing, China
[b]Jinling Institute of Technology, School of Software Engineering/Nanjing Software Institute, Nanjing, China
[c]Guangxi Normal University, College of Computer Science and Information Technology, Guilin, China

Abstract. Recently, a number of graph-based approaches have been proposed to detect salient regions in images. Although the graph is essential for these approaches, the graph construction method has not been studied in much detail. We propose a graph construction method that makes better use of multiple Gestalt principles. Specifically, spatial proximity, color similarity, and texture similarity between image regions are employed to choose different edges of the graph. Furthermore, the confliction among multiple Gestalt principles is solved using a primal rank support vector machines algorithm to compute the edge weights. Our experimental results on various salient region detection databases with comparisons to representative approaches demonstrate that the proposed graph construction method helps to detect salient objects in images. *© 2018 SPIE and IS&T* [DOI: 10.1117/1.JEI.27.5.051227]

Keywords: salient region detection; graph construction; Gestalt principle.

Paper 180058SS received Jan. 15, 2018; accepted for publication May 9, 2018; published online Jun. 7, 2018.

1 Introduction

The aim of salient region detection is to identify the visually distinctive objects in images. It is a fundamental problem and can be used as a preprocessing step in many applications of image processing and computer vision, including image segmentation,[1] image quality assessment,[2] image and video compression,[3] picture collage,[4] object detection and recognition,[5,6] object tracking,[7] region-of-interest extraction,[8,9] content-based image retrieval,[10] etc. Thus, salient region detection has gained much popularity.[11–25]

In recent years, a few graph-based approaches have been proposed to detect salient objects in images. Since a graph is capable of conveniently describing the relations between image regions, graph-based saliency detection approaches achieve top performances. Yang et al.[17] used a graph-based manifold ranking algorithm to detect salient objects in images. This work concentrates on the computation of the saliency-related queries and ranks the saliency values of different image regions based on the similarity between image regions and the queries. Lu et al.[16] proposed a method to learn optimal saliency-related seeds and diffuse the saliency information along a graph to get the final saliency detection results. Gong et al.[20] exploited an iterated graph-based optimization framework to propagate saliency value from simple image regions to difficult image regions.

Although the graph is the heart of graph-based approaches, very few works studied the graph construction methods in detail. In a common graph used in the above graph-based saliency detection approaches, spatially nearby points are connected as edges and the edge weights are computed based on the color features between connected image regions. Gestalt principles are implicitly employed in this graph. Gestalt principles enable the human vision to easily extract structured information from chaos.[26] This phenomenon is also called perceptual grouping, where visual stimulus elements are grouped together to form a percept based on a set of Gestalt principles, such as proximity and similarity.[27] In the commonly used graph for salient region detection, only one Gestalt principle, i.e., spatial proximity, is used to choose edges and one Gestalt principle, i.e., color similarity is used to compute edge weights. Yu et al.[24] explicitly exploited Gestalt principles and experimentally demonstrated that these principles are useful for salient region detection. In their work, two Gestalt principles (similarity and closure) are employed to compute edge weights. However, they still choose graph edges based on one Gestalt principle (spatial proximity). Zhang et al.[25] proposed a new graph-based optimization framework for salient region detection which incorporates multiple graphs and a popular cognitive property (rarity) about visual saliency. They constructed multiple graphs according to multiple Gestalt principles. For each graph, one Gestalt principle is exploited to choose graph edges. In both of the above works,[24,25] the confliction problem among Gestalt principles is not considered.

In this paper, we propose a graph construction method based on multiple Gestalt principles, i.e., spatial proximity, color similarity, and texture similarity, and solve the confliction problem among multiple Gestalt principles. In specific, we first oversegment the input image into multiple superpixels using a simple linear iterative clustering (SLIC) algorithm[28] and define these superpixels as the nodes of the constructed graph. We choose part of the edges of the constructed graph based on spatial proximity and connect each node with its spatial neighbors. We choose another part of the edges of the constructed graph based on color similarity and texture similarity: each node is connected to

*Address all correspondence to: Haikun Wei, E-mail: hkwei@seu.edu.cn

other nodes which have both similar color features and similar texture features. The edge weights are further computed based on all three Gestalt principles. Since different Gestalt principles are not simply added together in the human vision system and there are conflictions among these principles, in this work we learn the edge weights based on a primal rank support vector machines (PRSVM) algorithm[29] and solve the confliction problem among multiple Gestalt principles.

Based on the constructed graph, we further compute the saliency values of different nodes of the graph under a graph-based optimization framework. We test the proposed graph construction method on different salient region detection databases and prove that the proposed method outperforms 12 state-of-the-art models. The rest of this paper is organized as follows: the construction method of the graph is introduced in Sec. 2. The graph-based optimization framework is introduced in Sec. 3. The testing procedure and the comparison results are described in Sec. 4. And Sec. 5 concludes the paper.

2 Multiple Gestalt Principles-Based Graph

Let $G(\mathcal{V},\mathcal{E},\mathcal{W})$ denote the constructed graph for the input image I, where $\mathcal{V}$ represents a set of nodes, $\mathcal{E}$ represents a set of edges that each of them links a pair of nodes in $\mathcal{V}$, and $\mathcal{W}$ represents the corresponding weights of the edges. We first oversegment the input image I into n superpixels using the SLIC algorithm[28] and define these superpixels as the nodes $\mathcal{V}=\{v_1,v_2,\ldots,v_n\}$ of the constructed graph. For each node, we mainly extract three kinds of features: spatial location feature, color feature, and texture feature. For the spatial location feature, we compute the mean normalized row coordinate and mean normalized column coordinate for each node. For the color feature, we compute mean color values in CIELab color space[30] for each node. For the texture feature, we consider max responses from Leung–Malik filter bank[31] for each node.

2.1 *Choosing Edges Based on Multiple Gestalt Principles*

A commonly used graph for salient region detection chooses the edges based on spatial proximity.[17–22] In this work, we choose the edges of the graph based on multiple Gestalt principles, i.e., spatial proximity, color similarity, and texture similarity.

The Gestalt principle of spatial proximity means that the image regions that are spatially nearby are more likely to be grouped together to form a percept. We define an adjacency matrix $A=[a_{ij}]_{n\times n}$ to denote whether two nodes are spatial neighbors, i.e., two nodes share a boundary in the input image. If the nodes v_i and v_j are spatial neighbors, a_{ij} is 1; otherwise, a_{ij} is 0. Let $B\in\mathcal{V}$ denote a set of border nodes on the four borders of the input image. Inspired by spatial proximity, edges are chosen according to the following rules: (1) each node is connected to its spatial neighbors and to its spatial neighbors' neighbors which are spatial neighbors of any node in the first set of spatial neighbors; (2) any two nodes are connected if both of them are spatial neighbors of border nodes. We denote the above rules by R_1 and R_2, and the edge set $\mathcal{E}_1\in\mathcal{E}$ and $\mathcal{E}_2\in\mathcal{E}$ can be obtained as

$$\begin{aligned}R_1:\mathcal{E}_1&=\{(v_i,v_j)|v_i,v_j\in\mathcal{V},a_{ij}=1\}\\&\quad\cup\{(v_i,v_k)|v_k\in\mathcal{V},a_{jk}=1\}\\R_2:\mathcal{E}_2&=\{(v_i,v_j)|v_b,v_{b'}\in B,a_{ib}=1\ \&\ a_{jb'}=1\}.\end{aligned}\tag{1}$$

The Gestalt principles of similarity mean that image regions which are similar are more likely to be grouped together to form a percept. We employ both color similarity and texture similarity to choose another part of the edges. Inspired by color similarity and texture similarity, edges are chosen according to the following rule: each node is linked to other nodes which have similar features in both color space and texture space. We denote the above rule by R_3 and the edge set $\mathcal{E}_3\in\mathcal{E}$ can be obtained as

$$R_3:\mathcal{E}_3=\{(v_i,v_j)|\quad v_i,v_j\in\mathcal{V},D_{ij}^c<\eta_c\ \&\ D_{ij}^t<\eta_t\},\tag{2}$$

where D_{ij}^c denotes the Euclidean distance of CIELab color features between two nodes v_i and v_j. D_{ij}^t denotes the Chi-square distance of texture features between two nodes v_i and v_j. Two parameters η_c and η_t are used to choose neighbors.

2.2 *Computing Edge Weights Based on PRSVM*

In the commonly used graph for salient region detection,[17–22] one Gestalt principle, i.e., color similarity is usually used to compute edge weights of the graph. In this work, we employ multiple Gestalt principles to compute edge weights. A simple way to combine different Gestalt principles is to add the features measured according to these principles together. However, this strategy is not good because of the problem of Gestalt confliction.[26] Gestalt confliction means that the relative importance weights of each Gestalt principle are different and may conflict with each other.

In this section, we quantify the confliction among three Gestalt principles (spatial proximity, color similarity, and texture similarity) by learning the relative importance weights of different principles to get the corresponding edge weights. Specifically, we use a learning-to-rank algorithm, i.e., PRSVM algorithm to learn a ranking score $F(e_{ij})$ for an edge e_{ij} connecting two nodes v_i and v_j using the weighted sum of the features measured by three Gestalt principles

$$F(e_{ij})=\alpha_1 GP_{ij}^s+\alpha_2 GP_{ij}^c+\alpha_3 GP_{ij}^t.\tag{3}$$

In the above equation, GP_{ij}^s, GP_{ij}^c, and GP_{ij}^t denote the features between the nodes v_i and v_j according to spatial proximity, color similarity, and texture similarity, respectively. The parameters α_1, α_2, and α_3 represent relative importance weights of different Gestalt principles for salient region detection. Given any node $v_i\in\mathcal{V}$, the ranking score $F(e_{ij})$ should be larger than the ranking score $F(e_{ik})$, if the node v_i more tend to be grouped with the node v_j rather than the node v_k. Let "$(v_i,v_j)\succ(v_i,v_k)$" denote that "node v_i more tend to be grouped with v_j than v_k." The goal of the PRSVM algorithm is to learn a ranking function F that outputs ranking scores such that $F(e_{ij})>F(e_{ik})$ for any $(v_i,v_j)\succ(v_i,v_k)$.

According to Gestalt principles, if two nodes are close to each other in the image plane, these nodes should be grouped together and thus should have a large feature value of spatial proximity. Similarly, if two nodes have small color distance

or texture distance, they should have a large feature value of color similarity or texture similarity. In this work, we simply define GP^s_{ij}, GP^c_{ij}, and GP^t_{ij} as follows:

$$GP^s_{ij} = -D^s_{ij}, \quad GP^c_{ij} = -D^c_{ij}, \quad GP^t_{ij} = -D^t_{ij}. \tag{4}$$

We employ the Euclidean distance of CIELab color features between two nodes v_i and v_j to compute D^c_{ij} and employ the Chi-square distance of texture features between two nodes v_i and v_j to compute D^t_{ij}, as in Sec. 2.1. We employ the method proposed in Zhang et al.'s work[32] to compute the distance of spatial location features D^s_{ij}

$$D^s_{ij} = \sqrt{[\sin(\pi \cdot |x_i - x_j|)]^2 + [\sin(\pi \cdot |y_i - y_j|)]^2}. \tag{5}$$

In the above equation, x_i and y_i are the horizontal and vertical coordinates of a node v_i, which have been normalized to be in [0, 1]. This method to compute spatial distance helps to suppress the nonsalient background.

To learn the optimal relative importance weights of different Gestalt principles for salient region detection, we train the parameters $\alpha = [\alpha_1; \alpha_2; \alpha_3]$ on the saliency database and get the positive and negative training samples according to the following method. From the training image, we randomly choose pairs of nodes which belong to the same class (salient object or nonsalient background) as positive samples, denoted as v_+. The features of a positive sample $X(v_+)$ are constituted of the features measured according to spatial proximity, color similarity, and texture similarity between two nodes of the positive sample: $X(v_+) = [GP^s_{v_+}; GP^c_{v_+}; GP^t_{v_+}]$. We randomly choose pairs of nodes that belong to different classes (one node belongs to a salient object and another belongs to the nonsalient background) as negative samples, denoted as v_-. Similarly, the features of a negative sample $X(v_-)$ are constituted of the features measured according to spatial proximity, color similarity, and texture similarity between two nodes of the negative sample: $X(v_-) = [GP^s_{v_-}; GP^c_{v_-}; GP^t_{v_-}]$. The learning-to-rank objective is then to compute $F(e_{ij}) = \alpha^T X(e_{ij})$. The relative importance weights α of different Gestalt principles can be computed by solving the following optimization problem:

$$\min_{\alpha} \left(\frac{1}{2}\|\alpha\|^2 + C \sum_{(v_+, v_-) \in P} \max\{0, 1 - \alpha^T [X(v_+) - X(v_-)]\}^2 \right)$$
$$\text{s.t. } \|\alpha\|_1 = 1. \tag{6}$$

In this equation, P includes all the pairs of positive and negative samples. $C > 0$ is a penalty coefficient, representing the tolerance of the error. The larger C is, the smaller the tolerance would be. C is experimentally set as 0.01 in this work. Based on the above equation, the relative weights of different Gestalt principles can be learned. The ranking score of each edge $F(e)$ is then computed based on the learned α using Eq. (3). We further define the weight of each edge $W(e)$ according to the following equation:

$$W(e) = \exp\left[\frac{F(e)}{\sigma}\right]. \tag{7}$$

Based on this equation, an edge with a large ranking score should have a large weight and the weight value is between 0 and 1. The scale parameter σ controls the strength of the ranking score. The procedure of our proposed graph construction method for salient region detection is summarized in Algorithm 1.

Algorithm 1 Multiple Gestalt principles-based graph construction for salient region detection

Input: An input image I and required parameters.

1. Define the superpixels of the input image as the nodes of the constructed graph.
2. Choose different edges inspired by three Gestalt principles according to Eqs. (1) and (2).
3. Obtain the relative weights of different Gestalt principles based on Eq. (6) using the PRSVM algorithm and then compute the ranking score of each edge according to Eq. (3).
4. Compute the edge weights based on the ranking score of each edge using Eq. (7).

Output: A constructed graph $G(\mathcal{V}, \mathcal{E}, \mathcal{W})$.

3 Graph-Based Optimization Framework

Based on the constructed graph, we further detect salient objects in images using the following graph-based optimization framework

$$S^* = \arg\min_S \{\psi(S) + \mu\phi(S)\},$$
$$\psi(S) = \frac{1}{2}\sum_{i,j=1}^{n} (\mathcal{W}_{ij}\|S_i - S_j\|^2),$$
$$\phi(S) = \sum_{i=1}^{n} (\|S_i - Q_i\|^2). \tag{8}$$

In the term $\psi(S)$, S_i and S_j are saliency values for the nodes v_i and v_j separately. This term indicates that if two connected nodes in the constructed graph have a larger edge weight, these two nodes should have similar saliency values. In the term $\phi(S)$, Q_i denotes the query value of the node v_i: if v_i is a query $Q_i = 1$, and otherwise $Q_i = 0$. This term indicates that the final saliency value should not be too far away from the initial query assignment. $\mu \in (0, \infty)$ is a parameter, which represents the relative balance of the terms $\psi(S)$ and $\phi(S)$. The solution of Eq. (8) can be computed by setting the derivative to be zero. The resulted saliency value can be written as

$$S^* = (\mathcal{D} - \mathcal{W} + \mu I)^{-1} Q. \tag{9}$$

In this equation, $\mathcal{W}$ can be obtained based on the constructed graph. $\mathcal{D}$ is a diagonal matrix with the (i, i) element equal to the sum of the i'th row of $\mathcal{W}$ and I is a unit matrix. Q records the query assignment of different nodes. In this paper, we first separately choose each of the four image borders as the background queries to get four maps based on Eq. (9). These four maps are normalized and multiplied to get an initial result. Then, the foreground queries are acquired by thresholding the initial result according to the mean. The final salient region detection result is then computed using

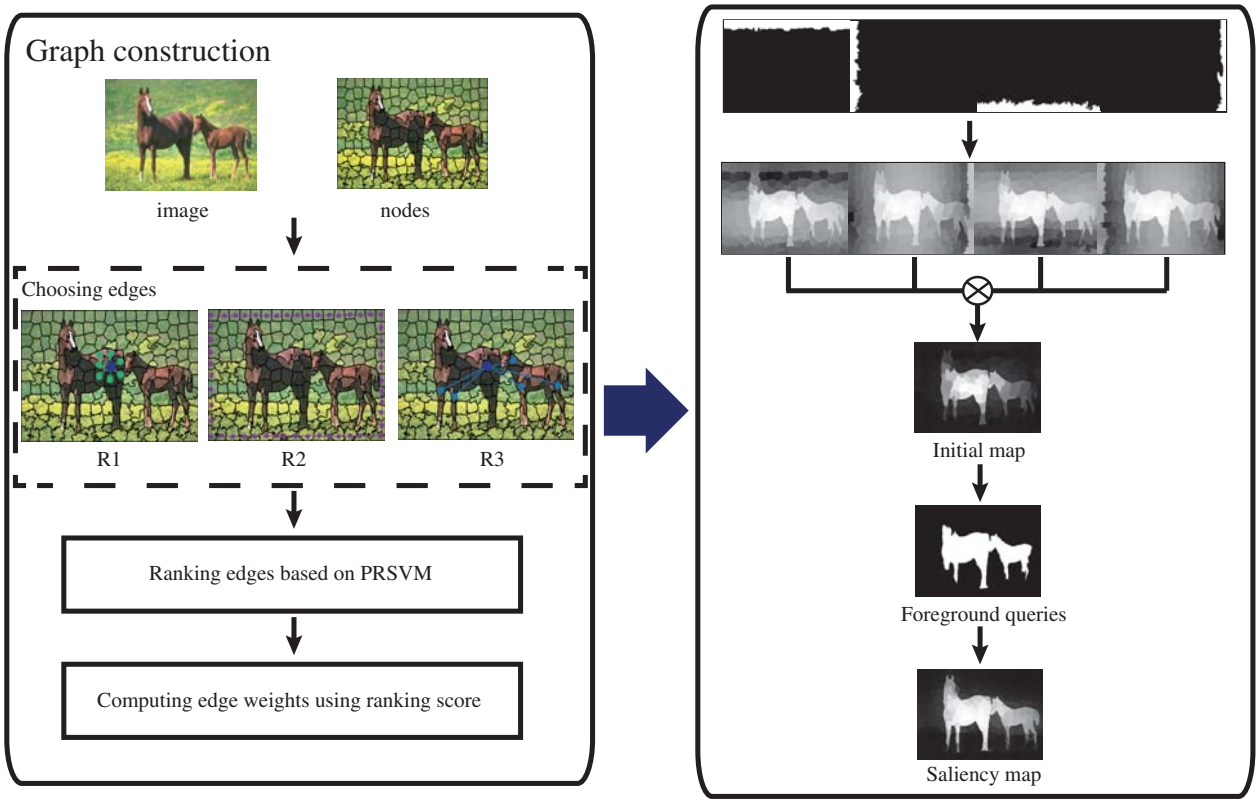

Fig. 1 Diagram of our method for salient region detection. The input image is first oversegmented into superpixels, which are the nodes of our constructed graph. Different edges are chosen inspired by three Gestalt principles according to Eqs. (1) and 2. Compute the ranking score of each edge based on PRSVM and the edge weights are then computed based on the ranking score of each edge. Each of the four image borders is set as the background queries separately to get four maps. An initial map is obtained by combining these maps and then further thresholded to get the foreground queries. The final saliency map is obtained based on the constructed graph and the foreground queries.

Eq. (9) again based on the foreground queries and the constructed graph. The complete procedure of our proposed salient region detection method can be seen in Fig. 1.

4 Experiment

4.1 *Experimental Setup*

4.1.1 *Databases*

We trained the relative importance weights $\alpha = [\alpha_1; \alpha_2; \alpha_3]$ of different features based on Gestalt principles on the extended complex scene saliency database (ECSSD).[33] This database contains 1000 natural images with complex patterns.

The experimental performance evaluation is conducted on four public salient region detection databases: THUS10K,[34] Pascal-S,[35] SOD,[36] and iCoseg.[37] The THUS10K database contains 10,000 natural images with pixel-level truth labelings and each image contains one salient object. The Pascal-S database contains 850 challenging natural images with salient object masks labeled by 12 subjects. We obtain binary ground truth by thresholding the masks at 0.5 as suggested in the work of Li et al.[35] The SOD database contains 300 challenging natural images from the Berkeley segmentation database.[38] The iCoseg database contains 643 natural images with one or multiple salient objects.

4.1.2 *Evaluation metrics*

We evaluate different approaches using the most commonly used evaluation metric for salient region detection: precision–recall curves. To get different precision–recall curves for different approaches, the saliency map is binarized at each threshold in the range [0:1:255] and the precision and recall values at each threshold are calculated by comparing the binary map with the truth.

In many situations, both high precision value and high recall value are required. Thus, we further report the area under the precision–recall curve (average precision, AP) as the measure of overall performance. The AP value for one approach is calculated as the mean of the areas under different precision–recall curves of different images on one database.

4.1.3 *Parameter setup*

We experimentally set parameters and use a fixed parameter configuration for all experiments on different databases in this work. There are four main parameters in the proposed method: two parameters (η_c and η_t) in Eq. (2), one parameter (σ) in Eq. (7), and one parameter (μ) in Eq. (9). We test different values for these parameters on the SOD database based on the metric of AP.

Table 1 Performance comparison with different η_c.

η_c	0.05	0.10	**0.15**	0.20	0.25
AP	0.706	0.727	**0.730**	0.724	0.715

Note: The optimal performance is highlighted in bold.

Table 2 Performance comparison with different η_t.

η_t	0.40	0.50	**0.60**	0.70	0.80
AP	0.719	0.724	**0.730**	0.728	0.721

Note: The optimal performance is highlighted in bold.

Table 3 Performance comparison with different σ.

σ	0.05	**0.10**	0.15	0.20
AP	0.723	**0.730**	0.725	0.717

Note: The optimal performance is highlighted in bold.

Table 4 Performance comparison with different μ.

μ	0.001	**0.01**	0.05	0.1
AP	0.665	**0.730**	0.725	0.719

Note: The optimal performance is highlighted in bold.

The parameter η_c is used to choose the neighbors in the color space. Larger η_c will make more nodes to be neighbors to each other in the color space. In this paper, we set η_c to be 0.15 according to Table 1. Similarly, the parameter η_t is used to choose the neighbors in the texture space and larger η_t will make more nodes to be neighbors to each other in the texture space. We experimentally set η_t to be 0.6 based on Table 2. Table 3 shows the effect when changing the value of the parameter σ. σ is a scale parameter to compute the edge weights and we set $\sigma = 0.1$ for all experiments. The parameter $\mu \in (0, \infty)$ controls the relative balance of different terms in Eq. (9). In this work, we set $\mu = 0.01$ based on Table 4.

4.2 *Performance Comparison*

In the experiments, we qualitatively and quantitatively compare our method with 12 state-of-the-art salient region detection approaches, including GB,[11] FT,[12] RC,[34] CB,[15] GR,[17] AM,[18] BD,[19] CL,[20] GP,[21] PM,[22] MST,[23] and GF.[25] Following previous works,[12,34] we chose different salient region detection approaches based on the following principles: recency (PM, MST, CL, GP, and GF) and variety. (FT is frequency tuned; RC employs regional contrast; CB is based on shape and context information; GB, GR, AM, BD, CL, GP, PM, and GF are graph based; and MST uses minimum spanning tree.)

4.2.1 *Quantitative comparison*

We first quantitatively compare our method with 12 state-of-the-art approaches based on the precision–recall curves shown in Fig. 2 and AP shown in Table 5.

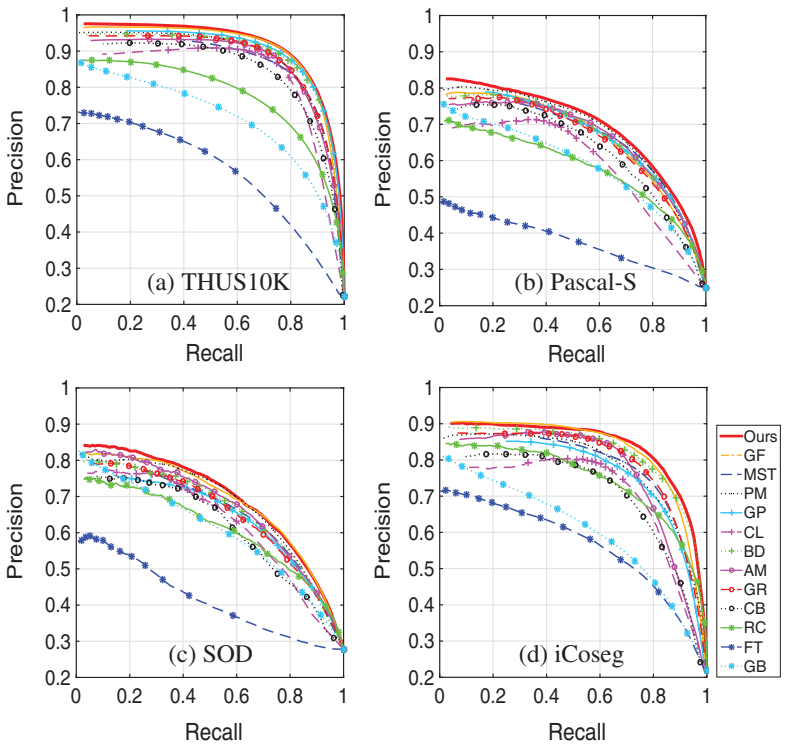

Fig. 2 The PR curves for different methods on THUS10K, Pascal-S, SOD, and iCoseg databases.

Table 5 Performance comparison based on AP.

AP	GB	FT	RC	CB	GR	AM	BD
MSRA10K	0.746	0.598	0.817	0.861	0.896	0.891	0.905
Pascal-S	0.631	0.405	0.631	0.678	0.706	0.717	0.713
SOD	0.644	0.433	0.655	0.637	0.691	0.705	0.699
iCoseg	0.653	0.611	0.787	0.738	0.820	0.802	0.849
AP	CL	GP	PM	MST	GF	Ours	—
MSRA10K	0.874	0.912	0.916	0.873	0.925	0.927	—
Pascal-S	0.656	0.713	0.731	0.688	0.719	0.742	—
SOD	0.672	0.684	0.722	0.678	0.737	0.730	—
iCoseg	0.755	0.796	0.829	0.801	0.857	0.861	—

THUS10K: Figure 2(a) and Table 5 show the quantitative performances of different approaches on THUS10K database with images containing one salient object. Recent graph-based methods, e.g., GF, PM, GP, CL, BD, AM, and GR show competitive detection performance. All these methods only consider one Gestalt principle to choose graph edges or compute edge weights. And no methods consider the confliction problem among multiple Gestalt principles. Our proposed method outperforms the above graph-based methods, which demonstrates the effectiveness of our approach to construct the graph based on multiple Gestalt principles which solves the confliction among these principles.

Pascal-S: Figure 2(b) and Table 5 show the PR curves and AP values of different approaches on the Pascal-S databases, which contain various challenging natural images with the ground truth labeled by 12 subjects. The performances of different methods as measured by PR curves and AP values are lower by a large margin in comparison to the THUS10K database. This phenomenon indicates that it is much harder to detect salient regions in challenging natural images. The proposed method outperforms other salient region detection methods on the Pascal-S database in terms of higher precision–recall values and max AP value, demonstrating that the proposed graph construction method helps to detect salient foreground objects in challenging natural images.

SOD: The SOD database is another difficult saliency detection database with challenging natural images. The quantitative comparison in Fig. 2(c) and Table 5 demonstrates that the proposed salient region detection method has comparable performance to other state-of-the-art methods on this database.

iCoseg: We also compare different methods on the iCoseg database, which contains natural images with one or multiple salient objects. Figure 2(d) and Table 5 show the PR curves and AP values of different approaches on the iCoseg database. It can be seen that the proposed method outperforms GB, GR, AM, BD, CL, GP, and PM which are also graph-based methods.

4.2.2 *Qualitative comparisons*

The qualitative comparison between our method and other representative approaches can be seen in Fig. 3, which shows different types of images to highlight the differences between these approaches. A natural scene image with a complex salient object is displayed in the first row. The complex building is uniformly highlighted by our method while most other approaches only highlight part of the building. The second row displays an image with a complex background pattern. The proposed method can well suppress the background region while a large amount of other approaches wrongly detect a part of the background. The last row displays an image with multiple salient objects. The proposed method can uniformly detect all of the salient objects and largely suppress the background region while a lot of other approaches only detect a portion of the salient objects or wrongly highlight part of the background. From these results, we can see that the proposed method outperforms other salient region detection approaches in generating saliency maps which are more consistent with the truth.

4.3 *Examination of Design Options*

The values of α learned on the ECSSD are shown in Table 6, which indicate the relative importance weights of different features measured according to different Gestalt principles. As we can see, all features contribute to the final weights and the Gestalt principle of color similarity has the largest contribution.

We further test the performances of different graphs, which use all or part of different Gestalt principles: spatial proximity, color similarity, and texture similarity on the SOD database to better understand the proposed method, shown in Fig. 4. The graphs based on only one Gestalt principle (spatial proximity, color similarity, or texture similarity) have the worst performances. After employing two Gestalt principles, the corresponding graphs outperform the graphs

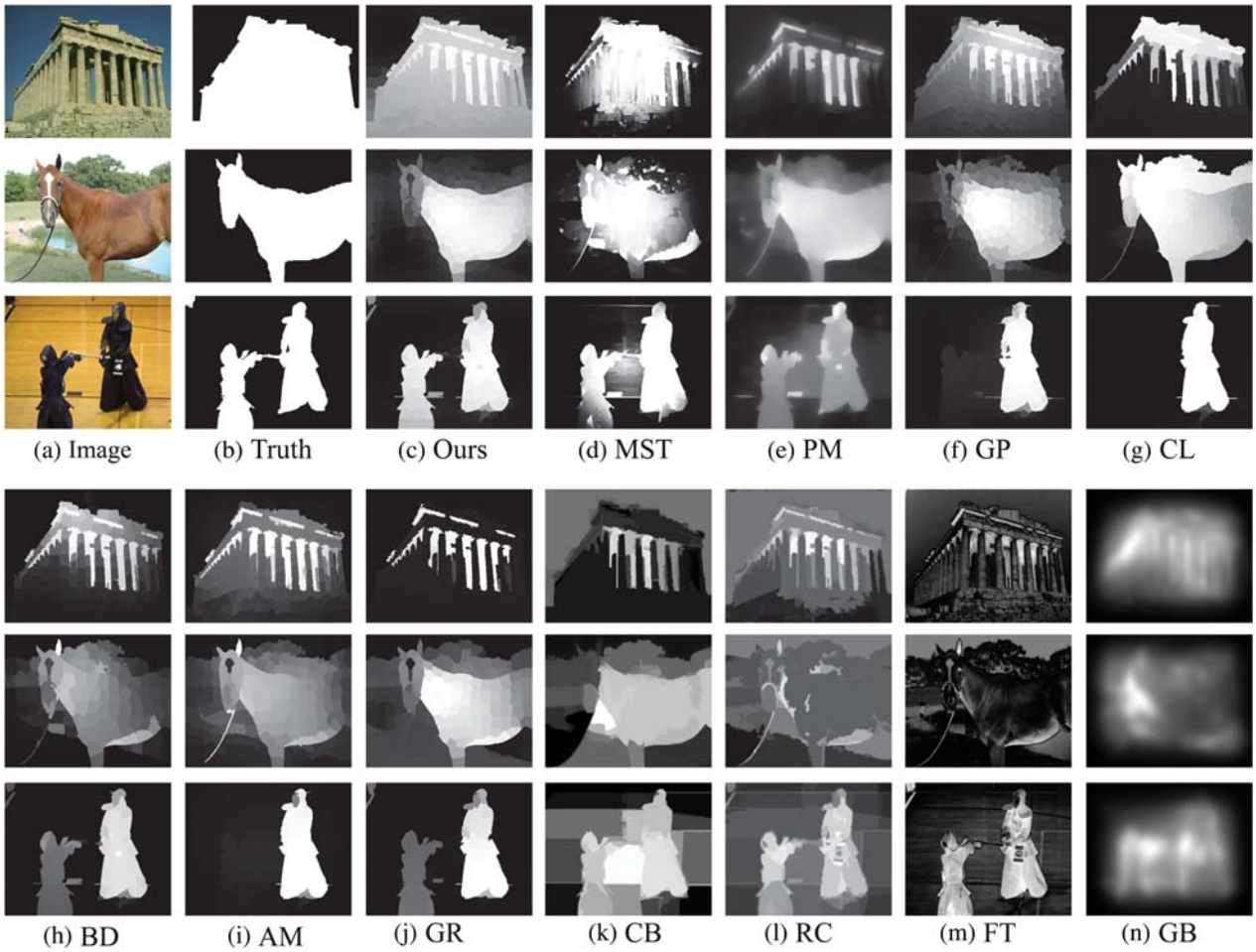

Fig. 3 Comparison of different salient region detection methods. The column (a) is the original image, the column (b) is the ground truth, and the remaining columns are results of the evaluated methods. Our method is the column (c) of the results.

Table 6 The learned values of α.

α_1	α_2	α_3
0.37	0.51	0.12

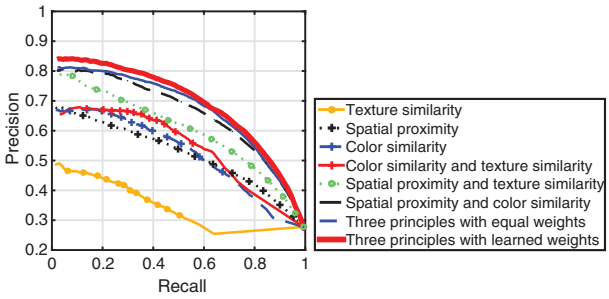

Fig. 4 The PR curves of different graphs on SOD database.

with only one principle. When all three Gestalt principles are employed with equal weights, the performance of the constructed graph further improves. And the graph using three Gestalt principles with learned weights has the best performance, demonstrating the effectiveness of the learning strategy of the proposed method.

5 Conclusion

In this work, we have proposed a graph construction method that makes better use of multiple Gestalt principles. To construct a graph, spatial proximity, color similarity, and texture similarity among image regions are employed to choose different edges. Furthermore, we solve the confliction among multiple Gestalt principles based on the PRSVM algorithm to compute the edge weights. Extensive experimental comparisons of various saliency databases demonstrate that the proposed method outperforms state-of-the-art approaches for salient region detection. In our future work, we will consider to exploit a nonlinear method to better solve the confliction problem among multiple Gestalt principles.

Acknowledgments

This work was supported by the National Natural Science Fund of China (Grant Numbers 61233011, 61374006, 61473086, 61463008, 61703100, 61773118, and 61773117); Major Program of National Natural Science Foundation of China (Grant Number 11190015); Natural Science Foundation of Jiangsu (Grant Numbers BK20131300 and BK20170692); 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. Q. Li, Y. Zhou, and J. Yang, "Saliency based image segmentation," in *Int. Conf. on Multimedia Technology*, pp. 5068–5071 (2011).
2. A. Li, X. She, and Q. Sun, "Color image quality assessment combining saliency and FSIM," *Proc. SPIE* **8878**, 88780I (2013).
3. C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.* **19**(1), 185–198 (2010).
4. S. Goferman, A. Tal, and L. Zelnik-Manor, "Puzzle-like collage," *Comput. Graphics Forum* **29**(2), 459–468 (2010).
5. R. Girshick et al., "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 580–587 (2014).
6. L. Zhang and Y. Zhang, "Airport detection and aircraft recognition based on two-layer saliency model in high spatial resolution remote-sensing images," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10**(4), 1511–1524 (2017).
7. Y. Wang and Q. Zhao, "Superpixel tracking via graph-based semi-supervised SVM and supervised saliency detection," in *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pp. 1–6, IEEE (2015).
8. L. Zhang, J. Chen, and B. Qiu, "Region-of-interest coding based on saliency detection and directional wavelet for remote sensing images," *IEEE Geosci. Remote Sens. Lett.* **14**(1), 23–27 (2017).
9. L. Zhang, J. Chen, and B. Qiu, "Region of interest extraction in remote sensing images by saliency analysis with the normal directional lifting wavelet transform," *Neurocomputing* **179**, 186–201 (2016).
10. G.-H. Liu, J.-Y. Yang, and Z. Li, "Content-based image retrieval using computational visual attention model," *Pattern Recognit.* **48**, 2554–2566 (2015).
11. J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in Neural Information Processing Systems*, pp. 545–552 (2006).
12. R. Achanta et al., "Frequency-tuned salient region detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1597–1604 (2009).
13. J. Zhang, J. Ding, and J. Yang, "Exploiting global rarity, local contrast and central bias for salient region learning," *Neurocomputing* **144**, 569–580 (2014).
14. M.-M. Cheng et al., "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(3), 569–582 (2015).
15. H. Jiang et al., "Automatic salient object segmentation based on context and shape prior," in *British Machine Vision Conf.*, Vol. **6**, pp. 1–7 (2011).
16. S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2790–2797 (2014).
17. C. Yang et al., "Saliency detection via graph-based manifold ranking," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3166–3173 (2013).
18. B. Jiang et al., "Saliency detection via absorbing Markov chain," in *IEEE Int. Conf. on Computer Vision*, pp. 1665–1672 (2013).
19. W. Zhu et al., "Saliency optimization from robust background detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2814–2821 (2014).
20. C. Gong et al., "Saliency propagation from simple to difficult," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2531–2539 (2015).
21. P. Jiang, N. Vasconcelos, and J. Peng, "Generic promotion of diffusion-based salient object detection," in *IEEE Int. Conf. on Computer Vision*, pp. 217–225 (2015).
22. Y. Kong et al., "Pattern mining saliency," *Lect. Notes Comput. Sci.* **9910**, 583–598 (2016).
23. W.-C. Tu et al., "Real-time salient object detection with a minimum spanning tree," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 2334–2342 (2016).
24. J.-G. Yu et al., "A computational model for object-based visual saliency: spreading attention along gestalt cues," *IEEE Trans. Multimedia* **18**(2), 273–286 (2016).
25. J. Zhang et al., "A novel graph-based optimization framework for salient object detection," *Pattern Recognit.* **64**, 39–50 (2017).
26. J. Wagemans et al., "A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization," *Psychol. Bull.* **138**(6), 1172–1217 (2012).
27. M. Wertheimer, "Laws of organization in perceptual forms," *Psychol. Forsch.* **4**, 301–350 (1923).
28. R. Achanta et al., "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2274–2282 (2012).
29. O. Chapelle and S. S. Keerthi, "Efficient algorithms for ranking with SVMS," *Inf. Retr.* **13**(3), 201–215 (2010).
30. B. Hill, T. Roger, and F. W. Vorhagen, "Comparative analysis of the quantization of color spaces on the basis of the CIELab color-difference formula," *ACM Trans. Graphics* **16**(2), 109–154 (1997).
31. T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision* **43**(1), 29–44 (2001).
32. J. Zhang et al., "A prior-based graph for salient object detection," in *IEEE Int. Conf. on Image Processing*, pp. 1175–1178 (2014).
33. Q. Yan et al., "Hierarchical saliency detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1155–1162 (2013).
34. M.-M. Cheng et al., "Global contrast based salient region detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 409–416 (2011).
35. Y. Li et al., "The secrets of salient object segmentation," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 280–287 (2014).
36. V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 49–56 (2010).
37. D. Batra et al., "iCoseg: interactive co-segmentation with intelligent scribble guidance," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3169–3176 (2010).
38. D. Martin et al., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *IEEE Int. Conf. on Computer Vision*, Vol. **2**, pp. 416–423 (2001).

Jinxia Zhang received her BS degree in the Department of Computer Science and Engineering, Nanjing University of Science and Technology (NUST), China, in 2009, and her PhD in the Department of Computer Science and Engineering, NUST, China, in 2015. She was a visiting scholar in the Visual Attention Lab at Brigham and Women's Hospital and Harvard Medical School from 2012 to 2014. She is currently a lecturer in the School of Automation, Southeast University. Her research interests include visual attention, visual saliency detection, computer vision, and machine learning.

Shixiong Fang received his BS degree in the School of Electrical and Electronic Engineering, Hubei University of Technology, China, in 2009, and his MS and PhD degrees in control theory and control engineering from Southeast University, China, in 2002 and 2009, respectively. He was a visiting scholar in European Organization for Nuclear Research from 2004 to 2006. He is currently a lecturer in the School of Automation, Southeast University. His research interests include computer vision and cyber physical systems.

Haifeng Zhao received his BE and PhD degrees in computer science and pattern recognition and intelligent systems at NUST (NJUST) in 2005 and 2012, respectively. He visited the Australian National University and NICTA from 2008 to 2010. He is currently a senior engineer at the School of Software Engineering, Jinling Institute of Technology, Nanjing, China. His main research interests include computer vision, pattern recognition, and human–computer interaction, especially on the scene understanding and robust reading. He is a member of IEEE, CCF, and JSAI.

Guang-Hai Liu is currently a professor in the College of Computer Science and Information Technology, Guangxi Normal University, in China. He received his PhD from the School of Computer Science and Technology, NUST. In 2011, he was engaged as an evaluation expert of science and technology project of Guangxi, China. His current research interests are in the areas of image processing, pattern recognition, and artificial intelligence.

Haikun Wei received his BS degree in the Department of Automation, North China University of Technology, China, in 1994, and his MS and

PhD degrees in the Research Institute of Automation, Southeast University, China, in 1997 and 2000, respectively. He was a visiting scholar at the RIKEN Brain Science Institute in Japan from 2005 to 2007. He is currently a professor in the School of Automation, Southeast University. His research interest is real and artificial in neural networks and industry automation.

Lihuan Chen received her BS degree in the School of Mechanical Engineering, Jiamusi University, China, in 1998, her MS degree in the Department of Automated Machines, St. Petersburg Polytechnic University, Russia, in 2003, and her PhD in the School of Mechanical Engineering, Southeast University, China, in 2014. She is currently a postdoctor in the School of Automation, Southeast University and an associate professor in the School of Intelligent Science and Control Engineering, Jinling Institute of Technology. Her research interests include intelligent control, MEMS control, computer vision, and machine learning.

Kanjian Zhang received his BS degree in mathematics from Nankai University, China, in 1994 and his MS and PhD degrees in control theory and control engineering from Southeast University, China, in 1997 and 2000. He is currently a professor in the School of Automation, Southeast University. His research is in nonlinear control theory and its applications, with particular interest in robust output feedback design and optimization control.