Gestalt-grouping based on path analysis for saliency detection

Lijuan Xu[1], Zhihang Ji[1,2], Laura Dempere-Marco[3], Fan Wang[1], Xiaopeng Hu[1,*]

[1] School of Computer Science and Technology, Dalian University of Technology, Dalian, 116024, China
[2] College of Information Engineering, Henan University of Science and Technology, Luoyang, 471003, China
[3] Department of Engineering, Faculty of Sciences and Technology, University of Vic-Central University of Catalonia, Vic (Barcelona), 08500, Spain
[*] Corresponding author e-mail address: xphu@dlut.edu.cn (X. Hu).

Abstract

Due to the arbitrary scales, uncertain distributions of objects and cluttered background in natural scenes, uniformly detecting salient regions remains a challenge. This paper first proposes a Gestalt-grouping connectedness method based on path analysis to reflect the topological relationship between image pixels. Inspired by the Gestalt principles of feature grouping, we apply a smoothest path-based distance metric to capture the similarity, local proximity and global continuity between image pixels. The distance is small if the image pixels belong to the same visual region and large otherwise. To identify salient regions in natural images, we then propose a path-based background saliency model that integrates both the topological connectedness and appearance dissimilarity. Experimental results demonstrate the advantage of applying the path-based background saliency model in uniformly highlighting salient regions in images with complex backgrounds.

Keywords: Gestalt-grouping, Smoothest path-based distance, Topological connectedness, Salient region detection

1. Introduction

Salient region detection, aiming to identify important or interesting locations in natural images, has attracted tremendous attention in the past decades. The detected salient regions are then preferentially allocated with computational resources for subsequent image analysis and processing. Saliency detection is broadly used in various fields including image classification [1], object recognition [2,3], image segmentation [4], adaptive compression [5] and content-aware image resizing [6], among others.

Existing psychological and biological studies have confirmed that local contrast (distinctiveness or rarity) is an influential factor in visual saliency [7]. Accordingly, many previous works have exploited the contrast between local neighbors for saliency detection [8,9]. However, recent studies demonstrate the role of global visual perception in the deployment of visual attention [10–12]. Without the awareness of the global structure, local based methods tend to assign high saliency to the edges or textures instead of uniformly highlighting salient objects [13,14]. Several attempts have been made to encapsulate such global information by exploiting the topological structure of an image for saliency detection [13–23]. In view of the assumption that the contrast to nearby pixels is much more significant than that to distant ones, feature contrast is inversely weighted by the spatial distances between pixels in the entire image [14,16–18]. Thus, the prominence of frequent features is alleviated by these global methods. However, the results demonstrate that these methods are sensitive to clutter. It is difficult to suppress small variances and highlight the whole salient regions uniformly. In [20], the enclosure topological relationship between figure and ground is introduced to model saliency. Although some successful results have been achieved with such surroundedness cue, locating salient regions remains difficult because there exist regions without closed outer contours in natural scenes. Jiang et al. formulate the saliency detection as a function of the time that it takes for the transient nodes to reach the absorbing nodes of an Absorbing Markov chain on an adjacent graph [22]. Yang et al. regard the saliency detection as a graph-based ranking problem by performing label propagation on a sparsely connected graph to characterize the overall differences between salient objects and background [23]. The topological information on the constructed graph represented in [22] and [23] is determined by the random walk theory. However, if there exist long-range smooth background regions near the center of the image, the random walker will be distracted, resulting in highlighted background regions. In [21], the saliency of image patches is defined as the shortest distance to the virtual background on the constructed graph. However, gradual changes and noise can generate non-uniform salient regions since small variances are accumulated along the shortest path. In [24,25], the minimum barrier distance transform is applied to estimate the boundary connectivity. Due to its non-continuity property and high-complexity, it is difficult to generate an accurate path to estimate the distance between graph nodes.

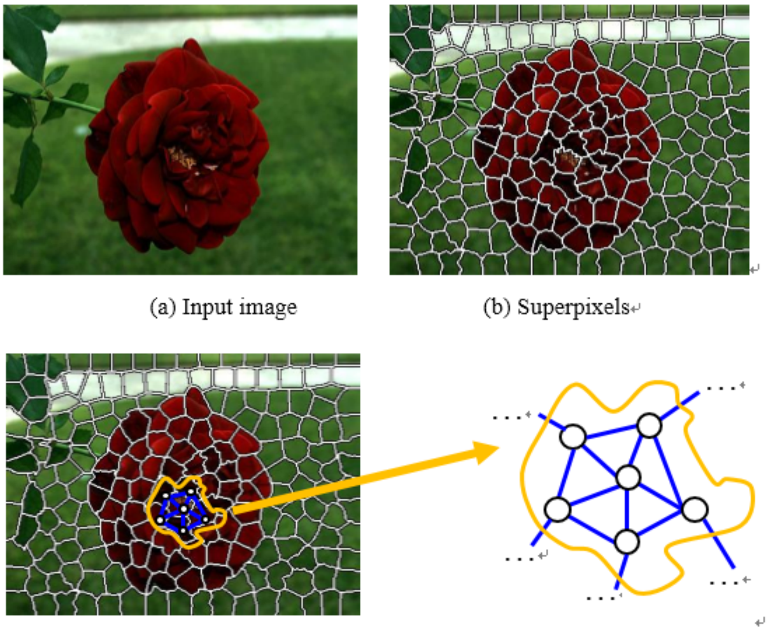
According to the Gestalt grouping principles in human perception, characterized by the laws of proximity, similarity and continuity [26, 27], the human visual system tends to perceive objects that are similar, close or connected without abrupt directional changes as a perceptual whole. More importantly, the proximity and continuity attributes have been considered as two basic structural components of perception representation in the deployment of visual attention [11,12]. Inspired by Gestalt psychology, we present a path-based distance metric, which integrates similarity, local proximity and global continuity information, to describe the relation between image elements. By generating the smoothest paths between each pair of nodes on the constructed undirected graph, the proposed method offers a way to incorporate both local and global information for visual representation. In our approach, the path distance is determined by a Laplacian analysis on the paths. The Multi-Dimensional Scaling (MDS) projection illustrates the promising results achieved by the path distance in uniformly clustering similar image elements and segmenting different ones regardless of arbitrary scales and uncertain distributions of the objects. We then apply the path distance to estimate the topological connectedness in the image and propose a path-based background method to model the saliency. The experimental results on state-of-the-art datasets demonstrate the high accuracy and robustness of the proposed method in detecting salient regions.

In addition, we consider two widely used saliency models including contrast-based and prior-based ones, to further test the performance of the proposed path distance in improving salient region detection. Contrast-based saliency methods usually use the pairwise Euclidean distance in a feature space to measure the contrast between image elements [14,16,18,28], and prior-based ones discover various priors (such as boundary prior [21–23,29,30], center prior [22,26], convex-hull prior [28], etc.) to highlight salient regions and weaken non-salient ones. Instead, we apply the proposed path distance in the two types of saliency models. The experimental results demonstrate the favor- able performance when the path distance is considered in contrast measurements and prior estimations.

The rest of the paper is organized as follows. Section 2 presents the basic components of our model: i.e. the generation of the smoothest path, and the formulation of the path distance. Section 3 describes its application in modeling saliency, and Section 4 shows the results obtained from the experiments conducted on several benchmark datasets. Finally, in Section 5, the main conclusions of our work are presented and discussed.

2. Gestalt-grouping based path distance

In this section, we propose a smoothest path to describe the proximity, similarity and continuity relationships between image elements as stated by the Gestalt principles of grouping. To this end, as illustrated in Fig. 1, we extract superpixels and use them as the basic image elements of our path-based approach to reduce the computational cost and preserve the boundaries of image objects.



(a) Input image

(b) Superpixels

(c) The graph

Fig. 1 (a) Example of an input image. The patches in (b) are superpixels. The construction of the weighted graph is shown in (c) where two adjacent superpixels are connected with a weight to specify their relationships.

2.1. Graph construction

As shown in Fig. 1, an undirected weighted graph G = (V, E) is constructed to represent the input image, where V is a set of nodes and E is a set of edges. In this paper, the nodes are visually homogeneous superpixels generated by the SLIC algorithm [31].

In the graph, each superpixel i is connected with its spatially adjacent neighbors to preserve their spatial relations (proximity) in the image. In addition, we further add edges between any two superpixels on the image boundary to increase the boundary connectivity of background regions with little effect on object regions [23,29]. Each edge is then assigned a weight ω_{ij} to describe the feature difference between two connected superpixels [22,27,32], which is defined as follows:

$$\omega_{ij} = \|\vec{x}_i - \vec{x}_j\|_2 \tag{1}$$

where $\vec{x}_i$ and $\vec{x}_j$ are the mean vectors of the color features (in the CIE- Lab color space $\{L, a, b\}$) as obtained from all the pixels belonging to each of the two superpixels i and j, respectively.

2.2. Path generation

By connecting locally adjacent superpixels in the image, the Gestalt grouping laws of proximity and local similarity are encoded by the edges on the constructed graph. However, these edges fail to reflect the law of continuity between nodes that are not directly connected. In this work, to explore such continuity hidden in the image, and inspired by the “minimal intra-cluster path” proposed by Fischer et al. [33], we generate the smoothest path for each pair of nodes on the weighted graph. In data clustering, the “minimal intra-cluster path” emphasizes the intra-cluster connectedness property, based on the observation that two objects which are assigned to the same cluster are either similar or there exists a mediating intra-cluster path without an edge with large cost. If two objects belong to the same cluster, the “minimal intra- cluster path” is defined as a path minimizing the largest edge cost among all paths connecting the pair of objects in a full graph, where every object in the cluster is connected with every other object. If two objects belong to different clusters, the path is not defined.

Although the “minimal intra-cluster path” in a complete graph follows the connectedness property within a given cluster [33], it is not suitable to describe the proximity and continuity between all pairs of image elements, as the local spatial information is not preserved in the complete graph. In this work, to circumvent this issue, we extend the “minimal intra-cluster path” on the weighted neighbor graph to generate the smoothest path by selecting a path without abrupt changes between each two nodes, i.e.

$$S_m(A,B) = \min_{p \subset P_{a,b}} \{ \max_{p[h],p[h+1] \in p} \{ w_{p[h],p[h+1]} \} \} \tag{2}$$

where $P_{a,b}$ denotes the set of all paths connecting nodes A and B on the graph, $p[h]$ denotes the hth node on the path p from A to B. The smoothest path p^* between A and B is thus defined as the path that minimizes the largest edge weight. Fig. 2 demonstrates an example of the path generation. It is worth noting that, due to a high computational cost, it is impossible to examine all paths that connect each pair of nodes by traversing the graph. However, it can be proved that the path between two nodes in a minimum spanning tree (MST) of a graph is one of the smoothest paths for that pair of nodes. In this work, we apply the Kruskal algorithm to generate the smoothest paths for all pairs of nodes in the constructed graph.

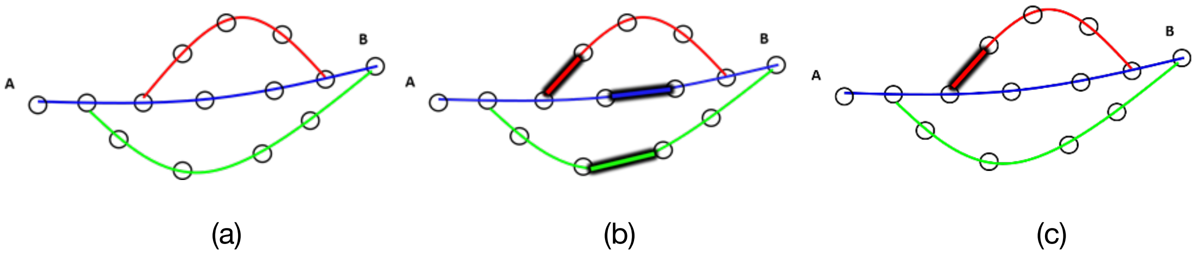


Fig. 2 Let us assume that there exist three possible paths between nodes A and B. The weights associated with each edge are indicated in (a). (a) Illustration of three possible paths between node A and B, signified by blue, red and green respectively. In (b), the edges that have the largest weight along their paths are highlighted by bold lines. In (c), the highlighted edge on the red path is selected because it has the minimal weight among those highlighted edges.

2.3. Laplacian analysis of the path

According to the definition of the smoothest path, the largest edge cost is selected to represent the relationship between two nodes. However, this representation is over-simplified and ignores some meaningful semantic image content usually indicated by sharp brightness changes, boundaries or discontinuities along the path. In this paper, we design a Laplacian method to identify significant changes of the path, which is subsequently included in the definition of the path-based distance.

Let us assume that the smoothest path p^* of A to B has $n+2$ nodes $\{A = N_0, N_1, N_2, \ldots, N_n, N_{n+1} = B\}$ and its corresponding feature vector sequence is $\{\overrightarrow{x}_0, \overrightarrow{x}_1, \overrightarrow{x}_2, \ldots, \overrightarrow{x}_n, \overrightarrow{x}_{n+1}\}$, $\overrightarrow{x}_i = \{f_{i,k}\}$, where $k = 1,2,3$ denotes their values in the CIE-Lab color space. For each node i, the discrete 1D implementation of the Laplace operator l becomes:

$$l(f_{i,k}) = f_{i+1,k} + f_{i-1,k} - 2f_{i,k} \tag{3}$$

We subsequently define $L(i)$ as,

$$L(i) = \sqrt{\sum_k l^2(f_{i,k})}, \text{ where } k = 1,2,3 \tag{4}$$

We set $f_{-1,k} = f_{0,k}$ and $f_{n+2,k} = f_{n+1,k}$. As each node on the path p^* links two edges, for each node $0 \leq i \leq n+1$, its corresponding edge (w') is defined as the edge with a larger weight, i.e.

$$\omega'(i) = \max(\|\overrightarrow{x}_i - \overrightarrow{x}_{i-1}\|_2, \|\overrightarrow{x}_i - \overrightarrow{x}_{i+1}\|_2) \tag{5}$$

First, we select m nodes with the largest L on the path (note that the complete path includes $n+1$ nodes). These selected nodes form a sequence

$$\{N_{m_1}, N_{m_2}, N_{m_3}, \ldots, N_{m_m}\} \tag{6}$$

along the path and, as a result, the path p^* is partitioned into $m+1$ sub-paths

$$\{\{N_0, \ldots, N_{m_1}\}, \{N_{m_1}, \ldots, N_{m_2}\}, \ldots, \{N_{m_m}, \ldots, N_{n+1}\}\} \tag{7}$$

The path-based distance is then defined as (see Fig. 3)

$$D(A, B) = \alpha \sum_{n=1}^{m} w'(m_n) + \sum_{\zeta=1}^{m+1} D'(\zeta) \tag{8}$$

In Eq. (8), the first term is the sum of the weights associated with the corresponding edges (see Eq. (5)) for the m selected nodes. It, thus, reflects sharp and obvious changes between image elements. Since L is evaluated for each node, and along a path the largest values are found in consecutive nodes (i.e. in pairs, which share the same corresponding edge), we set $\alpha = 0.5$ in this paper. The second term is defined to account for those areas showing smooth changes. These areas are represented by the $m+1$ sub-paths defined in Eq. (7). Algorithm I shows the pseudocode of the procedure used to calculate distance D' on such sub-paths.

In the experiments, the threshold θ is set to 0.95. By applying (8), we are able to obtain a refined path-based distance matrix $\mathbf{D}$ for all pairs of nodes in the graph.

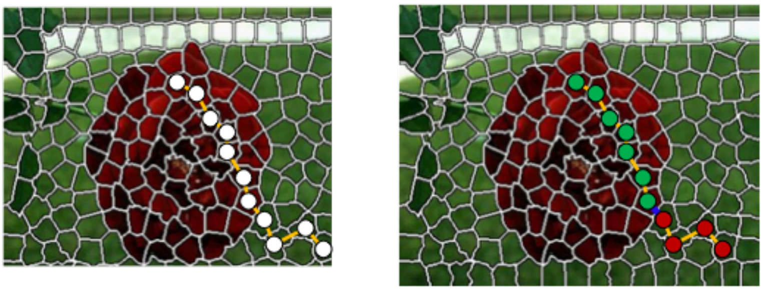


Fig. 3 An illustration of the Laplacian analysis on a path. (a) shows one path passing through two different image regions, and each region exhibits varying illumination. In (b), the blue edge partitions the path into two sub-paths, and is chosen as the corr-edge to reflect the sharp change between image regions. As there exist small differences along the sub-path in the same region (taken the sub-path with green nodes for example), we accumulate some important variations Extra-Dis to represent the pattern of slowly changing.

Algorithm I. Algorithm for calculating D' on the sub-path

Input: Each Sub-path ζ out of the $m+1$ defined in Eq. 7, $g_{i,k} = f_{i,k} - f_{i-1,k}$ for each node N_i of the sub-path ζ ($k = 1,2,3$), number of nodes n_ζ in the sub-path ζ , and the threshold θ .

```
for each sub-path $\zeta$ $(1 \le \zeta \le m+1)$,
    if $n_\zeta > 2$ , then
        for each $N_i \in \zeta$ , and $k = 1,2,3$
            if $g_{i,k} \ge 0 \rightarrow n^{+}(k) = n^{+}(k)+1$;
            if $g_{i,k} \le 0 \rightarrow n^{-}(k) = n^{-}(k)+1$;
        end;
    end;
    if $\left(\frac{n^{+}(k)}{n_\zeta}\right) \ge \theta \,||\, \left(\frac{n^{-}(k)}{n_\zeta}\right) \ge \theta$ , then
        for each $N_i \in \zeta$ $(1 \le N_i \le n_\zeta)$
            $D'(\zeta) = \max(D_{ij})$ , where $D_{ij}$ is the Euclidean distance in the Lab space between the
            feature vectors $\vec{x}_i$ and $\vec{x}_j$ associated with every pair of nodes $N_i$ and $N_j$ in path $\zeta$ .
        end;
    end;
end;
```

Output: $D'(\zeta)$.

2.4. Dimensionality reduction and visualization

The path-based distance matrix $\mathbf{D}$ obtained by using (8) is obtained in a high dimensional feature space. To reduce the dimensionality for visualization purposes, we apply the MDS method to construct a low dimensional embedding of the given vectors while preserving the original vector distances. Thus, all the image superpixels are projected onto a three-dimensional feature space (x, y, z) and normalized into [0,255]. in this paper. The normalized coordinates(x, y, z) are then used as the RGB values (i.e., R = x, G = y, and B = z) to fill each superpixel in the image. As a result, the pixels depicted by similar colors in the image describe close relations in the feature space.

Fig. 4(b) shows the visualization of the Gestalt-grouping based path distance on three natural images. Compared with the other three distances (Euclidean distance, Geodesic distance, and Minimum Barrier distance), it is evident that the proposed path-based distance metric faithfully follows the similarity, local proximity, and global continuity principles of Gestalt grouping, demonstrating favorable performance in semantically clustering and segmenting image elements regardless of illumination changes, image blurring and textures. The path-based distance between two nodes is small if they belong to the same region and large otherwise. Let us take the image with two swans as an example. By applying the proposed path-based distance, the grass and road regions are depicted by the same color (respectively) to reflect the law of continuity. The two swans, however, appear with different colors to signify that they depict two different objects. However, the Euclidean distance is not able to distinguish the two swans as independent objects since they share similar visual cues. In addition, the results generated by applying the Geodesic distance and the Minimum Barrier distance are much noisier when compared to that obtained by using our proposed path-based distance since it emphasizes the distinct value of the boundaries between image regions but does not overly accumulate noise along the path within each distinct object.

All in all, as illustrated in Fig. 5, the path-based distance presented in this work can be used with different purposes. On one hand, it can be used for visualization purposes to show the segmentation of images into different objects following the application of MDS to reduce the dimensionality of the data and the projection of such data onto a 3D color space. On the other hand, it can also be used for saliency evaluation. This is the main purpose of the path-based distance proposed in this paper. Moreover, such distance can be used to define new saliency methods – as is the case of the proposed Path-Based Background Saliency (PBS) method – but can also be incorporated into existing approaches such as contrast-based saliency detection methods or boundary priors based saliency detection methods to enhance their results. In this work, we explore all of these possible applications with special emphasis on the proposed PBS method described below.

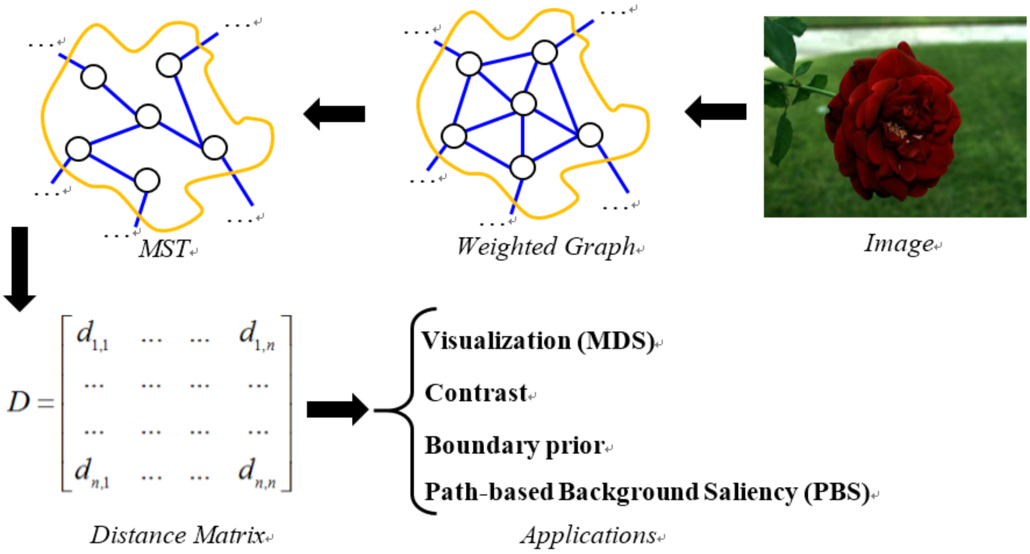

MST
Weighted Graph
Image
$D = \begin{bmatrix} d_{1,1} & \cdots & \cdots & d_{1,n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ d_{n,1} & \cdots & \cdots & d_{n,n} \end{bmatrix}$
Visualization (MDS)
Contrast
Boundary prior
Path-based Background Saliency (PBS)
Distance Matrix
Applications

Fig. 5 The framework of the Gestalt-grouping based path distance and its applications

3. Path-based background saliency (PBS)

In this section, the path-based distance is used to the estimate background saliency. To this end, two visual cues (connectedness and appearance) are defined and subsequently integrated to obtain such saliency.

3.1. Measuring the gestalt-grouping connectedness

In [21], Wei et al. reflect on the fact that most background regions can often be easily connected to image boundaries. This suggests an alternative view on the saliency of an image patch as the length of its shortest path to the image boundaries. Following this view, we define the saliency of an image patch by its connectedness to the background regions. When compared with the geodesic distance used in [21], the proposed path-based distance holds promise in representing the connectedness implicit in the image (shown in Fig. 4). The connectedness of a superpixel $\overrightarrow{x}_i$ to the background can be computed as follows,

$$C_G^t(i) = \min_{b^t \in B^t} (D(\overrightarrow{x}_i, b^t)) \tag{9}$$

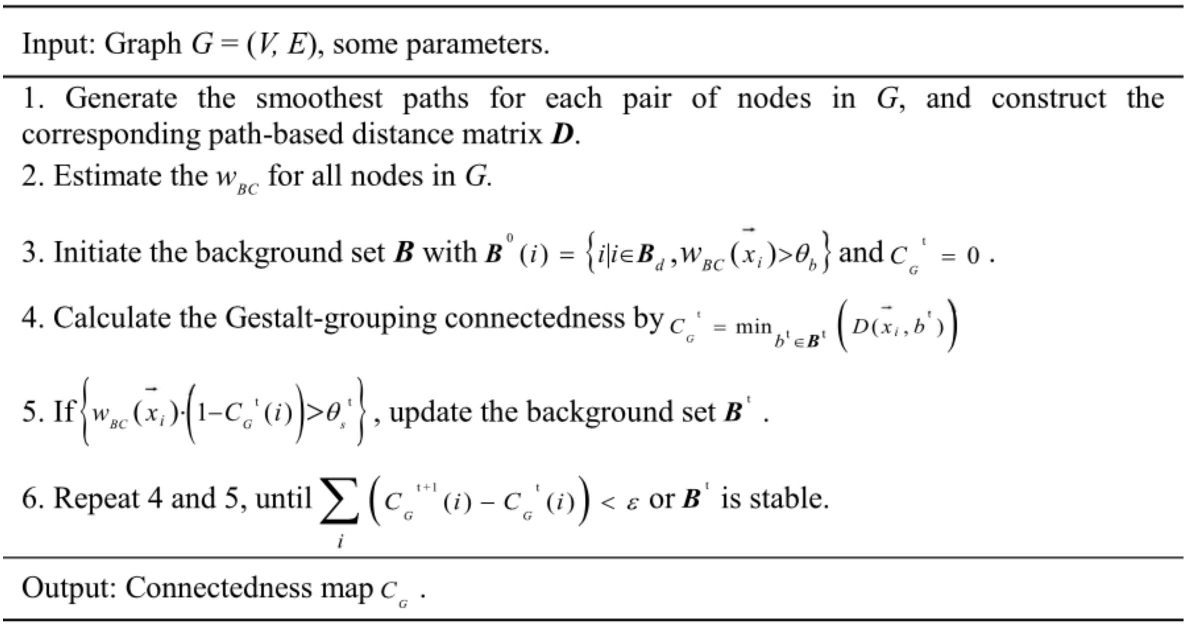
where $\mathbf{B}^t$ is the set of background superpixels, and C_G is subsequently normalized to lie in the interval $C_G \in [0,1]$. In contrast to existing methods, the superpixels on the image boundary are used as background seeds [22,23], and an iterative background growth algorithm is designed to generate a background set $\mathbf{B}$ until all the background pixels are included. To initiate the set $\mathbf{B}$ with robust background seeds, we estimate a boundary connectivity $w_{BC}(\overrightarrow{x}_i)$ for every superpixel i by applying the method described in [29]. However, instead of using the geodesic distance, we apply the proposed path-based distance to estimate the connectivity of all the superpixels to the image boundaries. Thus, we have,

$$\mathbf{B}^0(i) = \{i \,|\, i \in \mathbf{B}_d, w_{BC}(\overrightarrow{x}_i) > \theta_b\} \tag{10}$$

$$\mathbf{B}^t(i) = \{i \,|\, w_{BC}(\overrightarrow{x}_i) \cdot (1 - C_G^t(i)) > \theta_s^t\} \tag{11}$$

where $\mathbf{B}_d$ is the set of superpixels on the image boundary, θ_b is the mean value of $w_{BC}(\overrightarrow{x}_i)$ for all superpixels on the image boundary and θ_s^t is the mean of $w_{BC}(\overrightarrow{x}_i) \cdot (1 - C_G^t(i))$ for all the superpixels in rest of the image. The iterative process described by Eqs. (9-11) terminates when $\sum_i (C_G^{t+1}(i) - C_G^t(i)) < \varepsilon$ or $\mathbf{B}^t$ is stable. Algorithm II shows the pseudo-algorithm for calculating the Gestalt-grouping connectedness.

Algorithm II. Algorithm for getting the Gestalt-grouping connectedness based saliency map



```
Input: Graph $G = (V, E)$, some parameters.
1. Generate the smoothest paths for each pair of nodes in $G$, and construct the corresponding path-based distance matrix $\boldsymbol{D}$.
2. Estimate the $w_{BC}$ for all nodes in $G$.
3. Initiate the background set $\boldsymbol{B}$ with $\boldsymbol{B}^0(i) = \left\{ i | i \in \boldsymbol{B}_d, w_{BC}(\vec{x}_i) > \theta_b \right\}$ and $C_G^t = 0$.
4. Calculate the Gestalt-grouping connectedness by $C_G^t = \min_{b^t \in \boldsymbol{B}^t} \left( D(\vec{x}_i, b^t) \right)$
5. If $\left\{ w_{BC}(\vec{x}_i) \cdot \left( 1 - C_G^t(i) \right) > \theta_s^t \right\}$, update the background set $\boldsymbol{B}^t$.
6. Repeat 4 and 5, until $\sum_i \left( C_G^{t+1}(i) - C_G^t(i) \right) < \varepsilon$ or $\boldsymbol{B}^t$ is stable.
Output: Connectedness map $C_G$.
```

3.2. Measuring the appearance cue

The connectedness cue alone sometimes cannot produce satisfying results when the images have either a cluttered and scattered back- ground, or objects heavily touching the image boundary [24,25,29]. This paper also uses the appearance cue to highlight regions with high contrast against the image background regions.

Four image boundary regions, i.e. left-up, left-down, right-up, and right-down, are taken into consideration for computing the appearance cue. For each boundary region $c \in \{1,2,3,4\}$, we calculate its mean color in Lab space, i.e. $\bar{x}_c = [\bar{x}_L, \bar{x}_a, \bar{x}_b]$ and the color covariance matrix $Q_c = [q_{ij}]_{3\times 3}$. The appearance contrast a^i of superpixel i to the boundary region c is then computed by applying the Mahalanobis distance to its mean color

$$a_c^i = \sqrt{(x^i - \bar{x}_c) Q_c^{-1} (x^i - \bar{x}_c)^T}, c \in \{1,2,3,4\} \tag{12}$$

The results are subsequently normalized as follows:

$$a_c^i \leftarrow \frac{a_c^i}{\max_i a_c^i} \tag{13}$$

Finally, the appearance cue of superpixel i is defined as the weighted sum of a_c^i,

$$A(i) = \frac{\sum_{c=1}^{4} n_c a_c^i}{\sum_{c=1}^{4} n_c} \tag{14}$$

where n_c is the number of image pixels in boundary region c.

3.3. Integrating the connectedness and appearance cue

The Connectedness map C and the appearance map A are added pixelwise to form a saliency map M (i.e.$M = C + A$). This map M, which is also normalized, exploits the topology information from the Gestalt-grouping based path distance and the global appearance contrast from the color dissimilarity. In addition, a series of post-processing operations are subsequently applied to enhance the performance of the saliency map M based on the methods presented in [24,25]. Firstly, the center bias that accounts for the photographers' tendency to locate objects at or near the center of the image is considered [18,28]. The center bias is modeled by means of a Gaussian fall-off function as shown in (15).

$$C(i) = \exp\Big(-\frac{\|p_i(x) - \frac{W}{2}\|^2}{2\sigma_x^2} - \frac{\|p_i(y) - \frac{H}{2}\|^2}{2\sigma_y^2}\Big) \tag{15}$$

where W and H are the width and height of the image, and $p_i(x)$ and $p_i(y)$ are the mean values of the horizontal and vertical coordinates of superpixel i. The parameters σ_x, σ_y are set to $W/3$, $H/3$ respectively in this paper. We then pixel-wisely multiply M with the center prior map C.

Secondly, we use morphological smoothing opearations including reconstruction-by-dilation and reconstruction-by-erosion to smooth M while keeping the details of significant edges. The size of the square mask of the morphological operations is set to $\alpha\sqrt{\text{mean}(M)}$ as stated in [25], so that the operations are adaptive to the size of the salient regions.

Finally, we perform a nonlinear relaxation labeling operation [34] to increase the contrast between foreground and background and to improve the spatial consistency and structural coherence of the saliency map. Therefore, we achieve the PBS map by applying all above factors on M.

4. Experiments

4.1. Datasets

In order to evaluate the proposed PBS method, we apply the path- based distance metric for saliency detection in the following benchmark datasets: ASD, MSRA [35], SED2 [36], DUT_OMRON [23], ECCSD [18] and SOD.

ASD is a subset of MSRA and contains 1000 images with accurate human labeled segmentation masks provided by [30]. The dataset is chosen for its wide application in testing almost all saliency models. Nevertheless, it has several limitations. Most images in the dataset contain only one single salient object that is large and near the image center, and the background is usually simple and clean. In addition, the contrast between objects and background is strong. The other five datasets are more challenging. MSRA contains 5000 images with more complex background and lower contrast objects than ASD. The pixel-wise labeling images are obtained from [14]. SED2 has 100 natural images with exactly two salient objects in each image. DUT_OMRON contains 5168 images with one or more salient objects and exhibits relatively complex backgrounds. ECCSD contains 1000 semantically meaningful and structurally complex natural images acquired from the BSD dataset, PASCAL VOC and the Internet. Finally, SOD consists of 300 images with multiple objects of arbitrary scales and locations, and challenging backgrounds.

4.2. Evaluation metrics

The precision–recall curve (PR), area under the ROC curve (AUC), the mean absolute error (MAE) and the weighted F measure (F_β^w) [37] are used to evaluate the average performance of the proposed method. In a PR curve, the precision rate corresponds to the ratio of salient pixels which are correctly detected in the saliency map, while recall rate is the percentage of all detected salient pixels belonging to salient objects in ground truth. To generate a PR curve, we first normalize the saliency map into [0, 255], and produce a series of binary masks by segmenting the saliency map with a threshold changing from 0 to 255. The PR curve is obtained by comparing the binary masks with the ground truth. The curves obtained from all images on each dataset are then averaged to generate an overall PR curve. For the AUC, it is generated by calculating the area under the ROC curve.

For a more comprehensive comparison, the mean absolute error (MAE), supplied as a complement to the PR curve, is calculated to mea- sure how close a saliency map (S) is to the ground truth (G). The MAE criterion directly estimates the average per-pixel difference between the binary ground truth and the saliency map, which is more meaningful for applications such as object segmentation or image cropping.

$$MAE = \frac{1}{W \cdot H} \sum_{x=1}^{W} \sum_{y=1}^{H} |S(x,y) - G(x,y)| \tag{16}$$

where W and H are the width and height of the saliency map S, respectively, and G is the binary ground truth.

In addition, F_{β}^{w} is adopted to reliably evaluate the quality of a saliency map [38]. We use the code and default settings provided by the authors of [38], and set $\beta^2 = 1$.

$$F_{\beta}^{w} = (1+\beta^2)\frac{P^w \cdot R^w}{\beta^2 \cdot P^w + R^w} \tag{17}$$

where P stands for precision and R for recall.

4.3. Experimental results

In order to evaluate the performance of both, the path-based metric and the proposed PBS method, we have selected a battery of tests, which are discussed next. In first place, the path-based distance is incorporated into two existing saliency methods (see 4.3.1 and 4.3.2) and its performance is evaluated in comparison with the original methods on the ASD dataset. Then, we assess what is the impact of changing the distance metric on the proposed PBS method. To this end, the ASD is also considered. These evaluations allow us to fully assess the potential of the proposed path-based metric. Finally, the performance of the PBS method is assessed and compared to state-of-the-art methods on the complete set of benchmark datasets described in Section 4.1.

4.3.1. Contrast-based saliency model on ASD dataset

To evaluate the capability of the proposed path-based distance in estimating the contrast between different image elements, we consider a contrast-based saliency method. In the original works, image contrast is measured by pairwise Euclidean distance (D) in the feature space [14, 16,18,28], and the saliency of each pixel is defined as:

$$S(i) = \sum_{i \neq j} D(\overrightarrow{x}_i, \overrightarrow{x}_j) \cdot \exp\Big(-\frac{\|p_i - p_j\|^2}{2\sigma_p^2}\Big) \tag{18}$$

For each superpixel i, $\overrightarrow{x}_i$ is its color mean in the Lab color space, p_i is the center coordinate normalized to [0, 1] in the spatial domain, and σ_p controls the strength of spatial weighting. In this paper, we set $\sigma_p = 0.2$

It is worth noting that, as shown in Fig. 4(c), the pairwise Euclidean distance metric can be noise sensitive or semantically irrelevant. The Euclidean distance uses the similarity principle to measure the difference between image elements but, however, neglects the local proximity and global continuity hidden in the image. To overcome these problems, the proposed path-based distance can be used to define the contrast between image superpixels. The saliency of a superpixel is hence estimated by the sum of its weighted path-based distance (Eq. (18)) to all other superpixels in the image.

In the experiments, we set m in (Eq. (18)) to 4 and 6, respectively, to define the path-based distance. The obtained PR curves and AUC values are shown in Fig. 6. As can be seen, the results indicate that our path-based distance method has a better performance than methods based on the traditional Euclidean distance.

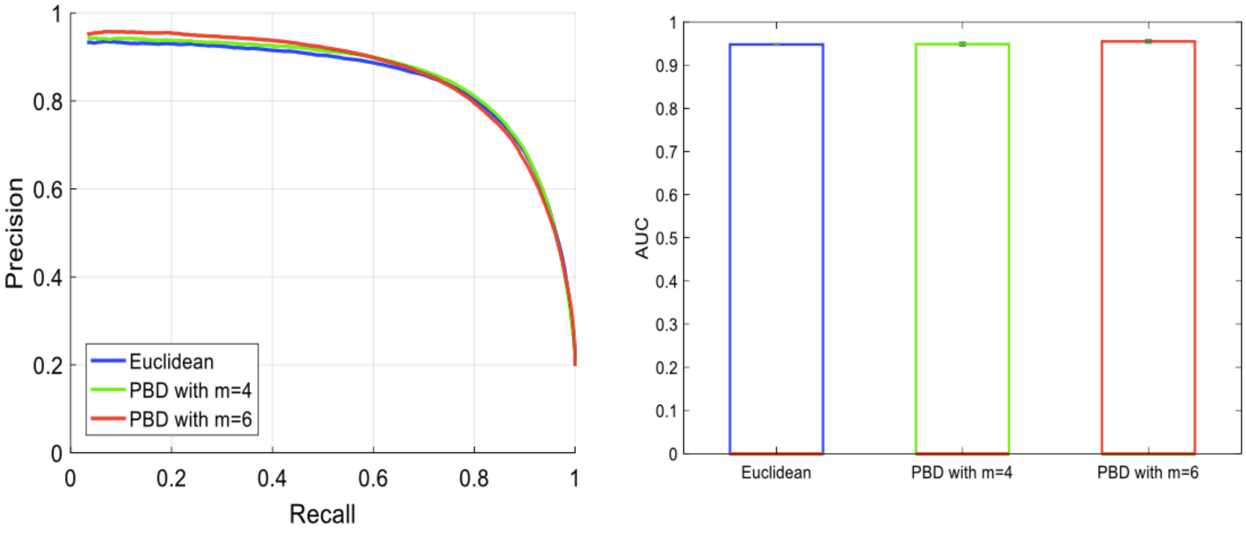


Fig. 6. Comparisons of PR curves and AUC values by applying Euclidean distance, path-based distance with m = 4, and m = 6, respectively, on the ASD dataset.

4.3.2. Prior-based saliency model on ASD dataset

Prior-based saliency models usually exploit various priors to enhance the performance of saliency detection. Some early studies use, for instance, the center prior to highlight the image center region based on the observation that the images taken by a photographer often place salient objects near the image center [18,20]. Similarly, the convex prior applies the convex hull of relevant points to estimate the location of the salient objects [28]. In [29], a robust boundary connectivity prior, which regards an image patch as background when it is heavily connected to the image boundary, is proposed to suppress the background. However, the calculation of this boundary connectivity depends on the performance of a prior image segmentation, which is a challenging task for images with cluttered background in natural scenes. In [29], the geodesic distance defined on the shortest path is applied to approximate this prior. However, it is not an optimal solution since small irrelevant weights are accumulated along the path, thus leading to the small-weight-accumulation problem [21], as shown in Fig. 4(d).

A general model for estimating the prior-based saliency can be derived from (Eq. (19)) by applying different priors as weights $w(i)$ for each superpixel i. In this experiment, we apply the center prior [18], the convex prior [28], the geodesic distance and the proposed path-based metric (based on the boundary connectivity prior presented in [29]) to calculate $w(i)$. Fig. 7 shows the experimental results obtained. The results demonstrate the superiority of our proposed method in estimating the boundary prior for salient region detection.

$$S(i) = w(i) \cdot \sum_{i \neq j} \| \overrightarrow{x}_i - \overrightarrow{x}_j \| \cdot \exp\left(- \frac{\|p_i - p_j\|^2}{2\sigma_p^2} \right) \tag{19}$$

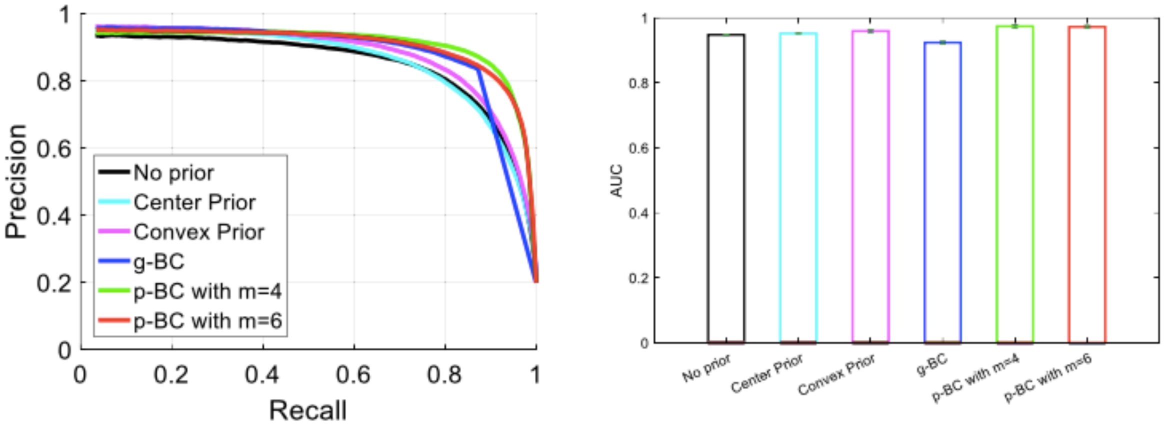


Fig. 7. PR curves and AUC values obtained by applying no prior with $w(i)$ = 1, center prior, convex prior, g-BC (boundary connectivity prior based on geodesic distance) prior and p-BC (boundary connectivity prior based on the path-based distance with m = 4 and 6 respectively) on the ASD dataset.

4.3.3. Comparisons of different distance metric with PBS model on ASD dataset

In this section, the proposed path-based distance metric is compared with the Geodesic distance [21], and the Minimum Barrier distance [24, 25] to estimate the boundary connectedness. To highlight the effects of the distance metrics on saliency estimation, we calculate the proposed path-based, the Geodesic and the Minimum Barrier distance, respectively, on MST, and then apply them to define the saliency in the proposed PBS framework. The comparison of the three-distance metrics in terms of five evaluation metrics on the ASD dataset are shown in Fig. 8. The experimental results indicate that the proposed path-based distance achieves the best performance in terms of the four evaluation metrics.

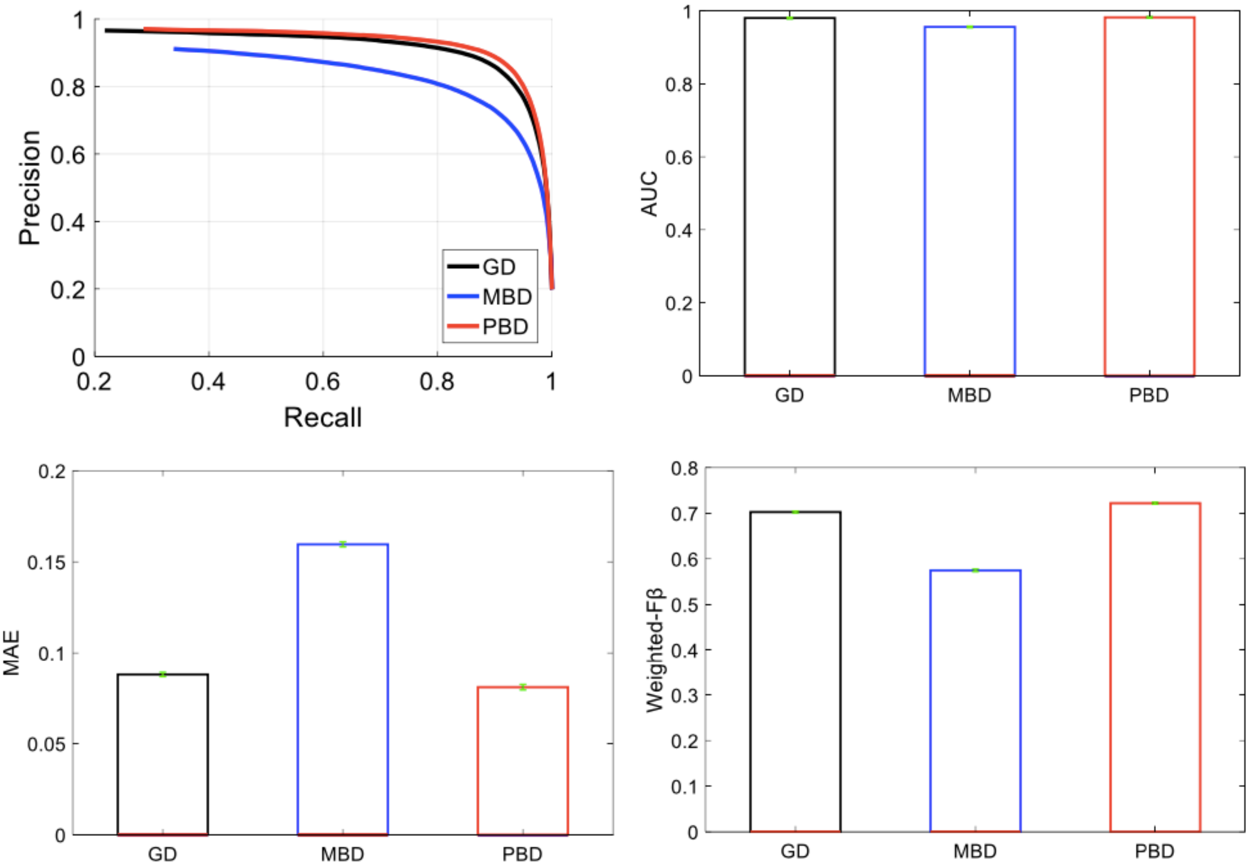


Fig. 8. PR curves, AUC, MAE values and F_β^w scores obtained by applying the GD, MBD and PBD in the PBS framework respectively on ASD dataset.

4.3.4. Comparison of the PBS model with state-of-the-art methods

In this section, the proposed path-based background saliency model is compared with other twenty methods on the six datasets described in Section 4.1 in order to evaluate the performance comprehensively. These methods include the most recent state-of-the-art approaches (GS_SP [21], HS [18], PCA [39], GMR [23], RBD [29], MC [22], DSR [40], SF [17], fastMBD [24] and mstMBD [25]) and, additionally, some other traditional ones (IT [8], LC [41], FT [35], MZ [42], SR [43], AC [44], GB [19], CA [9], SUN [45] and RC [14]) to reflect the diversity of the saliency models. For example, HS, RBD, SF, FT, CA and RC are contrast-based approaches, GS_SP, GMR, MC and GB are graph-based ones, PCA and DSR are transformation methods in linear or nonlinear domains, and SR is a spectrum-based method. In the experiments, we use the code from Achanta [35] for IT, FT and MZ, the code from Cheng [14] for RC, LC, SR, AC and GB, the saliency maps provided by the authors for GS_SP and SF on the ASD dataset, the code from Zhu [29] for

GS_SP and SF on the other five datasets, and the authors' own code for HS, PCA, GMR, RBD, MC, DSR, CA and SUN. Fig. 9 shows the comparisons of all methods in terms of four evaluation metrics on the ASD dataset. The experimental results indicate that the proposed PBS method outperforms all other methods in PR, AUC, MAE and F_β^w tests.

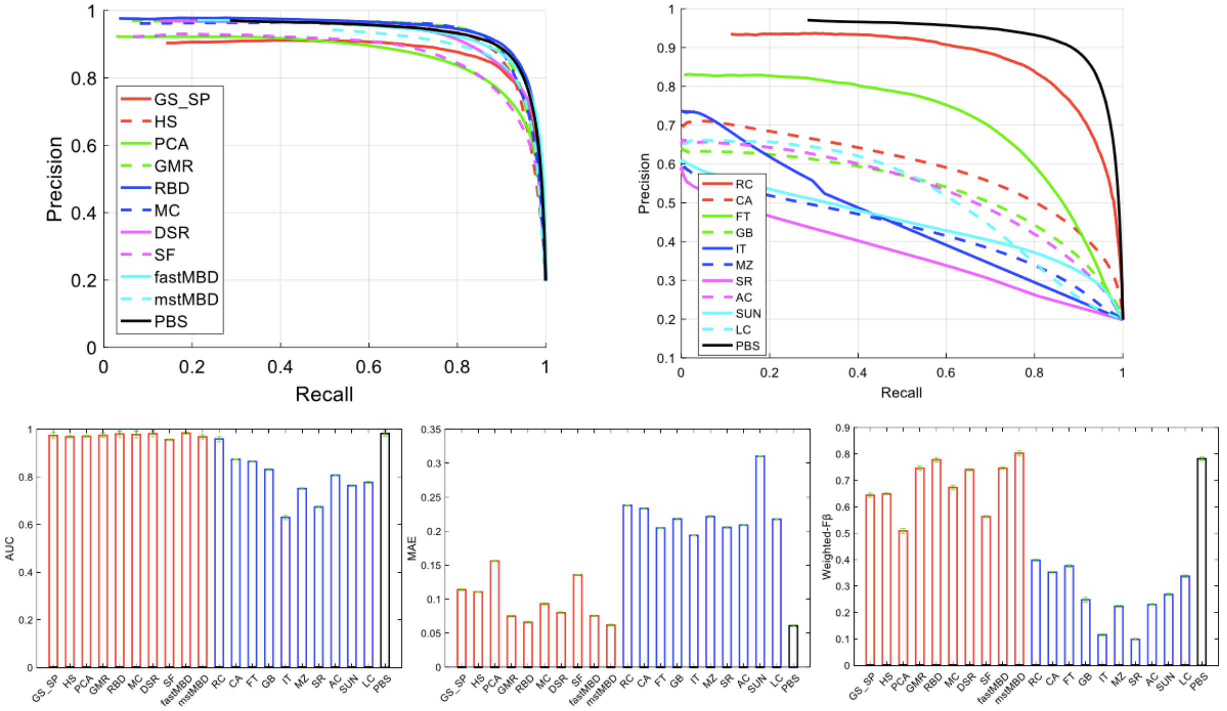


Fig. 9. Comparisons of PR curves, AUC, MAE and F_β^w for all methods on ASD dataset.

To further validate the effectiveness of the PBS method, we have applied the proposed algorithm on the MSRA, SED2, DUT_ORMON, ECCSD and SOD datasets, and then compare it with the ten state-of- the art approaches (i.e. GS_SP, HS, PCA, GMR, RBD, MC, DSR, SF, fastMBD and mstMBD). These approaches indeed demonstrate higher performance than the more traditional methods in Figs. 10 and 11 shows the experimental results on the five most challenging datasets.

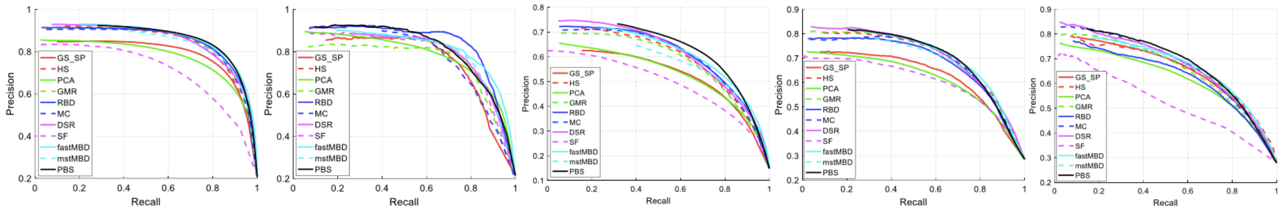


Fig. 10. Comparisons of PR curves, from left to right: MSRA, SED2, DUT_ORMON, ECCSD, and SOD.

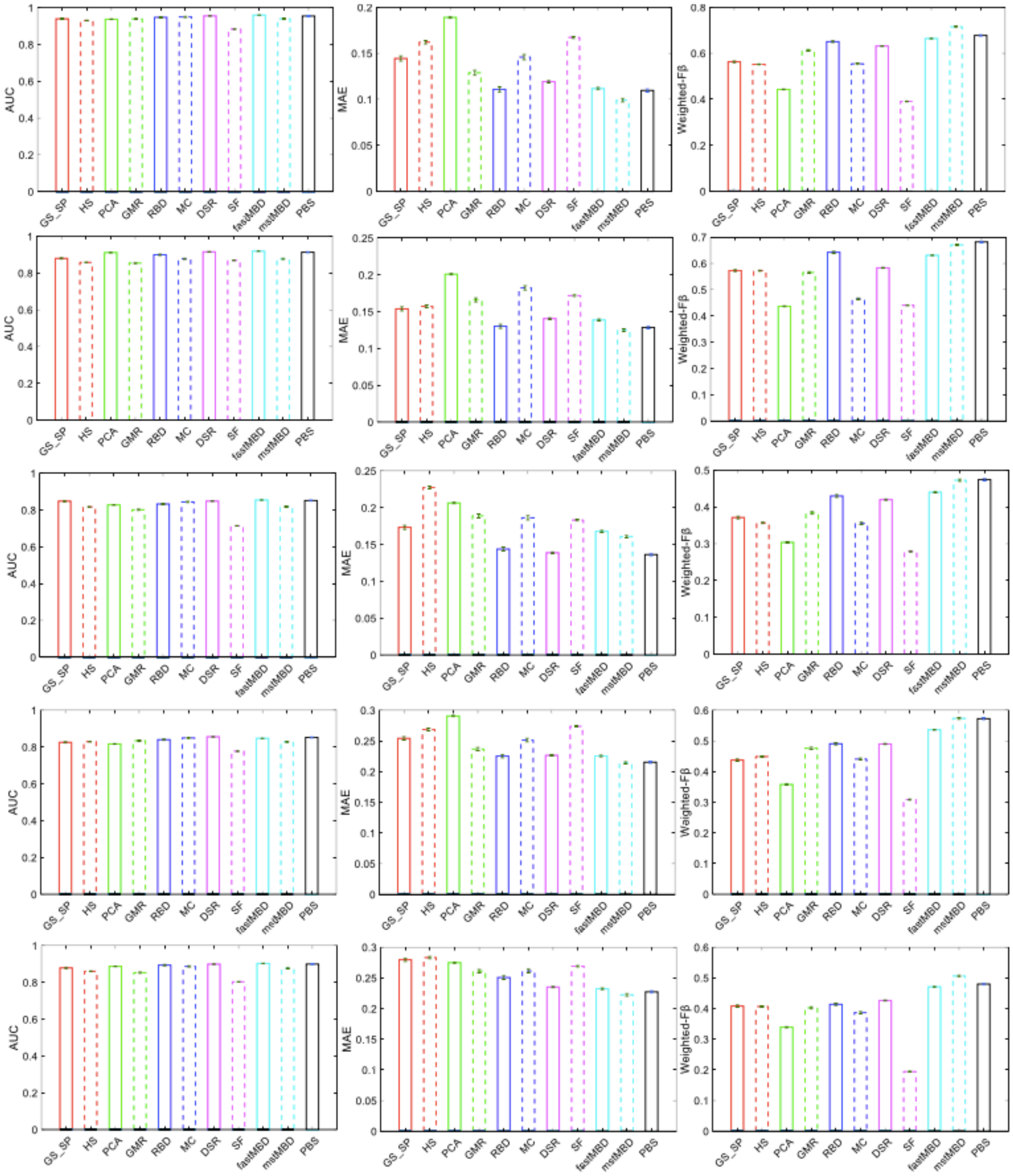


Fig. 11. Comparisons of AUC, MAE and F_β^w values on datasets, from top to bottom: MSRA, SED2, DUT_ORMON, ECCSD, and SOD.

We note that the proposed PBS model performs well on all datasets in terms of PR, AUC, MAE and F_β^w . When compared with the other path-based saliency models (GS_SP,RBD,mstMBD),our method is better than GS_SP on all datasets, and achieves comparable results to the state-of-the-art methods RBD and mstMBD, on some datasets (MSRA, SED2, SOD) and even better results in some other datasets (DUT_ORMON, ECCSD).

Finally, seven images (see Fig. 12) are selected to visually illustrate the capacity of the proposed method to homogeneously highlight salient regions. Such images contain objects of arbitrary shapes, cluttered backgrounds, blurred boundaries between objects and back- ground, or gradual illumination. By visually evaluating the saliency maps, one can establish the promising performance of the proposed PBS method to highlight salient regions uniformly. In contrast, the PCA method focuses mainly on the edges of the salient regions, while other methods show non-uniform salient regions or else, do not fully inhibit background details in the resulting saliency maps.

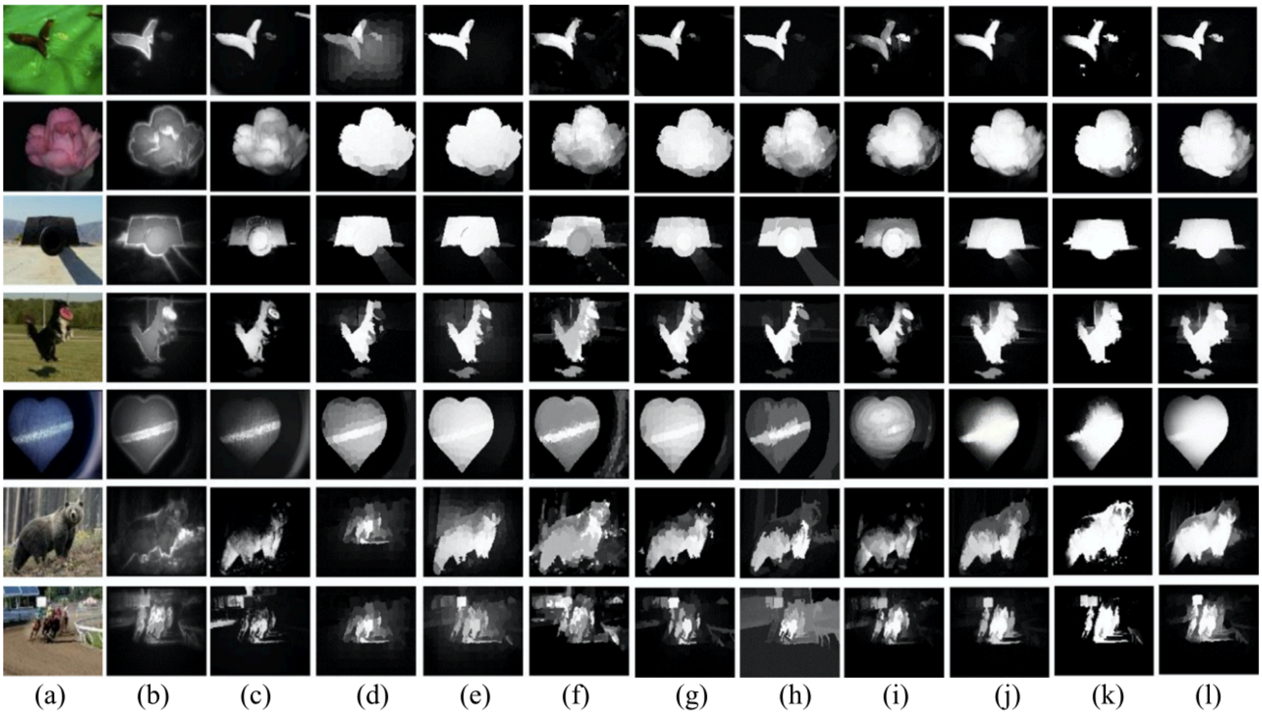


Fig. 12. Visual comparisons of different representative methods. From left to right: (a) original images and saliency maps obtained from (b) PCA, (c) SF, (d) GMR, (e) MC, (f) GS_SP, (g) RBD, (h)HS, (i) DSR, (j) fastMBD, (k) mstMBD and (l) PBS methods.

4.3.5. Limitations

In Fig. 13, we pinpoint three situations in which the PBS model fails to detect the salient objects in the images. The reason for such failure is that the contrast between the salient region and its surrounding background is not enough. Therefore, the smoothest path will falsely regard these different regions as the same one since consecutive dissimilarities between nodes along the path are very small. In this case, the background is incorrectly highlighted. One possible way to address such a problem may be to select more distinguishable features in order to better estimate the contrast, or alternatively, to incorporate some high-level priors such as object priors.

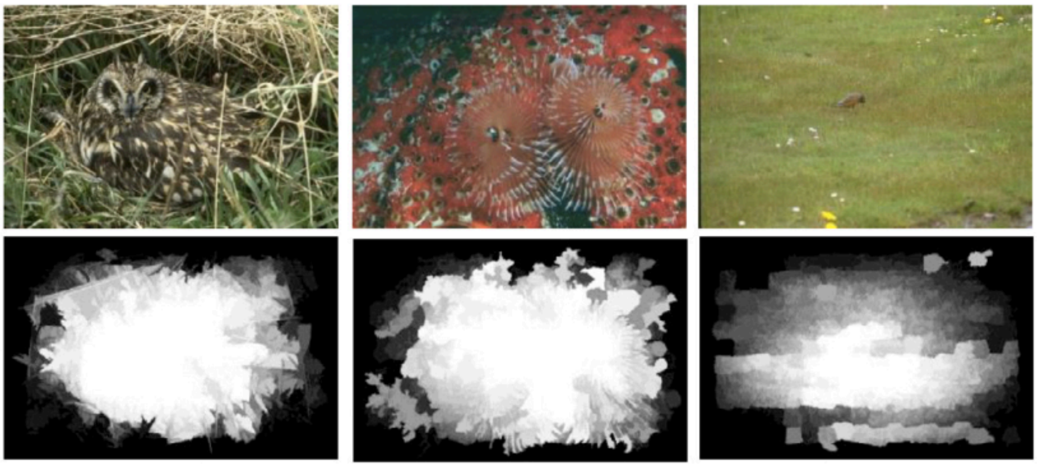


Fig. 13. Some failure cases where the contrast between salient objects and background are small.

5. Conclusions

Facing the challenge of unknown distributions, arbitrary shapes and cluttered background in natural images, this paper proposes a path distance metric to measure the relationship between image pixels and to model the saliency. The smoothest path, which follows the perceptual properties of similarity, proximity and continuity of Gestalt grouping, is generated on an undirected graph. The defined path distance metric, calculated by applying Laplacian analysis on the generated smoothest paths between each pair of nodes, shows promising results in segmenting different image elements and clustering similar ones. In addition, the proposed path-based distance metric can effectively solve the small- weight-accumulation problem in existing saliency methods modeled through geodesic distance or shortest distance. In the experiments, we first apply the path-based distance metric to existing saliency models, and the results demonstrate the improvement in performance. Then, we conduct a series of experiments by applying the proposed path-based background model on a series of challenging datasets. The experimental results demonstrate its capacity of accurately and uniformly highlighting salient regions regardless of arbitrary structures and uncertain distributions of image objects.

Compared with the Euclidean distance, geodesic distance and mini- mum barrier distance, the proposed distance defined on the smoothest path shows good performance in minimizing the intra-cluster differences and maximizing the inter-cluster ones. In our future work, we will further study its applications in the field of background modeling and non-local filtering related with detailed scene analysis and representation.

Acknowledgments

This research was supported by “National Natural Science Foun- dation of China” (No. 61272523, No. 61572103), “the National Key Project of Science and Technology of China” (No. 2011ZX05039-003- 4), and the Spanish national project LDM (TIN2013-40630-R).

References

[1] G. Sharma, F. Jurie, C. Schmid, Discriminative spatial saliency for image classification, in: CVPR, 2012.

[2] L. Wang, J. Xue, N. Zheng, G. Hua, Automatic salient object extraction with contextual cue, in: Computer Vision, 2011, pp. 105–112.

[3] U. Rutishauser, D. Walther, C. Koch, P. Perona, Is bottom-up attention useful for object recognition?, in: CVPR, 2004.

[4] V. Lempitsky, P. Kohli, C. Rother, T. Sharp, Image segmentation with a bounding box prior, in: Computer Vision, 2009, pp. 277–284.

[5] C. Guo, L. Zhang, A novel multi-resolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (2010) 185–198.

[6] G. Zhang, M. Cheng, S. Hu, R. Martin, A shape preserving approach to image resizing, in: Computer Graphics Forum, 2009, pp. 1897–1906.

[7] P. Derrick, L. Klinton, N. Ernst, Modeling the role of salience in the allocation of overt visual attention, Vis. Res. 42 (1) (2002) 107–123.

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

[9] S. Goferman. L. Zelnik-Manor, A. Tal, Context-aware saliency detection, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 1915–1926.

[10] J. Wolfe, T. Horowitz, What attributes guide the deployment of visual attention and how do they do it?, Nat. Rev. Neurosci. 5 (6) (2004) 495–501.

[11] L. Chen, Topological structure in visual perception, Science 218 (4573) (1982) 699–700.

[12] L. Chen, The topological approach to perceptual organization, Vis. Cognit. 12 (4) (2005) 553–637.

[13] A. Borji, M. Cheng, H. Jiang, et al., Salient object de tection: A benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722.

[14] M. Cheng, G. Zhang, N. Mitra, X. Huang, S. Hu, Global contrast based salient region detection, in: CVPR, 2011, pp. 409–416.

[15] Y. Qin, H. Lu, Y. Xu, et al., Saliency detection via cellular automata, in: CVPR, 2015, pp. 110–119.

[16] L. Duan, C. Wu, J. Miao, L. Qing, Y. Fu, Visual saliency detection by spatially weighted dissimilarity, in: CVPR, 2011, pp. 473–480.

[17] F. Perazzi, P. Krahenbuhl, Y. Pritch, A. Hornung, Saliency filters: Contrast based filtering for salient region detection, in: CVPR, 2012, pp. 33–740.

[18] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, in: CVPR, 2013, pp. 1155–1162.

[19] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, in: Advances in Neural Information Processing Systems, 2006, pp. 545–552.

[20] J. Zhang, S. Sclaroff, Saliency detection: A boolean map approach, in: ICCV, 2013.

[21] Y. Wei, F. Wen, W. Zhu, J. Sun, Geodesic saliency using background priors, in: ECCV, 2012.

[22] B. Jiang, L. Zhang, H. Lu, et al., Saliency detection via absorbing Markova chain, in: CVPR, 2013, pp. 1665–1672.

[23] C. Yang, L. Zhang, H. Lu, X. Ruan, M. Yang, Saliency detection via graph-based manifold ranking, in: CVPR, 2013, pp. 3166–3173.

[24] W. Tu, S. He, Q. Yang, et al., Real-time salient object detection with a minimum spanning tree, in: CVPR, 2016, pp. 2334–2342.

[25] J. Zhang, S. Sclaroff, Z. Lin, et al., Minimum barrier salient object detection at 80 fps, in: ICCV, 2015, pp. 1404–1412.

[26] W. Köhler, Gestalt Psychology: An Introduction to New Concepts in Modern Psychology, WW Norton & Company, 1970.

[27] A. Wolters, K. Koffka, Principles of Gestalt Psychology, 1936.

[28] C. Yang, L. Zhang, H. Lu, Graph-regularized saliency detection with convex-hull-based center prior, IEEE Signal Process. Lett. 20 (2013) 637–640.

[29] W. Zhu, S. Liang, Y. Wei, J. Sun, Saliency optimization from robust background detection, in: CVPR, 2014, pp. 2814–2821.

[30] H. Jiang, J. Wang, Z. Yuan, et al., Salient object detection: A discriminative regional feature integration approach, in: CVPR, 2013, pp. 2083–2090.

[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC Superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach.
Intell. 34 (2012) 2274–2282.

[32] K. Chang, T. Liu, H. Chen, et al., Fusing generic objectness and visual saliency for salient object detection, in: ICCV, 2011.

[33] B. Fischer, T. Zöller, J.M. Buhmann, Path based pairwise data clustering with application to texture segmentation, in: Proceedings of the Third International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition, 2001, pp. 235–250.

[34] M. Pelillo, The dynamics of nonlinear relaxation labeling processes, J. Math. Imaging Vision 7 (4) (1997) 309–323.

[35] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, in: CVPR, 2009.

[36] T. T. Liu, Z. Yuan, J. Sun, et al., Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.

[37] A. Borji, L. Itti, State-of-the-art in visual attention modeling, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 185–207.

[38] R. Margolin, L. Zelnik-Manor, A. Tal, How to evaluate foreground maps?, in: CVPR, 2014, pp. 248–255.

[39] R. Margolin, A. Tal, L. Zelnik-Manor, What makes a patch distinct?, in: Computer Vision and Pattern Recognition, 2013, pp. 1139–1146.

[40] X. Li, H. Lu, L. Zhang, et al., Saliency detection via dense and sparse reconstruction, in: CVPR, 2013, pp. 2976–2983.

[41] Y. Zhai, M. Shah, Visual attention detection in video sequences using spatiotem- poral cues, in: Proceedings of the 14th Annual ACM International Conference on Multimedia, 2006, pp. 815–824.

[42] Y. Ma, H. Zhang, Contrast-based image attention analysis by using fuzzy growing, in: Proceedings of the Eleventh ACM International Conference on Multimedia, 2003, pp. 374–381.

[43] X. Hou, L. Zhang, Saliency detection: A spectral residual approach, in: Computer Vision and Pattern Recognition, 2007, pp. 1–8.

[44] R. Achanta, F. Estrada, P. Wils, S. Süsstrunk, Salient region detection and segmentation, in: International Conference on Computer Vision Systems, Springer Berlin Heidelberg, 2008, pp. 66–75.

[45] L. Zhang, M. Tong, T. Marks, et al., SUN: A Bayesian framework for saliency using natural statistics, J. Vis. 8 (7) (2008) 32.