



# Visual saliency based on extended manifold ranking and third-order optimization refinement

Dongjing Shan<sup>a</sup>, Xiongwei Zhang<sup>a,\*</sup>, Chao Zhang<sup>b</sup>

<sup>a</sup> Army Engineering University of PLA, Nanjing 400035, China

<sup>b</sup> Key Laboratory of Machine Perception, Peking University, Beijing 100081, China

## ARTICLE INFO

### Article history:

Received 19 November 2017

Available online 4 September 2018

### Keywords:

saliency detection  
manifold ranking  
graphical model  
image segmentation

## ABSTRACT

Graph-based approaches for saliency detection have attracted much attention and been exploited widely in recent years. In this paper, we present a new method to promote the performance of existing manifold ranking algorithms. Initially, we use background weight map to provide seeds for manifold ranking; Next, we extend the traditional manifold ranking to second-order formula and add a weight mask to its fitting term. Finally, for further improvement of the performance, we establish a third-order smoothness framework to optimize the saliency map. In the experiments, we compare two versions (manifold ranking with and without optimization) of our model with seven previous methods and test them on several benchmark datasets. Different kinds of strategies are also adopted for evaluation and the results demonstrate that our method achieves the state-of-the-art.

© 2018 Elsevier B.V. All rights reserved.

## 1. Introduction

Recently, a number of graph-based approaches have been exploited and attracted much attention. J. Harel offered a graph-based bottom-up saliency model [1] based on markov chains, and Zhang et al. [2] proposed a model based on absorbing Markov chain (AMC), in which deep features extracted from fully convolutional networks are adopted to learn a transition probability matrix. In the work of Liu [3], a conditional random field (CRF) is learned to effectively combine a set of features for saliency detection. Also quadratic energy model has been adopted for saliency in [4], which is used as a alternative to the binary CRF. Li et al. [5] designed a regularized random walks ranking to formulate pixel-wised saliency maps from background and foreground estimations. Based on the random walk model, Kong et al. [6] introduced a pattern mining algorithm for seed selection and extended the random walk to 2-ring neighbors, but this method needs rough saliency maps created by other existing methods, also Jiang et al. [7] proposed a generic scheme to promote diffusion-based saliency detection algorithm by original ways to re-synthesize the diffusion matrix and construct the seed vector. Another widely used graph-based model is manifold ranking (MR). Yang et al. [8] utilizes the four boundaries of the image as background seeds, and extract foreground queries via manifold ranking to get the final saliency

map. Wang et al. [9] place an optimization framework [10] on the manifold ranking process, and then returned the generated map to MR for a better one. Tao propose an MR-based matrix factorization (MRMF) [11–13] method to model ranking problem in the matrix factorization framework and embeds query sample labels in the coefficients. Gong et al. [14] proposed a propagation algorithm employing teaching-to-learn and learning-to-teach strategies to optimize the propagation quality.

Although significant progress has been made, there remain some drawbacks for the previous MR methods. Firstly, most methods utilize the four boundaries of the image as background seeds, but it is not rigorous enough. Apparently, not all the boundary superpixels are background when the objects touch the image boundaries, and also it is difficult to recognize them to filter out outliers. Even the background superpixels in the boundaries are extracted accurately, they can not represent all the background information that exists in the whole image. Secondly, MR model is a tradeoff between two connected superpixels, this is not sufficient as the superpixels in 2-ring [6,8] neighbors or even 3-ring neighbors are relevant to the saliency of the certain superpixel. Thirdly, MR model propagates the seed labels to the rest of the superpixels in the image, the difference between the generated saliency map and the original labeled map is not weighted, so the saliency is constrained by the labeled map even in the non-seed positions.

In order to address the above issues, we propose our extended manifold ranking algorithm. At the first place, we utilize the boundary prior together with the four boundaries of the image to build background seeds. Next, we extend the traditional man-

\* Corresponding author. Tel.: +8613705157094.

E-mail address: [shandongjing@pku.edu.cn](mailto:shandongjing@pku.edu.cn) (D. Shan).

ifold ranking to second-order formula and add a weight mask to its fitting term (i.e. the second term of Eqn 2, and the first one is called first-order smoothness term in this paper) to make the saliency of non-labeled region more accurate. At last, we establish a minimization framework based on third-order smoothness to optimize the previous saliency map and generate the final one.

## 2. Related work

### 2.1. Manifold ranking

The graph-based manifold ranking model [5,8,9,15,16] exploits the intrinsic manifold structure of the image, and seeks to rank the graph nodes by given some labeled ones. It is used widely to calculate the rough saliency values of an image. Suppose the image is segmented into  $n$  superpixels by SLIC algorithm [17], and the features (e.g. the mean CIElab color values) are extracted on each superpixel. In order to establish a graph  $G = (V, E)$ , firstly, the features are denoted as a node set  $V = \{v_1, \dots, v_l, v_{l+1}, \dots, v_n\} \in \mathbb{R}^{n \times m}$ , in which the first  $l$  elements are labeled manually or by some priors,  $n$  is the total feature number, and  $m$  is the feature dimension. Usually,  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$  acts as an indication vector, where  $y_i = 1$  or  $0$  means the corresponding node  $v_i$  is a labeled seed or not, and many representative literatures [5,8,15] utilize image boundaries (see Fig. 4(a)–(d)) as the labeled seeds. Next, each node is connected to its 2-ring neighbors by undirected edges, and the graph edges  $E$  are weighted by an affinity matrix  $W = [w_{ij}]_{n \times n}$ , in which the element is defined as:

$$w_{ij} = \begin{cases} \exp\left[-\frac{\|v_i - v_j\|^2}{\delta^2}\right] & \text{if } v_i \text{ and } v_j \text{ are connected,} \\ 0 & \text{others.} \end{cases} \quad (1)$$

where  $\|\cdot\|$  denotes 2-norm of a vector, and  $\delta$  is controlling constant. Finally, let  $\mathbf{f}: V \rightarrow \mathbb{R}^n$  be a ranking function which assigns ranking values  $\mathbf{f} = [f_1, f_2, \dots, f_n]^T$  to the node set  $V$ , then it can be obtained by solving the following minimization problem:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \frac{1}{2} \left( \sum_{i,j=1}^n w_{ij} \left\| \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right\|^2 + \mu \sum_{i=1}^n \|f_i - y_i\|^2 \right) \quad (2)$$

where  $d_i$  is the  $i$ -th element of the degree matrix  $D = \text{diag}(d_1, \dots, d_n)$ , and it is defined as  $d_i = \sum_j w_{ij}$ .  $\mu$  is a controlling parameter which balances the smoothness and the fitting constraints. By setting the derivative of the formulation to be zero, the optimized solution can be written as:

$$\mathbf{f}^* = (I - \alpha S)^{-1} \mathbf{y} \quad (3)$$

where  $\alpha = \frac{1}{1+\mu}$ ,  $I$  is an identity matrix,  $S = D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ , and  $I - \alpha S$  constitutes a normalized Laplacian matrix [18]. But usually, the unnormalized Laplacian matrix  $(D - \alpha W)$  is adopted by the literatures, as the experiments demonstrate that it achieves better performance than the former one [8] (see Fig. 1(b)(c)), the ranking function is formulated as:

$$\mathbf{f}^* = (D - \alpha W)^{-1} \mathbf{y} \quad (4)$$

The paper [15] make more innovations and represent an image as a multi-scale graph with fine superpixels and coarse regions as nodes, then the nodes are ranked based on affinity matrices, finally the saliency map is calculated in a cascade scheme efficiently.

### 2.2. Optimization framework

The optimization framework [10] is used to smooth the raw saliency map by spacial consistency of the superpixels. in [10], the

background probability map and the background weighted contrast (foreground) map are computed successively. A superpixel's background probability  $\omega_i^{bg}$  is estimated by its connectivity to the image boundaries, and the process can be formulated as:

$$\text{BndCon}(p) = \frac{\text{len}_{\text{bnd}}(p)}{\sqrt{\text{Area}(p)}} \quad (5)$$

$$\omega_i^{bg} = 1 - \exp\left(-\frac{\text{BndCon}^2(p_i)}{2\delta^2}\right) \quad (6)$$

where the numerator term of Eq. 5 is the superpixel's length along the image boundaries, and the denominator represents its perimeter.  $\delta$  in Eq. 6 is set to 1 empirically. The superpixel's global contrast is defined as:

$$\omega\text{Ctr}(p) = \sum_{i=1}^N d_{\text{app}}(p, p_i) \omega_{\text{spa}}(p, p_i) \omega_i^{bg} \quad (7)$$

where  $d_{\text{app}}(p, p_i)$  is the appearance distance between the two superpixels,  $\omega_{\text{spa}}$  is the weight based on the space distance. The optimization framework is then defined as:

$$\sum_{i=1}^N \omega_i^{bg} s_i^2 + \sum_{i=1}^N \omega_i^{fg} (s_i - 1)^2 + \sum_{i,j} \omega_{ij} (s_i - s_j)^2 \quad (8)$$

where  $\omega_i^{fg}$  is the superpixel's foreground probability,  $\omega_{ij}$  is the weight between each pair of adjacent superpixels, and the last term is a tradeoff over the pair based on the weights, we call it first-order smoothness, as we will expend it to high-orders in our proposed method.

## 3. The proposed algorithm

In our algorithm, the image is first abstracted as a set of superpixels using SLIC method, and then each superpixel is characterized by the mean CIElab color and the Local Binary Pattern (LBP) features [19]. Thus, the image is symbolized as  $P = [p_1, p_2, \dots, p_n]$ , where  $p_i = [p_i^{\text{lab}}, p_i^{\text{LBP}}]^T$  means the catenated vector containing two kinds of features extracted from the  $i$ -th superpixel, and  $n$  is the superpixel number which is set to 200 in our method.

### 3.1. Extended manifold ranking

The paper [6] proposed an extended random walk algorithm, in which a quadratic Laplacian term was used to enforce further label consistency of nodes. Inspired by this instance, we design an extended manifold ranking algorithm as follows:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} \frac{1}{2} * \left( \sum_{i,j=1}^n w_{ij} \left\| \frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right\|^2 + \sum_{i=1}^n \left\| d_i \frac{f_i}{\sqrt{d_i}} - \sum_{j \in N(i)} w_{ij} \frac{f_j}{\sqrt{d_j}} \right\|^2 + \mu \sum_{i=1}^n \|f_i - y_i\|^2 \right) \quad (9)$$

where the first term is the same with the one in Eqn 2 and is a smoothness tradeoff between each pair of connected nodes, we define it as the first-order term of manifold ranking. The second term is a further smoothness between each node and its 2-ring neighbors  $N(i)$ , we call it the second-order term of manifold ranking. Additionally,  $f_i$ ,  $w_{ij}$ ,  $d_i$  and the last term (fitting term of manifold ranking) have the same meaning with the ones in Eqn 2,  $\mu$  is set to 0.01. Then we use the bottom, top, left and right image boundary as labeled seeds repeatedly and generate four saliency maps which can be integrated by the following process:

$$S_l(i) = S_b(i) \times S_t(i) \times S_l(i) \times S_r(i) \quad (10)$$

Download English Version:

<https://daneshyari.com/en/article/10146094>

Download Persian Version:

<https://daneshyari.com/article/10146094>

[Daneshyari.com](https://daneshyari.com)