Author's Accepted Manuscript

Multi-layer Graph Constraints for Interactive Image Segmentation via Game Theory

Tao Wang, Quansen Sun, Zexuan Ji, Qiang Chen, Peng Fu

Cite this article as: Tao Wang, Quansen Sun, Zexuan Ji, Qiang Chen and Peng Fu, Multi-layer Graph Constraints for Interactive Image Segmentation via Game Theory, *Pattern Recognition,* http://dx.doi.org/10.1016/j.patcog.2016.01.018

Multi-layer Graph Constraints for Interactive Image Segmentation via Game Theory

Tao Wang, Quansen Sun[*], Zexuan Ji[*], Qiang Chen, Peng Fu

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract

The combination of pixels and superpixels has been widely used for image segmentation, where the pixels and superpixels are segmented together. These combination methods can obtain more robust results by using more informative superpixel features. However, since the superpixel may not accurately capture the details for the small and slender regions, the results of these combination methods are often label inconsistent with the objects. Furthermore, these methods also fall into expensive time cost due to introducing more interactions between pixels and superpixels. To overcome the above problems, in this paper, we propose an interactive image segmentation method based on multi-layer graph constraints. The relationships between pixels/superpixels and labels are introduced into the conventional combination framework to further improve the segmentation accuracy. The segmentation model is constructed based on the estimation of probabilities of pixels and superpixels by a nonparametric learning framework. Then the probabilities of pixels and superpixels are updated iteratively by utilizing the game theory based optimization strategy. Experiments on challenging data sets demonstrate that the proposed method can obtain better segmentation results than the state-of-the-art methods.

Keywords Image segmentation, superpixel, multi-layer graph, nonparametric learning, game theory

Title Page

Article Information

Article Title: Multi-layer Graph Constraints for Interactive Image Segmentation via Game Theory

Authors: Tao Wang, Quansen Sun, Zexuan Ji, Qiang Chen, Peng Fu

Author affiliations: School of Computer Science and Engineering, Nanjing University of Science

[*] Corresponding author. Quansen Sun, Email: sunquansen@njust.edu.cn ; Zexuan Ji, E-mail: jizexuan@njust.edu.cn.

and Technology, Nanjing, 210094, China

Correspondence information

Corresponding author name: Tao Wang

Affiliation: School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China

Address: No. 200, Street Xiao Ling Wei, School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094,China.

Email address: wangtaoatnjust@163.com & 670101389@qq.com

Telephone number: (+86) 15850564521

1. Introduction

As a low-level computer vision task, image segmentation is very important for many high level applications [1-4] in computer vision. The problem of image segmentation can be interpreted as dividing an image into different regions, where pixels belonging to the same region should have consistent features. Though various image segmentation methods have been proposed, it is still a challenging problem due to the abroad types of images and large demands on users.

Generally, the segmentation methods can be classified into unsupervised, semi-supervised and full supervised approaches. Recently, the semi-supervised methods, in which the user is allowed to provide a few seeds to represent the label information, have gained much popularity. These methods are very practical in many applications, because the user can add his intention during the segmentation. During last decades, many interactive image segmentation methods have been proposed, such as graph cuts [5-10], random walk [11-14], shortest path [15-17] and so on. In graph cuts models, the image is represented by a weighted graph. Then the segmentation problem is translated to find a minimum cut in the constructed s/t graph via a maximum flow computation [18]. In random walk models, the labels of unseeded pixels are estimated based on the seed propagation on a relationship graph. In shortest path models, the distances of the paths between unseeded and seeded pixels are measured firstly. Then one unseeded pixel is assigned the foreground label if there is a shorter path from this pixel to a foreground seed than to any

background seeds. In general, the above methods are very sensitive to the seed's quantity and location. If enough seeds are provided, good results can be obtained. However, when the number of seeds is limited, it is hard to get satisfactory results and the user needs to pay more efforts. To solve the above problem, many methods [19-24] based on image regions obtained by unsupervised segmentation algorithms (superpixels) have been proposed, in which pixels within the same superpixel are considered to share the same labels. Compared with pixel-level based methods, the superpixel based methods can usually obtain more robust results by using more informative features. However, the results of these methods are generally affected by the qualities of superpixels. In fact, the situations that superpixels are not consistent with boundaries often arise in nature images. So the "hard" constraints between pixels and superpixels often lead to over-segmentation results.

Many methods [25-29] combine multiple segmentations produced by one unsupervised segmentation algorithm such as mean shift [30] with different parameters to overcome the shortcomings of the hard constraint. In the Robust P^n model [25], by using condition random filed in a principled manner, the hidden soft constraints between pixels and multiple superpixels are utilized for the segmentation. Furthermore, the quality of each superpixel is also measured to reduce the influence of "bad" superpixels. However, this method only explores how the superpixel impacts pixel labeling regardless of the inherent relationships between superpixels and pixels. Based on the Robust P^n model, many combination segmentation methods have been developed [27-29], where in their framework the pixels and superpixels are segmented together. In [27-28], the spectrum based models are utilized to segment the pixels and superpixels together. In [29], a nonparametric learning model is utilized to estimate the likelihoods of pixels and superpixels. These pixel-superpixel combination methods can obtain better results than classic interactive methods such as graph cuts and random walk with fewer seeds. However, since the superpixel may not accurately capture the details for the small and slender regions, the results of these combination methods are often label inconsistent with the objects. Furthermore, due to the introduction of more interactions between pixels and superpixels, these methods are also limited to the expensive algorithm cost. For example, the method in [29] has a heavy burden to solve the inversion of an over-large matrix.

To solve the above problems, in this paper, we propose an interactive image segmentation

method based on multi-layer graph constraints. The relationships between pixels/superpixels and labels are introduced to overcome the label inconsistent of the conventional combination methods. Then the multi-layer connections among the pixel-layer, superpixel-layer and label-layer are fused for the segmentation. A nonparametric learning framework is utilized to estimate the probabilities of pixels and superpixels, respectively. The optimization of the proposed segmentation energy function is considered as a game theory problem, where the probabilities of pixels and superpixels are updated iteratively until convergence. The proposed method can obtain high-quality segmentation results with more accurate object details, less user's interactions, and lower time cost. As shown in Fig. 1, in the first two columns only one pixel-seed is selected for each label, and in the last two columns the trimap is used for the challenging image with slender objects. It can be seen that satisfactory results can be produced shown in row 2nd.
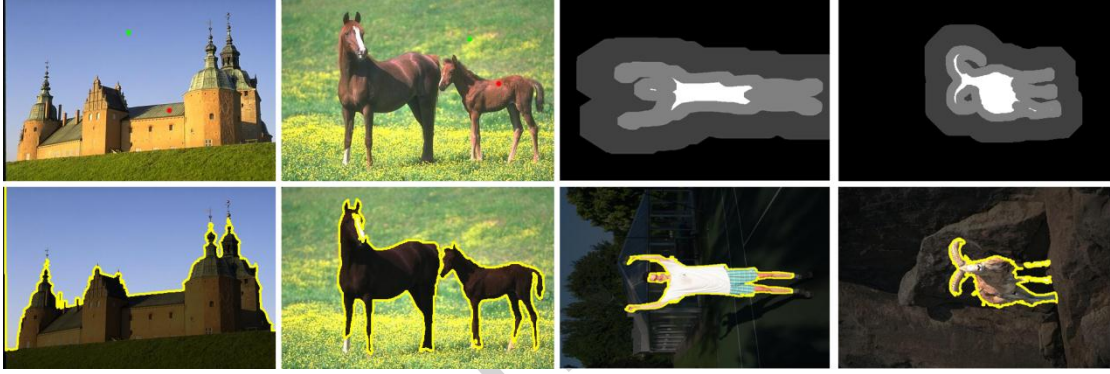

Fig. 1 Image segmentation results by using the proposed method. Columns 1st and 2nd show the segmentation results with only one pixel-seed for each label (red for foreground and green for background). Columns 3rd and 4th show the segmentation results for slender objects with trimaps.

2. Pixel-superpixel combination model

The problem of interactive segmentation can be regarded as estimating the labels of unseeded pixels based on seeded pixels. The similarities of all pairwise pixels are usually used to propagate the relationships between seeded and unseeded pixels. If one unseeded pixel has high similarity with the foreground seeds, then it is likely belonging to the foreground. Fig. 2 shows a neighboring relationship graph of pixels, where the neighboring pixels x_i and x_j share the weight ω_{ij}.
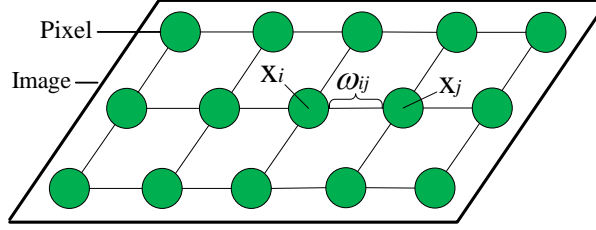
Fig. 2 The 4-neighborhood relationship graph for pixel-level based models, where x_i and x_j are neighborhood pixels and ω_{ij} represents their similarity.

The weight of pixels x_i and x_j is always defined as a typical Gaussian function:

$$\omega_{ij} = \begin{cases} \exp(-\beta \|c_i - c_j\|_2) & x_j \in \aleph_i \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where c_i denotes the intensity feature at pixel x_i and β is a constant that controls the strength of the weight. $\aleph_i$ represents the neighborhood of x_i. Then the energy function of the probabilities for pixels with respect to the label $l \in \{1,...L\}$ can be naturally defined as follows [11]:

$$E(\pi_l) = \frac{1}{2} \sum_{i,j=1}^{N_X} \omega_{ij}(\pi_{il} - \pi_{jl})^2 = \frac{1}{2} \vec{\pi_l}^T L \vec{\pi_l} \tag{2}$$

where N_X represents the number of all the pixels in the image, and π_{il} is the probability of the pixel x_i belonging to the label l. The right term is the matrix form of the middle term, where $\vec{\pi}_l = [\pi_{il}]_{N_X \times 1}$ and the Laplacian matrix $L = [l_{ij}]_{N_X \times N_X} = D - W$. $W = [\omega_{ij}]_{N_X \times N_X}$ and $D = diag([d_1,...,d_{N_X}])$, where $d_i = \sum_{j=1}^{N_X} \omega_{ij}$. Partitioning the pixels into two sets: X_M (labeled/seeds pixels) and X_U (unlabeled pixels), such that $X_M \bigcup X_U = X$ and $X_M \bigcap X_U = \varnothing$. Under the assumption the nodes in L and π are ordered such that seeded nodes are first and unseeded nodes are second [11], then the equation (2) can be decomposed into:

$$E(\pi_U) = \frac{1}{2} \begin{bmatrix} \vec{\pi}_M^T & \vec{\pi}_U^T \end{bmatrix} \begin{bmatrix} L_M & B \\ B^T & L_U \end{bmatrix} \begin{bmatrix} \vec{\pi}_M \\ \vec{\pi}_U \end{bmatrix} \tag{3}$$

where $\vec{\pi}_M = [\pi_i]_{N_{X_M} \times 1}$ and $\vec{\pi}_U = [\pi_i]_{N_{X_U} \times 1}$ correspond to the potentials of the seeded and unseeded pixels, respectively. $L_M = [l_{ij}]_{N_{X_M} \times N_{X_M}}$, $L_U = [l_{ij}]_{N_{X_U} \times N_{X_U}}$, and $B = [l_{ij}]_{N_{X_M} \times N_{X_U}}$ are three sub-blocks of the ordered L which correspond to the nodes' relationships between labeled pixels, unlabeled pixels, and unlabeled-labeled pixels, respectively. Since L is positive semi-definite

[11], the only critical points of $E(\pi_U)$ will be minimal. Differentiating the equation (3) with respect to π_U, the probabilities of unlabeled pixels belonging to the label l can be obtained:

$$L_U \pi_U^l = -B^T \pi_M^l \tag{4}$$

where $\pi_M^l = [\pi_{il}]_{N_M \times 1}$ and $\pi_{il} = \begin{cases} 1 & \text{if } Q(x_i) = l \\ 0 & \text{if } Q(x_i) \neq l \end{cases}$, $Q(x_i)$ denotes the label for the marked pixel x_i. Finally, the unlabeled pixels are assigned the labels with maximum probabilities.
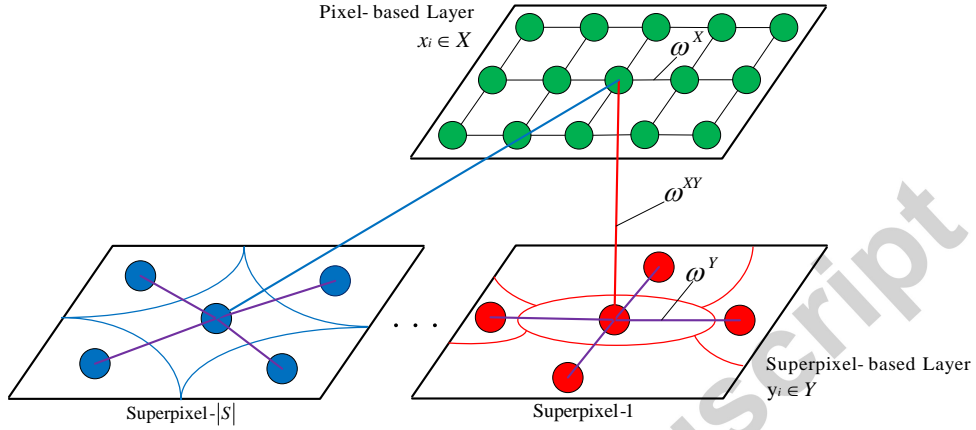

Fig. 3 The graph of multi-layer relationships between pixels and superpixels, where the top is the pixel-based layer, and the below is the superpixel-based layer. ω^X represents the pairwise weights between neighboring pixels, ω^Y represents the full connected weights of all superpixels and ω^{XY} represents the weights of inter-layer connections between pixels and superpixels. $|S|$ denotes the number of superpixel layers.

As the above description, the relationships between pixels are used to estimate the labels of the unmarked pixels in the pixel-level based methods. However, the similarities between pixels are not accurate enough, which makes the segmentation sensitive to the seeds' quantity and location. To improve the robustness of the segmentation, the superpixels obtained by unsupervised segmentation algorithms are successfully combined with pixels. The graph of multi-layer relationships between pixels and superpixels can be generally described in Fig. 3, where the top of the graph shows the pixel-based layer composed of all pixels in the image, and the below one shows the superpixel-based layer formed by one unsupervised algorithm with different parameters or different unsupervised algorithms.

The pairwise neighboring weights ω^X inside the pixels have been described in equation (1). Unlike the adjacent connections of pixels, the long-range connections between superpixels are usually considered, which is helpful in propagating the labeling cues of each region to the whole image areas. The full relationships between any two superpixels are typically defined as follows:

6

$$\omega_{ij}^Y = \exp(-\beta \|\bar{c}_i - \bar{c}_j\|_2) \tag{5}$$

where $\bar{c}_i$ is the mean color of the pixels in the superpixel y_i. Due to the fact that each pixel corresponds to only one superpixel in one over-segmentation, the weight ω_{ij}^{XY} ($= \omega_{ji}^{YX}$) between the pixel x_i and the superpixel y_j is described as follows:

$$\omega_{ij}^{XY} = \begin{cases} \exp(-\beta \|c_i - \bar{c}_j\|_2) & x_i \in y_j \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Based on the above connections, two quadratic cost functions are designed in [29] to estimate the likelihoods probabilities, one corresponds to the learning of pixel likelihoods and the other corresponds to the superpixel likelihoods:

$$\begin{cases} E(\pi_l^X) = \sum_{i,j=1}^{Nx} \omega_{ij}^X (\pi_{il} - \pi_{jl})^2 + \sum_{i=1}^{Nx} \lambda_i^X d_i^X (\pi_{il} - \pi_{il}^*)^2 + \sum_{i=1}^{Nx} \mu d_i^X (\pi_{il} - \bar{\pi}_{il})^2 \\ E(z_l^Y) = \sum_{i,j=1}^{Ny} \omega_{ij}^Y (z_{il} - z_{jl})^2 + \sum_{i=1}^{Ny} \lambda_i^Y d_i^Y (z_{il} - z_{il}^*)^2 + \sum_{i=1}^{Ny} \varepsilon d_i^Y (z_{il} - \bar{z}_{il})^2 \end{cases} \tag{7}$$

where $d_i^X = \sum_{j=1}^{Nx} \omega_{ij}^X$ and $d_i^Y = \sum_{j=1}^{Ny} \omega_{ij}^Y$, the two parameters λ_i^X and λ_i^Y are constants with high value for the seeded pixels/superpixels, and 0 otherwise. The seeded superpixels can be generally determined based on the seeded pixels, where superpixels containing the seeded pixels are considered as the seeds with the same label of the seeded pixels. However, in some cases the superpixels may contain pixel seeds with different labels and these inaccurate superpixels are excluded, and only the superpixels containing pixel seeds with the same label are utilized as the superpixel seeds. The pixel/superpixel–seed likelihoods π_{il}^* and z_{il}^* are 1 if the pixel/superpxiel x_i / y_i is the seed with the label l, and 0 otherwise. The estimated likelihood $\bar{\pi}_{il}$ is defined as the weighted average of the corresponding superpixel likelihoods at the pixel x_i, $\bar{\pi}_{il} = \sum_{j=1}^{Ny} \rho_{ij}^{XY} z_{jl}$ with $\rho_{ij}^{XY} = \omega_{ij}^{XY} / \sum_{j=1}^{Ny} \omega_{ij}^{XY}$. The estimated likelihood $\bar{z}_{il}$ is defined as the weighted average of the containing pixel likelihoods at the superpixel y_i, $\bar{z}_{il} = \sum_{j=1}^{Nx} \rho_{ij}^{YX} \pi_{jl}$ with $\rho_{ij}^{YX} = \omega_{ij}^{YX} / \sum_{j=1}^{Nx} \omega_{ij}^{YX}$. μ and ε are two adjustment coefficients. The first term is similar to equation (2) which is based on the relationships inside the pixel/superpixel layer. The second term imposes the constraint for seeds that each seed should be assigned to the initial label. The third term utilizes the relationships between pixels and superpixels. These two functions influence each

7

other and a nonparametric optimization method is used to obtain the likelihood probability of each pixel. For more details, please refer to [29].

Though the combination of pixels and superpixels can lead to more informative results, there still exist some problems which can be summarized as the following two aspects, one is the inconsistency with the label information caused by the inaccurate superpixel constraint especially in many small regions, and the other is the increasing algorithm burden due to the introduction of more interactions.

3. The proposed model

Generally, the interactive segmentation methods should satisfy the following two conditions, one is the aim of extracting accurate object details and the other is the segmentation should not be sensitive to the seeds. The combination of pixels and superpixels can improve the robustness to seeds. To further improve the segmentation accuracy, in this paper, we introduce the relationships between pixels/superpixels and labels into the conventional combination framework, and fuse the multi-layer connections among the pixel-layer, superpixel-layer, and label-layer together for the segmentation.

3.1 Construction of Graph

We construct an undirected graph $G = (V, E)$ with pixels, superpixels and labels as nodes $V = \{X, Y, L\}$, as shown in Fig. 4. The edges $E = \{E^X, E^Y, E^{XY}, E^{XL}, E^{YL}\}$ are the connections between pairwise nodes.
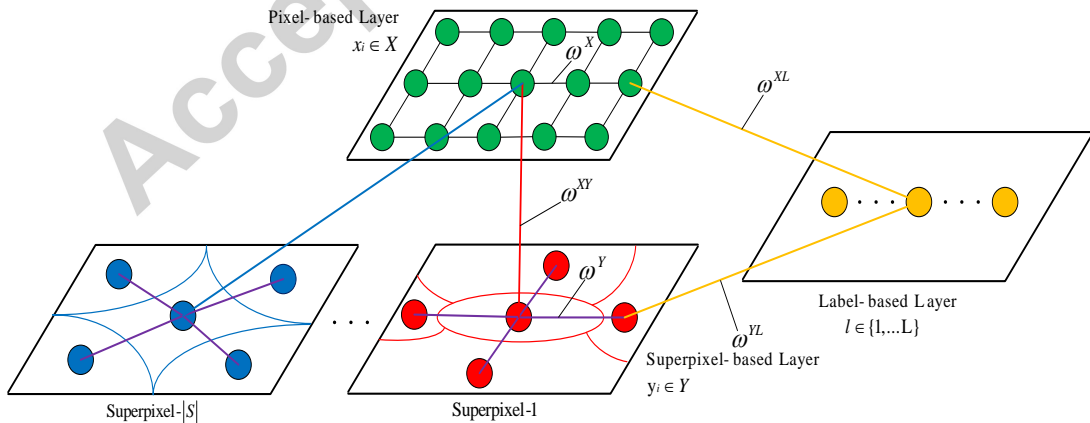
Fig. 4 An illustration of the proposed graph G. The vertex set V consists of three kinds of nodes, the first denotes the pixels $x_i \in X$ (green circles), the second denotes the multiple superpixels $y_i \in Y$ (red and blue circles) and the last denotes the labels $l \in \{1,...L\}$ (orange circles). The edge set E consists of five kinds of connections, where ω^X corresponds to the relationships between pixels, ω^Y corresponds to the relationships

between superpixels, ω^{XY} corresponds to the relationships between pixels and superpixels, ω^{XL} and ω^{YL} correspond to the relationships between pixels/superpixels and labels, respectively. $|S|$ represents the number of multiple superpixel layers.

1) E^{X}: the connections among neighborhood pixels. The weights between pixels ω^{X} were defined in equation (1);

2) E^{Y}: the full connections of superpixels. The weights between superpixels ω^{Y} were defined in equation (5);

3) E^{XY}: the connections between the pixel-based layer and the superpixel-based layer. It is used to describe the relationships of pixels to their corresponding superpixels. The weights between pixels and superpixels ω^{XY} were defined in equation (6);

4) E^{XL}: the connections between pixels and labels. Each pixel is linked with all label nodes with the weights ω^{XL}. The prior information of each label can be estimated by the seeds provided by the user. Any cluster algorithms can be chosen to model the seeds for each label, in this paper, we use the Gaussian Mixture Model (GMM) to learn this information. Then for each label $l \in \{1,...L\}$, the GMM with K components is utilized on the set $\{X_l^s, Y_l^s\}$ which consists of pixels and superpixels seeds with the label l. Therefore, we can get K Gaussian parameters set $\theta^l = \{\theta_1^l,...\theta_K^l\}$ and $\theta_k^l = \{v_k^l, \mu_k^l, \Sigma_k^l\}$ ($k \in \{1,...K\}$), where v_k^l, μ_k^l and Σ_k^l denote the mixture coefficient, the mean and the covariance for the *kth* Gaussian component, respectively. The weight ω_{il}^{XL} between the pixel x_i and the label l is defined as follows:

$$\omega_{il}^{XL} = \begin{cases} \max_{k \in \{1,...K\}} v_k^l \dfrac{\exp(-\dfrac{1}{2}[c_i - \mu_k^l]^{\mathrm{T}} (\Sigma_k^l)^{-1}[c_i - \mu_k^l])}{\sqrt{(2\pi)^{\dim} |\Sigma_k^l|}} & \text{if } x_i \notin X_l^s \\ \lambda & \text{otherwise} \end{cases} \qquad (8)$$

where dim is the dimension of c_i (3 for the color image) and λ is a constant with high value which is used to constraint the seeds belonging to the initial labels. Then we normalize the values of ω_{il}^{XL} for all unseeded pixels with the constraint $\sum_{l=1}^{L} \omega_{il}^{XL} = 1$. With these connections, it is possible to utilize the relationships between pixels and labels. Notably, some specific experiments with only one pixel-seed for each label are conducted to test the robustness to users' inputs. In these experiments, the GMM cannot model the seeds, and the weight ω_{il}^{XL} is simply defined as

9

the Euclidean distance between each pixel and the pixel-seed to approximately test the impact with/without the label constraint.

5) E^{YL} : the connections between superpixels and labels. The Gaussian models obtained above are utilized to estimate the relationships between superpixels and labels. The weight ω_{il}^{YL} between the superpixel y_i and the label l is defined as follows:

$$\omega_{il}^{YL} = \begin{cases} \max_{k \in \{1,\dots K\}} v_k^l \dfrac{\exp(-\dfrac{1}{2}[\overline{c}_i - \mu_k^l]^{\mathrm{T}} (\Sigma_k^l)^{-1} [\overline{c}_i - \mu_k^l])}{\sqrt{(2\pi)^{\dim} |\Sigma_k^l|}} & \text{if } y_i \notin Y_l^s \\ \lambda & \text{otherwise} \end{cases} \quad (9)$$

where Y_l^s is the set that consists of superpixel seeds with the label l. The normalization operation is also used for the unseeded superpixels. By these connections, it is possible to transfer the relationships between superpixels and labels. In the one seed experiments, the weight ω_{il}^{YL} is simply defined based on the Euclidean distance between each superpixel and the superpixel-seed to approximately test the impact with/without the label constraint.

3.2 Estimation of probability

Based on the definition of the multi-layer relationships among pixels, superpixels and labels in Section 3.1, we can simultaneously estimate all pixels and superpixels probabilities for each label based on the initial seeds.

We firstly define the cost function of pixels probabilities with respect to a label l as follows:

$$\begin{aligned} D(\pi_l^X) &= \zeta_{l,\mathrm{P}}^X + \zeta_{l,\mathrm{H}}^X + \zeta_{l,\mathrm{R}}^X \\ &= \sum_{i,j=1}^{Nx} \omega_{ij}^X (\pi_{il} - \pi_{jl})^2 + \sum_{i=1}^{Nx} \mu d_i^X (\pi_{il} - \overline{\pi}_{il})^2 + \sum_{i=1}^{Nx} \eta d_i^X (\omega_{il}^{XL} (\pi_{il} - 1)^2 + \sum_{l'=1, l' \neq l}^{L} \omega_{il'}^{XL} \pi_{il}^2) \end{aligned} \quad (10)$$

where the first term $\zeta_{l,\mathrm{P}}^X$ is the pixel-level constraint that two neighboring pixels should have the same label if their color features are similar. The second term $\zeta_{l,\mathrm{H}}^X$ is the soft superpixel consistency that a pixel probability should be similar to its corresponding superpixel probability. The third term $\zeta_{l,\mathrm{R}}^X$ is the constraint between pixels and the label l. If the pixel x_i has a large weight ω_{il}^{XL} with the label l, it should have a high probability π_{il} and this can be realized by the term $\omega_{il}^{XL} (\pi_{il} - 1)^2$. If x_i has a small weight ω_{il}^{XL} with the label l, it should have a low

10

probability π_{il} and this can be realized by the term $\sum_{l'=1,l'\neq l}^{L} \omega_{il'}^{XL} \pi_{il}^2$. The values of d_i^X and $\bar{\pi}_{il}$ have been defined in equation (7), where d_i^X is used to balance the influence with the first term.

The seeds $x_i \in X_l^s$ with the label l are imposed the hard constraint $\pi_{il}=1$ with the definition $\omega_{il}^{XL} = \lambda$ ($\lambda=\infty$) and the other seeds $x_i \in X_{l'}^s$ ($l' \neq l$) are also imposed the hard constraint $\pi_{il}=0$ with the definition $\omega_{il'}^{XL} = \lambda$ ($\lambda=\infty$). The parameter η is used to control the effect of the third term and the function $D(\pi_l^X)$ is degenerated to equation (7) if η is set to 0.

Then we define the cost function of superpixels probabilities with respect to a label l as follows:

$$
\begin{aligned}
D(z_l^Y) &= \zeta_{l,P}^Y + \zeta_{l,S}^Y + \zeta_{l,R}^Y \\
&= \sum_{i,j=1}^{N_Y} \omega_{ij}^Y (z_{il} - z_{jl})^2 + \sum_{i=1}^{N_Y} \varepsilon d_i^Y (z_{il} - \bar{z}_{il})^2 + \sum_{i=1}^{N_Y} \eta d_i^Y (\omega_{il}^{YL} (z_{il}-1)^2 + \sum_{l'=1,l'\neq l}^{L} \omega_{il'}^{YL} z_{il}^2)
\end{aligned}
\tag{11}
$$

where the first term $\zeta_{l,P}^Y$ is the superpixel-level constraint that two superpixels in the full neighborhood system should have the same label if their representative colors are similar. The second term $\zeta_{l,S}^Y$ is the estimated unary constraint that a superpixel probability should be similar to the weighted average of inner pixel probabilities. The third term $\zeta_{l,R}^Y$ is the constraint between superpixels and the label l. If the superpixel y_i has a large weight ω_{il}^{YL} with the label l, it should have a high probability z_{il} and this can be realized by the term $\omega_{il}^{YL} (z_{il}-1)^2$. If y_i has a small weight ω_{il}^{YL} with the label l, it should have a low probability z_{il} and this can be realized by the term $\sum_{l'=1,l'\neq l}^{L} \omega_{il'}^{YL} z_{il}^2$. The values of d_i^Y and $\bar{z}_{il}$ have been defined in equation (7), where d_i^Y is used to balance the influence with the first term. The seeds $y_i \in Y_l^s$ with the label l are imposed the hard constraint $z_{il}=1$ with the definition $\omega_{il}^{YL} = \lambda$ ($\lambda=\infty$) and the other seeds $y_i \in Y_{l'}^s$ ($l' \neq l$) are also imposed the hard constraint $z_{il}=0$ with the definition $\omega_{il'}^{YL} = \lambda$ ($\lambda=\infty$). The parameter η is used to control the effect of the third term and the function $D(z_l^Y)$ is degenerated to equation (7) if η is set to 0.

3.3 Optimization

The two cost functions $D(\pi_l^X)$ in Eq. (10) and $D(z_l^Y)$ in Eq. (11) can be reformulated as the matrix forms with respect to the probabilities $\vec{\pi}_l = [\pi_{il}]_{Nx \times 1}$ and $\vec{z}_l = [z_{il}]_{Ny \times 1}$:

$$D(\Pi_l^X) = \vec{\pi}_l^T (D^X - W^X) \vec{\pi}_l + \mu(\vec{\pi}_l - P^{XY} \vec{z}_l)^T D^X (\vec{\pi}_l - P^{XY} \vec{z}_l) + \eta(\vec{\pi}_l - O^X)^T D^X W_l^{XL} (\vec{\pi}_l - O^X)$$
$$+ \sum_{l'=1, l' \neq l}^{L} \eta \vec{\pi}_l^T D^X W_{l'}^{XL} \vec{\pi}_l \tag{12}$$

$$D(Z_l^Y) = \vec{z}_l^T (D^Y - W^Y) \vec{z}_l + \varepsilon(\vec{z}_l - P^{YX} \vec{\pi}_l)^T D^Y (\vec{z}_l - P^{YX} \vec{\pi}_l) + \eta(\vec{z}_l - O^Y)^T D^Y W_l^{YL} (\vec{z}_l - O^Y)$$
$$+ \sum_{l'=1, l' \neq l}^{L} \eta \vec{z}_l^T D^Y W_{l'}^{YL} \vec{z}_l \tag{13}$$

where $W^X = [\omega_{ij}^X]_{Nx \times Nx}$ and $W^Y = [\omega_{ij}^Y]_{Ny \times Ny}$. The matrixes $P^{XY} = [\rho_{ij}^{XY}]_{Nx \times Ny}$ and $P^{YX} = [\rho_{ij}^{YX}]_{Ny \times Nx}$ represent the relationships between pixels and superpixels. The diagonal elements of the matrixes $D^X = diag([d_1^X, ..., d_{Nx}^X])$ and $D^Y = diag([d_1^Y, ..., d_{Ny}^Y])$ show the degree of W^X and W^Y, respectively. The diagonal elements of the matrixes $W_l^{XL} = diag([\omega_{1l}^{XL}, ..., \omega_{Nxl}^{XL}])$ and $W_l^{YL} = diag([\omega_{1l}^{YL}, ..., \omega_{Nyl}^{YL}])$ are the relationships between pixels/superpixels and the label l. $O^X = [1]_{Nx \times 1}$ and $O^Y = [1]_{Ny \times 1}$.

Because the two cost functions $D(\pi_l^X)$ and $D(z_l^Y)$ are supplementary to each other, they can be solved based on the simultaneous convex optimization method in [29]. By utilizing simultaneous convex optimization, we can obtain the probabilities of pixels and superpixels $\vec{\upsilon}_l = [\vec{\pi}_l; \vec{z}_l]$ as follows (referring to appendix for detailed derivation):

$$\vec{\upsilon}_l = \eta H^{-1} \Omega [W_l^{XL} O^X; W_l^{YL} O^Y] \tag{14}$$

where the matrix $H = I - \Omega\Psi$. $\Psi = \begin{bmatrix} P^X & \mu P^{XY} \\ \varepsilon P^{YX} & P^Y \end{bmatrix}$ and the diagonal matrix

$$\Omega = \begin{bmatrix} \dfrac{1}{(1+\mu)I + \sum_{l=1}^{L} \eta W_l^{XL}} & \dfrac{1}{(1+\varepsilon)I + \sum_{l=1}^{L} \eta W_l^{YL}} \end{bmatrix}. \quad P^X = D^{X-1}W^X \text{ and } P^Y = D^{Y-1}W^Y.$$

From Eq. (14), it can be seen that H is a $(Nx + Ny) \times (Nx + Ny)$ matrix, where Nx is the number of all pixels and Ny is the number of all superpixels. Though optimal solution can be obtained by the simultaneous convex optimization, it is limited to the expensive cost of solving the

inversion of the matrix H.

In this paper, we consider the optimization of Eqs. (12-13) as a parallel game-theoretic decision making problem. The game is played out by a set of decision makers, which in our case corresponds to the probabilities of pixels $\vec{\pi}_l$ and the probabilities of superpixels $\vec{z}_l$. We assume P^1 is the set of strategies of the player 1, and P^2 is the set of strategies of the player 2, then players try to minimize their respective payoff functions $F^i(p^1, p^2)$ until finding the *Nash Equilibrium* (NE) [31, 32] of the system. Formally, a pair of strategies ($\bar{p}^1 \in P^1, \bar{p}^2 \in P^2$) constitutes a NE solution if

$$\forall p^1, p^2 \quad F^1(\bar{p}^1, \bar{p}^2) \leq F^1(p^1, \bar{p}^2) \; ; \; F^2(\bar{p}^1, \bar{p}^2) \leq F^2(\bar{p}^1, p^2) \tag{15}$$

If we move to the NE iteratively by taking t as the time index, this procedure can be viewed as [32]:

$$p_{t+1}^1 = \arg\min_{p^1 \in P^1} F^1(p^1, p_t^2) \; ; \; p_{t+1}^2 = \arg\min_{p^2 \in P^2} F^2(p_t^1, p^2) \tag{16}$$

1) Probabilities for pixels

For the probabilities of pixels $\vec{\pi}_l$ (player 1), we define:

$$\begin{aligned}
F^1(p^1, p^2) &= f_1(\vec{\pi}_l) + \mu f_{21}(\vec{\pi}_l, \vec{z}_l) \\
&= \vec{\pi}_i^T (D^X - W^X)\vec{\pi}_l + \eta(\vec{\pi}_l - O^X)^T D^X W_l^{XL}(\vec{\pi}_l - O^X) + \sum_{l'=1, l' \neq l}^{L} \eta \vec{\pi}_l^T D^X W_{l'}^{XL} \vec{\pi}_l \\
&\quad + \mu\left\{(\vec{\pi}_l - P^{XY}\vec{z}_l)^T D^X (\vec{\pi}_l - P^{XY}\vec{z}_l)\right\}
\end{aligned} \tag{17}$$

Differentiating $F^1(p^1, p^2)$ with respect to $\vec{\pi}_l$, and set to zero, we can get:

$$\frac{\partial F^1(p^1, p^2)}{\partial \vec{\pi}_l} = (D^X + \mu D^X - W^X + \eta D^X \sum_{l=1}^{L} W_l^{XL})\vec{\pi}_l - \eta D^X W_l^{XL} O^X - \mu D^X P^{XY} \vec{z}_l = 0 \tag{18}$$

Then from Eq. (18), we can obtain the value of $\vec{\pi}_l$ as:

$$\vec{\pi}_l = B^{-1}(\eta D^X W_l^{XL} O^X + \mu D^X P^{XY} \vec{z}_l) \tag{19}$$

where $B = D^X + \mu D^X - W^X + \eta D^X \sum_{l=1}^{L} W_l^{XL}$ is a symmetric matrix with size $N_X \times N_X$. The initial value of $\vec{z}_l$ is naturally set as follows:

$$z_{il} = \begin{cases} 1 & \text{if } y_i \in Y_l^s \\ 0 & \text{if } y_i \in Y_{l'}^s (l' \neq l) \\ 1/|L| & \text{otherwise} \end{cases} \tag{20}$$

where $|L|$ denotes the number of labels.

2) Probabilities for superpixels

For the probabilities of superpixels $\vec{z}_l$ (player 2), we define:

$$
\begin{aligned}
F^2(p^1, p^2) &= f_2(\vec{z}_l) + \varepsilon f_{12}(\vec{z}_l, \vec{\pi}_l) \\
&= \vec{z}_l^T(\mathrm{D}^Y - \mathrm{W}^Y)\vec{z}_l + \eta(\vec{z}_l - O^Y)^T \mathrm{D}^Y \mathrm{W}_l^{YL}(\vec{z}_l - O^Y) + \sum_{l'=1, l' \neq l}^{L} \eta \vec{z}_l^T \mathrm{D}^Y \mathrm{W}_{l'}^{YL} \vec{z}_l \qquad (21) \\
&\quad + \varepsilon\left\{(\vec{z}_l - \mathrm{P}^{YX}\vec{\pi}_l)^T \mathrm{D}^Y(\vec{z}_l - \mathrm{P}^{YX}\vec{\pi}_l)\right\}
\end{aligned}
$$

Differentiating $F^2(p^1, p^2)$ with respect to $\vec{z}_l$, and set to zero, we can get:

$$
\frac{\partial F^2(p^1, p^2)}{\partial \vec{z}_l} = (\mathrm{D}^Y + \varepsilon \mathrm{D}^Y - \mathrm{W}^Y + \eta \mathrm{D}^Y \sum_{l=1}^{L} \mathrm{W}_l^{YL})\vec{z}_l - \eta \mathrm{D}^Y \mathrm{W}_l^{YL} O^Y - \varepsilon \mathrm{D}^Y \mathrm{P}^{YX} \vec{\pi}_l = 0 \quad (22)
$$

Then From Eq. (22), we can obtain the value of $\vec{z}_l$ as:

$$
\vec{z}_l = \mathrm{C}^{-1}(\eta \mathrm{D}^Y \mathrm{W}_l^{YL} O^Y + \varepsilon \mathrm{D}^Y \mathrm{P}^{YX} \vec{\pi}_l) \qquad (23)
$$

where $C = \mathrm{D}^Y + \varepsilon \mathrm{D}^Y - \mathrm{W}^Y + \eta \mathrm{D}^Y \sum_{l=1}^{L} \mathrm{W}_l^{YL}$ is a symmetric matrix with size $N_Y \times N_Y$.

3) Convergence analysis

As pointed in [32], under the following constraints on F^1 and F^2:

(1) F^i is bounded in $p^i \in P^i$.

(2) F^i is continuously second order differentiable in $p^i \in P^i$.

(3) $\exists$ a closed neighborhood $u^i \subseteq P^i$ such that F^i is strongly convex in u^i.

There exists a locally stable NE solution, for any $p^1 \in U^1 \subseteq P^1$ and $p^2 \in U^2 \subseteq P^2$, the sequence of rational choices generated by the parallel decision making process converges and the limit point is a NE solution if the two parameters μ and ε corresponding to f_{21} and f_{12}, respectively, satisfy the following condition:

$$
\begin{aligned}
\tau = &\left\|\left[\left(\mu^{-1}\frac{\partial^2}{\partial p^1 \partial p^1}f_1(p^1) + \frac{\partial^2}{\partial p^1 \partial p^1}f_{21}(p^1, p^2)\right)^{-1}\left(\frac{\partial^2}{\partial p^1 \partial p^2}f_{21}(p^1, p^2)\right)\right] \right. \\
&\left. \times \left[\left(\varepsilon^{-1}\frac{\partial^2}{\partial p^2 \partial p^2}f_2(p^2) + \frac{\partial^2}{\partial p^2 \partial p^2}f_{12}(p^1, p^2)\right)^{-1}\left(\frac{\partial^2}{\partial p^2 \partial p^1}f_{12}(p^1, p^2)\right)\right]\right\| < 1
\end{aligned} \qquad (24)
$$

It can be seen that the proposed convex functions F^1 in Eq. (17) and F^2 in Eq. (21) satisfy the above three constraints. Then for any pixel i and its corresponding superpixel j, we compute the condition coefficient τ as follows:

14

$$\tau = \left\| \left[(\mu^{-1}(2d_i^X + 2\eta d_i^X \sum_{l=1}^{L} \omega_{il}^{XL}) + 2d_i^X)^{-1}(-2d_i^X \rho_{ij}^{XY}) \right] \right.$$
$$\left. \times \left[(\varepsilon^{-1}(2d_j^Y + 2\eta d_j^Y \sum_{l=1}^{L} \omega_{jl}^{YL}) + 2d_j^Y)^{-1}(-2d_j^Y \rho_{ji}^{YX}) \right] \right\|$$

(25)

$$= \left\| \frac{\rho_{ij}^{XY}}{\mu^{-1}(1 + \eta \sum_{l=1}^{L} \omega_{il}^{XL}) + 1} \times \frac{\rho_{ji}^{YX}}{\varepsilon^{-1}(1 + \eta \sum_{l=1}^{L} \omega_{jl}^{YL}) + 1} \right\|$$

Due to the fact that $\mu > 0$, $\eta > 0$, $\varepsilon > 0$, $\omega_{il}^{XL} \geq 0$, $\omega_{jl}^{YL} \geq 0$, $0 < \rho_{ij}^{XY} \leq 1$, $0 < \rho_{ji}^{YX} \leq 1$, we can get $\tau < 1$. So we can proof that the proposed optimization process converges and the limit point is a NE solution.

The proposed optimization algorithm works iteratively until convergence with the initial value of $\vec{z}_l$ in Eq. (20) which corresponds to the seeds. Compared with the simultaneous convex optimization, the advantage of the proposed method is the lower computational complexity while keeping almost the same result. As we mentioned before, the main time cost of the simultaneous convex optimization focuses on solving the inversion of the matrix H in Eq. (14) with the size $(N_X + N_Y) \times (N_X + N_Y)$, where N_X is the number of all pixels and N_Y is the number of all superpixels. The time complexity of solving the inversion of H is $O((N_X + N_Y)^3)$. The main time cost of the proposed method focuses on solving the inversion of the matrix B in Eq. (19) with size $N_X \times N_X$. It can be seen that B is a symmetric matrix ($B = B^T$), so its inverse matrix B^{-1} is also a symmetric matrix ($B^{-1} = (B^{-1})^T$). The time complexity of solving the inversion of B is $O(N_X^3 / 8)$. Though it is an iterative strategy of the proposed method, we can find that B is not associated with $\vec{z}_l$ from Eq. (19). Therefor B can be regarded as a fixed matrix and we only need to compute its inverse matrix once in all iterative steps. So the proposed method has a lower time complexity than the simultaneous convex optimization theoretically.

Consequently, the proposed algorithm for image segmentation can be summarized as follows:

Step 1. Initialization

Initializing the seeds, the size of the neighborhood, the controlling parameters (h_s, h_r) of mean shift algorithm to generate multiple over-segmentations, the coefficients μ, ε and η to control the relationships of pixel-superpixel and pixel/superpixel-label, the constants β and λ. Based on the seeds information, initializing the value of $\vec{z}_l$.

15

Step 2. Constructing multi-layer graph

- Step 2.1 Computing the relationships of pixels with Eq. (1).
- Step 2.2 Computing the relationships of superpixels with Eq. (5).
- Step 2.3 Computing the relationships between pixels and superpixels with Eq. (6).
- Step 2.4 Computing the relationships between pixels and labels with Eq. (8).
- Step 2.5 Computing the relationships between superpixels and labels with Eq. (9).

Step 3. Constructing segmentation model

- Step 3.1 Estimating the probabilities between pixels and labels with Eq. (10).
- Step 3.2 Estimating the probabilities between superpixels and labels with Eq. (11).

Step 4. Optimization

- Step 4.1 Updating the probabilities of pixels $\vec{\pi}_l$ with Eq. (19).
- Step 4.2 Updating the probabilities of superpixels $\vec{z}_l$ with Eq. (23).

Step 5. Checking the termination condition

- Step 5.1 Computing the current labeling: $f^t = \arg\max_l \pi_l$.

- Step 5.2 If f^t equals f^{t-1}, stop the iteration; otherwise, go to Step 4.

4. Experiments

Following [25], the multiple over-segmentations (superpixels) were generated by varying the parameters of the mean shift algorithm [30]. Two parameters (h_s, h_r) need to be manually set in the mean shift algorithm, where h_s is the scale in the spatial domain and h_r is the scale in the range domain. In our experiments, we respectively set $(h_s, h_r) = \{(10,7), (10,10), (10,15)\}$, and hence got three over-segmentation results as our superpixels. The constant β in equations (1) (5) and (6) was set to 60 and λ in equations (8) and (9) was set to a high value ($=10^5$) to impose each seed belonging to its initial label. The coefficients μ, ε and η are used to control the degrees in equations (10) and (11). We experimentally set these three coefficients μ, ε and η as 0.002, 0.2 and 0.001, respectively. The 8-neighborhood system was used for the pixel-level relationships.

We compared the performance of our method with the state-of-the-art methods [7] [11] [29] on the Berkeley segmentation data set [33] which consists of 500 natural images and Microsoft GrabCut database [34] which consists of 50 images with trimaps and ground truth segmentations. The "Lasso" form of trimaps provided by the Microsoft GrabCut database [34] contain the public seeds information, where pixels with gray value 255/64 correspond to the foreground/background seeds, respectively (examples were shown in the first row of Fig. 6). The Microsoft GrabCut database can evaluate the problem of extracting accurate object details, which has been widely

used for quantitative segmentation comparison with state-of-the-art methods. The error rates are utilized as the measurement of accuracy for segmentation, which is defined as the ratio of the number of wrongly labeled pixels to the total number of unlabeled pixels. The misclassified pixels are identified by the difference between the ground truth images and the segmentation results.

For a clear description, Table 1 lists all the parameter settings for the comparable methods, including random walk (RW) [11], extended graph cuts (GrabCut) [7] and nonparametric higher-order (NHO) [29]. Notably, graph cuts based methods can only solve the two-class segmentation problem. In the multi-class segmentation experiments, we implement the α-expansion algorithm [35] in GrabCut [7] to solve the multi-class problem.

Table 1 Conclusion of parameters for random Walk (RW) [11], GrabCut [7] and nonparametric higher-order (NHO) [29].

Method	Parameter	Value
RW [11]	The constant β in Eq. (1)	90
GrabCut [7]	The regularization parameter	50
	The number of Gaussian components	5
NHO [29]	(h_s, h_r) of mean shift algorithm [30]	$\{(10,7),(10,10),(10,15)\}$
	β in Eqs. (1, 5, 6)	60
	λ in Eq. (7)	10^5
	μ in Eq. (7)	0.002
	ε in Eq. (7)	0.2

We firstly demonstrated the quality of our method on the Microsoft GrabCut database [34]. Table 2 summarizes the error rates achieved by various methods. The mean, standard deviation (Std) and the average rank after the Friedman statistical test [38-39] (with a significance level of 0.05) of error rates for RW [11], GrabCut [7], NHO [29], the proposed segmentation energy function with the simultaneous convex optimization [29] (SCO), and our method are obtained from our implementation, and the average error rates of others are quoted directly from the best results reported in literature. Comparing with the state-of-the-art methods, we can observe that SCO and our method achieve very low error rate, which proofs the effectiveness of the proposed multi-layer graph based segmentation model. Comparing the results between SCO and our method, it can be seen that almost the same results are obtained, which proofs the effectiveness of the proposed optimization algorithm.

Table 2 Comparison of error rates of the proposed method and other methods with the Microsoft GrabCut database. Mean $\pm$ standard deviation (Std) (%) and the average rank after the Friedman statistical test [38-39] (with a

significance level of 0.05) of error rates for RW [11], GrabCut [7], NHO [29], SCO, and our method are obtained from our implementation, and the average error rates (%) of other methods are quoted directly from the best results reported in literature.

Methods	Error rate	
	Mean ± Std	Average rank
RW [11]	6.45 ± 4.8	4.48
GrabCut [7]	5.46 ± 4.2	3.66
NHO [29]	4.25 ± 3.7	2.94
SCO	**3.43 ± 2.8**	1.93
Our method	**3.44 ± 2.9**	1.99
GMMRF [36]	7.90 (reported in [13])	
Robust P^n model [25]	6.08 (reported in [29])	
LazySnapping [22]	6.65 (reported in [9])	
Constrained Random Walk [13]	4.08 (reported in [13])	
Convex Active Contours [37]	3.77 (reported in [9])	
Texture Aware model [9]	3.64 (reported in [9])	

The Friedman test [38-39] was conducted to compare the performances of multiple methods. This test exposes the average rank of each individual method as the output. In this statistical test, null hypothesis, H_0 affirms the equal behavior of the comparable methods. Hence, under H_0, each method possesses equal rank, which confirms that each method is equally efficient with others. The alternative hypothesis, H_1 endorses the difference in performances among the comparable methods. The average ranks for RW [11], GrabCut [7], NHO [29], SCO, and our method are listed in Table 2 as 4.48, 3.66, 2.94, 1.93 and 1.99, respectively. Moreover, Friedman test determines the chi-square (χ^2) value as 99.16 and the p-value as 1.48e-20. From chi-square (χ^2) distribution table, we find that the critical value for $(5-1) = 4$ degree of freedom with 0.05 significance level is 9.49. Since the chi-square value is greater than the critical value, H_0 is rejected and H_1 is accepted. Furthermore, the small p-value (close to zero) validates the rejection of H_0 and confirms the acceptance of H_1, which substantiates the significant difference in behavior among the comparable methods. Fig. 5 shows the comparison of the error rate for each image by applying RW [11] (green line), GrabCut [7] (blue line), NHO [29] (cyan line) and our method (red line). It should be noted that all these results are obtained based on the same seeds provided by the Microsoft GrabCut database, so it is reasonable to compare with the methods reported in literature.
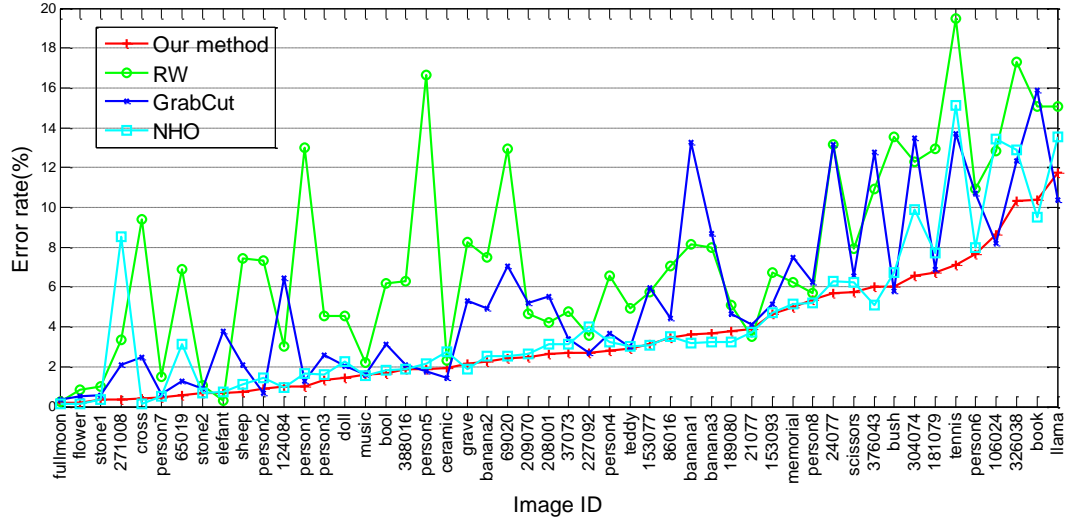
18

Fig. 5 Comparison of the error rate of each image in the Microsoft GrabCut database [34] of RW [11] (green line), GrabCut [7] (blue line), NHO [29] (cyan line) and our method (red line).

Fig. 6 illustrates the example segmentations on the Microsoft GrabCut database, which shows that our method produces high-quality segmentation results. For example, our method can detect the legs of the sheep with a few seeds in the first column. In the second column, we can find that the arms of the people can be well detected and the segmentation boundary is also very smooth in our method. Figs. 7-8 also show the segmentation results and error rates on the Microsoft GrabCut database by using our method. These quantitative and qualitative comparisons confirm the accuracy of our method.
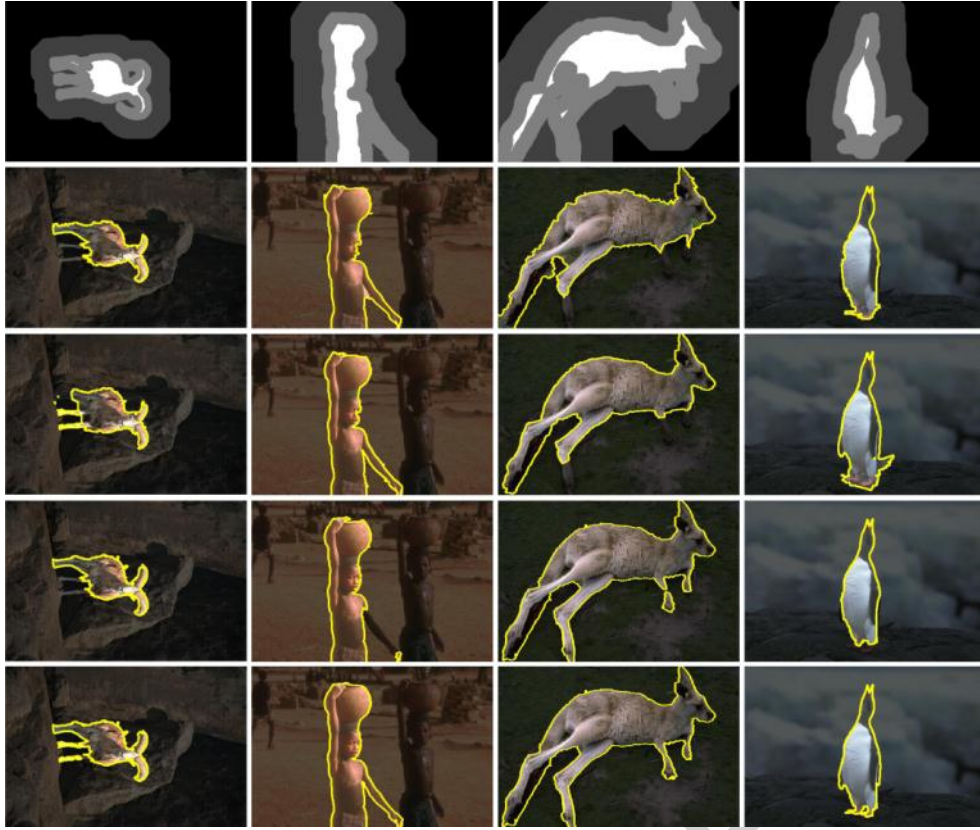
19

Fig. 6 Example segmentations using the trimaps on the Microsoft GrabCut database. Row 1st shows the test images with trimaps. Rows 2nd -5th show segmentation results obtained by RW [11], GrabCut [7], NHO [29] and our method, respectively.

Fig. 7 The segmentation results and error rates on the Microsoft GrabCut database using our method.

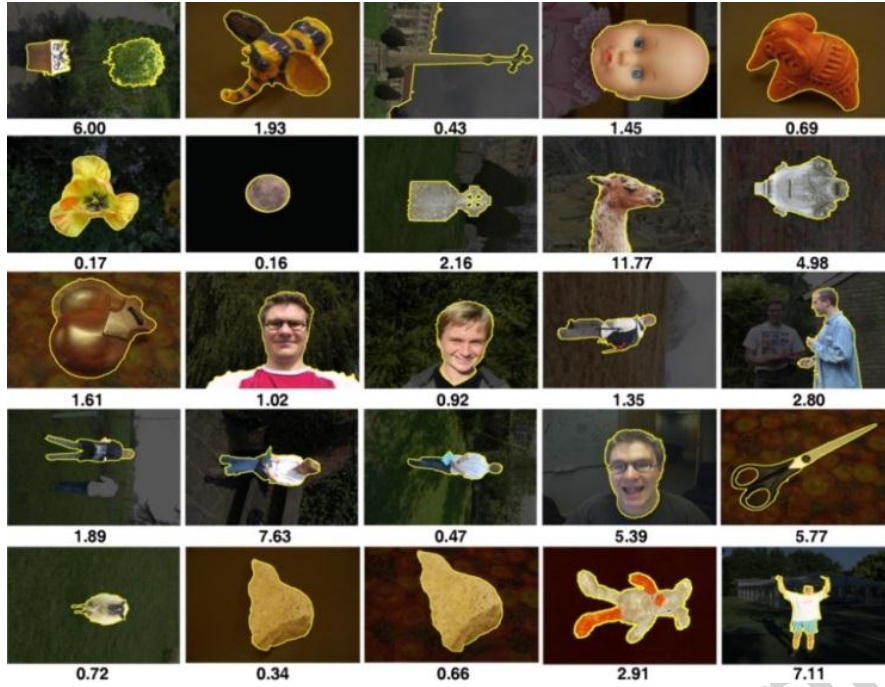| 6.00 | 1.93 | 0.43 | 1.45 | 0.69 |
0.17	0.16	2.16	11.77	4.98
1.61	1.02	0.92	1.35	2.80
1.89	7.63	0.47	5.39	5.77
0.72	0.34	0.66	2.91	7.11

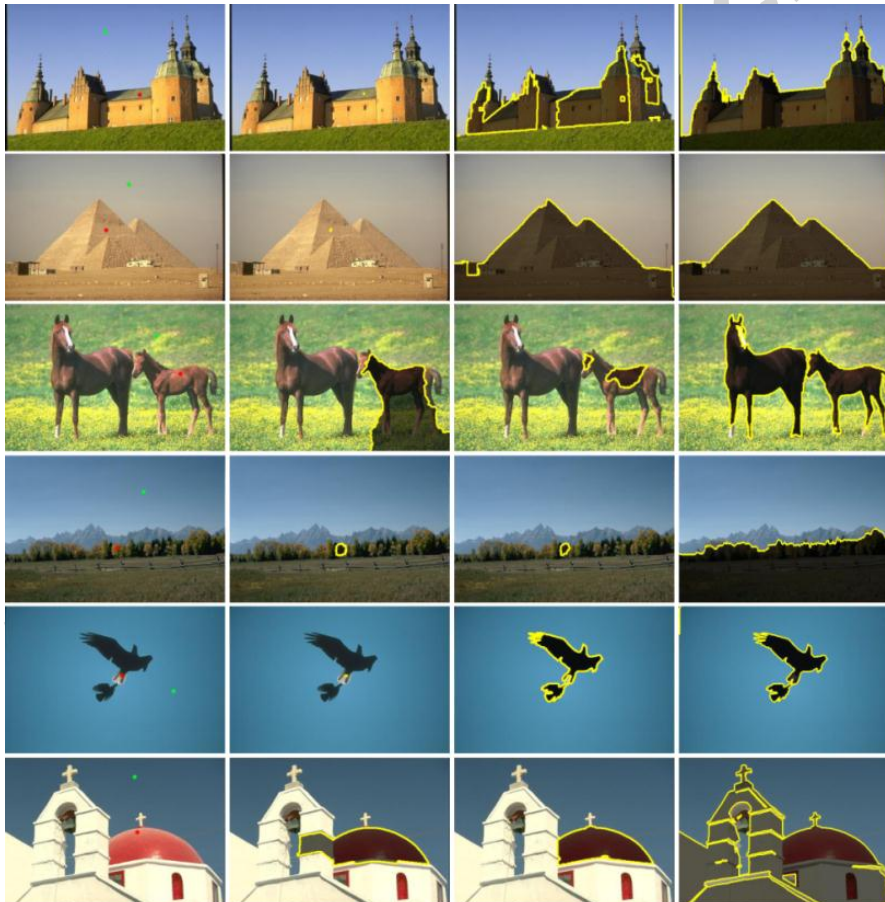Fig. 8 The segmentation results and error rates on the Microsoft GrabCut database using our method.

Fig. 9 Comparison of the segmentation results with only one pixel-seed for each label. The first column shows original images with one foreground pixel-seed (red) and one background pixel-seed (green). Columns 2-4 show the segmentation results by random walk [11] with pixel-level information, nonparametric higher-order [29] with pixel-superpixel information, and our method with multi-layer combination information.

21

(a)　　　(b)　　　(c)　　　(d)　　　(e)

Fig. 10 Comparison of our method with the state-of-the-art methods for two-class segmentation with a few scribbles. (a) Input images with scribbles for foreground (red line) and background (blue line). Segmentation results by (b) RW [11], (c) GrabCut [7], (d) NHO [29] and (e) our method.

Figs. 9-11 present the comparison of our method with the state-of-the-art interactive segmentation methods on the natural images selected from the Berkeley segmentation data set [33]. Figs. 9-10 show the two-class segmentation results. In Fig. 9, only one pixel-seed is selected for each label, where the first column shows the original images with one foreground pixel-seed (red) and one background pixel-seed (green), columns 2-4 show the segmentation results by applying RW [11] with pixel-level information, NHO [29] with pixel-superpixel information, and our method with multi-layer combination information. It can be seen that our method can obtain more satisfactory results even with one pixel-seed. The two-class segmentation with a few scribbles is shown in Fig. 10, where (a) is the input images with scribbles for foreground (red line) and background (blue line), and (b)-(e) are the results of RW [11], GrabCut [7], NHO [29] and our method, respectively. Fig. 11 shows the multi-class segmentations of the four methods, where (a)

is the input images with scribbles representing the red, green and blue labels. (b)-(e) are the segmentation results applied by RW [11], GrabCut [7], NHO [29] and our method, respectively. From (b), we can find that RW [11] gets bad segmentation results due to lacking enough seeds. Compared with GrabCut [7] in (c), NHO [29] obtains more satisfactory results even with a few seeds due to the utilizing of the superpixel cues. However, the results of NHO are not accurate enough especially for slender objects. Our method gets high-quality segmentation results with accurate object details, which benefits from the superpixel and label learning.

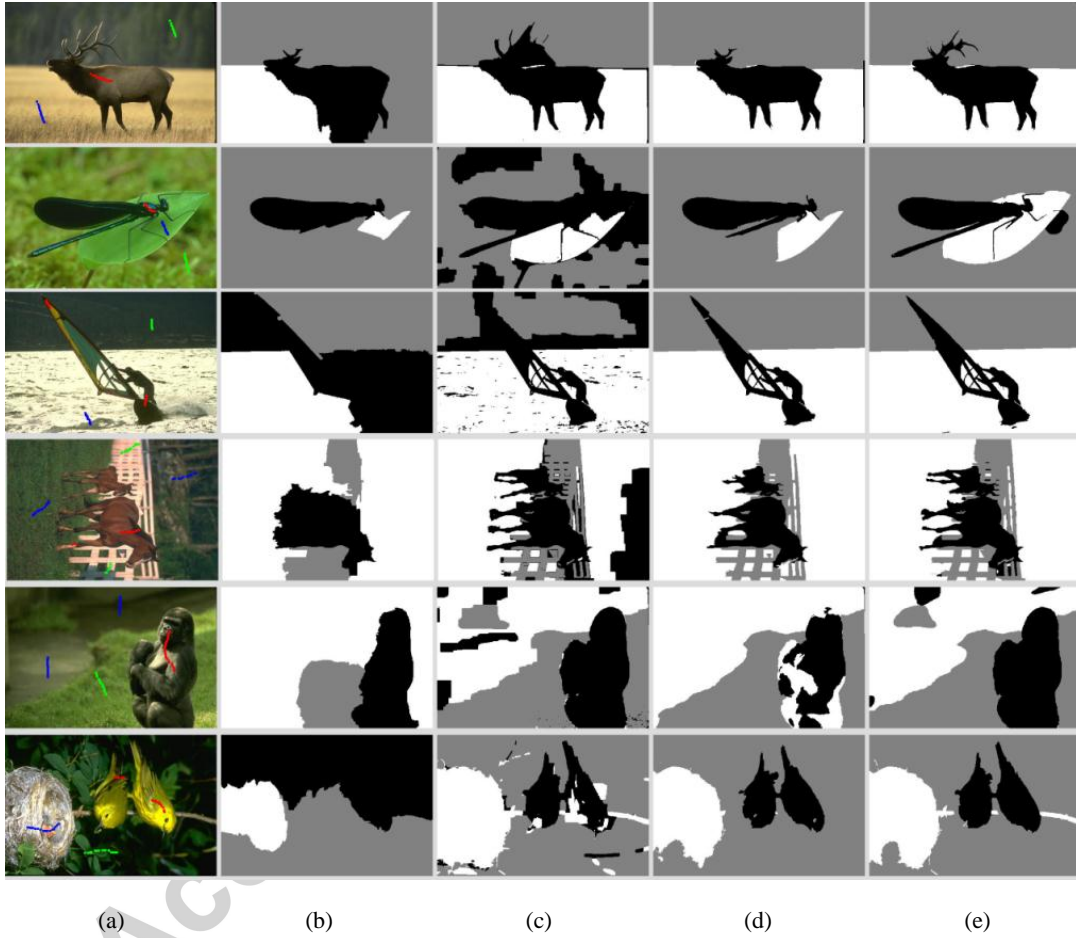|				
(a)	(b)	(c)	(d)	(e)

Fig. 11 Comparison of our method with the state-of-the-art methods for multi-class segmentation with a few scribbles. (a) Input images with scribbles representing the red, green and blue labels. (b)-(e) are the segmentation results applied by RW [11], GrabCut [7], NHO [29] and our method, respectively.

5. Discussion

5.1 Parameter Settings

The parameters μ, ε and η are used to control the influence of each term in Eqs. (10, 11). The parameter K in Eqs. (8, 9) represents the number of the Gaussian components in the GMM. In this section, we analyzed the effects on the segmentation results when varying these parameters.

Fig. 12 shows the segmentation results by varying the parameter η. With a larger η, the relationship between pixel/superpixel and label plays a more important part and the details in the objects can be preserved. However, as shown in (d), the boundaries are not smooth enough and it is hard to provide satisfactory segmentations. Comparatively smoother boundaries can be extracted with smaller η. However, as shown in (b), the segmentations may be over-smooth and the details around boundaries cannot be preserved well. So it is important to find an appropriate η to improve the accurate boundaries and reduce the over-smooth effect. The best segmentation result can be obtained when $\eta = 0.001$, which is shown in (c).
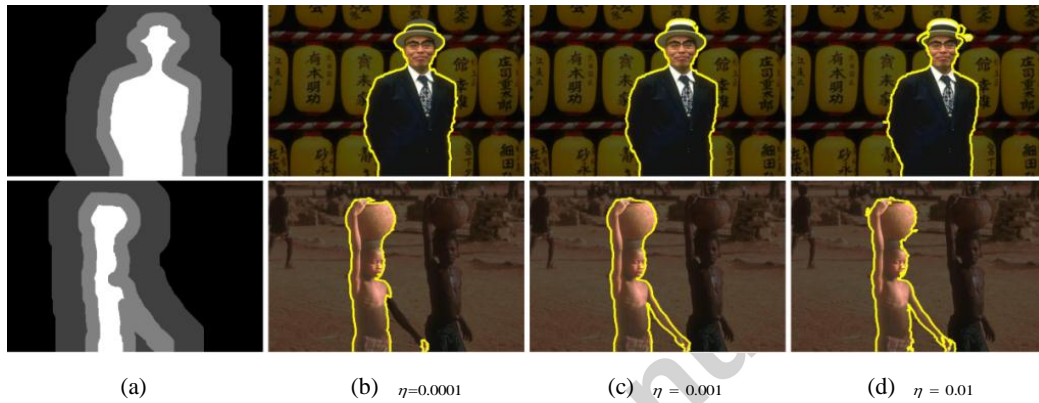

(a)   (b) $\eta=0.0001$ (c) $\eta = 0.001$ (d) $\eta = 0.01$

Fig. 12 Examples of the segmentations with respect to the variation of η in Eqs. (10-11). From the test images (a) with trimaps, (b)-(d) give the resulting segmentations according to η.

Table 3 Mean $\pm$ standard deviation (Std) (%) and the average rank after the Friedman statistical test [38-39] (with a significance level of 0.05) of error rates over all 50 images in the Microsoft GrabCut database with different parameters η, μ, ε and K.

	Test Values	0.0001	0.0005	0.001	0.005	0.01
η	Mean $\pm$ Std	4.15 ± 3.7	3.95 ± 3.6	3.44 ± 2.9	4.76 ± 3.7	5.71 ± 4.1
	Average rank	2.91	2.57	2.02	3.24	4.26
	Test Values	0.0002	0.001	0.002	0.01	0.02
μ	Mean $\pm$ Std	3.64 ± 2.8	3.47 ± 2.8	3.44 ± 2.9	4.05 ± 3.5	4.42 ± 3.8
	Average rank	3.50	2.73	2.41	2.79	3.57
	Test Values	0.02	0.1	0.2	0.5	1
ε	Mean $\pm$ Std	3.55 ± 2.8	3.45 ± 2.8	3.44 ± 2.9	3.45 ± 2.9	3.47 ± 2.9
	Average rank	3.75	3.10	2.63	2.70	2.82
	Test Values	2	3	4	5	
K	Mean $\pm$ Std	3.45 ± 2.9	3.44 ± 2.9	3.45 ± 2.9	3.55 ± 2.9	
	Average rank	2.53	2.64	2.59	2.24	

Table 3 shows the quantitative comparisons of the mean, standard deviation (Std) and the average rank after the Friedman test [38-39] (with a significance level of 0.05) of error rates with

different η, μ, ε and K over all 50 images in the Microsoft GrabCut database [34]. For the parameters η, μ and ε, Friedman tests determine the chi-square (χ^2) values as 56.37, 20.86 and 17.35, respectively, and the p-values as 1.68e-11, 3.37e-04 and 1.70e-03, respectively. Besides, at 0.05 significance level and $(5-1)=4$ degree of freedom, the chi-square values are larger than the critical value. Hence these two measures determine that H_0 is rejected and H_1 is accepted, which shows that the segmentations are sensitive to the parameters η and μ, and a little sensitive to the parameter ε. For the parameter K, Friedman test determines the chi-square (χ^2) value as 3.29 and the p-value as 0.35. From chi-square (χ^2) distribution table, we find that the critical value for $(4-1)=3$ degree of freedom with 0.05 significance level is 7.81. Since the chi-square value is smaller than the critical value, H_1 is rejected and H_0 is accepted, which shows that the segmentations are not sensitive to the parameter K. Furthermore, it can also be found that the lowest average error rates and average ranks are both obtained when $\eta=0.001$, $\mu=0.002$ and $\varepsilon=0.2$. In this paper, we experimentally set the parameters η, μ, ε and K to 0.001, 0.002, 0.2 and 3, respectively.

5.2 Sensitivity analysis

Similar to the evaluation of Kim et al. [29], we analyzed the sensitivity of our method with respect to seed quantity and placement. The standard segmentations were produced from the initial trimaps provided by the Microsoft GrabCut database. Then some seeds were randomly taken from 0.5% to 50% of total seed quantity. The perturbed segmentations were recomputed from these selected seeds and compared with the standard segmentations. The normalized overlap $a_o = \left| \dfrac{F_1 \bigcap F_2}{F_1 \bigcup F_2} \right|$ was used to measure the similarity of two segmentations, where F_1 and F_2 indicate the sets of pixels assigned as foreground in two segmentations. Table 4 shows the sensitivity check results, where column 1st is the test image IDs in the Microsoft GrabCut database and columns 2nd-6th show the normalized overlap a_o by varying the seed quantity as 50%, 30%, 10%, 1%, 0.5%, respectively, together with the number of pixel seeds (foreground/background). Mean and standard deviation (Std) of a_o when varying the seed quantity is shown in the last column. Fig. 13 shows the qualitative comparison of the standard segmentations with full seeds (Rows 1st and 3rd) and the perturbed segmentations with 0.5% seeds (Rows 2nd and 4th). From

these quantitative and qualitative results, it can be seen that the proposed method has a strong robustness to the seed quantity and placement, and can still provide robust segmentations even with 0.5% seeds.

Table 4 Sensitivity check with respect to seed quantity and location. Column 1st shows the test image IDs in the Microsoft GrabCut database. The normalized overlap a_o is computed to measure the similarity of two segmentations by varying the seed quantity as 0.5%, 1%, 10%, 30%, 50% of total seed quantity. The number of pixel seeds (foreground/background) is also shown together with the normalized overlap a_o. The last column shows the mean and standard deviation (Std) of a_o when varying the seed quantity.

Image ID	50%	30%	10%	1%	0.5%	Mean ± Std
37073	0.999 (9.4e03/2.0e04)	0.998 (5.6e03/1.2e04)	0.995 (1.9e03/3.9e03)	0.976 (188/394)	0.975 (94/197)	0.988 ± 0.011
65019	0.999 (1.3e04/2.2e04)	0.999 (8.0e03/1.3e04)	0.999 (2.7e03/4.4e03)	0.994 (266/438)	0.993 (133/219)	0.997 ± 0.003
69020	0.999 (9.4e03/3.3e04)	0.999 (5.6e03/2.0e04)	0.998 (1.9e03/6.5e03)	0.983 (188/654)	0.965 (94/327)	0.988 ± 0.015
86016	0.999 (1.1e04/9.4e03)	0.999 (6.7e03/5.6e03)	0.998 (2.2e03/1.9e03)	0.998 (222/188)	0.997 (111/94)	0.998 ± 0.001
124084	0.999 (2.4e04/2.9e04)	0.999 (1.4e04/1.7e04)	0.999 (4.7e03/5.8e03)	0.998 (474/576)	0.998 (237/288)	0.999 ± 7.5e-04
153077	0.999 (1.3e04/2.6e04)	0.996 (7.7e03/1.5e04)	0.990 (2.6e03/5.2e03)	0.968 (258/516)	0.940 (129/258)	0.978 ± 0.024
153093	0.999 (3.7e03/2.4e04)	0.999 (2.2e03/1.5e04)	0.998 (740/4.9e03)	0.984 (74/486)	0.972 (37/243)	0.991 ± 0.012
189080	0.999 (3.0e04/2.3e04)	0.999 (1.8e04/1.4e04)	0.997 (6.1e03/4.6e03)	0.997 (608/464)	0.995 (304/232)	0.997 ± 0.001
227092	0.999 (2.2e04/3.3e04)	0.999 (1.3e04/2.0e04)	0.999 (4.5e03/6.6e03)	0.999 (448/658)	0.989 (224/329)	0.997 ± 0.004
388016	0.999 (3.7e03/4.1e04)	0.999 (2.2e03/2.5e04)	0.999 (740/8.3e03)	0.996 (74/826)	0.994 (37/413)	0.998 ± 0.002

Fig. 14 shows the segmentation results of our method with only one pixel-seed for each label. Columns 1st and 3rd are the original images with the red seed for foreground and the green seed for background. From columns 2nd and 4th, it can be seen that our method can still obtain satisfactory results with only one pixel-seed. Consequently, these experiments show that the proposed method has a good robustness to user-inputs.
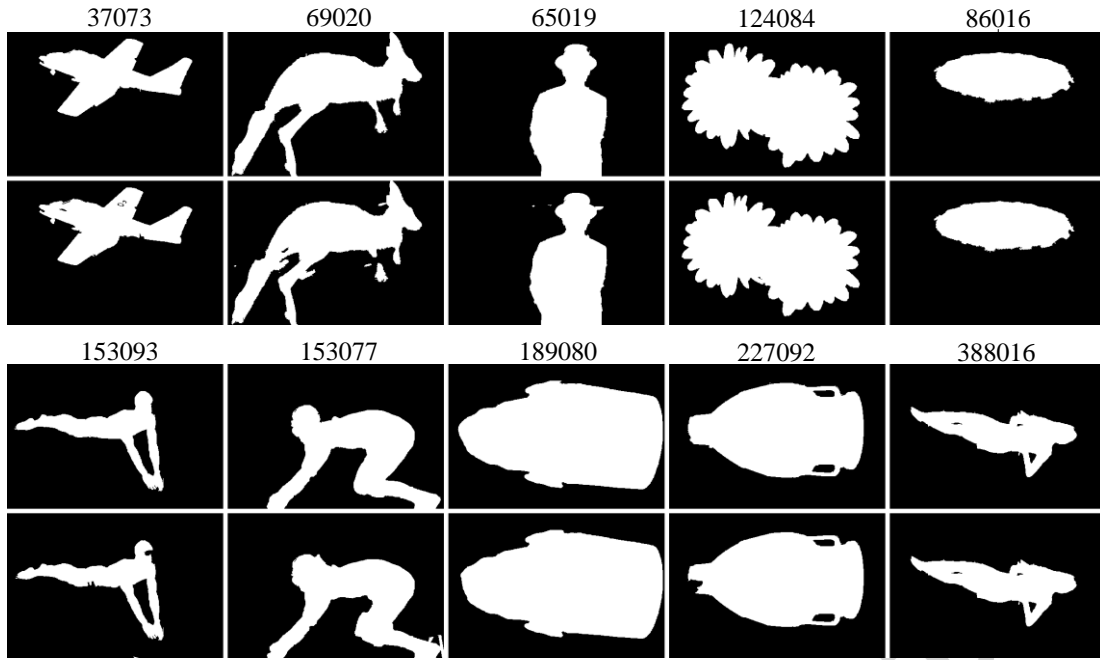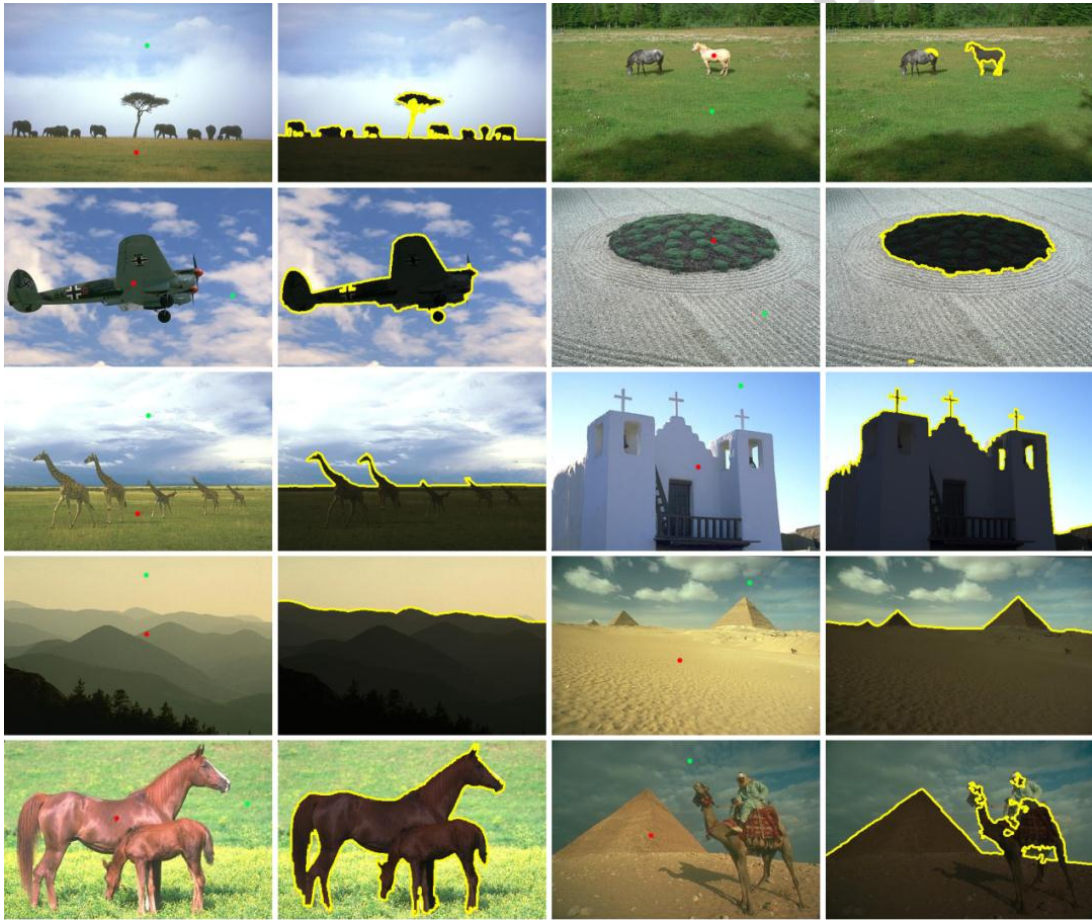
Fig. 13 Comparison of the standard segmentations with full seeds (Rows 1st and 3rd) and the perturbed segmentations with 0.5% seeds (Rows 2nd and 4th).

Fig. 14 The segmentation results of the proposed method with only one pixel-seed for each label. Columns 1st and 3rd are the original images with two pixel-seeds, where one pixel seed for foreground (red) and the other for background (green). Columns 2nd and 4th are the corresponding segmentation results.

5.3 Complexity consideration

The proposed algorithm works iteratively until convergence. The iteration step for each image in the Microsoft GrabCut database is shown in Fig. 15. The average iteration step of all the 50 test images is 6.44. Fig. 16 shows the comparison of running times for RW [11], GrabCut [7], NHO [29], SCO and our method on 20 test images with size 321×481 or 481×321 in the Microsoft GrabCut database on an Intel Xeon CPU running at 2.0 GHz in MATLAB. The time cost of NHO [29], SCO, and our method does not include the over-segmentation step which takes about 4.0 s for the mean shift algorithm [30] to generate three over-segmentations. It can be seen that the time costs of NHO, SCO, and our method are larger than the pixel-level based methods due to the utilizing of superpixel. The cost of SCO is a bit larger than NHO because of the computation of the relationship between pixel/superpixel and label. Our method has a lower time cost than SCO and NHO due to the effectiveness of the proposed optimization method based on game theory. As discussed above, the main time cost of these three algorithms focuses on solving the inversion of a large matrix. The time complexity of the proposed algorithm is $O(N_X^3 / 8)$ which is lower than those of the other two algorithms, i.e. $O((N_X + N_Y)^3)$. The average running times of the above five algorithms are shown in Table 5.
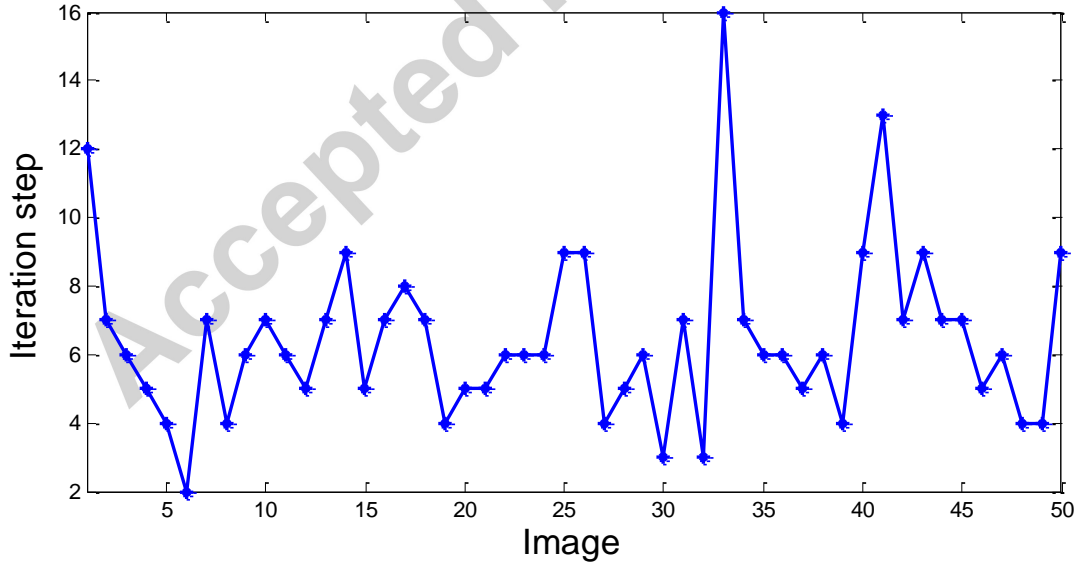

Fig. 15 The iteration step of the proposed optimization algorithm for each image in the Microsoft GrabCut database.
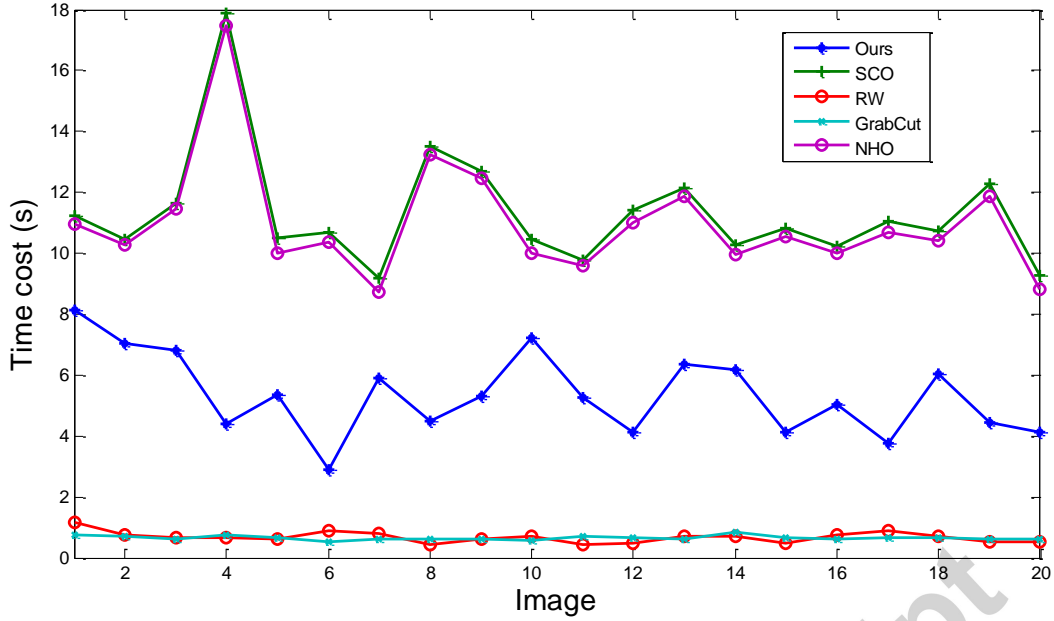
28

Fig. 16 Running times of RW [11], GrabCut [7], NHO [29], SCO, and our method on 20 test images in the Microsoft GrabCut database with size 321×481 or 481×321.

Table 5 Average running times of RW [11], GrabCut [7], NHO [29], SCO, and our method on 20 test images in the Microsoft GrabCut database with size 321×481 or 481×321.

	RW	GrabCut	NHO	SCO	Ours
Time (s)	0.67 ± 0.18	0.65 ± 0.06	10.98 ± 1.90	11.32 ± 1.90	5.35 ± 1.34

6. Conclusion

In this paper, we have proposed an interactive image segmentation method by fusing the multi-layer connections among pixel-layer, superpixel-layer and label-layer together. The relationships between pixels/superpixels and labels are utilized as the label constraints to further improve the segmentation accuracy. Then for the optimization of the proposed segmentation energy function, we update the probabilities of pixels and superpixels iteratively until convergence based on game theory. The performance of our method is tested on the challenging data sets. The segmentation results demonstrate that our method can produce high-quality segmentations with accurate object details. However, several adjusting parameters need to be manually set based on the experimental results, which would make the proposed algorithm less robust for the practical applications. Therefore, our future work will focus on how to automatically get the adaptive optimal parameters for each image.

Appendix

1. Derivation for Eq. (14)

Differentiating $D(\Pi_l^X)$ in Eq. (12) and $D(Z_l^Y)$ in Eq. (13) with respect to $\vec{\pi}_l$ and $\vec{z}_l$, respectively, and set to zero, we can obtain all the probabilities of pixels and superpixels $\vec{\upsilon}_l = [\vec{\pi}_l; \vec{z}_l]$ belonging to the label l:

$$\frac{\partial D(\Pi_l^X)}{\partial \vec{\pi}_l} = \vec{\pi}_l - P^X \vec{\pi}_l + \mu(\vec{\pi}_l - P^{XY} \vec{z}_l) + \eta W_l^{XL}(\vec{\pi}_l - O^X) + \sum_{l'=1, l' \neq l}^{L} \eta W_{l'}^{XL} \vec{\pi}_l = 0 \qquad (26)$$

$$\frac{\partial D(Z_l^Y)}{\partial \vec{z}_l} = \vec{z}_l - P^Y \vec{z}_l + \varepsilon(\vec{z}_l - P^{YX} \vec{\pi}_l) + \eta W_l^{YL}(\vec{z}_l - O^Y) + \sum_{l'=1, l' \neq l}^{L} \eta W_{l'}^{YL} \vec{z}_l = 0 \qquad (27)$$

where $P^X = D^{X-1} W^X$ and $P^Y = D^{Y-1} W^Y$. It can be jointly transformed into

$$\vec{\upsilon}_l = \Omega \Psi \vec{\upsilon}_l + \eta \Omega \left[W_l^{XL} O^X ; W_l^{YL} O^Y \right] \qquad (28)$$

or simply

$$(I - \Omega \Psi)\vec{\upsilon}_l = \eta \Omega \left[W_l^{XL} O^X ; W_l^{YL} O^Y \right] \qquad (29)$$

where the matrix $\Psi = \begin{bmatrix} P^X & \mu P^{XY} \\ \varepsilon P^{YX} & P^Y \end{bmatrix}$ and the diagonal matrix

$\Omega = \left[\dfrac{1}{(1+\mu)I + \sum_{l=1}^{L} \eta W_l^{XL}} \quad \dfrac{1}{(1+\varepsilon)I + \sum_{l=1}^{L} \eta W_l^{YL}} \right]$. Since $H = I - \Omega \Psi$ is positive definite, the

linear equation (29) can be solved and we can get the probabilities $\vec{\upsilon}_l$ of all pixels and superpixels:

$$\vec{\upsilon}_l = \eta H^{-1} \Omega \left[W_l^{XL} O^X ; W_l^{YL} O^Y \right] \qquad (30)$$

Acknowledgment

References

[1] M. Sonka, V. Hlavac, R. Boyle, Image processing, analysis, and machine vision, 2th ed., Pacific

Grove, 1999.

[2] Q. Chen, T. Leng, L. Zheng, L. Kutzscher, J. Ma, L. Sisternes, D. Rubin, Automated drusen segmentation and quantification in SD-OCT images, Medical image analysis 17 (8) (2013) 1058-1072.

[3] Z. Ji, J. Liu, G. Cao, Q. Sun, Q. Chen, Robust spatially constrained fuzzy c-means algorithm for brain MR image segmentation, Pattern Recognition 47 (7) (2014) 2454-2466.

[4] C Li, C Xu, C Gui, M.D. Fox, Distance regularized level set evolution and its application to image segmentation, IEEE Transactions on Image Processing 19 (12) (2010) 3243-3254.

[5] Y. Boykov, M. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, in: Proceedings of IEEE International Conference on Computer Vision, 2001, pp. 105-112.

[6] T. Wang, Z. Ji, Q. Sun, Q. Chen, S. Han, Image Segmentation based on Weighting Boundary Information via Graph Cut, Journal of Visual Communication and Image Representation 33 (c) (2015) 10-19.

[7] C. Rother, V. Kolmogorov, A. Blake, Grabcut: Interactive foreground extraction using iterated graph cuts, in: Proceedings of the ACM SIGGRAPH Conference, 2004, pp. 309-314.

[8] S. Han, W. Tao, D. Wang, X. Tai, X. Wu, Image segmentation based on GrabCut framework integrating multiscale nonlinear structure tensor, IEEE Transactions on Image Processing 18 (10) (2009) 2289-2302.

[9] H. Zhou, J. Zheng, L. Wei, Texture aware image segmentation using graph cuts and active contours, Pattern Recognition 46 (6) (2012) 1719-1733.

[10] T. Wang, Z. Ji, Q. Sun, S. Han, Combining Pixel-level and Patch-level Information for Segmentation, Neurocomputing 158 (c) (2015) 13-25.

[11] L. Grady, Random walks for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (11) (2006) 1768-1783.

[12] T. Kim, K. Lee, S. Lee, Generative image segmentation using random walks with restart, in: Proceedings of European Conference on Computer Vision, 2008, pp. 264-275.

[13] W. Yang, J. Cai, J. Zheng, J. Luo, User-friendly interactive image segmentation through unified combinatorial user inputs, IEEE Transactions on Image Processing 19 (9) (2010) 2470-2479.

[14] Z. Yuan, T. Lu, P. Shivakumara, A Novel Topic-Level Random Walk Framework for Scene Image Co-segmentation, in: Proceedings of European Conference on Computer Vision, 2014, pp. 695-709.

[15] X. Bai, G. Sapiro, A geodesic framework for fast interactive image and video segmentation and matting, in: Proceedings of IEEE International Conference on Computer Vision, 2007, pp. 1-8.

[16] A. Criminisi, T. Sharp, A. Blake, Geos: Geodesic image segmentation, in: Proceedings of European Conference on Computer Vision, 2008, pp. 99-112.

[17] J. Mille, S. Bougleux, L. Cohen, Combination of piecewise-geodesic paths for interactive segmentation, International Journal of Computer Vision 112 (1) (2014) 1-22.

[18] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9) (2004) 1124-1137.

[19] P. Arbeláez, L. Cohen, Constrained image segmentation from hierarchical boundaries, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8.

[20] X. He, R. Zemel, D. Ray, Learning and incorporating top-down cues in image segmentation, in:

31

Proceedings of European Conference on Computer Vision, 2006, pp. 338-351.

[21] A. Rabinovich, S. Belongie, T. Lange, J M Buhmann, Model order selection and cue combination for image segmentation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1130-1137.

[22] Y. Li, J. Sun, C. Tang, H. Shum, Lazy snapping. ACM Transactions on Graphics 23 (3) (2004) 303-308.

[23] P. Kohli, M. Kumar, P. Torr, P3 & beyond: Solving energies with higher order cliques, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-8.

[24] J. Shen, Y. Du, W. Wang, X. Li, Lazy random walks for superpixel segmentation, IEEE Transactions on Image Processing 23 (4) (2014) 1451-1462.

[25] P. Kohli, P. Torr, Robust higher order potentials for enforcing label consistency, International Journal of Computer Vision 82 (3) (2009) 302-324.

[26] T. Wang, J. Collomosse, Probabilistic motion diffusion of labeling priors for coherent video segmentation, IEEE Transactions on Multimedia 14 (2) (2012) 389-400.

[27] T. Kim, K. Lee, S. Lee, Learning full pairwise affinities for spectral segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1690-1703.

[28] Z. Li, X. Wu, S. Chang, Segmentation using superpixels: A bipartite graph partitioning approach, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 789-796.

[29] T. Kim, K. Lee, S. Lee, Nonparametric higher-order learning for interactive segmentation, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 3201-3208.

[30] D. Comaniciu, P. Meer, Mean shift: A robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 603-619.

[31] H.I. Bozma, J.S. Duncan, A game-theoretic approach to integration of modules, IEEE Transactions on Pattern Analysis and Machine Intelligence 16(11) (1994) 1074-1086.

[32] A. Chakraborty, J. S. Duncan, Integration of boundary finding and region-based segmentation using game theory, in: Proceedings of IEEE Workshop on Biomedical Image Analysis, 1994, pp. 13-22.

[33] Berkeley segmentation data set, Available at: /http://www.eecs.berkeley.edu/Research/ Projects/CS/vision/grouping/segbench/S.

[34] Msrc ground true data set. Available at: http://research.microsoft.com/en-us/um/cambridge/ projects/vi-sionimagevideoediting/segmentation/grabcut.htm

[35] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222-1239.

[36] A. Blake, C. Rother, M. Brown, P. Perez, P. Torr, Interactive image segmentation using an adaptive GMMRF model, in: Proceedings of Europenan Coference on Computer Vision, 2004, pp. 428-441.

[37] T. Nguyen, J. Cai, J. Zhang, J. Zhen, Robust interactive image segmentation using convex active contours, IEEE Transactions on Image Processing 21 (8) (2012) 3734-3743.

[38] M. Friedman, The use of ranks to avoid the assumption of normality implicit in the analysis of variance, Journal of the American Statistical Association 32 (1937) 674-701.

[39] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Annals of Mathematical Statistics 11 (1) (1940) 86-92.

Author Biography

Tao Wang received the B.E. in Computer Science from Nanjing University of Science and Technology (NUST), China, in 2012. Currently he is pursuing the Ph.D. degree in pattern recognition and intelligence system from NUST. His research interests include image segmentation, pattern recognition.

Zexuan Ji received the B.E. degree in Computer Science and Technology, and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2007 and 2012, respectively. Currently, he is a Lecturer and Postdoctoral fellow with the School of Computer Science and Engineering at the Nanjing University of Science and Technology. He visited the Shenzhen Institutes of Advanced Technology from eight months since Oct. 2009, and the School of Information Technologies, University of Sydney for one year since Nov. 2010. His current interests include medical imaging, image processing and pattern recognition.

Quansen Sun received the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2006. He is a professor in the Department of Computer Science at NUST. His current interests include pattern recognition, image processing, remote sensing information system, and medical image analysis.

Qiang Chen received B.E. degree in computer science and Ph.D. degree in Pattern Recognition and Intelligence System from Nanjing University of Science and Technology, China, in 2002 and 2007, respectively. Currently, he is a professor in the School of Computer Science and Technology at the Nanjing University of Science and Technology. His main research topics are image segmentation, object tracking, image denoising, and image restoration.

Peng Fu received the B.E. degree in Computer Science and Technology, and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2009 and 2015, respectively. Currently, he is a Lecturer and Postdoctoral fellow with the School of Computer Science and Engineering at the Nanjing University of Science and Technology. His research interests include image processing, noise estimation, and pattern recognition.

Highlights

- ➢ Multi-layer graph constraints are utilized for the interactive image segmentation.
- ➢ Labeling information is introduced into the conventional pixel-superpixel combinational model.
- ➢ The optimization based on game theory is proposed for the combinational energy functions.
- ➢ The proposed method can obtain better performance than the state-of-the-art methods.