Interactive Image Segmentation via Pairwise Likelihood Learning

Tao Wang[1], Quansen Sun[1]*, Qi Ge[2], Zexuan Ji[1]*, Qiang Chen[1], Guiyu Xia[1]
[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[2]School of Telecommunications and Information Engineering, Nanjing University of Posts and
Telecommunications, China
wangtaoatnjust@163.com, sunquansen@njust.edu.cn, geqi@njupt.edu.cn,
jizexuan@njust.edu.cn, chen2qiang@njust.edu.cn, xiaguiyu1989@sina.com

Abstract

This paper presents an interactive image segmentation approach where the segmentation problem is formulated as a probabilistic estimation manner. Instead of measuring the distances between unseeded pixels and seeded pixels, we measure the similarities between pixel pairs and seed pairs to improve the robustness to the seeds. The unary prior probability of each pixel belonging to the foreground F and background B can be effectively estimated based on the similarities with label pairs (F, F), (F, B), (B, F) and (B, B). Then a likelihood learning framework is proposed to fuse the region and boundary information of the image by imposing the smoothing constraint on the unary potentials. Experiments on challenging data sets demonstrate that the proposed method can obtain better performance than state-of-the-art methods.

1 Introduction

Image segmentation aims to separate the desired foreground object from the background. Segmented semantic regions or contours associated with real-word entities or scenes are the basis for further advanced image processing such as image editing, object tracking and recognition. Segmentation schemes can be classified into unsupervised [Comaniciu and Meer, 2002], semi-supervised [Wu et al., 2014] and fully supervised [Zhang et al., 2013; Zhao and Fu, 2015] approaches. Compared with the other two schemes, the advantage of semi-supervised schemes is the ability to add the users' intentions in order to obtain results meeting their demands. This paper considers semi-supervised (interactive) schemes for foreground-background segmentation.

During the last decades, many interactive segmentation approaches have been proposed, such as graph cut (GC) [Boykov and Jolly, 2001] and random walk (RW) [Grady, 2006]. In these approaches, unary and pairwise potentials are constructed for the segmentation. A unary potential measures the similarity of a pixel to the labels F and B, while the pairwise potential quantifies the similarity between pairs of
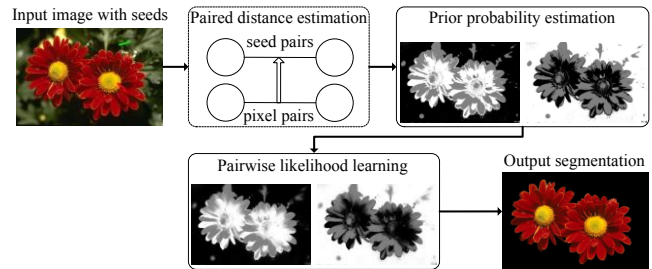

Figure 1: Overview of the segmentation framework.

pixels. Since the prior information of each class is provided by the user, the unary potential is generally quantified as the distances between unseeded pixels and seeded pixels via some clustering algorithms, such as Gaussian mixture model (GMM). If enough seeds are given, GMM can accurately estimate the potential distribution of each label. However, when the user's interaction is limited, it is hard to capture the accurate label information.

As an effective strategy to solve the above problems, methods based on superpixels have been proposed [He et al., 2006; Kohli et al., 2007]. However, superpixels tend not to emphasis sufficiently the proximity, and thereby generate isolated regions. In order to obtain reliable results, the relationships among pixels and superpixels are fused to enforce proximity and continuity [Kim et al., 2010; Kohli and Torr, 2009; Wang et al., 2016a; Wang et al., 2016b]. Both geometrical adjacency and long range cue are critical for the segmentation. However, more superpixel variables need to be defined and more relationships between pixels and superpixels need to be computed in these approaches, which increases the additional algorithm complexity. Furthermore, several parameters are utilized to control the influence of pixel-level and superpixel-level relationships, and the segmentation results are generally sensitive to these controlling parameters [Wang et al., 2016b].

To address the above problems, in this work, we propose an interactive image segmentation method based on probabilistic estimation. Fig. 1 shows the block diagram of the proposed algorithm. Instead of measuring the distances between pixels and seeds, we measure the similarities between pixel pairs and seed pairs to estimate the prior label

---

* Corresponding author

information. The contributions of the proposed approach are concluded as:

First, similarities between pixel pairs are utilized to capture the prior information. Pixel-pair-level relationships can be regarded as the higher-order information of pixel-level relationships, which can help to improve the robustness to user inputs. **Second**, instead of the original two labels F and B, four label pairs (F, F), (F, B), (B, F) and (B, B) are utilized for more accurate estimates of relationships between unseeded pixels and seeded pixels. **Third**, the unary potential of a pixel to the classes F and B can be effectively obtained based on the similarities with label pairs (F, F), (F, B), (B, F) and (B, B). **Last**, to further improve the segmentation accuracy, a likelihood learning framework is proposed to impose the smoothing constraint on unary potentials based on the pairwise similarities of pixels.

2 Related Work

Graph Cut: The interactive GC method was first proposed to segment the grayscale medical images [Boykov and Jolly, 2001]. Certain pixels are labeled by the user as foreground or background to provide the hard constraints, which helps to make the resulting segmentation adhere with the user's intention. Lazy snapping (LS) [Li et al., 2004] replaced pixels with superpixels to improve efficiency. A coarse-to-fine user interface is also designed to provide instant visual feedback. GrabCut [Rother et al., 2004] extended GC to color images by utilizing GMMs to model the F and B regions. Incomplete trimaps are also provided to simplify the user interaction through an iterative optimization process. Texture aware model (TAM) [Zhou et al., 2012] combined the color and texture information to overcome the difficulties in handling images containing textures. ACP-cut [Jian and Jung, 2016] utilized semi-supervised kernel matrix learning to better preserve the details around object boundaries. Moreover, the seed information is also propagated to achieve discriminative structure learning and reduce the computational complexity.

Random Walk: Grady *et al*. [Grady and Funkalea, 2004] first proposed the interactive RW method for medical image segmentation and extended it for general image segmentation in [Grady, 2006]. To overcome the weak boundary and texture problems of RW, Random walk with restart [Kim et al., 2008] constructed a generative segmentation model by utilizing the steady-state probability to reduce dependence on the seeds. Lazy random walk (LRW) [Shen et al., 2014] considered the global relationships between all the pixels and seeds to solve the superpixel segmentation problem in weak boundary and texture regions. Sub-Markov random walk (SMRW) [Dong et al., 2016] introduced the label prior by adding auxiliary nodes to further improve the segmentation accuracy especially in thin and elongated regions. Constrained random walk (CRW) [Yang et al., 2010] advocated the use of multiple intuitive user inputs to better reflect a user's intention. Laplacian Coordinates (LC) [Casaca et al., 2014] minimized the average of distances while better controlling anisotropic propagation of labels, which ensures better fitting on image boundaries as well as a pretty good neighborhood preservation.

Higher-orderModel: As shown in [Jian and Jung, 2016], the quality of these interactive methods is strongly affected by seed quantity and position, and they are likely to fail to keep global data coherence due to the bias caused by limited interactions. In order to improve the robustness to user inputs, many perceptual grouping methods have been proposed by utilizing superpixels to capture long range grouping cues. Robust P^n model (RP[n]) [Kohli and Torr, 2009] constructed parametric higher-order potentials based on multiple superpixels by utilizing higher-order condition random fields in a principled manner. Unlike RP[n] estimating the number of pixels in the superpixel that do not belong to the dominant label, in this work [Wang et al., 2016c], the sum of weights for pixels in the superpixel not taking the dominant label is measured, which helps to produce finer higher-order potentials. Nonparametric higher-order model (NHO) [Kim et al., 2010] utilized the pairwise relationship between pixels and their corresponding superpixels by a multi-layer graph. A higher-order cost function of pixel likelihoods is designed to enforce the label consistency inside the superpixels. To further improve the segmentation performance, Multi-layer graph constraints model (MGC) [Wang et al., 2016a] introduced the prior label information into the nonparametric higher-order model, and the multi-layer relationships among pixels, superpixels and labels are fused for the segmentation. As pointed out in [Kim et al., 2010], these superpixel-based methods are less sensitive to user inputs and produce high-quality segmentation results. However, the generation of multiple superpixels and the computation of higher-order relationships cause them limited to high algorithm complexity.

3 Image Segmentation by Pairwise Likelihood Learning

An input image I can be represented by a graph $G = (X, W)$, where $X = \{x_i\}_{i=1}^{N}$ is a collection of pixels, $W = [W_{ij}]_{N \times N}$ represents the relationships of pairwise pixels, and N is the number of pixels. The task of the segmentation is to classify each image element x_i as the label F or B. The core of the proposed approach is to estimate the marginal assignment probabilities $p(x_i \in F)$ and $p(x_i \in B)$ based on paired relationship metric and pairwise likelihood learning.

The similarity W_{ij} between pairwise pixels (x_i, x_j) is defined as a typical Gaussian function:

$$W_{ij} = \begin{cases} \exp(-\beta \|c_i - c_j\|_2) & if \ (x_i, x_j) \in \aleph \\ 0 & otherwise \end{cases} \quad (1)$$

where c_i and c_j denote the intensity feature at pixel x_i and x_j, respectively. $\beta > 0$ is a constant that controls the strength of the weight. $\aleph$ is the set of all neighboring pixel pairs in the image. It can be noticed that if two neighboring pixels have similar features, their weight is large, and vice versa.
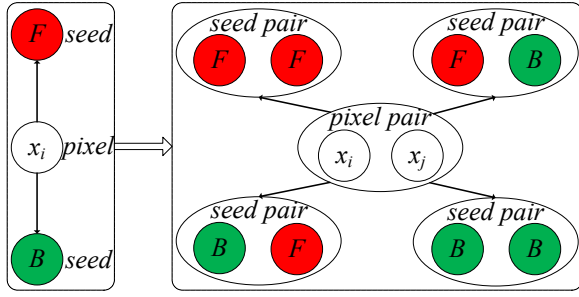
Figure 2: Sketch map of prior information estimation from user inputs:the left shows the distance measurement between pixels and seeds by the conventional methods, the right shows the paired relationship measurement between pixel pairs and seed pairs by the proposed method.

Figure 3: Comparison of the unary potentials, where (a) shows the test image with seeds, (b) and (c) show the probability maps of F and B, respectively, based on the pixel-level measurement by GMM, (d)-(e) show the probability maps of F and B, respectively, based on the proposed pixel-pair-level measurement.

3.1 Paired Distance Estimation

User inputs represent the prior label information, and the conventional methods measure the distances between pixels and seeds to estimate the unary assignment probabilities with labels (shown in Fig. 2 left). These methods are generally sensitive to user inputs and it is hard to obtain accurate results when the number of seeds is limited. To improve the robustness to user inputs, this paper considers the paired probability estimation by measuring the similarities between pixel pairs and seed pairs (shown in Fig. 2 right). Four types of seed pairs are involved, which respectively correspond to four label pairs (F, F), (F, B), (B, F) and (B, B).

The sets of seed pairs are generated based on the seeds. For simplicity, the k-means algorithm is utilized to cluster the seeds respectively from the foreground and background. Clustering centers of the foreground $\{c_i^F\}_{i=1}^{K_F}$ and the background $\{c_i^B\}_{i=1}^{K_B}$ can be obtained, where K_F and K_B are the number of clusters of the foreground and background, respectively.

The sets of seed pairs corresponding to the label pairs (F, F), (F, B), (B, F) and (B, B) are constructed as $S^{FF} = \{(c_i^F, c_j^F)\}_{i,j=1}^{K_F}$, $S^{FB} = \{(c_i^F, c_j^B)\}_{i=1,j=1}^{K_F,K_B}$, $S^{BF} = \{(c_i^B, c_j^F)\}_{i=1,j=1}^{K_B,K_F}$ and $S^{BB} = \{(c_i^B, c_j^B)\}_{i,j=1}^{K_B}$, respectively. For example, it can be obtained that $S^{FF} = \{(c_1^F, c_1^F), (c_1^F, c_2^F), (c_2^F, c_1^F), (c_2^F, c_2^F)\}$ and $S^{FB} = \{(c_1^F, c_1^B), (c_1^F, c_2^B), (c_2^F, c_1^B), (c_2^F, c_2^B)\}$ when $K_F = K_B = 2$.

For any pixel pair $(x_i, x_j) \in \aleph$, the distances d_{ij}^{FF}, d_{ij}^{FB}, d_{ij}^{BF} and d_{ij}^{BB} with label pairs (F, F), (F, B), (B, F) and (B, B) are defined as:

$$d_{ij}^{FF} = \max_{(c_m^F, c_n^F) \in S^{FF}} (\exp(-\beta \|c_i - c_m^F\|_2) \cdot \exp(-\beta \|c_j - c_n^F\|_2)) \quad (2)$$

$$d_{ij}^{FB} = \max_{(c_m^F, c_n^B) \in S^{FB}} (\exp(-\beta \|c_i - c_m^F\|_2) \cdot \exp(-\beta \|c_j - c_n^B\|_2)) \quad (3)$$

$$d_{ij}^{BF} = \max_{(c_m^B, c_n^F) \in S^{BF}} (\exp(-\beta \|c_i - c_m^B\|_2) \cdot \exp(-\beta \|c_j - c_n^F\|_2)) \quad (4)$$

$$d_{ij}^{BB} = \max_{(c_m^B, c_n^B) \in S^{BB}} (\exp(-\beta \|c_i - c_m^B\|_2) \cdot \exp(-\beta \|c_j - c_n^B\|_2)) \quad (5)$$

3.2 Prior Assignment Probabilities

The values of d_{ij}^{FF}, d_{ij}^{FB}, d_{ij}^{BF} and d_{ij}^{BB} are then normalized with the constraint $d_{ij}^{FF} + d_{ij}^{FB} + d_{ij}^{BF} + d_{ij}^{BB} = 1$ for any pixel pair
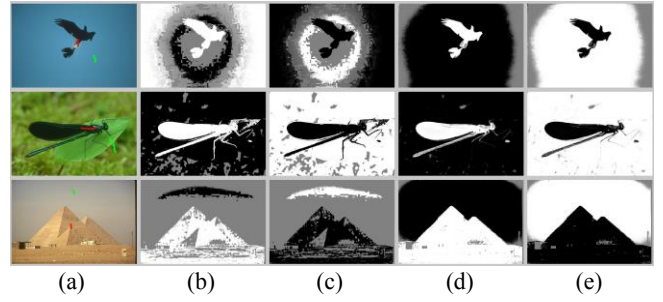
$(x_i, x_j) \in \aleph$. The pairwise assignment probabilities of the pixel pair (x_i, x_j) are defined as:

$$p(x_i \in F, x_j \in F) = d_{ij}^{FF} \quad (6)$$

$$p(x_i \in F, x_j \in B) = d_{ij}^{FB} \quad (7)$$

$$p(x_i \in B, x_j \in F) = d_{ij}^{BF} \quad (8)$$

$$p(x_i \in B, x_j \in B) = d_{ij}^{BB} \quad (9)$$

The unary assignment probabilities of the two pixels x_i and x_j can be estimated from these pairwise assignment probabilities. For pixel x_i, it has:

$$p(x_i \in F) = p(x_i \in F, x_j \in F) + p(x_i \in F, x_j \in B) \quad (10)$$

$$p(x_i \in B) = p(x_i \in B, x_j \in F) + p(x_i \in B, x_j \in B) \quad (11)$$

Then for pixel x_j, it has:

$$p(x_j \in F) = p(x_i \in F, x_j \in F) + p(x_i \in B, x_j \in F) \quad (12)$$

$$p(x_j \in B) = p(x_i \in F, x_j \in B) + p(x_i \in B, x_j \in B) \quad (13)$$

After the pairwise assignment probabilities of all pixel pairs in the set $\aleph$ are transformed into the unary probabilities, the value of $p(x_i)$ for each pixel $x_i \in X$ is normalized with the constraint $p(x_i \in F) + p(x_i \in B) = 1$.

Examples of prior probability estimation are shown in Fig. 3, where (a) shows the test image with seeds, (b)-(c) show the results obtained by GMM based on the pixel-level distance, and (d)-(e) show the results produced by the proposed method based on the pixel-pair-level relationship measurement. It can be seen that the pixel-level distances between pixels and seeds are not enough to discriminate the label information when the user inputs are limited. The pixel-pair-level relationships between pixel pairs and seed pairs can be regarded as the higher-order information of the pixel-level distances, and can produce more accurate estimation of the prior probability under the same seed information. Furthermore, instead of the two labels F and B, the distances to the four label pairs (F, F), (F, B), (B, F) and (B, B) make the relationships more discriminating.

The relationships between doublets can be naturally extended to higher-order relationships. More interactions can be considered in higher-order measurement, for example,

eight label groups are involved when measuring the distances between triplets. However, the number of label groups increases exponentially with an increase in order, which leads to higher algorithm complexity. To keep computational efficiency, in this paper, only the pixel-pair-level relationships are considered.

3.3 Pairwise Likelihood Learning

The above probability estimation only considers the region information regardless of the boundary information of the image. To further improve the segmentation accuracy, we impose the smoothing constraint on the unary prior potentials by a pairwise likelihood learning strategy.

Let $P^F = [p_i^F]_{N \times 1}$ and $P^B = [p_i^B]_{N \times 1}$ denote the prior probability vectors where $p_i^F = p(x_i \in F)$ and $p_i^B = p(x_i \in B)$. Let $\hat{P}^F = [\hat{p}_i^F]_{N \times 1}$ and $\hat{P}^B = [\hat{p}_i^B]_{N \times 1}$ denote the novel probability vectors after the likelihood learning, where $\hat{p}_i^F = \hat{p}(x_i \in F)$ and $\hat{p}_i^B = \hat{p}(x_i \in B)$ represent the probabilities of pixel x_i belonging to the labels F and B, respectively.

The likelihood learning process of $\hat{P}^F$ and $\hat{P}^B$ is defined as minimizing the following two cost functions:

$$E(\hat{P}^F) = E_{boundary}(\hat{P}^F) + E_{region}(\hat{P}^F)$$
$$= \sum_{(x_i, x_j) \in \aleph} W_{ij} \cdot (\hat{p}_i^F - \hat{p}_j^F)^2 + \lambda \sum_{x_i \in X} d_i \cdot \left[p_i^F \cdot (\hat{p}_i^F - \frac{1}{d_i})^2 + p_i^B \cdot (\hat{p}_i^F)^2 \right] \quad (14)$$

$$E(\hat{P}^B) = E_{boundary}(\hat{P}^B) + E_{region}(\hat{P}^B)$$
$$= \sum_{(x_i, x_j) \in \aleph} W_{ij} \cdot (\hat{p}_i^B - \hat{p}_j^B)^2 + \lambda \sum_{x_i \in X} d_i \cdot \left[p_i^B \cdot (\hat{p}_i^B - \frac{1}{d_i})^2 + p_i^F \cdot (\hat{p}_i^B)^2 \right] \quad (15)$$

where $E_{boundary}$ represents the boundary energy term, E_{region} represents the region energy term, $d_i = \sum_{j=1}^{N} W_{ij}$, and the parameter $\lambda = \alpha / (1 - \alpha)$ $(0 < \alpha < 1)$ is utilized to balance these two energy terms. $E_{boundary}$ constrains the neighboring pixels with high similarities to have similar likelihood probabilities. E_{region} constrains the likelihood estimation to keep consistent with the prior probabilities. A pixel x_i should be assigned a high $\hat{p}_i^F$ and a low $\hat{p}_i^B$ if its prior probability p_i^F is high, and vice versa. Both the region and boundary information are considered in the likelihood learning process, which makes the segmentation results keep regional connectivity and piecewise smooth.

Reformulate the cost functions in equations (14, 15) as the matrix forms:

$$E(\hat{P}^F) =$$
$$(\hat{P}^F)^T (D - W) \hat{P}^F + \lambda D \left[(\hat{P}^F - \frac{O}{D})^T \Omega^F (\hat{P}^F - \frac{O}{D}) + (\hat{P}^F)^T \Omega^B \hat{P}^F \right] \quad (16)$$

$$E(\hat{P}^B) =$$
$$(\hat{P}^B)^T (D - W) \hat{P}^B + \lambda D \left[(\hat{P}^B - \frac{O}{D})^T \Omega^B (\hat{P}^B - \frac{O}{D}) + (\hat{P}^B)^T \Omega^F \hat{P}^B \right] \quad (17)$$

where $D = diag([d_1, ..., d_N])$, $O = [1]_{N \times 1}$, $\Omega^F = diag(P^F)$, and $\Omega^B = diag(P^B)$. Differentiating $E(\hat{P}^F)$ and $E(\hat{P}^B)$ with respect to $\hat{P}^F$ and $\hat{P}^B$, respectively, and set to zero, we have:

$$\frac{\partial E(\hat{P}^F)}{\partial \hat{P}^F} = (D - W)\hat{P}^F + \lambda D(\Omega^F + \Omega^B)\hat{P}^F - \lambda \Omega^F O = 0 \quad (18)$$

$$\frac{\partial E(\hat{P}^B)}{\partial \hat{P}^B} = (D - W)\hat{P}^B + \lambda D(\Omega^F + \Omega^B)\hat{P}^B - \lambda \Omega^B O = 0 \quad (19)$$

Since $(\Omega^F + \Omega^B)$ is an identity matrix, $\Omega^F O = P^F$, $\Omega^B O = P^B$, and $\lambda = \alpha / (1 - \alpha)$, it can be derived:

$$\hat{P}^F = T P^F \quad (20)$$
$$\hat{P}^B = T P^B \quad (21)$$

where $T = \alpha(D - (1 - \alpha)W)^{-1}$. It can be seen that the likelihood probabilities $\hat{P}^F$ and $\hat{P}^B$ can be obtained by diffusing the prior probabilities P^F and P^B on the transfer matrix T. The multiplication of the inversion of a matrix by a single vector can be efficiently solved by the MATLAB division operator '\'.

The final segmentation can be produced by assigning pixels $\{x_i \mid \hat{p}(x_i \in F) > \hat{p}(x_i \in B)\}$ the foreground label, and pixels $\{x_i \mid \hat{p}(x_i \in B) > \hat{p}(x_i \in F)\}$ the background label.

4 Experimental Results

The proposed method was experimentally verified by comparing it with state-of-the-art approaches: GrabCut [Rother et al., 2004], RW [Grady, 2006], LC [Casaca et al., 2014], SMRW [Dong et al., 2016] and NHO [Kim et al., 2010] on the Berkeley segmentation data set[1] which contains 500 images with size 481×321 (or 321×481) and Microsoft GrabCut database[2] which contains 50 images with public seeds information. Parameters involved in the proposed scheme are set as follows: the constant β is fixed as 60, the number of clusters K_F and K_B are both set to 3, the controlling parameter α is set to 0.1, and the 4-neighborhood relationship of pixels is considered.

4.1 Qualitative Benchmark Results

Fig. 4 shows the comparison to state-of-the-art interactive segmentation methods. Fig. 4(a) shows the test images from the Berkeley segmentation data set with a few scribbles (red: foreground, green: background). Fig. 4(b)-(g) show the segmentation results of GrabCut, RW, LC, SMRW, NHO and the proposed method, respectively. It can be seen that GrabCut and RW cannot obtain satisfactory results when the number of seeds is limited. The RW extension methods, i.e. LC and SMRW, can obtain better results than RW. However, they are also sensitive to the seeds. The pixel-level information learnt from limited seeds is not enough to discriminate the foreground and background labels. Compared with these pixel-level-based approaches, the perceptual superpixel-based method NHO obtains more robust results by utilizing multiple superpixels to propagate long range grouping laws. However, the object details around the boundaries cannot be preserved well in NHO, especially in many thin and slender regions. It can be clearly seen that the proposed method produces the best results with accurate object details. The paired distance estimation can help to

---

[1] http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/S.

[2] http://research.microsoft.com/enus/um/cambridge/projects/visionimagevideoediting/segmentation/grabcut.htm

| | 12003 | 42049 | 61060 | 66075 | 113016 | 118035 | 161045 | 198087 | 201080 | 295087 |

(a) original   (b) GrabCut   (c) RW   (d) LC   (e) SMRW   (f) NHO   (g) Ours

Figure 4: Comparison to state-of-the-art approaches with a few scribbles. (a) Test image from the Berkeley segmentation data set with scribbles (red: foreground, green: background), (b)-(g) segmentation results of GrabCut, RW, LC, SMRW, NHO and the proposed method.

obtain accurate label prior information even with limited seeds, and the pairwise likelihood learning can help to produce smooth object boundaries.

4.2 Quantitative Benchmark Results

The error rate is utilized to quantitatively evaluate the segmentation accuracy, which is defined as the ratio of the number of wrongly labeled pixels to the total number of unlabeled pixels. Table 1 summarizes the mean, standard deviation and the average rank with the Friedman statistical test [Friedman, 1937] (with a significance level of 0.05) of error rates achieved by various methods on the Microsoft GrabCut database. Compared with state-of-the-art methods, it can be observed that the proposed method achieves the lowest error rate. Friedman test determines the χ^2 value as 64.69 and the p-value as 1.3e-12. From χ^2 distribution table, we find the critical value for $(6-1)=5$ degree of freedom with 0.05 significance level is 11.07. Since the χ^2 value is larger than the critical value, H_0 is rejected and H_1 is accepted, which substantiates the significant difference in behavior among the compared methods.

Methods	Error rate	
	Mean $\pm$ Std	Average rank
GrabCut	5.46 $\pm$ 4.2	3.8
RW	6.45 $\pm$ 4.8	5.0
LC	5.04 $\pm$ 3.8	3.6
SMRW	4.61 $\pm$ 3.2	3.6
NHO	4.25 $\pm$ 3.7	2.6
Ours	**3.49 $\pm$ 2.6**	**2.3**
GC [Boykov and Jolly, 2001]	6.60 (listed in [Kim et al., 2010])	
TAM [Zhou et al., 2012]	3.64 (listed in [Zhou et al., 2012])	
LS [Li et al., 2004]	6.65 (listed in [Yang et al., 2010])	
CRW [Yang et al., 2010]	4.08 (listed in [Yang et al., 2010])	
LRW [Shen et al., 2014]	5.70 (listed in [Wang et al., 2016a])	
RP^n [Kohli and Torr, 2009]	6.08 (listed in [Kim et al., 2010])	
MGC [Wang et al., 2016a]	3.79 (listed in [Wang et al., 2016a])	

Table 1: Mean $\pm$ standard deviation (Std) (%) and the average rank of error rates for the compared methods on the Microsoft GrabCut database.

Percent seeds	50%	30%	10%	1%
a_o (%)	99.8 ± 0.1	99.7 ± 0.5	99.4 ± 0.7	98.2 ± 1.2

Table 2: Mean $\pm$ standard deviation (%) of normalized overlap a_o with 50%, 30%, 10% and 1% percent seeds.
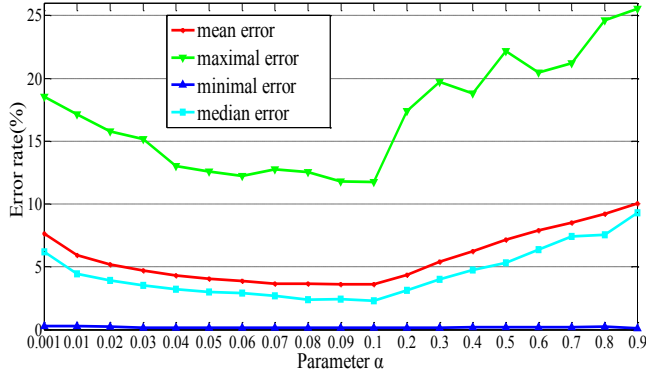

Figure 5: Error rates (%) of the Microsoft GrabCut database for a varying values of the parameter α.

4.3 Sensitivity analyzing

We analyze the sensitivity of the proposed method with respect to seed quantity and placement. The standard segmentations are achieved from the initial trimaps provided by the Microsoft GrabCut database. The initial seeds are then randomly taken from 50% to 1% of total seed quantity. The perturbed segmentations are recomputed from these selected seeds and compared with the standard segmentations. The normalized overlap $a_o = |F_1 \bigcap F_2 / F_1 \bigcup F_2|$ is used to measure the similarity of two segmentations [Kim et al., 2010], where F_1 and F_2 indicate the sets of pixels in two segmentations. Table 2 summarizes the mean and standard deviation of a_o on the Microsoft GrabCut database with 50%, 30%, 10% and 1% percent seeds. It can be seen that even with 1% seeds, the proposed method can still obtain satisfactory results.

4.4 Parameter settings

The parameter α is utilized to control the influence of the region and boundary energies. Fig. 5 shows the quantitative evaluation of the error rates on the Microsoft GrabCut database with different values of α. It can be observed that the lowest error rate is obtained when $\alpha = 0.1$. From the variation of the error rates, we can find that the segmentation results are not sensitive to α in the range of 0.02 to 0.3.

Fig. 6 shows the average error rates over all 50 images in the Microsoft GrabCut database with different number of clusters K_F and K_B. From the variation range of the error rates, it can be seen that the results are not sensitive to the number of clusters. The lowest average error rate can be obtained as 3.29% when $K_F = 5$ and $K_B = 4$. However, the algorithm complexity is higher with more number of clusters. To obtain accurate segmentation results while keeping low algorithm cost, in this paper, K_F and K_B are both set to 3.
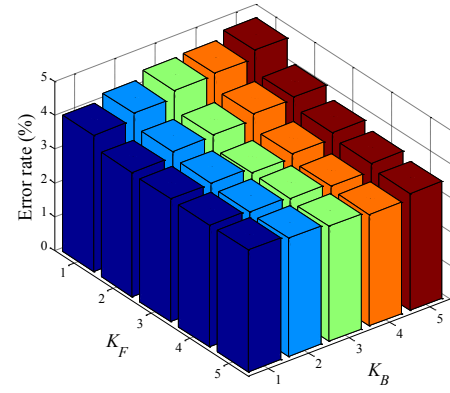

Figure 6: Average error rates (%) over all 50 images in the Microsoft GrabCut database with different number of clusters K_F and K_B.

Methods	GrabCut	RW	LC	SMRW	NHO	Ours
Time (s)	0.7	0.8	3.2	5.1	15.1	3.5

Table 3: Average running times of the compared methods on all 20 images with size 321×481 in the Microsoft GrabCut database.

4.5 Runtimes

Table 3 lists the average running times of GrabCut, RW, LC, SMRW, NHO and the proposed method on all 20 test images with size 321×481 in the Microsoft GrabCut database on an Intel Xeon CPU running at 2.0 GHz in MATLAB. The proposed method takes about 3.5 s to segment an image with size 321×481.

5 Conclusion

In this work we presented an interactive image segmentation method by a probabilistic framework consisting of unary potential estimation and likelihood learning. To improve the robustness to the seeds, the distances between pixel pairs and seed pairs are measured to obtain the prior label information. To further improve the segmentation accuracy, the region information and boundary information of the image are combined by a likelihood learning framework. The similarities of pairwise neighboring pixels are utilized to impose the smoothing constraint on the unary potentials. The qualitative and quantitative comparisons with state-of-the-art interactive approaches demonstrate the superior performance of the proposed method.

Acknowledgments

References

[Boykov and Jolly, 2001] Yuri Y. Boykov, and Marie-Pierre Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In Proceedings of IEEE International Conference on Computer Vision, pages 105-112, 2001.

[Casaca et al., 2014] Wallace Casaca, Luis Gustavo Nonato, and Gabriel Taubin. Laplacian Coordinates for Seeded Image Segmentation. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 384-391, 2014.

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

[Dong et al., 2016] Xingping Dong, Jianbing Shen, Ling Shao, and Luc Van Gool. Sub-Markov random walk for image segmentation. IEEE Transactions on Image Processing, 25(2):516-527, 2016.

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

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

[Grady and Funkalea, 2004] Leo Grady, and Gareth Funkalea-Lea. Multi-label Image Segmentation for Medical Applications Based on Graph-Theoretic Electrical Potentials. In Proceedings of Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, pages 230-245, 2004.

[He et al., 2006] Xuming He, Richard S. Zemel, and Debajyoti Ray. Learning and incorporating top-down cues in image segmentation. In Proceedings of European Conference on Computer Vision, pages 338-351, 2006.

[Jian and Jung, 2016] Meng Jian, and Cheolkon Jung. Interactive Image Segmentation Using Adaptive Constraint Propagation. IEEE transactions on image processing, 25(3):1301-1311, 2016.

[Kim et al., 2008] Tae Hoon Kim, Kyoung Mu Lee, and Sang Uk Lee. Generative image segmentation using random walks with restart. In Proceedings of European Conference on Computer Vision, pages 264-275, 2008.

[Kim et al., 2010] Tae Hoon Kim, Kyoung Mu Lee, and Sang Uk Lee. Nonparametric higher-order learning for interactive segmentation. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 3201-3208, 2010.

[Kohli et al., 2007] Pushmeet Kohli, M. Pawan Kumar, and Philip H. S. Torr. P3 & beyond: Solving energies with higher order cliques. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 1-8, 2007.

[Kohli and Torr, 2009] Pushmeet Kohli, and Philip H. S. Torr. Robust higher order potentials for enforcing label consistency. International Journal of Computer Vision, 82(3):302-324, 2009.

[Li et al., 2004] Yin Li, Jian Sun, Chi-Keung Tang, and Heung-Yeung Shum. Lazy snapping. ACM Transactions on Graphics, 23(3):303-308, 2004.

[Rother et al., 2004] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In Proceedings of the ACM SIGGRAPH Conference, pages 309-314, 2004.

[Shen et al., 2014] Jianbing Shen, Yunfan Du, Wenguan Wang, and Xuelong Li. Lazy random walks for superpixel segmentation. IEEE Transactions on Image Processing, 23(4):1451-1462, 2014.

[Wang et al., 2016a] Tao Wang, Quansen Sun, Zexuan Ji, Qiang Chen, and Peng Fu. Multi-layer graph constraints for interactive image segmentation via game theory. Pattern Recognition, 55:28-44, 2016.

[Wang et al., 2016b] Tao Wang, Zexuan Ji, Quansen Sun, Qiang Chen, Xiaoyuan Jing. Interactive multi-label image segmentation via robust multi-layer graph constraints. IEEE Transactions on Multimedia, 18(12):2358-2371, 2016.

[Wang et al., 2016c] Tao Wang, Qiang Chen, Zexuan Ji, and Quansen Sun. Label Propagation and Higher-order Constraint based Segmentation of Fluid-associated Region in Retain SD-OCT Images. Information Sciences, 358:92-111, 2016.

[Wu et al., 2014] Jiajun Wu, Yibiao Zhao, Jun-Yan Zhu, Siwei Luo, Zhuowen Tu. MILCut: A Sweeping Line Multiple Instance Learning Paradigm for Interactive Image Segmentation. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 256-263, 2014.

[Yang et al., 2010] Wenxian Yang, Jianfei Cai, Jianmin Zheng, and Jiebo Luo. User-friendly interactive image segmentation through unified combinatorial user inputs. IEEE Transactions on Image Processing, 19(9):2470-2479, 2010.

[Zhang et al., 2013] Ke Zhang, Wei Zhang, Yingbin Zheng, Xiangyang Xue. Sparse reconstruction for weakly supervised semantic segmentation. In Proceedings of International Joint Conference on Artificial Intelligence, pages 1889-1895, 2013.

[Zhao and Fu, 2015] Handong Zhao, and Yun Fu. Semantic single video segmentation with robust graph representation. In Proceedings of International Joint Conference on Artificial Intelligence, pages 2219-2225, 2015.

[Zhou et al., 2012] Hailing Zhou, Jianmin Zheng, and Lei Wei. Texture aware image segmentation using graph cuts and active contours. Pattern Recognition, 46(6):1719-1733, 2012.