PAPER

# An Enhanced Affinity Graph for Image Segmentation

**Guodong SUN**[†a)], ***Member***, **Kai LIN**[†], **Junhao WANG**[†], ***and*** **Yang ZHANG**[††], ***Nonmembers***

**SUMMARY** This paper proposes an enhanced affinity graph (EA-graph) for image segmentation. Firstly, the original image is over-segmented to obtain several sets of superpixels with different scales, and the color and texture features of the superpixels are extracted. Then, the similarity relationship between neighborhood superpixels is used to construct the local affinity graph. Meanwhile, the global affinity graph is obtained by sparse reconstruction among all superpixels. The local affinity graph and global affinity graph are superimposed to obtain an enhanced affinity graph for eliminating the influences of noise and isolated regions in the image. Finally, a bipartite graph is introduced to express the affiliation between pixels and superpixels, and segmentation is performed using a spectral clustering algorithm. Experimental results on the Berkeley segmentation database demonstrate that our method achieves significantly better performance compared to state-of-the-art algorithms.

***key words:*** *image segmentation, superpixels, sparse reconstruction, enhanced affinity graph, spectral clustering*

## 1. Introduction

Image segmentation is one of the fundamental yet challenging tasks in computer vision, which has great potential in high-level applications [1], [2]. In the literatures, the unsupervised spectral clustering algorithms become popular in the field of image segmentation. The core of this problem is how to create a reliable graph [3]–[7].

Clearly, the effectiveness of graph-based segmentation algorithms depends on the construction of affinity graph, which depends on the neighborhood topology and pairwise affinities between nodes (pixels or superpixels). As pointed out in [8], the *GL*-graph is constructed by fusing *local*-graph and sparse *global*-graph, while can obtain better segmentation results compared with a local graph as shown in [9], and a global graph as shown in [10]. However, the *local*-graph and *global*-graph are not easily obtained because the layouts of superpixels are diverse at difference scales. In addition, a global-graph built with all superpixels is a dense graph, resulting in large computation costs.

To solve these problems, an enhanced affinity graph is proposed to accurately capture both short- and long-range grouping cues of the superpixels at different scales. The several sets of superpixels are firstly obtained by over-segmentation at different scales. Then, the affinities among neighborhood superpixels are computed to build the local affinity graph, and it can accurately capture short-range grouping cues. Meanwhile, the global affinity graph is obtained by sparse reconstruction among global superpixels to search for long-range grouping cues. The local and global affinity graph are combined to obtain an enhanced affinity graph (EA-graph). Finally, a bipartite graph is introduced to express the affiliation between pixels and superpixels, and a spectral clustering algorithm is used to divide the superpixels at different scales. Experiments are carried out on the Berkeley segmentation database (BSD) with four different criteria, namely, PRI, VoI, GCE, and BDE. The results show that our method achieves significantly better performance compared to state-of-the-art methods, and it also can produce meaningful results with $k = 2$ for most images, while is more realistic applications. As shown in Fig. 1, the visual comparison results are shown with the SAS [9] and *GL*-graph [8].

The organization of this paper is as follows: in Sect. 2 we discuss the different models of affinity graph. In Sect. 3 we introduce the construction process of the proposed

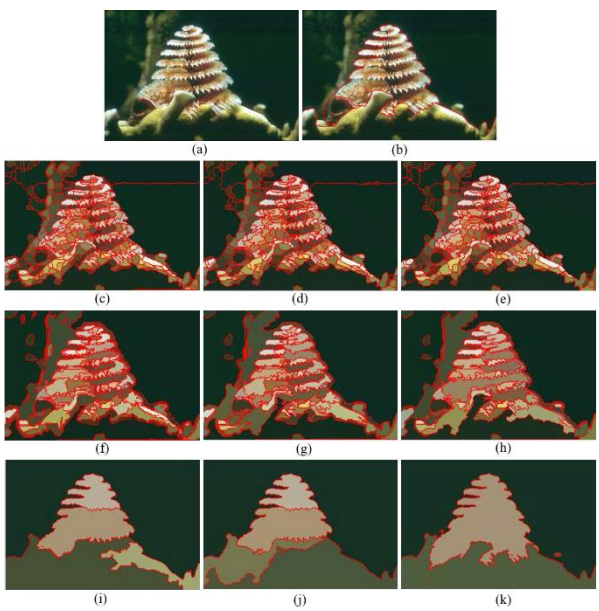


**Fig. 1** Superpixels-based segmentation. (a) Original image, (b) ground truth, (c-h) multi-scale superpixels. (i) SAS, (j) *GL*-graph, and (k) the proposed EA-graph. Our result can better segment the target object compared to other methods

Manuscript received September 21, 2018.
Manuscript revised December 17, 2018.
Manuscript publicized February 4, 2019.

[†]The authors are with Hubei University of Technology, Wuhan, Hubei, China.

[††]The author is with Nanjing University, Nanjing, Jiangsu, China.

a) E-mail: sgdeagle@163.com

DOI: 10.1587/transinf.2018EDP7322

EA-graph. In Sect. 4 we present experimental results compared to state-of-the-art methods on the BSD, and the conclusion is described in Sect. 5.

## 2. Related Works

The conventional graph-cut models are insufficient to achieve a desirable segmentation, since local neighborhood relationships between data nodes is only considered [8], [9], [11], [12], *e.g.*, the $\varepsilon$-neighborhood graph, the $k$-nearest neighborhood graph and the fully connected graph, *etc*. Therefore, many studies about propagating local grouping cues across long-range spatial connections are gradually proposed to improve the performance of segmentation. For instance, Cour et al. [12] proposed the multiscale spectral method to capture fine- and coarse-level details of the large images. Li et al. [9] proposed a graph-cut method based on superpixels to replace the pixels as graph nodes by pre-segmenting the image into small regions.

Specially, Cheng et al. [13] proposed a multi-task low-rank affinity pursuit method to fuse multiple type of image features. This method infers a unified affinity graph to encode the segmentation of the image by seeking the sparsity consistent low-rank affinities from the joint decompositions of multiple feature matrices into pairs of sparse and low-rank matrices. Experimental results show that this method can achieve a good segmentation result.

Wang et al. [10] proposed a graph-cut method to construct an affinity graph using $\ell_0$-sparse representation. This method encodes the connection information between superpixels with the non-zero representation coefficients, and the affinity of connected superpixels is derived by the corresponding representation error. This method also has fine properties of long range and sparsity on superpixels of different scales, and enables propagation of grouping cues between superpixels of different scales based on a bipartite graph [9]. Experimental results show that the model of $\ell_0^-$ sparse reconstruction can achieve good segmentation performance.

Wang et al. [8] proposed a novel sparse global/local affinity graph (*GL*-graph) over superpixels to capture both short- and long-range grouping cues. This method over-segments the input image into superpixels at different scales and divides superpixels adaptively into small-, medium-, and large-sized sets. Medium-sized superpixels are used to achieve global grouping through a sparse representation of superpixels' features by solving a $\ell_0$-minimization problem with orthogonal matching pursuit (OMP). Small- and large-sized superpixels are then used to achieve local smoothness through an adjacent graph. Finally, a bipartite graph is also introduced to enable propagation of grouping cues between superpixels of different scales. Experimental results show that the *GL*-graph yields very competitive quantitative segmentation results.

Our method follows a similar, yet not identical, strategy as the SAS [9], $\ell_0$-graph [10], and *GL*-graph [8] algorithms. i.e., building a bipartite graph over multiple superpixels and pixels, then using Tcuts for image segmentation. The main difference between SAS, $\ell_0$-graph, and our method is the construction model of the affinity graph. In SAS, the neighborhoods of the superpixels are only used to build adjacent graph, and the affinity among superpixels is computed by the Gaussian weighted Euclidean distance in the color feature space. In $\ell_0$-graph, the global affinity graph of superpixels is only used, but in our method, we build an EA-graph combining local affinity graph and global affinity graph over multiscale superpixels, and it can capture both short- and long-range grouping cues of the superpixels. In particular, the best difference between *GL*-graph and our method is the dictionary construction of local/global affinity graph. In *GL*-graph, the dictionary of local affinity graph is built with small- and large-sized superpixels. The dictionary of global affinity graph is built with medium-sized superpixels. In our method, the dictionary of local affinity graph is built with the neighborhoods of the given superpixels. The dictionary of global affinity graph is built by searching for the $L$ superpixels which are similar to the given superpixel in global space. Therefore, our method preserves the intrinsic properties of the image without prior experience and can accurately describe local and global information.

## 3. Enhanced Affinity Graph

### 3.1 Superpixels Generation and Feature Extraction

As pointed out in [9], superpixels generated by employing different algorithms with varying parameters can effectively encode complex image structures for segmentation. The original images can be over-segmented into sub-regions with consistent color and texture features. As showed in Fig. 2, we use the methods and parameters as same as the SAS [9] to over-segment the original image into superpixels of different scales (*e.g.* 5 scales in the Fig. 2) with the Mean Shift (MS) [14] and the Felzenszwalb-Huttenlocher (FH) graph-based method [4]. Then, visual features of each superpixel are computed to obtain a discriminative affinity graph. Considering that mLab (mean value in Lab color space) is applied to approximate human vision and its L component closely matches human perception of lightness. LBP (Local Binary Pattern) is reputed to encode micro-texture and robust to monotone light changes [8]. We extract the mLab and LBP features of the superpixels to build an enhanced affinity graph. The parameters of these features will be introduced in Sect. 4.2.

### 3.2 Enhanced Affinity Graph Construction

The affinity graph describes the affinity among superpixels, playing an important role in image segmentation. However, the conventional affinity graph only calculates the affinity between the superpixel and its neighboring superpixels to build a local affinity graph, as shown in Fig. 3 (a). It only makes use of the local information of the image, and ignores regions in the segmentation. To improve the

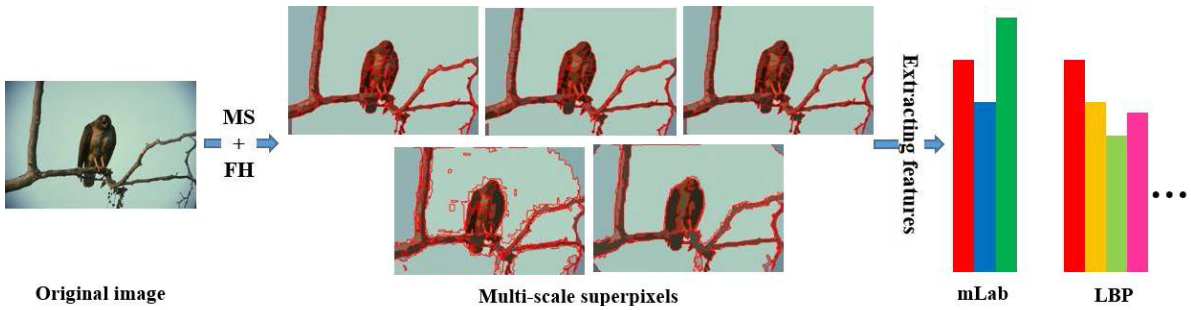


**Fig. 2** Illustration of extracting multi-features over multi-scale superpixels

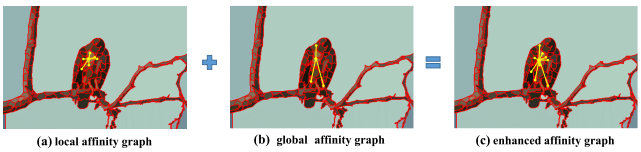


**Fig. 3** The construction process of enhanced affinity graph

segmentation performance, an enhanced affinity graph is built by adding global information based on the local affinity graph, as shown in Fig. 3. The construction process of the enhanced affinity graph is divided into local and global affinity graph construction.

### 3.2.1 Building a Local Affinity Graph

Let $SI_l = \{s_i^l\}_{i=1}^{N_l}$, $l = 1, 2, 3, \cdots$ denote the collection of the superpixels generated by over-segmenting the input image $I$ in Sect. 3.1, where $l$ and $N_l$ represent the scales $l$ of superpixels and the number of superpixels at scale $l$, respectively. For each superpixel $s_i^l$, we can extract mLab and LBP features vectors $(x_i^l)^{mLab} \in R^3$ and $(x_i^l)^{LBP} \in R^{64}$ at scale $l$, respectively. Therefore, we can obtain the associated feature matrix $[x_1^l, \cdots, x_{N_l}^l] \in R^{m \times N_l}$ ($m = 3$ for mLab and $m = 64$ for LBP, see Sect. 4.2) at sacle $l$.

To compute the affinity accurately between adjacent superpixels, we use the neighborhood superpixels of the superpixel $s_i^l$ to obtain a local affinity graph, unlike the local affinity graph in [8], which is obtained with small- and large-sized superpixels. For each feature matrix $[x_1^l, \cdots, x_{N_l}^l]$, the specific construction steps of the corresponding local affinity graph are as follows:

a) The neighborhood dictionary $D_i^l = [x_{a_k}^l]_{k=1}^{n_i}$ of the superpixel $s_i^l$ is first constructed and unitized, where $a_k$ denotes the index of the neighborhood superpixels, and $x_{a_k}^l$ denotes the feature vectors of the neighborhood of the superpixel $s_i^l$ at scale $l$.

b) The superpixel $s_i^l$ is computed as a linear combination of elements in $D_i^l$. In practice, this solution can be transformed into the following optimization problem:

$$\hat{c}^i = \arg\min_{c_i} \left\{ \left\| x_i^l - D_i^l c^i \right\|_2^2,\ c^i \in R^{n_i \times 1},\ c_{a_k}^i \geq 0 \right\} \quad (1)$$

where $\hat{c}^i$ denotes the optimal combination coefficient obtained by using the mLab or LBP feature of the superpixel $s_i^l$, and $c_{a_k}^i \geq 0$ denotes that each element in the $\hat{c}^i$ is non-negative.

c) Calculate the reconstruction error $r_{ij}^l$ between superpixel $s_i^l$ and its neighborhood superpixel $s_j^l$ at scale $l$. The $r_{ij}^l$ is defined as

$$r_{ij}^l = \left\| x_i^l - \hat{c}_j^i x_j^l \right\|_2 \quad (2)$$

where $\hat{c}_j^i$ denotes the corresponding coefficient of the neighborhood superpixel $s_j^l$ in the vector $\hat{c}^i$

d) Calculate the affinity $w_{ij}^l$ between superpixel $s_i^l$ and its neighborhood superpixel $s_j^l$ as follows.

$$w_{ij}^l = \begin{cases} 0 & if\ i = j \\ 1 - (r_{ij}^l + r_{ji}^l)/2 & if\ i \neq j \end{cases} \quad (3)$$

e) Finally, we define $W_l = (w_{ij}^l)$ as the local affinity graph at scale $l$.

### 3.2.2 Building a Global Affinity Graph with Orthogonal Matching Pursuit

To avoid the subjectivity of global affinity graph ($\ell_0$-graph) obtained by medium-sized superpixels in [8], we build the global dictionary $\tilde{D}^l = [x_i^l]_{i=1}^{N_l}$, $l = 1, 2, 3 \cdots$ with all superpixels, and then we search for the $L$ superpixels which are similar to the given superpixel $s_i^l$ with $\tilde{D}^l$ in the global scope. Meanwhile, superpixel $s_i^l$ is computed as a linear combination of the $L$ superpixels. The solution of sparse reconstruction can be transformed into the following optimization problem.

$$\min \|c^i\|_0 \quad s.t.\ x_i^l = \tilde{D}^l c^i,\ c_i^i = 0 \quad (4)$$

where $c^i$ denotes the sparse reconstruction coefficient of superpixel $s_i^l$ in the global dictionary $\tilde{D}_l$, and $\|\cdot\|_0$ denotes the $\ell_0$ norm, which counts the number of nonzero elements in a vector.

However, the Eq. (4) which searches the sparsest solution of linear equations is NP-hard, which is hard to get results directly. In practice, we make use of orthogonal matching pursuit (OMP) to seek an approximation of the sparsest solution with the $\ell_0$-norm to solve Eq. (4). Considering that the reconstruction coefficient among superpixels

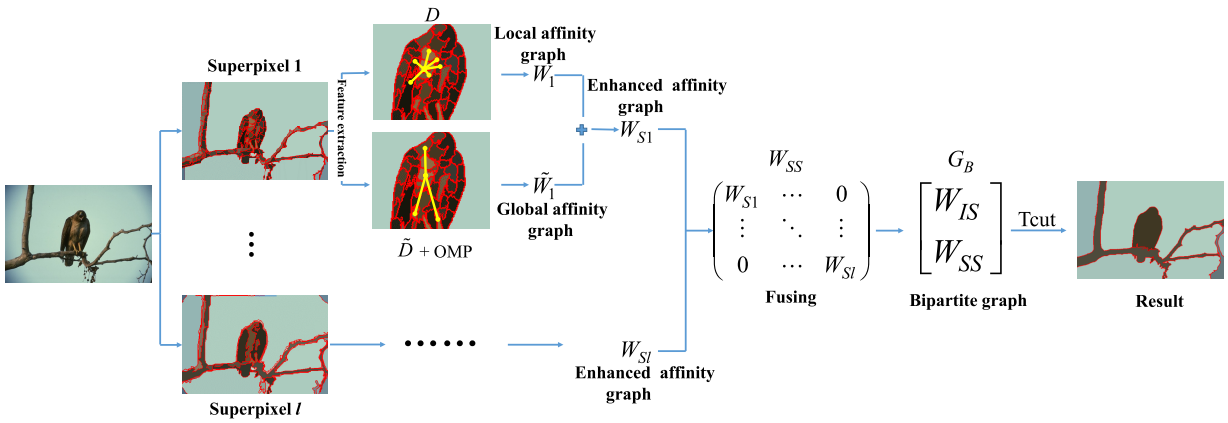


**Fig. 4** The overall framework of the proposed EA-graph

must be non-negative, the constraint condition $c_j^i \geq 0,\ j \neq i$ is added based on the OMP algorithm. The Eq. (4) is transformed into the following optimization problem:

$$\tilde{c}^i = \arg\min_{c_i}\left\{\left\|x_i^l - \tilde{D}^l c^i\right\|_2^2,\ \left\|c^i\right\|_0 \leq L,\ c_i^i = 0,\right.$$
$$\left. c_j^i \geq 0\ (j \neq i)\right\} \tag{5}$$

where the parameter $L$ controls the sparsity of the reconstruction, and $L$ is the maximal number of coefficients for each input data atom $x_i^l$.

Once the sparse representation coefficient $\tilde{c}^i$ has been computed, the global affinity graph $\tilde{W}_l = (\tilde{w}_{ij}^l)$ is computed by the same method as Eqs. (2) and (3).

### 3.2.3 Building and Fusing Enhanced Affinity Graphs of Different Scales

To comprehensively consider the advantages of the local and global affinity graph, the global affinity graph can be used to enhance the local affinity graph to obtain the enhanced affinity graph $W_{Sl}$. Since $W_l$ and $\tilde{W}_l$ are calculated by similar methods, they have the same status at the digital level and can be added directly to get $W_{Sl}$. The calculation method is as follows:

$$W_{Sl} = W_l + \tilde{W}_l \tag{6}$$

According to the above steps, the enhanced affinity graph $W_{Sl}^{mLab}$ and $W_{Sl}^{LBP}$ can be obtained with the mLab and LBP features respectively, and then $W_{Sl} = (w_{ij}^{Sl})$ is obtained by using the same method as the $\ell_0$-graph [10] to fuse the enhanced affinity graphs of different features. The calculation method is as follows.

$$w_{ij}^{Sl} = \sqrt{(w_{ij}^{mLab})^2 + (w_{ij}^{LBP})^2} \tag{7}$$

Finally, to fuse the enhanced affinity graphs of different scales, we plug each scale enhanced affinity graph $W_{Sl}$ into a block diagonal multiscale affinity matrix $W_{SS}$ as follows:

$$W_{SS} = \begin{pmatrix} W_{S1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & W_{Sl} \end{pmatrix},\ l = 1, 2, 3, \cdots \tag{8}$$

**Algorithm 1**: EA-graph

**Input**: Original image $I$ and the number of segements $k$.
**Output**: A $k$-way segmentation of $I$.

1: Image $I$ is over-segmented by MS and FH algorithm to obtain several sets of superpixels with different scales.
2: Extracting the mLab and LBP features of each superpixel.
3: Building local affinity graph of mLab and LBP by using feature affinity among neighborhood superpixels, respectively.
4: Building global affinity graph of mLab and LBP using sparse recon-struction among all superpixels.
5: The local affinity graph and the global affinity graph are superimposed to obtain the enhanced affinity graphs $W_{Sl}^{mLab}$ and $W_{Sl}^{mLab}$, respectively, and then the $W_{Sl}^{mLab}$ and $W_{Sl}^{LBP}$ are fused by the Eq. (7) to obtain $W_{Sl}$.
6: Fuse the enhanced affinity graph $W_{Sl}$ of all scales to obtain $W_{SS}$.
7: The enhanced affinity graph $W_{SS}$ is transformed into a bipartite graph $G_B = \{U, V, B\}$, and $G_B$ is segmented by Tcuts spectral clustering algorithm.

where $W_{Sl}$ denotes the enhanced affinity graph at scale $l$.

## 3.3 Bipartite Graph Construction and Partition

To achieve the final segmentation of the image, the enhanced affinity graph is transformed into a bipartite graph [9] and segmented by Tcuts spectral clustering algorithm [11]. The bipartite graph can incorporate the affiliation information between pixels and superpixels. The specific steps are as follows: First, we build a bipartite graph $G_B = \{U, V, B\}$ over the pixels set $I$ and superpixels set $SI$, where $U = I \cup SI$, $V = SI$, and $B = [W_{IS}; W_{SS}]$, where $W_{IS} = (b_{ij})_{|I|\times|V|}$ denotes the affiliation matrix between pixels and superpixels.

$$b_{ij} = \begin{cases} \gamma & pixel\ i \in superpixel\ s_j \\ 0 & otherwise \end{cases}, \ \gamma = 0.01 \tag{9}$$

Then, the Tcuts method yields a partition of the bipartite graph $G_B$ into $k$ clusters. More precisely, it provides the bottom $k$ eigenpairs $\{\lambda_i, f_i\}_{i=1}^{k}$ of the following generalized eigenvalue problem over superpixels only:

$$L_V = \lambda D_V f \tag{10}$$

where $L_V = D_V - W_V$, $D_V = diag(B^{\mathrm{T}}\mathbf{1})$, $W_V = B^{\mathrm{T}} D_U^{-1} B$, and $D_U = diag(B\mathbf{1})$

### 3.4 The Overall Framework of the Algorithm

Based on the enhanced affinity graph among superpixels, the overall framework of the image segmentation algorithm proposed in this paper is shown in Fig. 4. The specific calculation process is shown in **Algorithm 1**.

## 4. Experiments and Analysis

To verify the feasibility and effectiveness of the proposed EA-graph. Firstly, we introduce the database and quantitative evaluation criteria used in experiments. Then, the influence of different features and sparsity $L$ is analyzed for image segmentation performance. Finally, the proposed EA-graph is compared with state-of-the-art algorithms. All experiments are carried out under the condition of Windows 7, CPU of Intel Xeon E5-2640 @2.40GHz, 64GB RAM and MATLAB R2017a.

### 4.1 The Database and Evaluation Criteria

All experiments are carried out on the Berkeley Segmentation Database (BSD) [15], which includes 300 natural images with a size of $481 \times 321$ pixels. Each image is manually annotated by different human subjects, and at least four ground truths are available for each image.

To further quantitatively evaluate the effectiveness of the proposed EA-graph, four standard measurements are used for quantitative evaluation: The probabilistic rand index (PRI) [16], The variation of information (VoI) [17], The global consistency error (GCE) [18], and The boundary displacement error (BDE) [19]. The PRI computes the ratio of pixel pairs whose labels are consistent between the segmentation result and the ground truth. The VoI computes the amount of information loss/gain between the compared images. The GCE measures the extent to which two segmentations are mutually consistent. The BDE computes the average displacement error of boundary pixels between two segmentation results. The segmentation is better if PRI is larger and the other three criteria are smaller, when compared to the ground truths.

### 4.2 Results of Fusing Different Visual Features

Since the different features have a direct impact on the segmentation result, we extract the mLab and LBP features of different dimensions and combine them to measure the effectiveness on fusing different visual features. The parameters in all experiments are adjusted to the best. The results on fusing different visual features are shown in Table 1.

**Table 1** Quantitative comparison of different features

| Feature | PRI↑ | VoI↓ | GCE↓ | BDE↓ |
|---|---|---|---|---|
| mLab | 0.8388 | 1.7604 | 0.1983 | 11.7402 |
| LBP(32) | 0.8238 | 1.7828 | 0.2068 | 12.9158 |
| LBP(64) | 0.8221 | 1.7565 | 0.2033 | 12.9694 |
| LBP(128) | 0.8237 | 1.7559 | 0.2028 | 13.4286 |
| LBP(256) | 0.8230 | 1.7524 | 0.2039 | 14.0314 |
| mLab+LBP(32) | 0.8398 | 1.7474 | 0.19778 | 11.4651 |
| mLab+LBP(64) | 0.8452 | 1.7051 | 0.1850 | 11.2190 |
| mLab+LBP(128) | **0.8459** | **1.6774** | **0.1845** | **11.2638** |
| mLab+LBP(256) | 0.8445 | 1.7013 | 0.1854 | 11.2419 |

**Table 2** Quantitative comparison of different sparsity $L$

| Sparsity($L$) | PRI↑ | VoI↓ | GCE↓ | BDE↓ |
|---|---|---|---|---|
| $L$=2 | 0.8428 | 1.7305 | 0.1903 | 11.2816 |
| $L$=3 | **0.8459** | **1.6774** | **0.1845** | **11.2638** |
| $L$=4 | 0.8411 | 1.7344 | 0.1913 | 11.4604 |
| $L$=5 | 0.8419 | 1.7318 | 0.1902 | 11.4248 |
| $L$=6 | 0.8416 | 1.7366 | 0.1911 | 11.4851 |
| $L$=7 | 0.8410 | 1.7409 | 0.1906 | 11.4857 |
| $L$=8 | 0.8400 | 1.7409 | 0.1919 | 11.5376 |

When only a single feature is selected, the segmentation performance is significantly lower than that of the combined feature. The result indicates that a single feature is not sufficient to express the affinity and difference among superpixels. When two features are fused, the segmentation result is affected by the difference dimensions of LBP feature. Specifically, if the dimension of the LBP is too small, the statistics of LBP feature are too concentrated to highlight the differences between superpixels; on the contrary, if the dimension of the LBP is too large, the statistics of the LBP features are too scattered to highlight the similarity between superpixels. In summary, the combination between mLab and 128-dimensions LBP is more helpful to improve the performance of the proposed EA-graph.

### 4.3 Results of Different Sparsity $L$

When the sparse reconstruction among superpixels is computed using Eq. (5), the sparsity $L$ directly determines the number of reconstructed atoms. To verify the influence of parameter $L$, we set different $L$ for comparison to find the best $L$. The same parameters are used for all experimental procedures except that the sparsity $L$ is changed. The results corresponding to different $L$ are shown in Table 2.

The segmentation result is affected by the sparsity $L$. If the sparsity $L$ is too small or too large, the segmentation performance of our method will be weakened. Specifically, if the sparsity $L$ is too small, the global affinity graph does not sufficiently capture the global similarity information among superpixels; on the contrary, if the sparsity $L$ is too large, the similarity information captured by the global affinity graph is too redundant. In summary, considering the accuracy and efficiency of the proposed EA-graph, the sparsity $L$ is set to 3.

**Table 3** Quantitative comparison of the proposed EA-graph with state-of-the-art algorithms on BSD

| Methods | PRI↑ | VoI↓ | GCE↓ | BDE↓ |
|---|---|---|---|---|
| Ncut[3] | 0.7242 | 2.9061 | 0.2232 | 17.15 |
| JSEG[23] | 0.7756 | 2.3217 | 0.1989 | 14.40 |
| MNCut[12] | 0.7559 | 2.4701 | 0.1925 | 15.10 |
| NTP[20] | 0.7984 | 2.1130 | 0.2171 | 13.580 |
| TBES[24] | 0.8000 | 1.7600 | N/A | N/A |
| SDTV[21] | 0.7758 | 1.8165 | **0.1768** | 16.24 |
| LFPA[6] | 0.8146 | 1.8545 | 0.1809 | 12.21 |
| UCM[22] | 0.8100 | 1.6800 | N/A | N/A |
| Context-sensitive(mLab)[25] | 0.7937 | 3.9174 | 0.4165 | **9.9046** |
| Cotransduction(mLab+LBP)[26] | 0.8083 | 2.3644 | 0.2681 | 14.1972 |
| TPG[27] | 0.8227 | 1.7696 | N/A | N/A |
| FusionTP[28] | 0.7771 | 3.3089 | 0.3654 | 13.2428 |
| SAS[9] | 0.8319 | 1.6849 | 0.1779 | 11.2900 |
| $\ell_0$-Graph[10] | 0.8355 | 1.9935 | 0.2297 | 11.1900 |
| *GL*-Graph[8] | 0.8384 | 1.8012 | 0.1934 | 10.6633 |
| Our method | **0.8459** | **1.6774** | 0.1845 | 11.2638 |
| SAS ($\boldsymbol{k=2}$) | 0.6197 | 2.0119 | 0.1106 | 42.2877 |
| $\ell_0$-Graph ($\boldsymbol{k=2}$) | 0.6270 | 2.0299 | 0.1050 | **23.1298** |
| GL-Graph ($\boldsymbol{k=2}$) | 0.6298 | 2.0349 | 0.1120 | 41.8435 |
| Our method ($\boldsymbol{k=2}$) | **0.6459** | **1.9431** | **0.0924** | 28.3079 |

**Fig. 5** Visual results of our method with $k = 2$

### 4.4 Comparison with State-of-the-Art Algorithms

We report quantitative comparison with state-of-the-art algorithms: Ncut [3], Normalized Tree Partitioning (NTP) [20], Multi-scale Ncut (MNcut) [12], Saliency Driven Total Variation (SDTV) [21], Learning Full Pairwise Affinity (LFPA) [6], Ultrasound Contour Map (UCM) [15], JSEG [22], Texture and Boundary Encoding-based Segmentation (TBES) [23], Context-sensitive [24], Cotransduction [25], Tensor Product Graph (TPG) [26], Fusion with TPG [27], SAS [9], $\ell_0$-graph [10], and *GL*-graph [8]. The parameters of all methods are adjusted to the optimal, and the quantitative results are shown in Table 3, where we highlight in bold the best result for each qualitative criterion.

As shown in Table 3, we can find that our method ranks in the first place with PRI and VoI, and is competitive with others in terms of GCE and BDE, in particular the gain is significant for PRI and VoI. However, these scores are computed by manually adjusting the $k$ of each image to select the best results, which has no practical meaning in the specific application. Thus, to demonstrate the obvious advantage of our method related to $k$ in practical applications, we compare the average scores of SAS, $\ell_0$-Graph, *GL*-Graph, and our method by fixing $k = 2$ for all images on the BSD. Those scores are shown in the lower part of Table 3, and we show more visual segmentation results of the proposed method with $k = 2$ as shown in Fig. 5. Table 3 shows that our method is superior to other related algorithms with PRI, VoI, and GCE, in particular the gain is significant for PRI. In Fig. 5, our method can first segment the target object when the parameter $k$ is set to 2.

The main reason is that our method combines the local and global information among the superpixels, which helps to separate the foreground and background of the image. The advantages of our method are better reflected in the following 3 cases. 1) the detected object is tiny (as seen in the first row); 2) multiple objects are needed to segment in the same image (as seen in the middle row); 3) the color of background and object are quite similar (as seen in the last row).

## 5. Conclusion

We proposed an enhanced affinity graph for unsupervised image segmentation, which can capture local and global

information between superpixels. The proposed method mainly uses superpixels with different scales as the segmentation nodes to build the affinity graph model. Meanwhile, the sparse reconstruction of the global dictionary is used to enhance the local affinity graph to obtain the enhanced affinity graph, and the clustering is performed with Tcuts. Experimental results on the BSD database show that our method has good segmentation performance and competitive advantage, in particular PRI and VoI rank first in comparison with other state-of-the-art methods.

## Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (Grant 51775177, Grant 51675166), and the Natural Science Foundation of Jiangsu Province (Grant BK20150016).

### References

[1] H.Y. Zhu, F.M. Meng, J.F. Cai, and S.J. Lu, "Beyond pixels: a comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," J. Vis. Commun. Image Represent., vol.34, pp.12–27, Jan. 2016. DOI: 10.1016/j.jvcir.2015.10.012

[2] S. Zhang, W. Jiang, and S. Satoh, "Multilevel thresholding color image segmentation using a modified artificial bee colony algorithm," IEICE Trans. Inf. & Syst., vol.E101-D, no.3, pp.2064–2071, Aug. 2018. DOI: 10.1587/transinf.2017EDP7183

[3] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.8, pp.888–905, Aug. 2000.

[4] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," Int. J. Comput. Vis., vol.59, no.2, pp.167–181, Sept. 2004.

[5] C. Fowlkes, D.R. Martin, and J. Malik, "Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Madison, WI, USA, pp.II-54–II-64, June 2003. DOI: 10.1109/CVPR.2003.1211452

[6] T.H. Kim, K.M. Lee, and S.U. Lee, "Learning full pairwise affinities for spectral segmentation," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., San Francisco, CA, USA, pp.2101–2108, June 2010. DOI: 10.1109/CVPR.2010.5539888

[7] B. Wang and Z. Tu, "Affinity learning via self-diffusion for image segmentation and clustering," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Providence, RI USA, pp.2312–2319, June 2012. DOI: 10.1109/CVPR.2012.6247942

[8] X.F. Wang, Y.X. Tang, S. Masnou, and L.M. Chen, "A global/local affinity graph for image segmentation," IEEE Trans. Image Process., vol.24, no.4, pp.1399–1411, April 2015. DOI: 10.1109/TIP.2015.2397313

[9] Z. Li, X.-M. Wu, and S.-F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Washington, DC, USA, pp.789–796, June 2012.

[10] X. Wang, H. Li, C.-E. Bichot, S. Masnou, and L. Chen, "A graph-cut approach to image segmentation using an affinity graph based on $\ell_0$-sparse representation of features," Proc. 20th IEEE Int. Conf. Image Process., Melbourne, Australia, pp.4019–4023, Sept. 2013.

[11] U. von Luxburg, "A tutorial on spectral clustering," Statist. Comput., vol.17, no.4, pp.395–416, Dec. 2007.

[12] T. Cour, F. Bénézit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., San Diego, CA, USA, pp.1124–1131, June 2005. DOI: 10.1109/CVPR.2005.332

[13] B. Cheng, G. Liu, J. Wang, Z. Huang, and S. Yan, "Multi-task low-rank affinity pursuit for image segmentation," Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Barcelona, Spain, pp.2439–2446, Nov. 2011. DOI: 10.1109/ICCV.2011.6126528

[14] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.5, pp.603–619, May 2002. DOI: 10.1109/34.1000236

[15] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.5, pp.898–916, May 2011. DOI: 10.1109/TPAMI.2010.161

[16] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.6, pp.929–944, June 2007. DOI: 10.1109/TPAMI.2007.1046

[17] M. Meilă, "Comparing clusterings: An axiomatic view," Proc. 22nd Int. Conf. Mach. Learn., New York, NY, USA, pp.577–584, Aug. 2005. DOI: 10.1145/1102351.1102424

[18] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," Proc. 8th IEEE Int. Conf. Comput. Vis., pp.416–425, July 2001.

[19] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," Proc. 7th Eur. Conf. Comput. Vis., pp.408–422, April 2002. DOI: 10.1007/3-540-47977-5_27

[20] J. Wang, Y. Jia, X.-S. Hua, C. Zhang, and L. Quan, "Normalized tree partitioning for image segmentation," Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Anchorage, AK, USA, pp.1–8, June 2008. DOI: 10.1109/CVPR.2008.4587454

[21] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," Proc. IEEE 12th Int. Conf. Comput. Vis., Kyoto, Japan, pp.817–824, Sept./Oct. 2009. DOI: 10.1109/ICCV.2009.5459296

[22] Y. Deng and B.S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.8, pp.800–810, Aug. 2001. DOI: 10.1109/34.946985

[23] S.R. Rao, H. Mobahi, A.Y. Yang, S.S. Sastry, and Y. Ma, "Natural image segmentation with adaptive texture and boundary encoding," Proc. 9th Asian Conf. Comput. Vis., pp.135–146, Sept. 2009. DOI: 10.1007/978-3-642-12307-8_13

[24] X. Bai, X. Yang, L.J. Latecki, W. Liu, and Z. Tu, "Learning context-sensitive shape similarity by graph transduction," IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.5, pp.861–874, May 2010. DOI: 10.1109/TPAMI.2009.85

[25] X. Bai, B. Wang, C. Yao, W. Liu, and Z. Tu, "Co-transduction for shape retrieval," IEEE Trans. Image Process., vol.21, no.5, pp.2747–2757, May 2012.

[26] X. Yang, L. Prasad, and L.J. Latecki, "Affinity learning with diffusion on tensor product graph," IEEE Trans. Pattern Anal. Mach. Intell., vol.35, no.1, pp.28–38, Jan. 2013. DOI: 10.1109/TPAMI.2012.60

[27] Y. Zhou, X. Bai, W. Liu, and L.J. Latecki, "Fusion with diffusion for robust visual tracking," Advances in Neural Information Processing Systems 25, pp.2978–2986, Curran Associates, Red Hook, NY, USA, 2012.

**Guodong Sun** is a professor in the School of Mechanical Engineering at the Hubei University of Technology. He received his BS degree in energy and power engineering and his PhD degree in mechanical and electronic engineering from Huazhong University of Science and Technology in 2002 and 2008, respectively. His current research interests include machine vision and imaging processing.

**Kai Lin** received the B.S. degree from Hubei University of Technology, Wuhan, China, in 2016. He is currently pursuing the master degree at the School of Mechanical Engineering, Hubei University of Technology, Wuhan, China. His current research directions are machine vision and machine learning.

**Junhao Wang** received the B.S. degree from Hubei University of Technology, Wuhan, China, in 2017. He is currently pursuing the master degree at the School of Mechanical Engineering, Hubei University of Technology, Wuhan, China. His current research directions are machine vision and machine learning.

**Yang Zhang** received the B.S. degree and the M.S. degree from Hubei University of Technology, Wuhan, China, in 2014 and 2017. He is currently pursuing the Ph.D. degree in the National Key Laboratory for Novel Software Technology, Department of Computer Science and Technology, Nanjing University, Nanjing, China. His current research interests are machine learning and computer vision.