AN ADAPTIVE AFFINITY GRAPH WITH SUBSPACE PURSUIT FOR NATURAL IMAGE SEGMENTATION

Yang Zhang[1],*Huiming Zhang*[1],*Yanwen Guo*[1,2,*], *Kai Lin*[3], *Jingwu He*[1]

[1] National Key Laboratory for Novel Software Technology, Nanjing University
[2] Science and Technology on Information Systems Engineering Laboratory
[3] School of Mechanical Engineering, Hubei University of Technology

ABSTRACT

Graph-based segmentation methods have become a major trend in computer vision. Due to the advantages of assimilating different graphs, a multi-scale fusion graph have a better performance than a single graph with single-scale. However, it is not reliable to determine a principle of graph combination. In this paper, we propose an adaptive affinity graph with subspace pursuit (AASP-graph) for natural image segmentation. The input image is first over-segmented into superpixels at different scales. An improved affinity propagation clustering method is proposed to select global nodes of these superpixels adaptively. Then, a ℓ_0-graph at each scale is obtained by a sparse representation of global nodes based on subspace pursuit. The adjacency-graph is finally built upon all superpixels of each scale and updated by the ℓ_0-graph. Experimental results on the Berkeley segmentation database show the effectiveness of the proposed AASP-graph in comparison with state-of-the-art approaches.

Index Terms— Image segmentation, graph, subspace pursuit, sparse representation, affinity propagation

1. INTRODUCTION

As a fundamental and challenging task in computer vision, image segmentation is a process of decomposing an image into independent regions, which plays an important role in many high-level applications [1]. Unsupervised methods receive much attention because they require no prior knowledge. In the literature, many unsupervised segmentation methods have been intensively studied [2]. Some representative works of them rely on constructing a reliable affinity graph [3, 4] for the representation of image content, achieving superior performance.

Clearly, for these affinity graph-based methods, the segmentation performance significantly depends on the effectiveness of the constructed affinity graph, with particular emphasis on the neighborhood topology and pairwise affinities between nodes (*i.e.*, pixels or superpixels). As pointed out in GL-graph [4], the graph built by connecting local nodes and global nodes can facilitate a better segmentation result compared with a single graph (*i.e.*, adjacency-graph [3]) as shown in Fig. 1. However, the local and global nodes are not easily defined because superpixel features vary at different scales. In addition, a global affinity graph built with global nodes is a dense graph, incurring high computation cost.

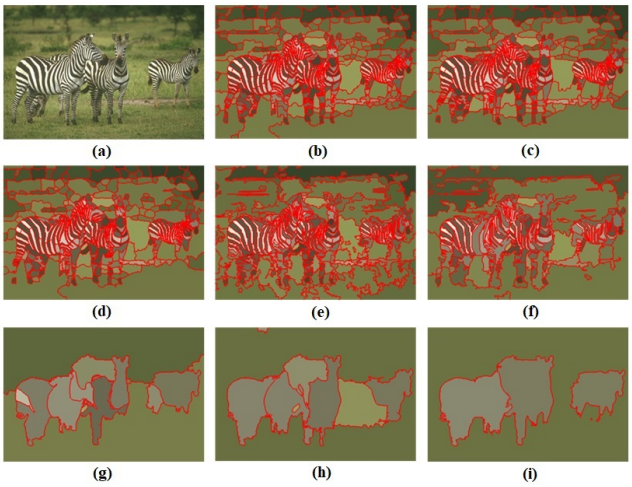

Fig. 1. Segmentation using superpixels. (a) Input image. (b-f) Superpixels generated by over-segmenting the image. (g-i) Segmentation results by adjacency-graph [3], GL-graph [4], and the proposed AASP-graph, respectively.

To solve these two problems, an *adaptive affinity graph with subspace pursuit* (AASP-graph) is proposed to combine the local and global nodes of superpixels at different scales based on affinity propagation. The superpixels of an input image are obtained by over-segmenting the input image at different scales. And then, the affinity propagation clustering (APC) is employed to select global nodes of these superpixels adaptively. Moreover, the subspace pursuit is used to obtain a ℓ_0-graph at each scale via a sparse representation of the mLab features of global nodes. All superpixels at each scale are used to build an adjacency-graph, which are further updated by the ℓ_0-graph. We introduce a bipartite graph to map the re-

*Corresponding author. E-mail address: ywguo@nju.edu.cn.

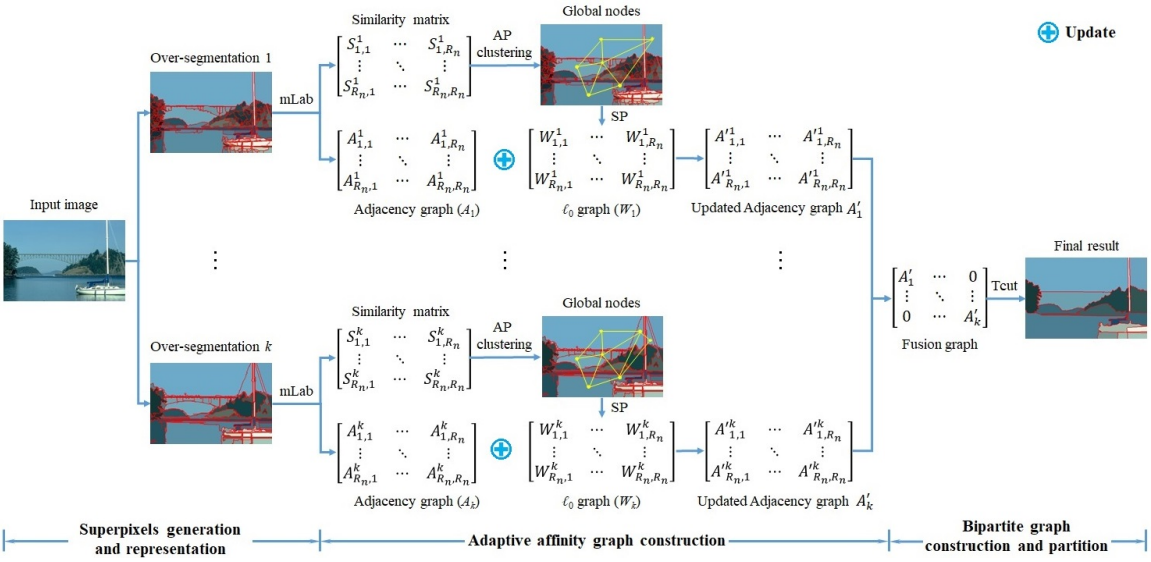

Fig. 2. The overall framework of image segmentation based on the proposed AASP-graph. After over-segmenting the input image as the SAS method, we obtain superpixels at k different scales. Global nodes of superpixels are sorted out through APC and then build the ℓ_0-graph by SP. The adjacency-graph is constructed by all superpixels and it is updated by ℓ_0-graph at each scale. The updated graphs are fused to obtain the final result through the Tcut.

lationship between the original image pixels and superpixels and enable propagation of grouping cues across superpixels at different scales. Intensive experiments are conducted on the Berkeley segmentation database with four metrics, including PRI, VoI, GCE, and BDE for quantitative comparisons.

This work makes the following contributions. First, we construct an AASP-graph to adaptively combine different graphs with the sparsity and a high discriminative power for natural image segmentation. Second, after reducing noise, we apply the APC to decide the global nodes, accurately mining the feature distribution of superpixels. Third, the results show the effectiveness of our method compared with the state-of-the-art approaches.

2. RELATED WORKS

The unsupervised methods segment images without any human intervention. As we focus on unsupervised image segmentation in this paper, a review of related unsupervised methods is only introduced in this section. A more review of the image segmentation process can be found in [1].

The essence of image segmentation can be regarded as a clustering problem, which groups the pixels into local homogenous regions. Some clustering-based methods such as k-means, mean-shift (MS) [5], and region merging [6], are the typical examples. Besides clustering-based methods, the graph-based method has become one of the most popular image segmentation methods which can be regarded as image perceptual grouping and organization methods. Graph-based methods are based on the fusion of the feature and spatial information, such as normalized cut (Ncut) [7], ℓ_0-graph [8], and multi-scale Ncut (MNcut) [9], etc.

More recently, Wang *et al.* [10] introduced the normalized tree partitioning and the average tree partitioning to optimize normalized and average cut over a tree for image segmentation. Especially, Li *et al.* [2] proposed a region adjacency cracking method to adaptively loosen the color labeling constraints. Moreover, a heuristic four-color labeling algorithm was used to establish a uniform appearance of those regions with global consistency. In this method, affinity propagation clustering is applied to crack the adjacent constraint, which is suitable for the task that does not know the distribution of the regional feature beforehand.

Wang *et al.* [4] proposed a GL-graph over superpixels to enable propagation of different grouping cues between superpixels of different scales based on a bipartite graph [3]. This method divides superpixels into small-, medium-, and large-sized sets based on their area. Small- and large-sized superpixels are defined as local nodes which are used to achieve local smoothness through an adjacency-graph. Medium-sized superpixels regarded as global nodes are used to obtain global grouping through a sparse representation by solving a ℓ_0-minimization problem. So, the combination of the global graph and the local graph over superpixels can obtain a better segmentation result.

3. METHOD

The overall framework of the proposed AASP-graph approach for image segmentation is shown in Fig. 2. The

proposed approach primarily consists of three components: superpixels generation and representation, adaptive affinity graph construction, and fusion graph partition.

3.1. Superpixels generation and representation

Various and multiscale visual patterns of a natural image can be captured through superpixels generated by different methods with different parameters. In Fig. 2, an input image is over-segmented into superpixels at k different scales (*e.g.*, 5 scales in Fig. 1) as the SAS [3] in the sequel. And then, the features of each superpixel are calculated to obtain a discriminative and adaptive affinity graph. Specifically, in this work, color feature is characterized by using mean value in the CIE L*a*b* space (mLab). Unlike RGB, Lab color space approximates human vision and its L component closely matches human perception of lightness.

For graph-based image segmentation, the basic principle is to approximate each superpixel in the feature space into a linear combination of other superpixels, which are regarded as neighbors, and their affinities are calculated from the corresponding representation error [4]. Let SI_k=$\{si_i\}_{i=1}^{R_n}$ be the superpixels of an input image I at scale k, where R_n is the number of superpixels. Formally, such an approximation can be written as:

$$d_i = Dc_i, \quad c_{ii} = 0 \tag{1}$$

where $c_i \in R^n$ is the sparse representation of superpixels, and $d_i \in R^m$ over the dictionary D which is a matrix representation of superpixels. The constraint $c_{ii} = 0$ prevents the self-representation of d_i.

3.2. Adaptive affinity graph construction

Selecting global nodes of superpixels: In general, all natural images contain the noise, which influences the accuracy of pairwise similarities and further hurts the affinity propagation. To solve this problem, we use an improved kernel density estimation (IKDE) to estimate color features of natural images [12]. After IKDE processing, an exponential smoothing is applied to further reduce the noise of the features. The IKDE can improve the reliability of features as explained in Section 4.2. As shown in Figs. 1(b) to 1(f), the features of superpixels, especially the areas, are various at different scales. The global nodes can not be easy to be defined in natural images. To address this issue, we apply APC [13] with the mLab feature f_i of superpixels to find global nodes. It can be noticed that APC is very suitable for the task that we know nothing about the feature distribution of the superpixels beforehand [2]. The APC can adaptively decide the global nodes of superpixels according to the relationship of the given mLab features. We will give a more detailed discussion about the reliability of APC in Section 4.2.

Building a ℓ_0-graph: To capture long-range grouping cues, we utilize Eq. (1) to approximate each global node from others in the mLab features. It is proved that the sparsest solution of Eq. (1) measured in the sense of ℓ_0-norm is unique and conveys the most meaningful information of a signal. So, the sparse solution can be regarded as followings:

$$min||c_i||_0 \quad s.t. \quad d_i = Dc_i, \quad c_{ii} = 0 \tag{2}$$

where $||\cdot||_0$ represents the ℓ_0-norm, which calculates the number of nonzero values in the vector.

Hence, we can solve Eq. (2) using subspace pursuit (SP) [14], which is a simple and fast greedy method to seek an approximation of the sparsest solution.

$$\tilde{c_i} = argmin_{c_i} \left\{ ||f_i - M_W c_i||_2^2, ||c_i||_0 \leq LC, c_{ii} = 0 \right\} \tag{3}$$

where the parameter LC is the maximal number of coefficients for each input mLab feature vector f_i, which controls the sparsity of the matrix-representation M_W. The default value of LC is set as three according to dimension of mLab features. Once the above representation is implemented, the affinities coefficient $W_{i,j}$ of their ℓ_0-graph W between superpixels si_i and si_j $(i, j = 1, 2, ..., R_n)$ can be computed as follows:

$$W_{i,j} = \begin{cases} 1 & if \quad i = j \\ 1 - (r_{i,j} + r_{j,i})/2 & if \quad i \neq j \end{cases} \tag{4}$$

where

$$r_{i,j} = ||f_i - c_{i,j} f_j||_2^2 \tag{5}$$

Building an adjacency-graph: As for all superpixels at each scale, every superpixel is connected to its adjacent superpixels, denoted as adjacency-graph. To facilitate combination between the adjacency-graph and the ℓ_0-graph, we unify the similarities calculation principles. Let M_A be the matrix-representation of all its adjacent neighbors, we attempt to represent f_i as a linear combination of elements in M_A. In practice, we solve the following optimization problem:

$$\tilde{c_i^*} = argmin_{c_i^*} ||f_i - M_A c_i^*||_2 \tag{6}$$

If a minimizer $\tilde{c_i^*}$ has been obtained, the affinities coefficient $A_{i,j}$ of the affinities A between a superpixel and its graph are calculated as in Eqs. (4) and (5).

Building a fusion graph: Under the above unified similarities calculation principles, the ℓ_0-graph W is used to replace the adjacency-graph A at global nodes for combining local graph and global graph to obtain the updated adjacency-graph A' at a certain scale. Furthermore, to fuse all scales of superpixels, we plug each scale affinity matrix A'_k corresponding to its AASP-graph into a block diagonal multiscale affinity matrix A_{SS} as follows:

$$A_{SS} = \begin{pmatrix} A'_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A'_k \end{pmatrix} \tag{7}$$

Algorithm 1 AASP-graph for natural image segmentation

Input: Source image I, scale k, group k_T.
Output: Pixel-wise labels.

1: Over-segment the image I by the MS and FH methods and obtain superpixels of k different scales;
2: Select global nodes of superpixels based on IKDE and APC;
3: Build the ℓ_0-graph by the mLab features extracted from global nodes of superpixels with SP;
4: Construct the adjacency-graph by the mLab features extracted from all superpixels;
5: Update the adjacency-graph by ℓ_0-graph at each scales;
6: Fuse updated graphs at k different scales are computed to obtain the final segmentation result (pixel-wise labels) through Tcut with the group k_T.

The above matrix A_{SS} gathers all the informative intra-scale similarities for grouping. Furthermore, a bipartite graph is constructed to enable propagation of long-range grouping cues across scales, as introduced in the next subsection.

3.3. Bipartite graph construction and partition

To map the relationships between pixels and superpixels and enable propagation of grouping cues across superpixels at different scales, a bipartite graph is built to describe the relationships of pixels to superpixels and superpixels to superpixels. Formally, let $G = \{U, V, B\}$ denote the above bipartite graph with node set $U \cup V$, where $U := I \cup SI$ and $V := SI$. The across-affinity matrix is defined as $B = \begin{bmatrix} A_{IS} \\ A_{SS} \end{bmatrix}$, where $A_{IS} = (a_{ij})_{|I|\times|SI|}$ are the relationships between pixels and superpixels with $a_{ij} = 0.001$, if a pixel i belongs to a superpixel j; $a_{ij} = 0$, otherwise.

Given the above graph G, the task is to partition it into k_T groups, where k_T is manually specified. Various techniques can be applied to fulfill the task, such as spectral clustering algorithms [15]. Spectral clustering capture essential clustering structure of a graph utilizing the spectrum of graph Laplacian. In this case, the bipartite graph is unbalanced, which is suitable for solving it with the Transfer cuts (Tcut) algorithm. Based on superpixels, the proposed adaptive affinity graph, and the bipartite graph, the overall AASP-graph scheme for natural image segmentation is summarized in **Algorithm 1**.

4. EXPERIMENTS AND ANALYSIS

To verify the performance of the proposed AASP-graph, we first introduce the Berkeley Segmentation Database and evaluation metrics used in this section, and then show the results of our approach and its variations. Finally, the results are presented compared with the state-of-the-art methods.

Table 1. Quantitative comparison for different modules using in AASP-graph baseline.

APC	IKDE	LKM	PRI↑	VOI↓	GCE↓	BDE↓
		√	0.8442	1.6530	0.1741	14.6827
	√	√	0.8442	1.6532	0.1740	14.6807
√		√	0.8442	1.6529	0.1739	14.6761
√	√	√	**0.8446**	**1.6485**	**0.1737**	**14.6416**

4.1. Experimental setup

All experiments are carried out on the Berkeley Segmentation Database (BSD) [16], which includes 300 images and the ground truth data. Each image has a fixed size of 481×321 pixels. To evaluate our method, there are four standard measurements: the probabilistic rand index (PRI) [17], the variation of information (VoI) [18], the global consistency error (GCE) [16], and the boundary displacement error (BDE) [19]. The setting of the scale k and group k_T are the same as those in the GL-graph [4]. The experiment is conducted under the PC condition of 3.40GHz of Intel Xeon E5-2643 v4 processor, 64G RAM, Ubuntu 16.04 OS and Matlab 2018a.

4.2. Results of different modules using in AASP-graph

To see how AASP-graph is affected by different modules, we report the scores of combinations of different modules using in AASP-graph baseline. Specially, if the baseline does not select the APC, we use the k-means with 2 clustering number for fair comparison. The Lite k-means (LKM) [15] is used in Tcut method, respectively. All other parameters are consistent, including the number of segments for each image.

The results of different methods using in the baseline are shown in Table 1. Our method is stable for the selection of global nodes and robust to the noise in the mLab feature space. In addition, there are no obvious changes in the performance of our method before and after using the IKDE algorithm. So, the IKDE does not affect the original structure of mLab feature while filtering noise. This approach can further improve the reliability of mLab features.

4.3. Comparison with state-of-the-art methods

We also report quantitative comparison with Ncut [7], MNcut [9], NTP [23], JSEG [20], HIS-FEM [21], heuristic better and random better (H_+R_Better) [2], co-transduction [24], higher-order correlation clustering (HO-CC) [25], LFPA [26], Fusion TPG [22], SAS [3], ℓ_0-graph [8], and GL-graph [4]. The results are shown in Table 2, where we highlight in bold the best result for each qualitative criterion. We directly take their evaluations reported in publications for fair comparison.

From the Table 2, it can be noticed that our method has a better performance compared with the state-of-the-art meth-

Fig. 3. Visual comparison obtained with the SAS, ℓ_0-graph, GL-graph, and AASP-graph. Two columns of the comparison results are shown here. From left to right, input images, the results of the SAS, ℓ_0-graph, GL-graph, and AASP-graph are presented. The results of AASP-graph are visually better, in particular often more accurate.

Table 2. Quantitative results of the proposed method with state-of-the-art approaches on BSDS300 dataset.

Methods	PRI↑	VOI↓	GCE↓	BDE↓
Ncut[7]	0.7242	2.9061	0.2232	17.15
MNCut[9]	0.7559	2.4701	0.1925	15.1
JSEG[20]	0.7756	2.3217	0.1989	14.4
HIS-FEM[21]	0.7769	2.3067	0.2215	10.66
FusionTPG[22]	0.7771	3.3089	0.3654	13.2428
NTP[23]	0.7984	2.113	0.2171	13.58
H_+R_Better[2]	0.8073	1.826	0.2079	12.16
Co-transduction[24]	0.8083	2.3644	0.2681	14.1972
HO-CC[25]	0.8140	1.743	N/A	10.377
LFPA[26]	0.8146	1.8545	0.1809	12.21
SAS[3]	0.8319	1.6849	0.1779	11.29
ℓ_0-Graph[8]	0.8355	1.9935	0.2297	11.19
GL-Graph[4]	0.8384	1.8012	0.1934	10.6633
AASP-Graph	**0.8446**	**1.6485**	**0.1737**	14.6416

ods, in particular the gain is significant for PRI, VoI, and GCE (ranks the first in PRI, GCE, and VoI). In particular, our method only uses pixel color information, which fails to capture enough contour cues of segmenting images. In addition, our method follows a similar, but not identical, strategy as the SAS, ℓ_0-graph, and GL-graph. Instead of using only adjacent neighborhoods of superpixels in SAS and affinity graph of superpixels in ℓ_0-graph, we build an AASP-graph combining ℓ_0-graph and adjacency-graph, which allows the constructed graph to have the characteristics of a long-range neighborhood topology. The main differences between GL-graph and our method are the global nodes selection and their graph construction. In GL-graph, the superpixels are simply divided into three parts according to their area. In the AASP-graph, the global nodes are selected adaptively by the APC and further used to build the ℓ_0-graph through SP.

Figure 3 shows various segmentation results obtained with the SAS, ℓ_0-graph, GL-graph, and our AASP-graph. The visual results suggest that our approach generates the visually pleasing results in most cases. We attribute this to the adaptive combination of the global graph and the local graph, which helps to accurately separate the foreground and background of natural images.

5. CONCLUSION

An AASP-graph is proposed to obtain good results for natural image segmentation. The method uses superpixels of different scales as segmentation primitive. Besides, the improved APC is applied to adaptively select global sets, which are used to build ℓ_0-graph with an SP method to update the adjacency-graph of all superpixels. Experimental results on a large number of images show the good performance and high efficiency of the proposed method. We also compare our method with the state-of-the-art, and our method achieves competitive results, such as ranking the first in PRI, VoI, and GCE.

6. ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (Grant 61772257 and 61672279), and the Fundamental Research Funds for the Central Universities 020214380058.

7. REFERENCES

[1] H. Zhu, F. Meng, J. Cai, and S. Lu, "Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," *JVCIR*, vol. 34, pp. 12–27, 2016.

[2] K. Li, W. Tao, X. Liu, and L. Liu, "Iterative image segmentation with feature driven heuristic four-color labeling," *PR*, vol. 76, pp. 69–79, 2018.

[3] Z. Li, X. Wu, and S. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *CVPR*, 2012, pp. 789–796.

[4] X. Wang, Y. Tang, S. Masnou, and L. Chen, "A global/local affinity graph for image segmentation," *TIP*, vol. 24, no. 4, pp. 1399–1411, 2015.

[5] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *TPAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[6] S. Li and D. Wu, "Modularity-based image segmentation," *TCSVT*, vol. 25, no. 4, pp. 570–581, 2015.

[7] J. Shi and J. Malik, "Normalized cuts and image segmentation," *TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[8] X. Wang, H. Li, C. Bichot, S. Masnou, and L. Chen, "A graph-cut approach to image segmentation using an affinity graph based on ℓ_0-sparse representation of features," in *ICIP*, 2013, pp. 4019–4023.

[9] T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *CVPR*, 2005, pp. 1124–1131.

[10] J. Wang, H. Jiang, Y. Jia, X. Hua, C. Zhang, and L. Quan, "Regularized tree partitioning and its application to unsupervised image segmentation.," *TIP*, vol. 23, no. 4, pp. 1909–1922, 2014.

[11] A. Saglam and N.-A. Baykan, "Sequential image segmentation based on minimum spanning tree representation," *PRL*, vol. 87, pp. 155–162, 2017.

[12] B. Foster, U. Bagci, Z. Xu, B. Dey, B. Luna, W. Bishai, S. Jain, and D.-J. Mollura, "Segmentation of PET images for computer-aided functional quantification of tuberculosis in small animal models.," *TBE*, vol. 61, no. 3, pp. 711–724, 2014.

[13] B.-J. Frey and D Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[14] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *TIT*, vol. 55, no. 5, pp. 2230–2249, 2009.

[15] D. Cai and X. Chen, "Large scale spectral clustering via landmark-based sparse representation," *TCYB*, vol. 45, no. 8, pp. 1669–1680, 2015.

[16] 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," in *ICCV*, 2001, pp. 416–423.

[17] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *TPAMI*, vol. 29, no. 6, pp. 929–944, 2007.

[18] M. Meilă, "Comparing clusterings—an information based distance," *JMVA*, vol. 98, no. 5, pp. 873–895, 2007.

[19] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí, "Yet another survey on image segmentation: Region and boundary information integration," in *ECCV*, 2002, pp. 408–422.

[20] M. Donoser, M. Urschler, M. Hirzer, and H. Bischof, "Saliency driven total variation segmentation," in *ICCV*, 2009, pp. 817–824.

[21] S. Yin, Y. Qian, and M. Gong, "Unsupervised hierarchical image segmentation through fuzzy entropy maximization," *PR*, vol. 68, pp. 245–259, 2017.

[22] Y. Zhou, X. Bai, W. Liu, and L.-J. Latecki, "Fusion with diffusion for robust visual tracking," in *NIPS*, 2012, pp. 2978–2986.

[23] J. Wang, Y. Jia, X. Hua, C. Zhang, and L. Quan, "Normalized tree partitioning for image segmentation," in *CVPR*, 2008, pp. 1–8.

[24] X. Bai, B. Wang, C. Yao, W. Liu, and Z. Tu, "Co-transduction for shape retrieval," *TIP*, vol. 21, no. 5, pp. 2747–2757, 2012.

[25] S. Kim, C.-D. Yoo, S. Nowozin, and P. Kohli, "Image segmentation using higher-order correlation clustering," *TPAMI*, vol. 36, no. 9, pp. 1761–1774, 2014.

[26] T.-H. Kim, K.-M. Lee, and S.-U. Lee, "Learning full pairwise affinities for spectral segmentation," *TPAMI*, vol. 35, no. 7, pp. 1690–1703, 2013.