Boundary-Aware Superpixel Segmentation Based on Minimum Spanning Tree

Li Xu††††, Member, Bing Luo†††, Student Member, and Zheng Pei††††, Nonmember

SUMMARY In this paper, we propose a boundary-aware superpixel segmentation method, which could quickly and exactly extract superpixel with a non-iteration framework. The basic idea is to construct a minimum spanning tree (MST) based on structure edge to measure the local similarity among pixels, and then label each pixel as the index with shortest path seeds. Intuitively, we first construct MST on the original pixels with boundary feature to calculate the similarity of adjacent pixels. Then the geodesic distance between pixels can be exactly obtained based on two-round tree recursions. We determinate pixel label as the shortest path seed index. Experimental results on BSD500 segmentation benchmark demonstrate the proposed method obtains best performance compared with seven state-of-the-art methods. Especially for the low density situation, our method can obtain the boundary-aware oversegmentation region.

We divide the existing methods into two classes. The first class is Euclid-distance based similarity metric, such as Ncut [4], SLIC [5], LSC [6] and SNIC [7]. The second class is geodesic distance based similarity metric, such as RPS [8] and FH [9]. The former evaluates similarity metric based on weight Euclid-distance between color and location features, which is sensitive to low contrast region and results into high under-segmentation error. Meanwhile, Euclid-distance based method cannot handle non-sphere cluster and sensitive to irregular object shape. The latter evaluates similarity among pixels by geodesic shortest path. Pixels in the same object region will have similar geodesic distance to seeds, which lead into the same superpixel labels. Hence, those methods are robust to variance of object shape. FH [9] models superpixel segmentation as the merge process of MST. Due to the definition of proper refinement of segmentation, it is sensitive to pixel intensity and can not explicitely control superpixel number. RPS [8] searches shortest path between adjacent junction pair to fit the object boundary. However, the uniformly grid seed and the constraint by connecting vertical and horizontal junctions result into high under-segmentation error, especially for the fine structure object. In contrast, we calculate geodesic distance based on MST, which not only significantly reduces candidate paths between any two pixels, but also avoids the inaccurate merging process and preserves the fine structure of irregular object.

In this paper, a novel geodesic distance-based superpixel segmentation method is proposed, of which the basic idea is to measure the distance among pixels along minimum spanning tree. The pixel will be labeled as the seed indication with shortest path. Similarity measure based on a tree structure significantly reduce the irrelevant potential paths from pixel to seeds, which lead to a low time complexity for similarity evaluation. The structure edge [10] is utilized to construct boundary-aware tree structure, which make the MST to fit the object boundary. The low-level information, such as color, is sensitive to noise and blurred edge. We use the object boundary map to weight the color distance, which could obtain the MST with object structure property. Then we introduce two round recursive searching to label image pixels quickly. For low dense superpixel, the proposed method could well preserve the object structure information. Some subjective results for our method are shown in Fig. 1. Experimental results on BSD500 segmentation benchmark demonstrate that our method provides better segmentation performance compared with seven state-of-the-art methods.
2. Proposed Method

2.1 Problem Formulation

Given an image \( I \), we build an undirected weighted graph \( G = (\mathcal{V}, \mathcal{E}) \) based on the 4-neighborhood of pixel. Define vertices as image pixel set \( \{p_i\} \) and a virtual node \( O \), i.e., \( \mathcal{V} = \{p_i\} \cup \{O\} \). The virtual node \( O \) is a sink node to connect with superpixel seeds set \( S = \{s_1, \ldots, s_K\} \), which is a subset of \( \mathcal{V} \). \( K \) denotes the number of seeds. The edges consist of two sets, i.e., adjacent edges between image pixels and dummy edges from seeds to sink, \( \mathcal{E} = \{(p_i, p_j)|p_i, p_j \text{ is a neighborhood of } p_j\} \cup \{(p_i, O)|p_i \in S\} \). We model superpixel segmentation as a geodesic label assignment problem [1], [11], i.e., the superpixel label of a pixel \( p \) is assigned as the seed indication who is along the shortest path from \( p \) to sink \( O \) on graph \( \mathcal{V} \):

\[
d(p, O) = \min_{p_i=p, p_2, \ldots, p_n=O} \sum_{i=1}^{n-1} e(p_i, p_{i+1}) \tag{1}
\]

s.t. \((p_i, p_{i+1}) \in \mathcal{E}\)

where \( e(p_i, p_{i+1}) = e_c(p_i, p_{i+1}) \cdot e_b(p_i, p_{i+1}) \) is consisted of color distance \( e_c \) and boundary distance \( e_b \) between image adjacent pixels. For dummy edges from seeds to sink, we set \( e(p_i, p_{i+1}) = 0 \) uniformly.

\[
L(p) = \arg \min_{p \in S} d(p, O), \quad S = \{s_1, \ldots, s_K\} \tag{2}
\]

Figure 2 shows the graph construction for image pixels. There is a shortest path from pixel \( p \) to sink \( O \), which pass through seed \( s_m \). In terms of pixel-level nodes, optimizing Eqs. (1) and (2) need calculate huges of candidate paths, which is time consuming. It should develop a fast and efficient algorithm to accelerate this label optimization.

2.2 Geodesic Label Based on Minimum Spanning Tree

Fortunately, MST is a connected acyclic graph. There is one and only one path to connect any of two nodes, which significantly reduce candidate paths from any pixel node to the sink node [12]. The virtual node \( O \) is sink node and the weight of all dummy edges are 0. We delete the sink \( O \) to construct the MST on image pixels, which does not change the result. Furthermore, the distance among pixels can be recursively updated to avoid repeating calculation. Based on the observation, we construct MST for graph \( G \) based on Prim’s algorithm [13]. We denote MST for image \( I \) as \( G_T = (\mathcal{V}_T, \mathcal{E}_T) \), where \( \mathcal{V}_T = [\mathcal{V} \setminus \{O\}] \).

2.3 Optimization for Node Labels

In this section, we introduce two round recursive operations to obtain the label optimization exactly. Figure 3 shows the algorithms for bottom-up recursive process and top-down recursive process. Specifically, given seed nodes \( S = \{s_1, \ldots, s_K\} \) for superpixels, we initialize the distance value of each node to the nearest seed as \( \infty \) and label indication 0. Meanwhile, enforce the distance value of seed as 0 and label indication as itself, i.e., \( s_k \). Then, superpixel label of each node could be recursively calculated on the MST.

For the bottom-up version, the recursion starts from leaf nodes to the root and labels each node as follows:

\[
L(p_{i+1}) = \arg \min_{p \in \mathcal{C}(p_{i+1})} \{D(p_i) + e(p_i, p_{i+1}), D(p_{i+1})\} \tag{3}
\]

s.t. \((p_i, p_{i+1}) \in \mathcal{E}_T\)

where \( D(p) \) is a table which records the minimum distance value from node \( p \) to the nearest seed. In the bottom-up recursion, \( p_{i+1} \) is the \( p_i \)’s parent node and \( C(p_{i+1}) \) denotes the set of \( p_{i+1} \)’s child nodes. It suggests that node \( p_{i+1} \)’s label is decided by its child nodes, who also have the smallest accumulated distance with the nearest seed. For the seed, it will preserve the corresponding label. After the
bottom-up recursive process, all the ancestor node of seed will be labeled by the nearest seed indication, as shown in Fig. 3. Hence, we need a top-down recursive process to label nodes along the path from leaf nodes to the nearest seed.

For the top-down version, the recursion starts from the root node to leaf nodes as follows:

$$L(p_{i+1}) = \arg\min_{p_{i+1} \in \mathcal{C}(p_i)} \{D(p_i) + e(p_i, p_{i+1}), D(p_{i+1})\}$$

s.t. $(p_i, p_{i+1}) \in \mathcal{E}_T$ (4)

In this recursion, $p_{i+1}$ is $p_i$’s child node and $\mathcal{C}(p_i)$ denotes the set of $p_i$’s child nodes. It suggests that node $p_{i+1}$’s label is decided by the label of its parent node, who has the smallest accumulated distance to the nearest seed. After the top-down recursion, all child nodes of seeds will be labeled and pixel labeling has been completed, as shown in Fig. 3.

2.4 Post-Processing

Different with SLIC-like algorithms [5], [6], our method clusters pixels into connected components. However, some small superpixels may be generated among geodesic label assignment. Figure 5 illustrates the reason for trivial superpixels formulation. Some seeds near the object boundary tend to aggregate few pixels as trivial superpixels. Especially, for the noise pixel labeled as seeds, it will be isolated as a single region due to the high-local contrast. To deal this problem, we merge superpixels which are less than one tenth of desired superpixel size to their adjacent superpixels.

3. Experiments

3.1 Dataset and Compared Methods

We test our method on BSD500 segmentation dataset, which contains 500 natural images annotated boundary maps by human beings. Boundary recall (BR), under segmentation error (UE) and achievable segmentation accuracy (ASA) [14] are used to evaluate superpixel segmentation performance. We compare seven state-of-the-arts: LSC [6], SNIC [7], SLIC [5], NC [4], TP [15], RPS [8] and FH [9].

3.2 Objective Results

Figure 4 shows the segmentation performance along with the increasing superpixel number for all compared methods. As can be seen, our method outperforms other methods in terms of BR. This demonstrates that superpixels extracted by our method have more higher boundary fitness with ground truth boundary. This benefits from the geodesic property of MST, which could capture the object boundary and preserve the topology structure for objects. In terms of UE as well as ASA, our method is comparable to the state-of-the-art algorithms. LSC [6] obtain the lowest UE and highest ASA due to its regular superpixel, which yields less segmentation leakage from groundtruth boundary. However, our method can extract accurate object region in lower superpixel density, such as Fig. 6.

3.3 Subjective Results

We first show some results with different granularity for our method in Fig. 6. Our method preserves the object’s boundary and topology structure, especially for the low superpixel granularity, such as 10 superpixels. Then, Some
superpixel segmentation results for compared methods are shown in Fig. 7. Left are the original images and results of compared methods are sequently listed from up to bottom and left to right: our method, LSC [6], S NIC [7], SLIC [5], NC [4], TP [15], RPS [8] and FH [9]. It can be seen that our method obtains better segmentation performance under the complex background and blur object boundary. For example, our method preserves the narrow and long object structure completely, such as the duck’s neck. Meanwhile, it also distinguishes between the duck’s breast and surrounding, which have a blur boundary. For the wolf image, the proposed method extract the rear of the object, which is a difficult task for other algorithms. LSC [6] yields high under segmentation error for the rear superpixel segmentation. This is because that the regular constraint weakens the weak boundary preserving under low contrast area. The same result is shown in the orangutan image. As shown in third row, LSC [6] loses the left face of the orangutan and our method preserves the irregular face region with high brightness, which is due to the boundary-aware property of MST.

3.4 Time Complexity

We perform 200 superpixel segmentation for $481 \times 321$ image and compare the running times of all methods on an Intel 2.33GHz CPU with 4GB RAM desktop. The results are shown in Table 1. SNIC [7] obtains the fastest implementation due to their C++ code. Our method is comparable to the state-of-the-arts although the code is implemented in Matlab without any optimization. In future, we will improve our code based on C++ implementation.

4. Conclusions

This paper proposes a boundary-aware superpixel segmentation method, which could quickly and exactly extract superpixel with a non-iteration framework. We first construct MST on the original pixels to represent the similarity between adjacent pixels. Then the geodesic distance of pixels can be exactly obtained based on two-round recursions. The pixel label is determined as the shortest path seed index. Experimental results on BSD500 segmentation benchmark demonstrate the proposed method obtains best performance compared with seven state-of-the-art methods.

Acknowledgments

This work is supported by the Typical Spectrum of Radio Signal szjj2016-092, the Postdoctoral Station at Xihua University Based on Collaboration Innovation Center of Sichuan Automotive Key Parts, the National Natural Science Foundation 61372187, 61473239.

References

[1] B. Luo, H. Li, T. Song, and C. Huang, “Object segmentation from long video sequences,” Proc. 23rd ACM International Conference on Multimedia, pp.1187–1190, ACM, 2015.
[2] J.R. Uijlings, K.E. Van De Sande, T. Gevers, and A.W. Smeulders, “Selective search for object recognition,” International Journal of Computer Vision, vol.104, no.2, pp.154–171, 2013.
[3] G. Mori, X. Ren, A.A. Efros, and J. Malik, “Recovering human body configurations: Combining segmentation and recognition,” IEEE Conference on Computer Vision and Pattern Recognition, pp.326–333, 2004.
[4] J. Shi and J. Malik, “Normalized cuts and image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.8, pp.888–905, 2000.
[5] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Mach. Intell., vol.34, no.11, pp.2274–2282, 2012.
[6] Z. Li and J. Chen, “Superpixel segmentation using linear spectral clustering,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1356–1363, 2015.
[7] R. Achanta and S. Süsstrunk, “Superpixels and polygons using simple non-iterative clustering,” IEEE Conference on Computer Vision and Pattern Recognition, pp.4895–4904, 2017.
[8] H. Fu, X. Cao, D. Tang, Y. Han, and D. Xu, “Regularity preserved superpixels and supervoxels,” IEEE Trans. Multimed., vol.16, no.4, pp.1165–1175, 2014.
[9] P.P. Felsenszwalb and D.P. Huttenlocher, “Efficient graph-based image segmentation,” International Journal of Computer Vision, vol.59, no.2, pp.167–181, 2004.
[10] P. Dollár and C.L. Zitnick, “Fast edge detection using structured forests,” IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.8, pp.1558–1570, 2015.
[11] B. Luo, H. Li, F. Meng, Q. Wu, and C. Huang, “Video object segmentation via global consistency aware query strategy,” IEEE Trans. Multimed., vol.19, no.7, pp.1482–1493, 2017.
[12] Q. Yang, “Stereo matching using tree filtering,” IEEE Trans. Pattern Anal. Mach. Intell., vol.37, no.4, pp.834–846, 2015.
[13] R.C. Prim, “Shortest connection networks and some generalizations,” Bell System Technical Journal, vol.36, no.6, pp.1389–1401, 1957.

[14] J. Shen, Y. Du, W. Wang, and X. Li, “Lazy random walks for superpixel segmentation,” IEEE Trans. Image Process., vol.23, no.4, pp.1451–1462, 2014.

[15] A. Levinshtein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, and K. Siddiqi, “TurboPixels: Fast superpixels using geometric flows,” IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.12, pp.2290–2297, 2009.