LETTER

Superpixel Segmentation Based on Global Similarity and Contour Region Transform*

Bing LUO^{†a)}, Member, Junkai XIONG[†], Li XU[†], and Zheng PEI[†], Nonmembers

SUMMARY This letter proposes a new superpixel segmentation algorithm based on global similarity and contour region transformation. The basic idea is that pixels surrounded by the same contour are more likely to belong to the same object region, which could be easily clustered into the same superpixel. To this end, we use contour scanning to estimate the global similarity between pixels and corresponded centers. In addition, we introduce pixel's gradient information of contour transform map to enhance the pixel's global similarity to overcome the missing contours in blurred region. Benefited from our global similarity, the proposed method could adhere with blurred and low contrast boundaries. A large number of experiments on BSDS500 and VOC2012 datasets show that the proposed algorithm performs better than traditional SLIC.

key words: superpixel segmentation, global similarity, contour, SLIC

1. Introduction

Superpixel segmentation benefits many computer vision task because of its local regions consistency and perceptual characterization. It has been widely used as a basic element to accelerate and improve some succeed computer vision tasks, such as 3D-reconstruction [1], tracking [2], object detection [3], [4], semantic segmentation [5], [6] and depth estimation [7].

Current superpixel segmentation algorithms are mainly based on graph theory and gradient. The former models an image as a graph, where pixels and edges are vertices and adjacent relationship of graph. Superpixel segmentation is to obtain several unrelated subgraphs by dividing the graph. Graph-based algorithms include Normalized Cuts (NC) [8], Felzenswalb and Huttenlocher (FH) [9] and Entropy Rate Superpixel (ERS) [10], etc. The gradient-based algorithm first generates the initial rough superpixels, and then uses gradient descent method to optimize the superpixels until convergence. Gradient-based algorithms include Watershed [11], Mean Shift [12], Turbopixel [13] and SLIC [14], etc.

SLIC can acquire compact segmentation regions of similar color quickly. However, since only the color and spatial position are considered, the segmentation accuracy is degraded at the weak edge where colors of foreground and

Fig. 1 Some segmentation results. (a) Ours. (b) SLIC. The number of superpixel is 100, 300 and 500.

background are very similar. For a blurred edge, the spatial distance will heavily decreases the segmentation accuracy when the color measure cannot significantly distinguish object boundaries.

To this end, this letter proposes a new contour region transformation superpixel segmentation algorithm based on global similarity. Intuitively, pixels surrounded by the same contour are more likely to belong to the same object region, while pixels separated by contours are more likely to belong to different object regions. Therefore, by adding additional contour information, it is possible to quickly acquire the surrounding relationship between pixels' global similarity. However, because the blurred object edge may be lost via the existing edge detection algorithm, there is a discontinuity in the edge. Thus we use an edge-based distance transform map to overcome influence of missing contours. The proposed algorithm performs better than SLIC under the same parameter conditions, which could be easily extended to other existing superpixel methods. As shown in Fig. 1, the segmentation results of the proposed algorithm at the peaks and flowers have higher boundary adherence in spite of low region contrast and blurred object boundary. This benefits from our global similarity via contour surrounding consistency.

2. Proposed Method

The basic idea is to obtain the structural similarity of pixels by adding global auxiliary information such as edges to improve the segmentation performance. First, based on binary edge map, regions-based pixel's similarity could be calcu-

Manuscript received August 16, 2019.

Manuscript revised November 6, 2019.

Manuscript publicized December 3, 2019.

[†]The authors are with Center for Radio Administration Technology Development, Xihua University, Chengdu 610039, China.

*This work is supported by the National Natural Science Foundation of China (No. 61801398, 61806028) and "Young Scholars" program of Xihua University.

a) E-mail: mathild1987@163.com

DOI: 10.1587/transinf.2019EDL8153

Fig. 2 According to the process, the following are original image, edge image, four scan images, four connected images and two region images.

lated by scanning the map horizontally and vertically. Pixels located in the same regions have stronger similarities. While pixels located in different regions have weaker similarities. Second, the edges obtained from blurred region tends to be discontinuous, which cause the inaccurate similarity calculation at these breakpoints. To overcome this problem, we introduce a gradient direction similarity via an edge-based distance transform map. Hence, the pixels on the same side of the breakpoints have the similar gradient directions, while the ones on the different sides have opposite gradient directions. This helps the algorithm to distinguish the different object regions more easily.

2.1 Global Similarity Measurement Based on Edge

Intuitively, edge information could facilitate the extraction of structural similarities between pixels. Two pixels are more similar if they are surrounded by the same edge. Otherwise, they are far more dissimilar. We use canny [15] operator to extract a binary edge map. Then a regional scanning algorithm [16] could integrate the pixels on the same edge side into a same region to calculate the structural similarity. Specifically, assuming that a ray scans the edge map from left to right, the pixel's label $l(i)$ is inverted between $\{0, 1\}$ once the ray intersects the edge. p_i represents current pixel and p_{i-1} represents the corresponding previous pixel. $m(i)$ denotes the pixel edge map. The label of the initial pixel is 0, i.e., $l(p_0) = 0$. It will become 1 when crossing the edge to go into a object. Then the label will invert to 0 after going through the other side of the object edge. A total of four scan maps are obtained, the two of rows are scanned from left to right and from right to left, and the two columns are scanned from top to bottom and bottom to top, respectively.

$$l(p_i) = \begin{cases} 1 - l(p_{i-1}) & \text{if } m(p_i) \in \{0, 1\} \\ l(p_{i-1}) & \text{otherwise} \end{cases} \quad (1)$$

Then the seed fill algorithm [17] could fill two column and two row scan maps to obtain two column and two row connected graphs. The row connected graphs are merged to obtain a horizontal region map, and the column connected graphs are combined to obtain a vertical region map. The

Fig. 3 (a) Edge image. (b) Distance transform image. (c) Horizontal gradient image. (d) Vertical gradient image. Similar gradient direction tends to similar color. (e) Explanation for gradient direction.

regional similarity can be calculated using these two region maps. Set the horizontal region as $\Phi = \{\phi_1, \dots, \phi_{n_h}\}$, and the vertical region as $\Psi = \{\psi_1, \dots, \psi_{n_v}\}$, where n_h and n_v are the number of horizontal regions and vertical regions, respectively. For any horizontal region ϕ_i , set a vector $H = \{h_1, \dots, h_{n_v}\}$, where $h_j = 1$ means ϕ_i and ψ_j intersect, otherwise it is 0. The similarity between ϕ_i and ϕ_j is defined as follows

$$d(\phi_i, \phi_j) = \frac{1}{\sum_{k=1}^{n_v} h_j(k)} \sum_{k=1}^{n_v} h_i(k) \cdot h_j(k) \quad (2)$$

Where $h_i(k)$ indicates whether or not horizontal region ϕ_i crossed by vertical region ψ_k . We can obtained $d(\phi_i, \phi_j)$ as the same way. Then, we compute the horizontal region and vertical region similarity matrix, i.e., $S_h \in R^{n_h \times n_h}$ and $S_v \in R^{n_v \times n_v}$. Let $M(i, j)$ denotes the similarity between pixels p_i and p_j . Then $M(i, j) = (S_{h_{ac}} + S_{h_{ca}} + S_{v_{bd}} + S_{v_{db}})/4$, where p_i is referred to horizontal region ϕ_a and vertical region ψ_b . p_j is referred to horizontal region ϕ_c and vertical region ψ_d .

2.2 Gradient of Edge's Distance Transformation Map

The edge-based global similarity can discriminate pixels in object and background. However, for some discontinuous edges caused by blurred boundary, there will be some incomplete edge fitting, which will result inaccurate scanning regions. Therefore, how to solve the label distribution near discontinuous edge is still an important issue.

To this end, we construct an edge-based distance transform map, i.e., the minimum distance value from the pixel to the nearest edge. It will increase gradually from the edge to the object center. We extract its gradient map and utilize gradient direction and magnitude to represent the global relationship between the pixel and the edge. Pixels with similar gradient directions are more likely to belong to the same region. We show the gradient direction near the break points on different sides of object edge on Fig. 3 (e). Intuitively, pixel p and q that are surrounded by the same edge tend to have the similar gradient direction, which are more likely to belong to the same region. While pixels on both sides of the same edge are opposite in direction and are more likely to belong to different regions, such as p and p' . Near the edge

breakpoint, the gradients on both sides are opposite and introduce the additional discrimination information. Figure 3 explains the generation of gradient maps. Figure 3 (a) is the binary edge map. The distance transform map is shown in Fig. 3 (b) by calculating the shortest distance between pixel and its nearest edge. The horizontal gradient G_x and vertical gradient G_y are obtained by performing on (b), as shown in (c) and (d), respectively.

2.3 Superpixel Segmentation of Global Similarity

We construct a superpixel segmentation method based on the global structural similarity. Specifically, the proposed method can be constructed as the following energy minimization problem

$$L(\mathbf{u}, \boldsymbol{\mu}) = \min_{\mathbf{u} \in \mathbb{S}} \sum_k \sum_i M(i, c(i)) |\mathbf{u}_i - \boldsymbol{\mu}_{c(i)}|_2. \quad (3)$$

Where $\mathbf{u}_i$ represent the feature vectors of i -th pixel. $c(i) \in \{1, 2, \dots, k\}$ represents center referred to the i -th pixel and $\boldsymbol{\mu}_{c(i)}$ denotes its corresponding feature vector. We calculate $M(i, c(i))$ as the global similarity between the regions corresponding to pixel p_i and center $p_{c(i)}$, which is referred to Sect. 2.1. Let $\mathbf{u}_i = [l, a, b, x, y, v_x, v_y]$, where the color and spatial position $[l, a, b, x, y]$ reflects the local features of the pixel, while gradient $[v_x, v_y]$ reflects the direction and magnitude, i.e., the global relationship between the pixel and the edge.

We sample K initial clustering centers on the grid with τ pixels interval, where $\tau = \sqrt{\frac{N}{k}}$ and N is the number of pixels in image. The optimization is performed as the weighted k-means, which is used in SLIC [14]. Then the pixel's labels and superpixel centers are alternatively optimized via step-by-step iteration. Specifically, we obtain the label assignment of image pixels by calculating minimum distance between pixel and center. Then, the centers are updated by averaging the space feature of corresponding member pixels. Iterating the above steps until convergence. Hence, pixel p_i 's label is assigned as the label of center $c(i)$ with minimum distance to pixel p_i .

3. Experiment and Analysis

The experiments in this letter are based on the MATLAB platform and tested on VOC2012 and BSDS500 datasets. We compare with SLIC [14], Simple Non-iterative Clustering (SNIC) [18], TP [13], Regularity Preserved Superpixels (RPS) [19] and ERS [10]. We set our method the same parameters as SLIC, i.e., 20 for the balance value of color and spatial position. We uses Canny operator in matlab with $threshold = 0.4$. Other compared methods are performed with the default parameters.

Figures 4 and 5 show the objective results on VOC2012 and BSDS500 datasets, respectively. We compare the methods via the indicators [20] such as Achievable Segmentation Accuracy (ASA), Boundary Recall (BR), Under-segmentation Error (UE). The number of superpixels ranges

Fig. 4 Quantitative evaluation on the VOC2012 dataset. (a) ASA. (b) BR. (c) UE.

Fig. 5 Quantitative evaluation on the BSDS500 dataset. (a) ASA. (b) BR. (c) UE.

Fig. 6 Running time. (a) and (b) are tested on VOC2012 dataset. (c) and (d) are tested on BSDS500 dataset. (a) and (c) are histograms of time and image number for our algorithm in the case of 100 superpixels.

from 100 to 1000, with an interval of 100. It can be seen that ERS [10] utilizes local adjacent graph to construct pixel's similarity, which obtains a global optimization and leads into the best segmentation performance. Our method ranks in the top two. Our method, SLIC [14] and SNIC [18] obtain higher under-segmentation error compared with ERS [10] due to the regular sampling strategy. Meanwhile, our method has better boundary adherence than SLIC [14] and SNIC [18], which is due to the global structure similarity between pixel and centers to help discriminate objects and background. RPS [19] and TP [13] have inferior segmentation performance due to their higher regular constraints.

We also compare the running time on Pascal VOC2012 and BSD500 and show the results in Fig. 6 (b) and (d). It can be seen that our method is more time-consuming than SLIC. This is mainly due to the large number of scanning regions, which is time-consuming in calculating regional similarity.

Fig. 7 Some subjective results. (a) Ours. (b) SLIC. (c) ERS. The number of superpixel is 100, 300 and 500.

We list the average time curves of compared methods for all images on the whole datasets. While, we list the time median of the proposed method. This is because that our method uses a same Canny threshold on the datasets, which results into a large number of edges and scanning regions on few images. Hence, average time can not effectively represent the performance of our algorithm. We also list the distribution of time consuming. We can see that the execution time on most images is 0.5 s, which is less than ERS [10], TP [13] and RPS [19].

Some subjective results are shown in Fig. 7. Under the same compactness parameters, our method can better segment some weak object boundaries than SLIC such as the middle hill in second image. Meanwhile, our method extracts the more accurate plane contour compared with SLIC. This is because that compactness constrain in SLIC [14] will decrease the boundary adherence without the proposed edge-based global structure similarity. In addition, ERS [10] obtain the higher boundary adherence due to its graph optimization framework. However, its regular constraint on balancing term is not explicit and results into the irregular segmentation, especially in the smooth regions.

4. Conclusions

In this letter, we propose a new contour region transformation superpixel segmentation algorithm based on global similarity. Global information such as image contours can aid in superpixel segmentation, i.e. pixels that are surrounded by the same contour are more likely to belong to the same region, while pixels separated by the same contour are more likely to belong to different regions. We scan the image edges to estimate similarity of different region pixels, which benefits the global similarity measure between pixel and center. In addition, in order to solve the edge discontinuity caused by the blurred edge of the image, we introduce pixel's gradient information of the edge distance transform map to enhance the discrimination of difference regions. Pixels locate in the same region tend to have the similarity gradient directions. Experiments on the BSDS500 and

VOC2012 datasets show that the proposed algorithm performs better than some state-of-the-arts, which validates the efficacy of our method.

References

- [1] A. Bódis-Szomorú, H. Riemenschneider, and L. van Gool, "Superpixel meshes for fast edge-preserving surface reconstruction," IEEE Conference on Computer Vision and Pattern Recognition, pp.2011–2020, 2015.
- [2] F. Yang, H. Lu, and M.-H. Yang, "Robust superpixel tracking," IEEE Trans. Image Process., vol.23, no.4, pp.1639–1651, 2014.
- [3] J. Yan, Y. Yu, X. Zhu, Z. Lei, and S.Z. Li, "Object detection by labeling superpixels," IEEE Conference on Computer Vision and Pattern Recognition, pp.5107–5116, 2015.
- [4] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders, "Selective search for object recognition," International Journal of Computer Vision, vol.104, no.2, pp.154–171, 2013.
- [5] C. Cadena and J. Košecká, "Semantic segmentation with heterogeneous sensor coverages," IEEE International Conference on Robotics and Automation, pp.2639–2645, 2014.
- [6] F. Meng, H. Li, Q. Wu, B. Luo, and K.N. Ngan, "Weakly supervised part proposal segmentation from multiple images," IEEE Trans. Image Process., vol.26, no.8, pp.4019–4031, 2017.
- [7] M. Liu, M. Salzmann, and X. He, "Discrete-continuous depth estimation from a single image," IEEE Conference on Computer Vision and Pattern Recognition, pp.716–723, 2014.
- [8] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Trans. Pattern Anal. Mach. Intell., vol.22, no.8, pp.888–905, 2000.
- [9] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient graph-based image segmentation," International Journal of Computer Vision, vol.59, no.2, pp.167–181, 2004.
- [10] M.-Y. Lui, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," CVPR 2011, pp.2097–2104, 2011.
- [11] F. Meyer, "Color image segmentation," International Conference on Image Processing and Its Applications, pp.303–306, 1992.
- [12] Y. Cheng, "Mean shift, mode seeking, and clustering," IEEE Trans. Pattern Anal. Mach. Intell., vol.17, no.8, pp.790–799, 2002.
- [13] A. Levinstein, A. Stere, K.N. Kutulakos, D.J. Fleet, S.J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric ows," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.12, pp.2290–2297, 2009.
- [14] 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.2274–2281, 2012.
- [15] L. Ding and A. Goshtasby, "On the canny edge detector," Pattern Recognition, vol.34, no.3, pp.721–725, 2001.
- [16] F. Meng, H. Li, Q. Wu, B. Luo, C. Huang, and K.N. Ngan, "Globally measuring the similarity of superpixels by binary edge maps for superpixel clustering," IEEE Trans. Circuits Syst. Video Technol., vol.28, no.4, pp.906–919, 2018.
- [17] W.G.M. Geraets, A.N. van Daatselaar, and J.G.C. Verheij, "An efficient filling algorithm for counting regions," Computer Methods and Programs in Biomedicine, vol.76, no.1, pp.1–11, 2004.
- [18] 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.
- [19] 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.
- [20] M. Wang, X. Liu, Y. Gao, X. Ma, and N.Q. Soomro, "Superpixel segmentation: A benchmark," Signal Processing: Image Communication, vol.56, pp.28–39, 2017.