

# Improved accuracy of superpixel segmentation by region merging method

Song ZHU, Danhua CAO (✉), Yubin WU, Shixiong JIANG

School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan 430074, China

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2015

**Abstract** Superpixel as an important pre-processing technique has been successfully used in many vision applications. In this paper, we proposed a region merging method to improve superpixel segmentation accuracy with low computational cost. We first segmented the image into many accurate small regions, and then progressively agglomerated them until the desired region number was reached. The region merging weight was derived from a novel energy function, which encourages the superpixel with color consistency and similar size. Experimental results on the Berkeley BSDS500 data set showed that our region merging method can significantly improve the accuracy of superpixel segmentation. Moreover, the region merging method only need 50 ms to process a  $481 \times 321$  image on a single Intel i3 CPU at 2.5 GHz.

**Keywords** image processing, image segmentation, superpixels, region merging

## 1 Introduction

Superpixel algorithms group neighboring pixels into perceptually meaningful homogeneous regions, the so-called superpixels [1]. Superpixel segmentation has gradually become a useful preprocessing step in many computer vision applications, such as image segmentation [2], object recognition [3] and object localization [4]. There are some general properties required by various vision applications as follows:

1) As a pre-processing step, superpixels should be fast to compute and there should be fewer parameters in the algorithm for simple use.

2) As the boundary is very important feature for recognizing objects and superpixel segmentation has the

risk of losing meaningful edges, the boundaries of the superpixels should adhere well to the object boundaries.

The goal of this paper is to improve the superpixel segmentation accuracy with low computational cost. Unlike other methods, which directly segment the image into a desired number of superpixels, we proposed a two-step method called region merging superpixel, namely RMSP. We first grouped pixels into many accurate small regions, and then progressively merged them according to a novel energy function. Through this function, the superpixels are with homogeneous color and similar size (a common property of many superpixel methods). RMSP can improve the segmentation performance because the accurate boundaries from the initial segmentation are preserved during the region merging process. The computational cost of the region merging method is cheap because we start the merging from initial small regions, but not from pixels.

Recently, Achanta et al. [5] have introduced a simple linear iterative clustering (SLIC) method which produces regular superpixels with a low computational complexity. They compared SLIC with other popular methods [1,6,7] on the Berkeley benchmark data set in Ref. [5]. The experimental results showed that SLIC outperforms other methods before 2010 under the standard metric of boundary recall and under-segmentation error. Recently, entropy rate clustering (ERS) proposed by Liu et al. [8] has achieved better results than SLIC but with much lower speed.

Besides, another important segmentation technique is region merging. The main difference among various region merging methods is the merging criterion. In general, regions are merged according to the mean color value distance [9,10], statistical properties [11,12] and spatio-temporal similarity [13]. Most algorithms of merge region method are according to local decisions, which do not have some desirable global properties. Whereas, algorithms in Refs. [14,15] intend to obtain the minimum overall region color variance by optimizing the global Mumford-Shah

energy function during the region merging process. In this paper, we proposed a new energy function with global properties rather than the traditional Mumford-Shah energy function.

## 2 Region merging method

Almost in all superpixel algorithms, the segmentation is highly accurate when the desired size of the superpixels is small. Inspired by this, we can first segment the image into many accurate small superpixels, and then merge them iteratively until the desired number is reached. Through the region merging method, the segmentation performance can be improved because the highly accurate boundaries from the initial segmentation are preserved.

### 2.1 Region merging model

Our region merging method was based on optimizing a new energy function, which consists of two terms,  $E_{\text{data}}$  and  $E_{\text{size}}$ , in Eq. (1). The segmentation of an observed image  $I$  into a fixed number of regions is obtained by minimizing this function. Minimizing  $E_{\text{data}}$  encourages the region with homogeneous color, and minimizing  $E_{\text{size}}$  encourages the region with similar size.

$$E = \frac{E_{\text{data}}}{\max(E_{\text{data}}) - \min(E_{\text{data}})} + \beta \frac{E_{\text{size}}}{\max(E_{\text{size}}) - \min(E_{\text{size}})}. \quad (1)$$

As the meanings of the two terms in the function are different, we used the max-min method to normalize them into an appropriate range. Before optimizing the function, we estimated the maximum and minimum values according to the specific definitions of  $E_{\text{data}}$  and  $E_{\text{size}}$ .  $\beta$  is a positive weight parameter combines the two terms. The bigger  $\beta$ , the superpixels we get are more size similar. However, too large  $\beta$  will degenerate the segmentation accuracy. In this paper, we set  $\beta$  experimentally in Section 3.1.

The data term,  $E_{\text{data}}$ , defined in Eq. (2), measures the overall color diversity index of image  $I$  on each region (superpixel)  $R_k$ .  $P_k$  is the color probability distribution of region  $R_k$ .  $n_k$  is the size of region  $R_k$ . In this paper, we model a region as a set of independent and identically distributed pixels. The probability distribution of a region is estimated by a discrete color histogram.  $H(P)$  is a diversity index function.

$$E_{\text{data}} = \sum_{k=1}^K n_k H(P_k), \quad (2)$$

$$H(P) = \sum_{b=1}^B f(P_b).$$

We introduced two diversity indices which are widely

used in ecology and economics [16]. If we set the function  $f(x) = -x \log(x)$ ,  $H(P)$  becomes the Shannon-Wiener index. If we set the function  $f(x) = 1 - x^2$ ,  $H(P)$  becomes the Gini-Simpson index. The maximum and minimum values of the  $H(P)$  can be estimated according to the Appendix. When each region has only one color,  $E_{\text{data}}$  reaches minimum, and when each region has a uniform color probability distribution,  $E_{\text{data}}$  reaches maximum.

The size term,  $E_{\text{size}}$ , defined in Eq. (3), measures the overall size similarity, which is used to control the superpixel shape. The definition is inspired by the diversity index. When all the sizes of the regions are equal,  $E_{\text{size}}$  reaches the minimum.

$$E_{\text{size}} = - \sum_{k=1}^K f(n_k). \quad (3)$$

To minimize the energy function, a greedy method is used during the region merging process, in which we search for the merger with the smallest increasing energy. When we merge two adjacent regions,  $R_1$  and  $R_2$ , the energy value defined in Eq. (1) will change. The increasing value  $\Delta E(R_1, R_2)$  derived from Eqs. (1)–(3) as follows.

$$\Delta E(R_1, R_2) = \frac{\Delta E_{\text{data}}(R_1, R_2)}{\max(E_{\text{data}}) - \min(E_{\text{data}})} + \beta \frac{\Delta E_{\text{size}}(R_1, R_2)}{\max(E_{\text{size}}) - \min(E_{\text{size}})},$$

$$\Delta E_{\text{data}}(R_1, R_2) = (n_1 + n_2)H(P_{12}) - n_1H(P_1) - n_2H(P_2),$$

$$\Delta E_{\text{size}}(R_1, R_2) = -f(n_1 + n_2) + f(n_1) + f(n_2). \quad (4)$$

If we set  $f(x) = 1 - x^2$ ,

$$\Delta E_{\text{data}}(R_1, R_2) = \frac{n_1 n_2}{n_1 + n_2} \sum_{b=1}^B (P_{1b} - P_{2b})^2.$$

$\Delta E_{\text{data}}$  can be interpreted as the size-weighted Euclidean distance of the color probability distribution. So a region tends to merge its adjacent region with the similar probability distribution. If we set  $f(x)$  as a concave function,  $\Delta E_{\text{size}}(R_1, R_2)$  is monotonically increasing with  $n_2$  when  $n_1$  fixed. So a region tends to merge its adjacent region with a small size.

### 2.2 Computational complexity analysis

For implementation of the greedy region merging algorithm, image is modeled as a region adjacency graph (RAG) [9]. RAG is an undirected graph  $G(V, F)$  with vertices  $v \in V$ , the set of regions to be segmented, and edges  $(v_i, v_j) \in F$  corresponding to pairs of neighboring vertices. Each graph edge weight is set to be the increasing value  $\Delta E$ , which is defined in Eq. (4). As we always search the

minimum edge weight, so we need a priority queue which can be implemented efficiently by a heap [17].

In the beginning, constructing the heap takes  $O(|F|)$  time. At each merging step, the edge with the minimum weight is popped out in  $O(\log_2(|F|))$  time. As two regions are merged, all the edge weights of the node resulting from the merging with its connected neighboring nodes change and must be recalculated. The positions of the changed weight edges in the heap must be updated, requiring time  $O(\log_2(|F|))$  for each update. So the total time for each merging step is  $O(d \cdot \log_2(|F|))$ , where  $d$  denotes the degree of the new node produced by merging.

Assume that in the very beginning, there are  $N$  regions and we perform the region merging for  $M$  times ( $M \leq N$ ). The RAG is usually a sparse graph when used for region merging based image segmentation, so we approximate that the degree of each node in the RAG is a small constant  $d$ , then the edge number  $|F|$  is equal to  $d \cdot N$ .

In summary, the computational complexity of the region merging is  $O(d \cdot M \cdot \log_2(d \cdot N))$ . Note that as we start the merging from initial small regions, but not from pixels,  $N$  and  $M$  are not large (usually only the square root of the image size). So the greedy region merging method is fast in practice.

### 3 Experimental results

To evaluate the boundary adherence performance of superpixel algorithms, we used standard metrics on the Berkeley BSDS500 data set, as used in most recent superpixel papers [5,8]. The standard metrics are boundary recall (BR), under-segmentation error (UE) and achievable segmentation accuracy (ASA). All experiments are done on a single Intel i3 CPU at 2.5 GHz.

Because RMSP needs an initial segmentation as an input, and the effects of segmentation accuracy improve-

ment are similar by any input. In this paper, SLIC [5] was used and the desired superpixel number was set to be 3200 for initial segmentation.

#### 3.1 Parameter sensitivity

In the RMSP algorithm, we create a  $bin \times bin \times bin$  discrete color histogram in the RGB color space. Each channel is uniform quantization. Here, we studied the impact of the quantization degree on the segmentation quality. Figure 1 presents the performance variation for different  $bin$  numbers. We can see that the segmentation quality is not sensitive when  $bin \geq 8$ . The smaller the bin number, the faster the speed in the region merging process, so we set  $bin = 8$  in subsequent experiments.

Similarly, we studied the impact of the superpixel size similarity on the segmentation quality. Figure 2 shows the performance variation for different parameter  $\beta$ . In terms of BR, the BR monotonically decreases with the value of  $\beta$ , but when  $\beta \geq 64$ , the difference is slight. In terms of UE and ASA, the curves almost overlap each other, so the UE and ASA are not sensitive with  $\beta$ . To make the superpixel size as similar as possible without significant segmentation accuracy degeneration, we set  $\beta = 64$ .

#### 3.2 Comparison of different color diversity indices in the energy function

In our energy function, we do not restrict to use a particular index. Here, we compare the effect of two famous diversity indices. One is  $f(x) = -x \log(x)$ , which represents the Shannon-Wiener index and the other one is  $f(x) = 1 - x^2$ , which represents the Gini-Simpson index. When  $f(x) = 1 - x^2$ , we tuned  $\beta = 4$  in a similar way as described in Section 3.1. As shown in Fig. 3, all the three metrics are slightly better when using the Shannon-Wiener index, which declares that Shannon-Wiener index is more appropriate to measure region color consistency. So Shannon-Wiener index is applied in subsequent experiments.

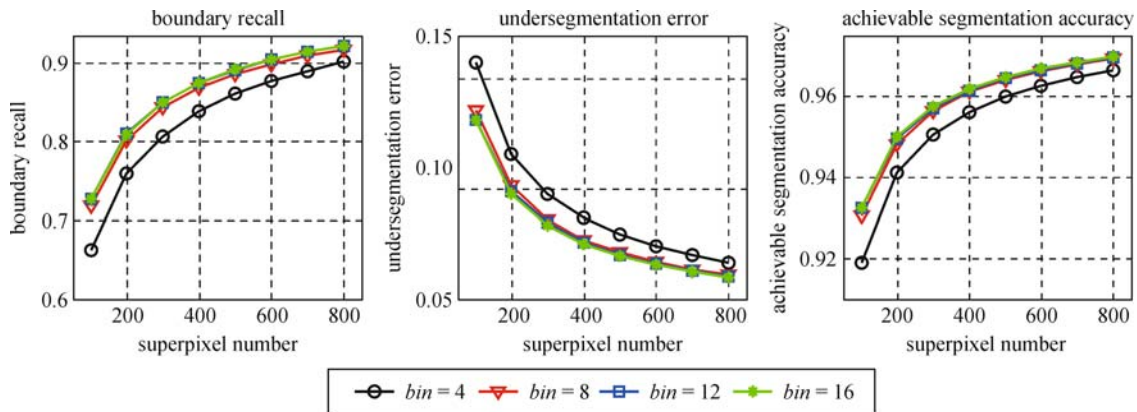
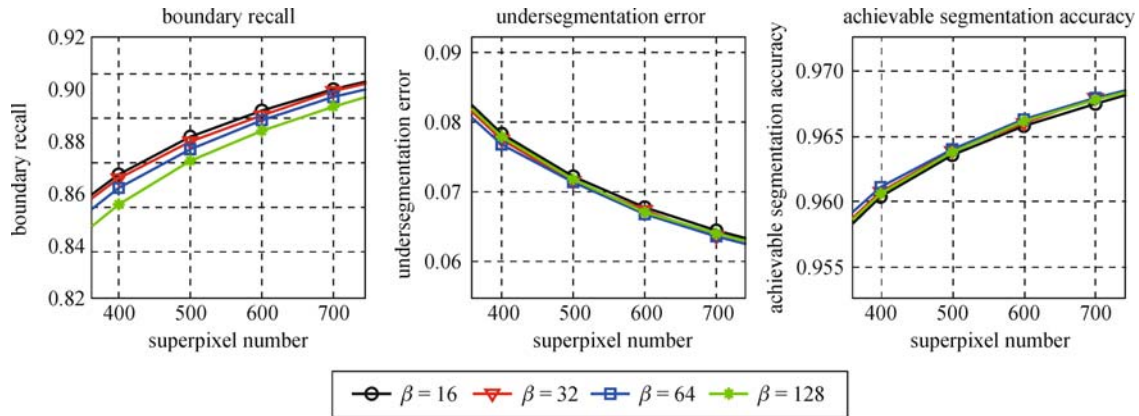
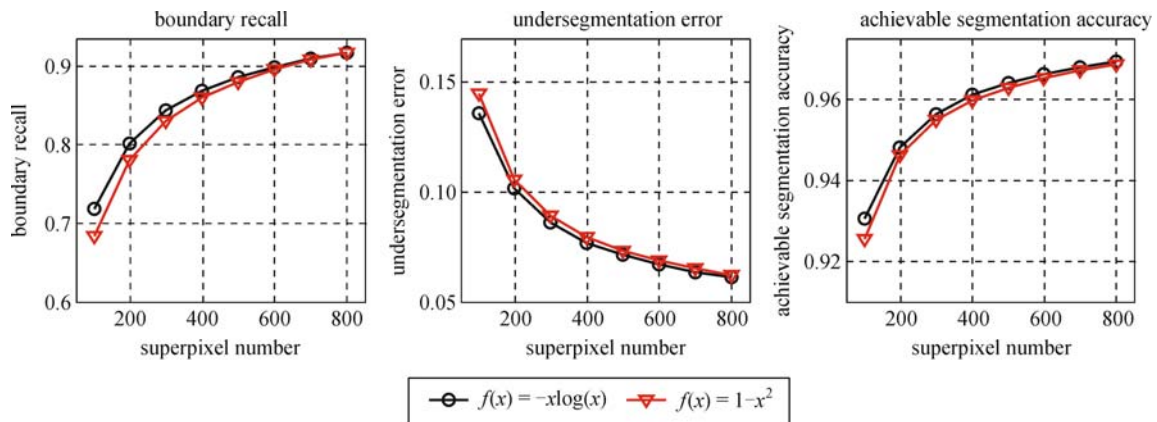


Fig. 1 Segmentation performance with different  $bin$  numbers. We set the function  $f(x) = -x \log(x)$  and  $\beta = 64$



**Fig. 2** Segmentation performance with different values of the parameter  $\beta$ . Color histogram is computed by 8 bins per channel. We set the function  $f(x)=-x\log(x)$



**Fig. 3** Segmentation performance with two different functions. Color histogram is computed by 8 bins per channel. When using  $f(x)=-x\log(x)$ , we set  $\beta = 64$ . When using  $f(x)=1-x^2$ , we set  $\beta = 4$

### 3.3 Improving superpixel segmentation performance by RMSP

In this section, we investigated the impact of region merging on segmentation quality. We compared RMSP (using region merging) with SLIC and ERS (both not using region merging) in Fig. 4. The SLIC algorithm is the state-of-the-art superpixel segmentation algorithm and has already been compared with other algorithms before 2010 in Ref. [5]. ERS [8] is the most accurate algorithm we have known. Parameters in the two algorithms are set to the default values in the original papers. SLIC-RMSP is obviously better than the other two methods. Comparing SLIC-RMSP with SLIC, we can find that the segmentation performance improves greatly. Moreover, although SLIC is less accurate than ERS, the combined method SLIC-RMSP is much more accurate.

In addition, the region merging algorithm runs very fast which needs only 50 ms to process a  $481 \times 321$  image on a single Intel i3 2.5 GHz CPU, while SLIC need about 150 ms and ERS need more than 1500 ms. So it is worth using

the region merging algorithm to improve the segmentation accuracy.

### 3.4 Comparison of the region merging algorithms

In this section, we compared RMSP with three other region merging methods including the region merging version of the Mumford-Shah energy function [15] (marked as MS), Binary Partition Tree proposed by Salembier [10] (marked as BP) and the Independent Identically Distributed-Kullback Leibler tree proposed by Calderero [12] (marked as IIDKL).

Figure 5 shows the comparison results. We added a baseline performance obtained by segmenting the image directly by SLIC. When comparing SLIC with other region merging methods, we can see that all the region merging methods significantly improve the boundary adherence performance, especially when the desired superpixel number is fewer. This improvement benefits from the highly accurate initial boundaries which are preserved after region merging. Among all the four region merging



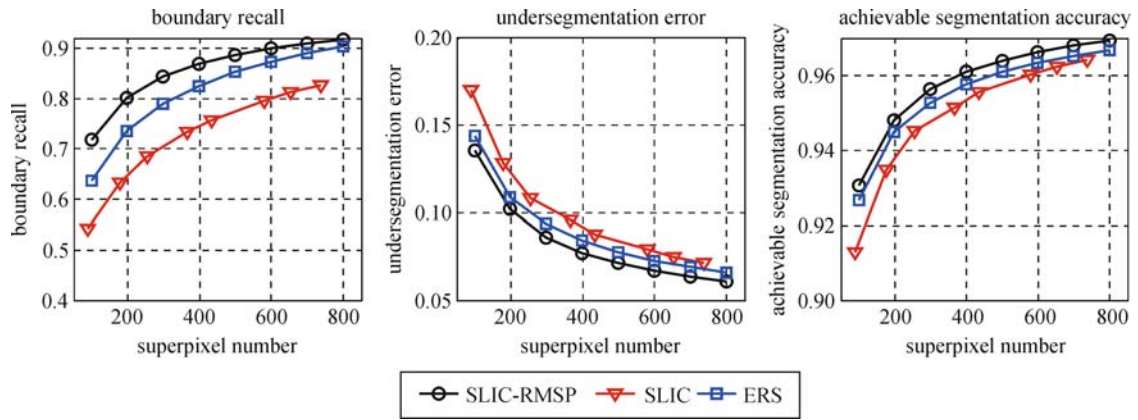


Fig. 4 Segmentation performance comparison between SLIC-RMSP, SLIC and ERS

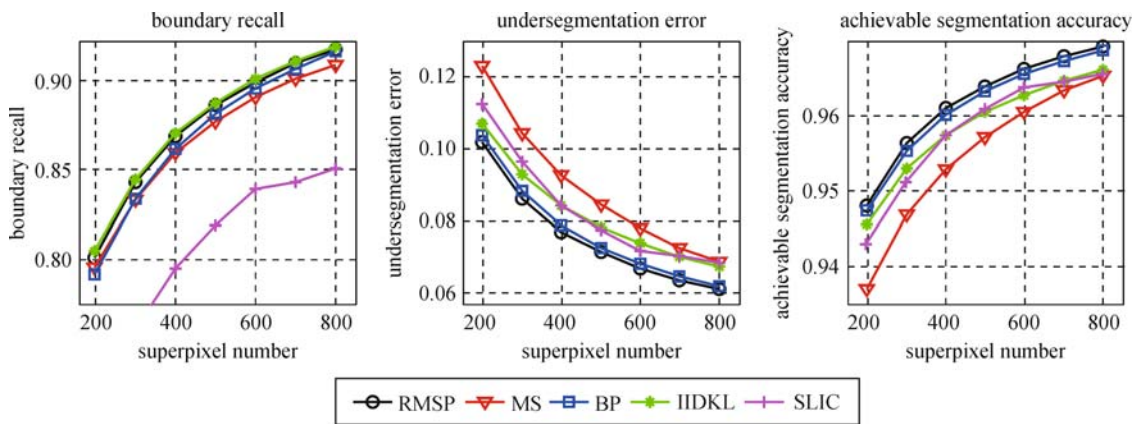


Fig. 5 Segmentation performance of different region merging methods

methods, our RMSP performs best. It is the contribution of the data weight based on the color diversity index, which is defined in Eq. (4).

As shown in Fig. 6, only RMSP can make the region size similar and avoid serious under-segmentation.

To further compare region size similarity between different region merging methods, we show the region size distributions in Fig. 7. According to the experimental results, only RMSP avoids making the superpixel too big or too small, which is resulted from the size weight based on the size similarity energy, which is defined in Eq. (4).

## 4 Conclusions

In this paper, we proposed a region merging method RMSP. We first segmented the image into many accurate small regions, and then progressively agglomerated them into the desired superpixels according to a novel energy function. Different from other region merging methods, RMSP can make the merged superpixels with similar size because the new energy function not only encourages the region with homogeneous color, but also encourages the regions with similar size. From the experimental results,

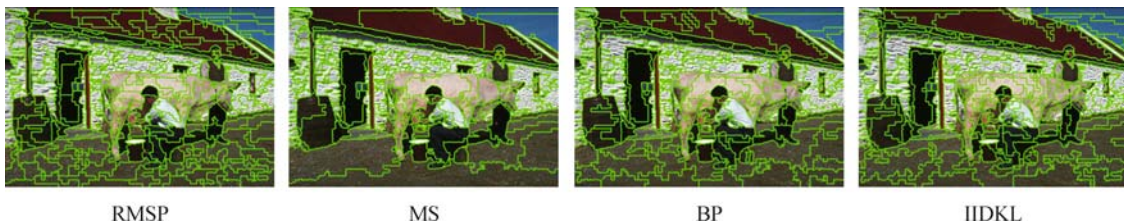
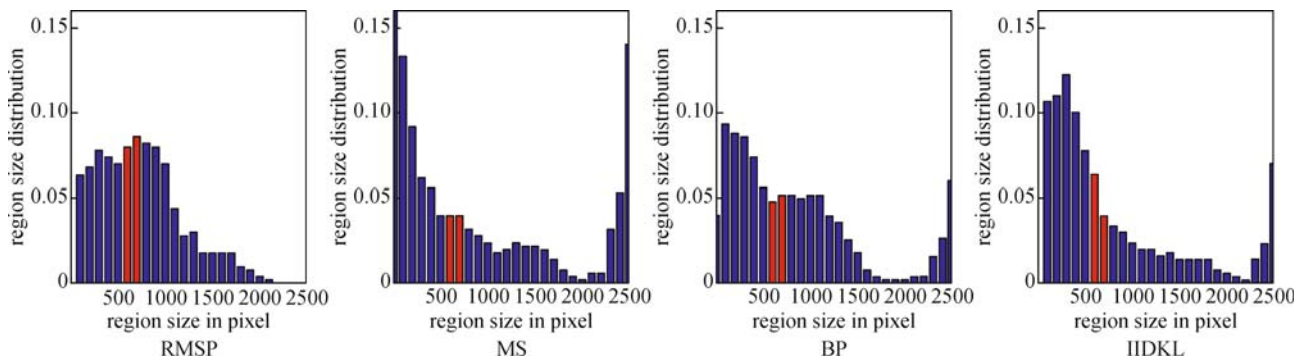


Fig. 6 Visual comparison of superpixels produced by various region merging methods. The superpixel number is 200



**Fig. 7** Region size distributions of different region merging methods. The average size bins are marked red. The superpixel number is 200

we found that RMSP is an effective method to improve the superpixel accuracy with low computational cost. The segmentation accuracy is not sensitive with parameters in RMSP.

## Appendix

Extreme points of the convex/concave function defined in this paper

Let's solve the extreme points of the following function where  $C$  is a positive constant and  $f(x)$  is a convex/concave function.

$$E(x_1, x_2, \dots, x_K) = \sum_{k=1}^K f(x_k), \quad (\text{A1})$$

$$\text{s.t. } \sum_{k=1}^K x_k = C \text{ and } x_1, x_2, \dots, x_K \geq 0.$$

We define a new function by a Lagrange Multiplier  $\alpha$  as follows.

$$F(x_1, x_2, \dots, x_K) = \sum_{k=1}^K f(x_k) + \alpha \left( \sum_{k=1}^K x_k - C \right). \quad (\text{A2})$$

The extreme points can be obtained by solving the partial derivative equations as follows.

$$\begin{aligned} \partial f(x_1) / \partial x_1 + \alpha &= 0, \\ \partial f(x_2) / \partial x_2 + \alpha &= 0, \\ &\dots \\ \partial f(x_K) / \partial x_K + \alpha &= 0, \\ \sum_{k=1}^K x_k - C &= 0. \end{aligned} \quad (\text{A3})$$

As  $f(x)$  is a convex/concave function, the derivative  $\partial f(x) / \partial x$  is monotonically increasing/decreasing, there is

only one solution, that is,  $x_1 = x_2 = \dots = x_K = C/K$ . The extreme value is  $Kf(C/K)$ . When  $f(x)$  is a convex function, it is the minimum value. When  $f(x)$  is a concave function, it is the maximum value.

Another extreme points are on the boundary, that is,  $x_i = C$  and  $x_1 = \dots = x_{i-1} = x_{i+1} = \dots = x_K = 0$ . The extreme value is  $f(C) + (K-1)f(0)$ . When  $f(x)$  is a convex function, it is the maximum value. When  $f(x)$  is a concave function, it is the minimum value.

## References

- Ren X, Malik J. Learning a classification model for segmentation. In: Proceedings of IEEE International Conference on Computer Vision. 2003, 10–17
- Kohli P, Ladický L, Torr P H S. Robust higher order potentials for enforcing label consistency. International Journal of Computer Vision, 2009, 82(3): 302–324
- Pantofaru C, Schmid C, Hebert M. Object recognition by integrating multiple image segmentations. In: Proceedings of European Conference on Computer Vision. 2008, 5304: 481–494
- Fulkerson B, Vedaldi A, Soatto S. Class segmentation and object localization with superpixel neighborhoods. In: Proceedings of IEEE International Conference on Computer Vision. 2009, 670–677
- Achanta R, Shaji A, Smith K, Lucchi A, Fua P, Süsstrunk S. SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274–2282
- Levinshtein A, Stere A, Kutulakos K N, Fleet D J, Dickinson S J, Siddiqi K. TurboPixels: fast superpixels using geometric flows. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2290–2297
- Veksler O, Boykov Y, Mehrani P. Superpixels and supervoxels in an energy optimization framework. In: Proceedings of European Conference on Computer Vision. 2010, 6315: 211–224
- Liu M Y, Tuzel O, Ramalingam S, Chellappa R. Entropy-rate clustering: cluster analysis via maximizing a submodular function subject to a matroid constraint. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1): 99–112
- Haris K, Efstratiadis S N, Maglaveras N, Katsaggelos A K. Hybrid

image segmentation using watersheds and fast region merging. *IEEE Transactions on Image Processing*, 1998, 7(12): 1684–1699

10. Salembier P, Garrido L. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Transactions on Image Processing*, 2000, 9(4): 561–576
11. Nock R, Nielsen F. Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(11): 1452–1458
12. Calderero F, Marques F. Region merging techniques using information theory statistical measures. *IEEE Transactions on Image Processing*, 2010, 19(6): 1567–1586
13. Moscheni F, Bhattacharjee S, Kunt M. Spatio-temporal segmentation based on region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(9): 897–915
14. Ben Ayed I, Mitiche A. A region merging prior for variational level set image segmentation. *IEEE Transactions on Image Processing*, 2008, 17(12): 2301–2311
15. Beaulieu J M, Goldberg M. Hierarchy in picture segmentation: a stepwise optimization approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(2): 150–163
16. Wikipedia. Diversity index, 2014, [http://en.wikipedia.org/wiki/Diversity\\_index](http://en.wikipedia.org/wiki/Diversity_index)
17. Cormen T H, Leiserson C E, Rivest R L, Stein C. *Introduction to Algorithms*. 2nd ed. Cambridge, Mass: MIT Press, 2001



**Song Zhu** is a Ph.D. candidate in the School of Optical and Electronic Information, Huazhong University of Science and Technology (HUST). He received his B.E. degree in the School of Optical and Electronic Information, HUST, 2010. His research interests include 3D computer vision, image segment and pattern recognition.



**Danhua Cao** is a professor in the School of Optical and Electronic Information, Huazhong University of Science and Technology. She received her Ph.D. degree in electronic physics and devices, HUST, 1993; B.E. degree in measuring and control technology and instrumentations, HUST, 1987. She is the permanent member of the Professional Committee of Opto-Electronic Technology in the Chinese Optical Society. Her research interests include optoelectronic sensing and signal processing, machine vision algorithm and systems.



**Yubin Wu** is an associate professor in the School of Optical and Electronic Information, Huazhong University of Science and Technology (HUST). He received his M.E. degree in optical engineering from Institute of Optics and Electronics of the Chinese Academy of Sciences, 1987; B.E. degree in optical instruments, HUST, 1984. His research interests include optoelectronic sensing and signal processing, machine vision, and the development of high-tech products.



**Shixiong Jiang** is a Ph.D. candidate in the School of Optical and Electronic Information, Huazhong University of Science and Technology (HUST). He received his B.E. degree in the School of Optical and Electronic Information, HUST, 2011. His research interests include machine vision, 3D reconstruction and pattern recognition.