Journal of Network Intelligence ©2020 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **5**, Number **4**, November **2020**

Annotation Method to Improve the Mapping Between Image Features and High Level Semantic Expression

[1]Juan Wang, [1]Wenmin Song, [1]Xiuyan Sun, [2]Linlin Tang, and [3]Jyh-haw Yeh

[1]Laiwu Vocational and Technical College
[2]Harbin Institute of Technology, Shenzhen
[3]Department of Computer Science, Boise State University, Idaho, USA
Email: hittang@126.com

Received August 2020; revised October 2020

Abstract. *In order to improve the accuracy of scene annotation, this paper proposes a new annotation method based on improving the image segmentation effect to improve the mapping between image features and high-level semantic expression. Experiments on classical image data sets, the advantages of the improved image segmentation algorithm and the effectiveness of scene annotation algorithm using image segmentation are verified from two aspects of visual observation and numerical evaluation.*
Keywords: Scene Annotation, Image Segmentation, Parameter Optimization, Multi-scale Feature, Mapping

1. **Introduction.** Image annotation not only has more and more applications in many fields, but also should extract and analyze key information in the image more autonomously and accurately. In nowadays, image annotation has been applied to the fields of architecture and engineering design, art collection, geographic information and remote sensing system, intellectual property rights, medical diagnosis and recognition[1], and plays an increasingly important role in image retrieval by virtue of autonomy and processing efficiency[2]. With the development of computer technology in all aspects of people's life, people's demand for image retrieval is also increasing. Image annotation can not only help people to add labels or descriptions to the image, but also can better capture information in image, avoiding the omission of information and ambiguity of information. Now people begin to give more keywords and descriptions to images, such as shooting time, place, objects in the scene, etc., which greatly promotes the demand of people to label the computer independently. On the one hand, image annotation can save a lot of manpower and time, on the other hand, it can form a more unified standard and improve the overall efficiency of annotation, so as to provide a basis for better use of image information. At the same time, it also facilitates more detailed management of images, such as classification and use of medical images, information extraction in remote sensing detection, object recognition and duplicate checking in trademarks, etc., which can make good use of image annotation to improve efficiency and value of image processing[3]. At the same time, the development of image annotation algorithm can promote the development of other disciplines and a variety of fields provide the basis and development power.

Scene annotation is a kind of image annotation. Improving the accuracy is conducive to improving the resolution of images, which will greatly facilitate people's life and communication. Scene annotation is to label all the pixels in the image reasonably, so as to realize classification according to the advanced semantic features of the image[4]. Scene annotation is the basis for people to construct new images and scene design, which can provide convenience for people to display or express with images. The realization of scene annotation is also the process of image segmentation, which makes it convenient for people to select part of the image from the image to use in other fields. Such as image matting, part of the color changes in the image, image shape replacement, etc., can be implemented on the basis of scene annotation algorithm. Through the scene annotation, people can process the image more conveniently and make the communication between people more accurate and intuitive.

2. **Related Work.** The algorithm proposed here is an improvement on TBES algorithm[5] to improve the two aspects mentioned in the introduction.

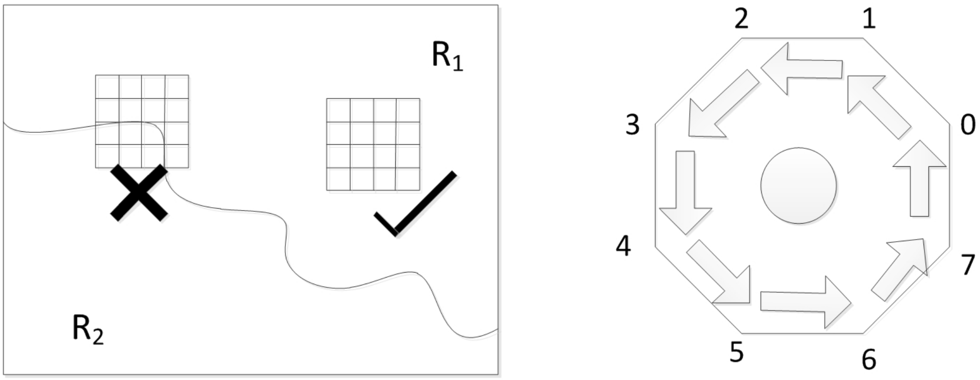
2.1. **Brief Introduction of TBES Algorithm.** Texture and Boundary Encoding Compression (TBES) is a method based on lossy compression. It was first proposed by Rao et al in 2009[5], and improved and improved in the following time. Compared with Compression Texture Method (CTM) algorithm[6], the main advantages of TBES algorithm are in following two aspects: first, the color features of the region used by TBES are completely obtained from the information of pixels in the region, which can represent the features in the region. At the same time, TBES also uses the boundary features of each region, encodes the boundary features and color features of the region respectively, and uses the sum of the final two encoding lengths to represent the block region. In this way, region merging errors caused by similar color features can be avoided. Second, in the algorithm of comparing the coding length changes between the fused region and the two regions before fusion, the TBES algorithm takes into account the relationship in the set, and fully considers the problem of feature overlap after merging the two regions.

In the process of feature extraction, such as extracting the color features of region $R1$ in Figure 1 a), the features of pixels completely belonging to the region are extracted by mask in the graph, and the positions of pixels extracted twice can not coincide, and finally multiple column vectors are formed. Figure 1 b) shows the coding method of the region boundary contour. The specific implementation method is to code the position relationship of any two pixels as the region boundary, and the coding is carried out according to the direction shown in the figure. In order to be able to calculate the amount of computation, only the direction of the first boundary is recorded, and subsequent direction is calculated by the encoded difference value, so that a large number of 0 or 1 can appear, thus realizing sparse representation.

Based on the use of boundary feature coding, the criteria for evaluating whether two regions need to be merged need to reflect this change. As described in the following formula (1), whether the two regions should be merged is determined by combining the color features and contour features of region R_i and region R_j.

$$\begin{aligned} \Delta L_{w,\varepsilon}\left(R_i,R_j\right) &= L^S_{w,\varepsilon}\left(\Re\right) - L^S_{w,\varepsilon}\left(\left(\Re\backslash\left\{R_i,R_j\right\}\right)\cup\left\{R_i\cup R_j\right\}\right) \\ &= L_{w,\varepsilon}\left(R_i\right) + L_{w,\varepsilon}\left(R_j\right) - L_{w,\varepsilon}\left(R_i\cup R_j\right) + \tfrac{1}{2}\left(B\left(R_i\right)+B\left(R_j\right)-B\left(R_i\cup R_j\right)\right) \end{aligned} \quad (1)$$

Here, $L_{w,\varepsilon}\left(R_i\right)$ represents coding length of color features in extracted region R_i , and $B(R_i)$ is coding length of the boundary features surrounding the region R_i , and its calculation method is shown in formula (2). $P\left[\Delta o = i\right]$ is the number of boundaries with



(a) Image Color Feature Extraction (b) Coding Method of Contour

FIGURE 1. Feature Coding of TBES Algorithm

direction i, and $\#\left(\Delta o_t = i\right)$ is the weight of boundaries with direction i.

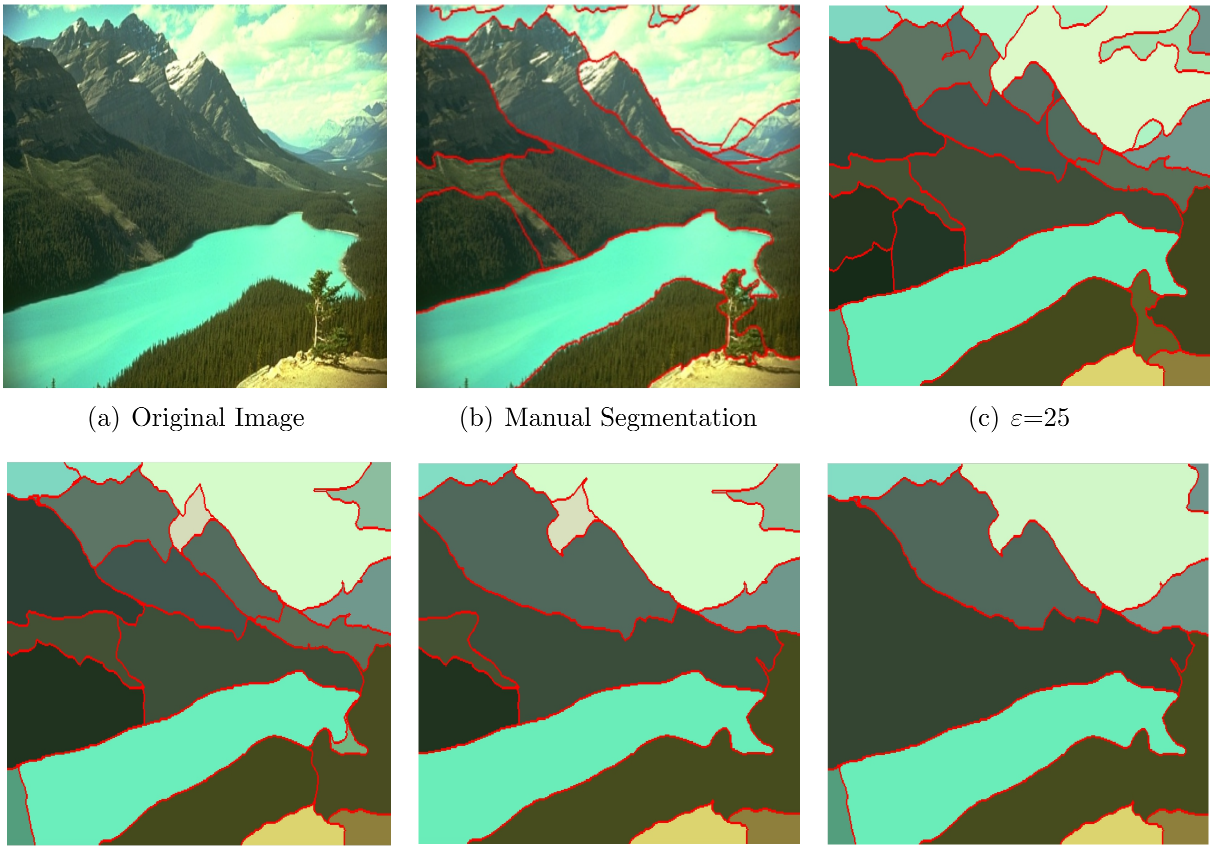
$$B\left(R\right) = -\sum_{i=0}^{7} \#\left(\Delta o_t = i\right) \log_2\left(P\left[\Delta o = i\right]\right) \tag{2}$$

By calculating the value $\Delta L_{w,\varepsilon}\left(R_i, R_j\right)$ to determine whether the two regions are fused, this algorithm can achieve the fusion of the first segmentation based on graph theory, and it can achieve relatively overall image segmentation results. Compared with CTM algorithm, the feature utilization of the algorithm is more abundant and accurate. It not only uses the color features in the region, but also uses the contour features of region, so it can more accurately judge the effect of the combination of two regions. In the process of using boundary and color features, this algorithm eliminates the common features in the two regions, such as the boundary features at the junction, so that the features of the merged region are more comparable to those of the two regions.

2.2. **Improved Model Basis.** Image segmentation is the basis of image annotation. Accurate image segmentation results can provide better materials for the subsequent image processing process. According to the results of image segmentation, the realization efficiency of the final image annotation will be different. In order to achieve a better image segmentation effect, we need to adjust the corresponding parameters according to the different images. As can be seen from Figure 2, the results of image segmentation will be greatly different with different initial parameters. According to the results of artificial segmentation, choosing appropriate parameters can achieve a effect of image segmentation closer to the results of artificial segmentation. Based on this, color features and shape contour features of segmentation results are better used to achieve the result of image annotation.

3. **Our Improved Method.** The main goal of image segmentation is to provide better materials for subsequent image recognition or image processing. To achieve this goal, it is necessary to select the optimal parameters and features for image segmentation. TBES (KFTBES) based on Kernel Fisher algorithm is to optimize result of image segmentation [7], so that it can be used as a basis to meet other research objectives. The main research direction of this topic is image annotation, the goal of image segmentation is to identify the whole thing in the image better. The process can be described as follows:

1. **Preprocessing** According to the standard of image segmentation (PRI and VOI) and the results of image segmentation, the image in training set is selected as the



(a) Original Image (b) Manual Segmentation (c) ε=25

(d) ε=5 (e) ε=100 (f) ε=175

FIGURE 2. Results of Artificial Image Segmentation and Segmentation Results with Different Parameters

optimal fusion parameter ε, and images in training set are divided into 8 categories according to different ε. The kernel function type used in SVM classification algorithm is determined by experiments. The Gaussian kernel function is used in the algorithm of this subject, and the projection direction and selection criteria of each category are calculated.

2. **Determination of graphic segmentation parameters** For a new input image, the initial image segmentation parameters of the image should be determined firstly. Then, according to the contour features of the image in the gray mode, the kernel Fisher algorithm is used to classify the image, so as to determine the image segmentation parameter ?. The initial number of blocks is 200.
3. **Initial image segmentation based on graph theory** The key part of image segmentation algorithm based on graph theory is to construct the weight matrix between pixels, which is the basis of clustering pixels in the image. In this algorithm, the gradient changes of lightness, color and texture features between adjacent pixels are used to calculate the possibility matrix C with boundary between any two pixels. Then the final weight matrix is determined according to the position relationship between any two pixels and the boundary possibility from one pixel to another. Then, according to weight matrix and the basic idea of NC algorithm, the pixels are clustered into the image segmentation results with the initial number of blocks. Among them, the gradient changes of lightness, color and texture features are compared and calculated in the form of color histogram, and the possibility that

the pixel is used as the boundary in 8 directions is detected. The final possibility matrix C records the direction and value that is most likely to be the boundary.

4. **Region fusion based on compressed coding** On the basis of initial segmentation, the process of merging adjacent regions is the same as that in TBES algorithm. According to the fusion standard ε, the final image segmentation result is obtained.

4. **Experiments and Analysis.** In order to verify the feasibility and effectiveness of the image segmentation algorithm proposed in this paper, we experiment and compare the algorithm and related algorithms on the general training set.

Berkeley segmentation dataset (BSD) data set is a data set mainly for image segmentation group proposed by Computer Vision Laboratory of Berkeley University. It is composed of 500 images, each image has three or four artificial segmentation results. By comparing with artificial segmentation, the effect of image segmentation algorithm can be tested. For this data set, 200 images are used as test set, and the rest are used as training set to preprocess the segmentation parameters. At the same time, this project tests the implementation effect of the same kind of correlation algorithm, and the implementation results are shown in Figure 3.

As can be seen from Figure 3, KFTBES algorithm is closer to the boundary contour of the original image compared with other algorithms, and the result of image segmentation is more inclined to the overall description of the object or scene in the image. For example, in the case of large number of blocks, Ms algorithm still segments the human body incorrectly, so the segmentation result is difficult to represent the scene or object information. CTM, F & H algorithm is similar to this, in the segmentation process, the basic contour information of the image is destroyed, while NC algorithm is more accurate when the number of blocks is large. This is the conclusion drawn from the intuitive observation. According to the two evaluation criteria of image segmentation, ours evaluates the results of image segmentation by numerical form. The data in Table 1 reflects the similarity difference between image segmentation results and manual segmentation results, where PRI represents the similarity between segmentation results and manual segmentation, and VoI represents the differences between segmentation results and manual segmentation. At this time, the Gaussian kernel function σ= 900 in KFTBES algorithm.

Table 1. Experimental Results of Several Image Segmentation Algorithms on BSD

Index/Method	MS	NC	F&H	CTM	TBES	KFTBES
PRI (higher is better)	0.7914	0.7139	0.7792	0.7624	0.8000	0.8021
VoI (lower is better)	1.9146	2.6693	2.4404	2.3005	1.7004	1.6989

From the above data in Table1, it can be seen that our proposed algorithm is better than classical image segmentation algorithm, and slightly better than the original algorithm. This data set contains a variety of segmentation results, which can be more comprehensive and accurate analysis and comparison of segmentation algorithms. From the experimental results, we can see that the algorithm used in this paper is more suitable as the basis of scene annotation, and can better describe the outline information of the object or scene to be labeled.

5. **Conclusions.** The effect of the optimization algorithm proposed in this paper performs better than the classical image segmentation algorithm, and also slightly better than the original. In addition, it can set the initial parameter adjustment standard according

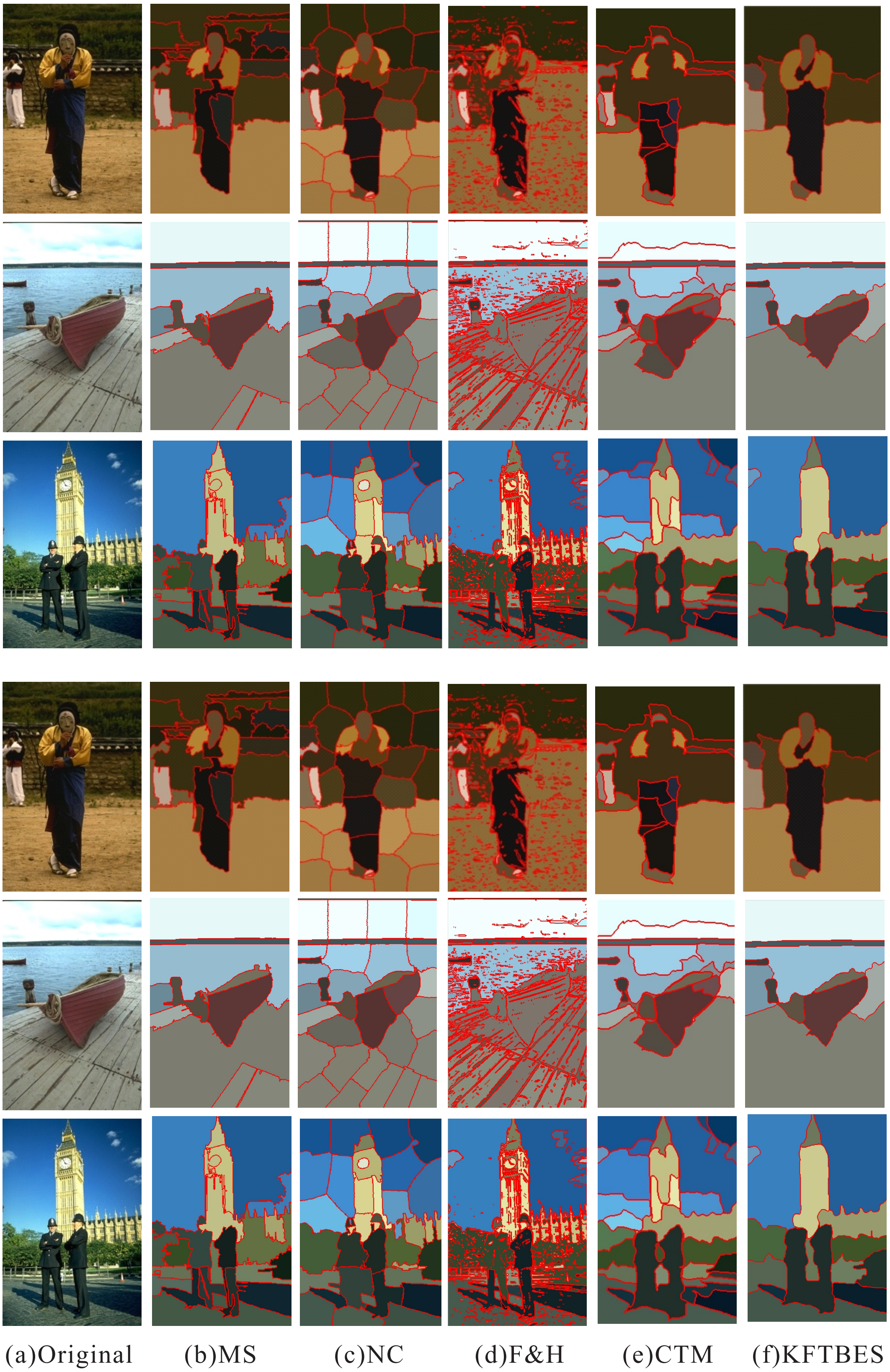


FIGURE 3. Experimental Results of Several Image Segmentation Algorithms on BSD

to the experimental data, and can set the corresponding initial parameter standard according to the experimental objectives and research significance, which is more conducive to realize the subsequent scene annotation process.

Acknowledgment. This work was supported by Shenzhen Science and Technology Plan Fundamental Research Funding JCYJ20180306171938767 and Shenzhen Foundational Research Funding JCYJ20180507183527919.

REFERENCES

[1] Jia-Yi Zhao, Jin Gong, Si-Teng Ma, Zhe-Ming Lu, Shu-Chuan Chu and John F. Roddick, Finger Vein Recognition Scheme Based on Convolutional Neural Network Using Curvature Gray Image, *Journal of Network Intelligence*, 4(3): 114-123, August 2019.

[2] Akgul C B, Rubin D L, Napel S, et al. Content-based Image Retrieval in Radiology: Current Status and Future Directions[J]. *Journal of Digital Imaging*, 24(2): 208-222, 2011.

[3] Toyama K, Vronay D, Anandan P. Image Exchange with Image Annotation: U.S. Patent 7,068,309[P]. 2006-6-27.

[4] Yitzhaky Y, Peli E. A Method for Objective Edge Detection Evaluation and Detector Parameter Selection[J]. *Pattern Analysis and Machine Intelligence*, 25(8): 1027-1033, 2003.

[5] Rao S R, Mobahi H, Yang A Y, et al. Natural Image Segmentation with Adaptive Texture and Boundary Encoding[M]. *Springer Berlin Heidelberg*, pp. 135-146, 2010.

[6] Yang A Y, Wright J, Ma Y, et al. Unsupervised Segmentation of Natural Images via Lossy Data Compression[J]. *Computer Vision and Image Understanding*, 110(2): 212-225, 2008.

[7] Liu J, Zhao F, Liu Y. Learning Kernel Parameters for Kernel Fisher Discriminant Analysis[J]. *Pattern Recognition Letters*, 34(9): 1026-1031, 2013.