Research on Image Region Classification Based on Automatic Segmentation

Liao YuanLai

Heyuan Polytechnic, HeYuan 517000, china zsblyl@163.com

Abstract

To deal with rough image region segmentation results, a common method of bag-ofwords in statistical text is proposed in the paper. The method splits regional objects to numerous small image blocks, from which rough semantic concepts of regional objects are fetched; then, through the application of multi-instance learning idea and computation of type confidence degree of each rough semantic concept, the impacts of type errors on such concepts can be effectively eliminated, and thus feature semantic concepts of various regional object type are obtained.

Keywords: region classification, image processing, Non-photorealistic rendering

1. Introduction

The existing non-photorealistic rendering technologies focus mostly on simulation results of artistic treatment. Schofield [1] pointed out that non-photorealistic rendering actually can be defined to construct images with inseparable and graphical sub units like brush marks which are bigger than image pixel [2-4]. From the perspective of visual style of hand drawing, artists finally complete jobs with strokes on by one whenever they're doing creations [5-6]. Hence, the rendering algorithm based on stroke layout receives increasing attentions from researchers and has become an important topic in the field of non-photorealistic rendering [7]. When strokes are used to overlay the whole picture, describing illumination with different colored strokes can render the shape of feeling and using strokes of different size can deliver different shape features of the depicted object. With special rendering ways like special kind of strokes or only sketches of object outline, different materials like furs, grassland or leaves can be portrayed [8-10].

According to methods for extracting image semantics, high-level semantic models include three methods: semantic attribute, local semantic concept, as well as semantic object. Semantic attribute method depicts the entire image with a group of global visible characteristics of it, *e.g.* Tamura features [11]. The semantic information is not bound with local object or region, instead associated with global construction and scene structure. Some approaches [12] introduce one middle-level semantic expression and use it as transit between image attribute and scene classification, as to fill up the semantic gap between low-level features and high-level semantics. This type of technique defines firstly a local semantic concept dictionary, included in which the local semantic concept usually defines a few local objects like blue sky, grey sky, snowy mountain, snowless mountain *etc.* Through learning of those concepts, the distribution of them in every scene type can be acquired [13-14].

The other is semantic object method [15]. This type segments one image to many partitions with independent significance. Detect and recognize all existing objects in each region; next based on those objects, describe the whole image or classify the whole scenario to find that the goal of this method is in complete agreement with the required input of non-photorealistic rendering method. The major difficulty of this method is in the necessity of accurately segmenting image regions at earlier stage, while image region

segmentation is itself a pending issue till now. Vogel [16] cut image to 10*10 subregions; then, classify those sub-regions to avoid segmentation of image region; however, since the split sub-areas are not semantic objects with meanings, so the solution can't reach better image region classifying result [17-18].

For semantic object method, we use bag-of-words method in the statistical text analysis. By combining multi-instance learning concept, we raised a semantic object method for image regions of rough segmentation [19-20]. For training images by rough segmentation, each regional object after separation is cut into smaller image blocks [21-22]. Based on those images, rough semantic concepts of various regional objects are fetched; then based on the confidence degree of those rough semantic concepts, feature semantic concepts of object type in each region, so as to implement effectively the classification of objects in image zones [23-24].

2. Classifying Method of Multi-Instance Image Regions

Traditional semantic object methods require firstly regional segmentation of images, regarding each region as an independent object; then classify objects in each region. Since image region segmentation algorithms are of inaccuracy, together with nature diversities, many different types of objects are included in one area. In the traditional training methods, these hybrid objects can be considered as object of one type, objects of the same type marked for training. Take Figure 1 for instance. Plentiful rocks exist in the seawater; but image segmentation can't separate them and thus the whole segmentation region is labeled as seawater type. Rocks there can result in mistakes during the training. For traditional semantic object methods [25], if it's not possible to accurately break up objects of a variety in those regions, those regions will become false examples in the training process and therefore affect the ultimate training results.



Figure 1. Marked Object Contains the Rock Types of Objects in the Sea

In multi-instance learning methods, each packet waiting for classification is consisted of a few samples. If in the package exists at least one correct sample, it's called positive package. If all samples in the package is wrong, it's called negative package. At present, the most common multi-instance learning method is Diverse Density method [26]. The method calculates in the feature space the point which is the closest to each positive packet and the farthest from each negative point, which is considered as characteristic point of the type. When one testing pack is determined to which it belongs, it's only required to compute if the distance between the testing pack and the feature point of the type is smaller than one threshold value. The method proposed in the paper splits one regional object into a few image blocks. The actual type of some image blocks is identical to the regional object, and the actual type of some is different from the regional object. Unlike general multi-instance learning problems, most image blocks have similar type with the original regional object, *i.e.* most samples in the pack are correct samples. But in general multi-instance learning problems, a great amount of samples in the positive packet are false samples. To adapt to the feature of the problem, we introduce one multi-instance learning method which utilizes secondary independent clustering, which guarantees the precision and also enhances the speed of multi-instance learning.

The proposed training method here includes two steps: rough semantic concept extraction and feature semantic concept extraction. In the fetching process of rough semantic concepts, image blocks of objects separated from each region, the rough semantic concept of the region is produced. This semantic concept mixes some false characteristics; so it can't be used directly to classify new regional objects. After rough semantic concepts of all regional objects are fetched, on the basis of these concept, many feature semantic concepts are generated. In the feature semantic concept, rough semantic concepts are excluded from blended faulty features, and the annotated type can be precisely conveyed.

3. Extraction of Rough Semantic Concepts

Firstly, divide each region object to many 5x5 image blocks, overlapping pixels of 2 rows or 2 columns in neighboring image blocks. It is shown in Figure 2. Any one regional object R can be expressed as $R = \{b_o, b_1, ..., b_n\}$, where, $b_j \in R$ is an image block that is cut out. The type of annotation of regional object as L(R).



Figure 2. 5 x 5 Image Block with 2 Rows or 2 Rows of Overlapping Pixels

After segmentation of region object is completed, the number of acquired image blocks is enormous; so it's necessary to simplify its scale. According to location feature and color feature of those image blocks, we can use Gaussian distance to perform autonomous clustering, naming some obtained clustering centers as the rough semantic concept of original regional object. The type of such rough semantic concept is marked similar with the type of the original regional object. It is shown in Figure 3.

Use f to represent rough semantic concept and its annotated type is tagged L(f). From one region object $R = \{b_o, b_1, ..., b_n\}$, $\{f_0, f_1, ..., f_m\}$ can be withdrawn as rough semantic concept of the regional object. If one image block b_j (j = 0, 1, ..., n) belongs to rough semantic concept f_i (i = 0, ..., m), Then it is written as $b_j \in f_i$, after clustering, since some image blocks of wrong types exist in the regional object, the actual type of rough semantic concept acquired after clustering of those image blocks differs from the annotated type. Based on the following equation, we can get the credibility of each rough semantic concept generated during the clustering:

$$C(f_i) = \left(\sum_{b_j \in f_i} \frac{1}{d(b_j, f_i) + 1}\right) \times \frac{V_{f_i}}{V_{avr}}$$
(1)

Take example of sea in Figure 4, in the rough semantic concept produced by ocean, there's at least one marine feature which can be used for classifier learning.



Figure 3. Rough Semantic Concept Extraction





Figure 4. The Image Blocks with Different Semantic Properties are Different from the Rough Semantic Concepts in the Same Region

4. Extraction of Feature Semantic Concepts

Suppose one rough semantic concept S and its actual type L_0 ; but regional objects belonging to L_1 are falsely marked as $L(S) = L_1$ type; perform autonomous clustering of all rough semantic concepts fetched from all regional objects. S will turn to the clustering center where the rough semantic concept of each actual type which is L_0 reaches and that S is separated from the type L_1 . It is shown in Figure 5.

Make total regional object $\{R_0, R_1, ..., R_t\}$; and all rough semantic concepts $\{f_0, f_1, ..., f_m, f_{m+1}, ..., f_s\}$ which are fetched from regional objects; then carry out autonomous clustering of those rough semantic concepts; the generated clustering center is called feature semantic concept, as F. Assume the fetched feature semantic concept

 $\{F_0, F_1, ..., F_r\}$. If one rough semantic concept $f_i (i = 0, ..., s)$ belongs to feature semantic concept $F_k (k = 0, 1, ..., r)$, it marks $f_i \in F_k$. For each F_k , choose randomly two rough semantic concepts $f_i, f_i \in F_k$, they should have similar actual type, *i.e.* the type $L(F_k)$ of feature semantic concept F_k , but their annotated type $L(f_i)$ and $L(f_i)$ are not similar. Since it's not possible to know beforehand the actual type of each rough semantic concept, it is necessity to find out the type of feature semantic concept, with the following expression:

$$P(L(F_k) = L) = \sum_{L(f_j) = L} \frac{C(f_j)}{D(f_j, F_k) + 1}, \quad k = 0, 1, \dots, r, \quad j = 0, 1, \dots, s$$
(2)

In the process of the experiment, eight types of regions $\{L_0, R_1, ..., L_7\}$ are used. So, for each feature semantic concept, it is required to calculate the probability of the eight types, and selected the type that has the greatest possibility, and label it as the type of the semantic concept of the feature.

$$L(F_{k}) = \arg \max_{P}(L_{j}), \quad k = 0, 1, \cdots, r, \quad j = 0, 1, \cdots, 7$$
(3)



Figure 5. Extraction of Feature Semantic Concepts

5. Experiment Design and Discussion

Experiments used the LabelMe spatial database in the envelope directory as a training image sample library. The image library contains a total of 2688 color images of 256 x 256. It is divided into coast, forest, highway, inside-city, mountain, open-country, street and tall-building eight kinds of global scenarios

5.1. Training Stage

Images in LabelMe library have been largely segmented manually; also each regional object after segmentation is manually annotated. *i.e.* sea, mountain, sky, building, traffic. Studies reveal that to understand contents in complicated scenarios does not require

recognizing all detail objects in the scenes, which means, in the image library some manually annotated detail objects are not too helpful to understand the global scenes of images. Hence, the objective of the method here is to identify accurately main regional objects in the images. After preliminary screening and classification of objects' tags, manually annotated regional objects are divided into eight kinds, *i.e.* sky, building, water, stone, plant, sand, field, road. Objects with little space in the image are removed. It is shown in Figure 6.

The method in the paper has no specialized restrictions for image segmentation; in other words, if training image library utilizes machine segmentation algorithm for image partition, the technique here is applicable as well. The reason for applying training image library in manual segmentation regions rests with manual segmentation tends to mix together regions with close semantics. However, those regions vary a lot from each other in terms of visible features like color and texture, which is corresponding to different semantic concept, increasing the difficulty in using traditional methods. For the proposed algorithm here, the difficulty is equivalent to the problem encountered when the machinery automatic segmentation algorithm is employed. But, since manual segmentation and marking demand huge workload, the segmentation precision is greatly restricted by operator's subjective wills. Therefore, in the training process here, we use manual segmentation image to generate feature model. In the testing procedure, it still needs to use automatic segmentation method to execute regional segmentation of images which are ready for testing.



Figure 6. Preliminary Screening and Classification of Image Data

5.2. Testing Stage

When it requires to make regional classification of testing images, traditional Mean-Shift method should be firstly used to make image cutting; then compute the Gaussian distance between each segmented region and each semantic concept; next based on results, classify the segmented regions.

Likewise, split regional objects into image blocks of 5x5 and calculate respectively the Gaussian feature distance between those image blocks and each feature semantic concept; then vote according to the following strategies as to judge the class of those regional objects:

(1) Compute the closest feature semantic concept of each image block and label image block as the type of the feature semantic concept; in the entire regional object, the type which is marked the most times is type of the regional object;

(2) Calculate characteristic distance between each image block and the nearest feature semantic concept; then compute the average feature distance from the whole regional object to different object type; the closest type is the one of the regional object;

(3) Estimate the feature distance between each image block and all feature semantic concepts as to get the feature semantic concept which is the most approximate to the feature distance of the overall regional object, the type of which is the type of the regional object.

In the experiment, we found that the first two voting strategies realized favorable accuracy rate of results, except the third one which had bad results. That is because the third strategy attempted to regard the whole regional object as the same one accurate semantic concept and thus overlooked other feature semantic concepts in the entire region which may belong to the same type, leading to less reliable results than the other two. It is shown in Table 1.

Table 1. Cross Validation Results Using Three Different Voting Strategies (%)

	Strategy 1	Strategy 2	Strategy 3
Include building objects	79.53	80.05	74.96
Not include building	87.01	87.76	81.48
objects			

In the testing, since automatic segmentation algorithm is adopted, some wrong regional objects may occur, which might be classified into one of the eight objects. To classify them won't have impacts on the accuracy of categorizing the whole scene. Again SVM classifier is utilized to test a variety of separated regional objects.

Experimental results are seen in Figure 7. The method performs better for natural objects which have evenly distributed texture and fewer forms, *i.e.* with fewer feature semantic concepts, *e.g.* water, sand, field, road; Namely 96.8%, 99.5%, 96.6%, 97.2%; also it reaches good accuracy rate of classifying for natural objects like sky, water, stone, plant. Namely 89.6%, 82.9%, 85.7%, which has more shapes, *i.e.* more semantic feature concepts. But the accuracy rate of classifying objects like building is very low. Namely 67.7%.



Figure 7. Cross Validation Results Using Lib-SVM for Various Terrain

Why the accuracy rate differs is because building belongs to man-made works so that its feature is greatly different from natural things. Natural things are usually featured with color and texture; while man-made things are often characteristic of shape and edge. During the implementation of the proposed method, clustering the divided image blocks to fetch rough semantic concepts uses only location and color features. During the segmentation, no specific treatment is made in accordance to regional object's shape and boundary. Therefore, in this case, it's not appropriate for the classification of artificial things. That's why the classifying result of building is worse.

6. Conclusion

With the use of a new algorithm, traditional semantic object methods depend less on the accuracy of image region segmentation. Before the introduction of better cutting method, it's possible to get better accuracy rate of classifying regional objects with bagof-words method and multi-instance learning idea, providing better prerequisite inputs for choosing proper non-realistic rendering parameters as per different types of the region.

References

- [1] S. Schofield, "Non-photorealistic Rendering: A Critical Examination and Proposed System", Ph.D. thesis, School of Art and Design, (1994).
- [2] "Middlesex University".
- [3] B. Hong, X. Guangmei, F. Songhe and X. De, "Research progress in the technology of automatic image annotation", Computer science, vol. 7, (2011), pp. 35-40.
- [4] W. Meng and L. Jianping, "Image segmentation algorithm based on edge detection and automatic seeded region growing", Journal of Xi'an University of Posts and Telecommunications, vol. 6, (2011), pp. 16-19.
- [5] Y. Huaiyi, "Support vector machine for multi object image segmentation algorithm simulation research", Computer simulation, vol. 12, (2011), pp. 223-226.

- [6] J. Kong, W. Jianan, W. X. Gu and M. Yin, "Based on regional automatic seeded region growing method of color image segmentation algorithm", Journal of Northeast Normal University (Natural Science Edition), vol. 4, (2008), pp. 47-51.
- [7] S. M. Li and H. Guoqiang, "Segmentation method based on K-means clustering algorithm of image region", Computer engineering and applications, vol. 16, (2008), pp. 163-167.
- [8] L. Xianglou, D. Ping, Y. Long and Z. Ming, "Automatic segmentation method based on skin color of eye area", Automation and instrumentation, vol. 5, (**2013**), pp. 140-142.
- [9] G. Yanwen, W. Linbo, Chen and J. Guoping, "Sample driven semi auto image set before background segmentation", Computer aided design and computer graphics, vol. 6, (2013), pp. 794-801.
- [10] J. Lixing and H. Jin, "Automatic image annotation based on ensemble classification algorithm", Automation journal, vol. 8, (2012), pp. 1257-1262.
- [11] B. S. Funch, "The Psychology of Art Appreciation", Museum Tusculanum Press, (1997).
- [12] H. Tamura, S. Mori and T. Yamawaki, "Texture features corresponding to visual perception", IEEE Transaction on Systems, Man and Cybernetics, vol. 8, no. 6, (1978), pp. 460-473.
- [13] M. Varma and A. Zisserman, "Texture classification: are filter banks necessary", Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2003), pp. 691-698.
- [14] J. Portilla and E. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients", International Journal of Computer Vision, vol. 40, no. 1, (**2000**), pp. 49-70.
- [15] J. Y. Gao and H. Luo, "Statistical modeling and conceptualization of natural images", Pattern Recognition, vol. 38, (2005), pp. 865-885.
- [16] J. Vogel and B. Schiele, "Natural scene retrieval based on a semantic modeling step", Lecture Notes in Computer Science. Heidelberg: Springer, vol. 3115, (2004), pp. 207-215.
- [17] Z. Rong and W. J. Wang, "Baixuefei. Automatic selection of training samples of SVM color image segmentation method based on", Computer science, vol. 11, (2012), pp. 267-271.
- [18] W. Bin and G. X. Bo, "The automatic image segmentation method based on level set relay", Journal of software, vol. 5, (2009), pp. 1185-1193.
- [19] L. Dongxing, H. Xincheng and D. Qinjun, "Automatic extraction algorithm based on region of interest and SVM", Computer simulation, vol. 1, (2014), pp. 365-369.
- [20] R. L. Yang and G. Xinbo, "Based on weighted fuzzy c-means clustering fast image automatic segmentation algorithm", Journal of image and graphics, vol. 12, (2007), pp. 2105-2112.
- [21] L. Lijun, X. Zhidong and Q. D. Wang, "Automatic segmentation of virtual human color images based on SVM", Journal of Huazhong University of Science and Technology (Natural Science Edition), vol. 1, (2006), pp. 42-45.
- [22] L. Qingwen and C. Wufan, "Automatic segmentation of medical images based on mutual information entropy difference measure", Chinese science E: Information Science, vol. 6, (**2006**), pp. 657-667.
- [23] K. Yongzhen, J. W. Zhang, J. Sun, Z. Yi and Z. Xiaozhou, "Combined with support vector machine (SVM) and C mean clustering image segmentation", Computer Application, vol. 9, (2006), pp. 2081-2083.
- [24] L. Y. Jianjun, L. Jianrong and J. Aike, "Automatic segmentation of medical image based on adaptive region", Journal of Beijing University of technology, vol. 8, (2010), pp. 1124-1129.
- [25] A. Bosch, X. Munoz and R. Marti, "Review: which is the best way to organize/classify images by content", Image and Vision Computing, vol. 25, (2006), pp. 778-791.
- [26] O. Maron and A. L. Ratan, "Multiple-instance learning for natural scene classification", Proceedings of the Fifteenth International Conference on Machine Learning, (1998), pp. 341-349.

Author



Liao YuanLai, he received his master's degree of engineering in Guangdong University of Technology. He is a lecturer in Heyuan Polytechnic. He is in the research of Software Engineering, Image processing, Image quality assessment and Data mining. International Journal of Multimedia and Ubiquitous Engineering Vol.11, No.4 (2016)