Region Labeling in Natural Scene Images

Kyawt Kyawt Htay
Falculty of Computer Science
Myanmar Institute of Information Technology (MIIT)
Mandalay, Myanmar
kyawtkyawthtay2014@gmail.com

Nyein Aye
Principal of Computer University (Hpa-An)
Computer University (Hpa-An)
Kayin State, Myanmar
nyeinaye@gmail.com

***Abstract*— Semantic region labeling is important in the fields of low, mid and high level computer vision. This paper proposes an approach for region labeling in natural scene images using over-segmented regions as a basic unit. This system consists of three phases: segmentation, feature extraction and labeling. The paper uses Marker-controlled Watershed Segmentation (MCWS) algorithm for segmented regions generation. The aim of this system focuses labeling on nine semantic concept classes using color, texture features and 3-layer Feed Forward Neural Network (FFNN) classifier. The system performance is evaluated on the use of public MSRC 9-class dataset.**

Keywords— Marker-controlled Watershed Segmentation, Color, Texture, Feed Forward Neural Network (FFNN)

I. INTRODUCTION

There are two types of objects in image recognition applications: well-defined objects and natural objects. Recognizing objects in natural scene images is more complicated than objects of well-defined. The well-defined objects are mostly found in research of machine vision and automated target recognition applications. Natural objects are found mostly in remote sensing and scene analysis applications. There are many labeling systems in early worked on individual pixels, but later efforts on patches or super-pixels often achieve higher efficiency and consistency than working on individual pixels. The earlier works have drawbacks such as complexity increases rely on having an individual pixel to make inference tractable. There are several difficult problems to recognize objects in natural scene images. Firstly, this needs to find candidates for objects or objects of interest in the image and localize them. This problem is called Multi-class image segmentation or multi-class image labeling. This is a well-known problem in the area of computer vision and object recognition applications. Multiclass image labeling methods attempt to classify all pixels and aim at concurrent multi-class object recognition in an image. The early works operate on pixel level and utilize super pixels as a starting point. The author [1] uses super pixels as the basic unit and construct a classifier on the histogram of local features. The paper detected each super pixel and regularizes the classifier by aggregating histograms in the neighborhood of each super pixel. The proposed model by the authors [2], images are subdivided into a grid of 10x10 blocks which are classified into a class that are belonging to a class, out of predefined nine local-concept classes. Image regions are represented by a concatenation of 84-bin HSI color histograms, 72-bin edge-direction histograms, and 24 features of global holistic gist gray-level co-occurrence matrix for automatic concept classification of color feature is present. Image regions are represented by a concatenation of a 32-bin intensity histogram for automatic concept classification of color has been removed. Two support vector machine (SVM) classifiers were trained by using low-level feature information. The classification performance for color present is 71.7% on image region level and the performance drops to 65.7% on classifying gray-scaled regions. This paper focuses to solve the problem of simultaneously segmenting an image into its constituent semantic regions and automatically labeling each region into a class, out of a pre-determined set of classes.

II. PROPOSED SYSTEM

A. Segmentation

The system uses Marker-controlled Watershed algorithm for segmented regions generation. The algorithm consists of preprocessing, watershed transform and post-processing steps. This algorithm prewitt edge mask is used to produce gradient magnitude image. This operator uses 3 X 3 masks to find edges. The morphological dilation and erosion operations are basic in segmentation. The size and shape of structuring element is important to decide the number of pixels added or removed from the objects in an image. The algorithm uses disk-shaped flat structuring element for morphological operations. Finally, Otsu's thresholding method uses for background marker. By using the modified algorithm, the system produces labeled regions within a few seconds.

B. Feature Extraction

Feature extraction is an important step in region labeling. For color, 32+32+32=96 color features are extracted by using Color Histogram (CH) in RGB space. For texture, three texture features in four orientations 3x4=12 texture features are extracted by using Gray Level Co-occurrence Matrix (GLCM). The 96 color and 12 texture features extracted from each region are given as a feature set input to FFNN classifier for region labeling.

C. Labeling

In this system, two 3-layer FFNN classifiers are implemented by using color features alone and combination of color and texture features. Each network has its own network architectures. In each output layer, the two 3-layer FFNN classifiers have 9 neurons since the number of semantic classes is nine: aero-plane, cow, bicycle, building, face, car, sky, vegetation and unknown.

III. REGION MERGING

This paper, region merging approach is proposed. In the region merging approach, Region Adjacency Graph (RAG) is used. Region merging is a post-processing technique that merges regions adjacent with the same color values. In the proposed region merging approach, there are two predicates:

- The first predicate is minimum pair for each region.
- The second one is color homogeneity distribution of 3-layer FFNN classifier output image.

For first predicate, minimum edge pairs for each region are discovered by using Region Adjacency Graph (RAG). For second predicate, 3-layer (FFNN) classifier output images are used to obtain color homogeneity distribution. The region pair satisfies these two predicates are actually merged. The region pair does not satisfy these two predicates are ignored.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This paper uses the Microsoft Research Cambridge object recognition database [3]. This system uses images for training and testing from MSRC 9-class dataset. Accuracy of our approach on region labeling is measured according to ground truth images of MSRC 9-class dataset.

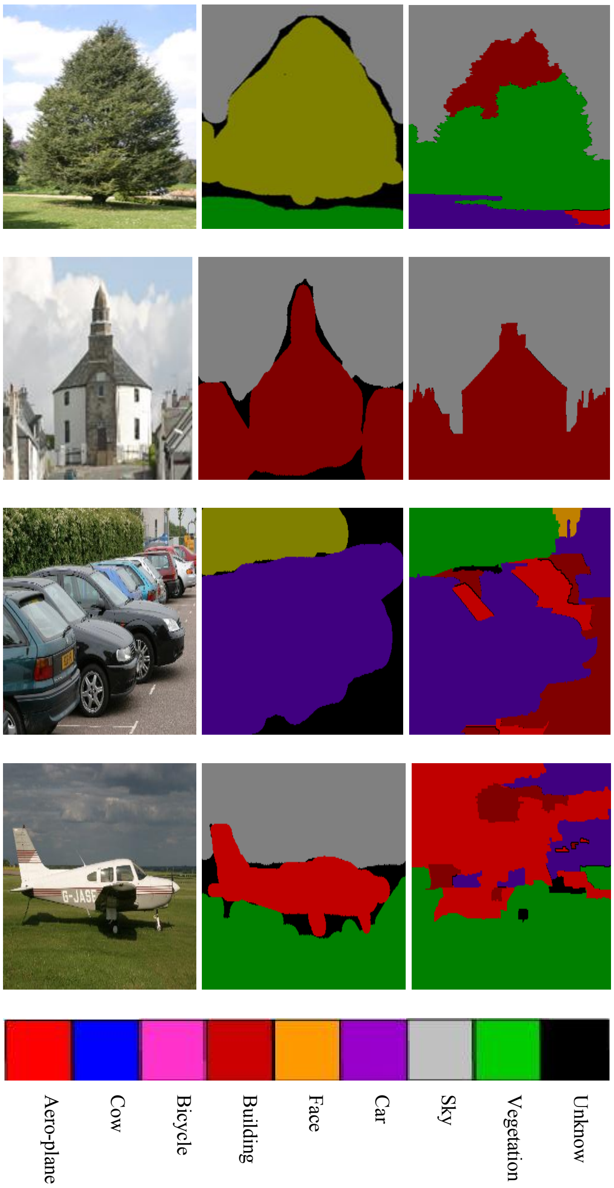

Fig. 1. Result Images of Our Method Compare with Ground Truth Images

The above figure shows the result images of our method compare with the ground truth images of MSRC 9-class dataset. First column represents original images, second column its ground truth images and third column represents the result images of our method.

TABLE I. COMPARISON OF OUR METHOD ON STATE-OF-THE-ART MSRC RESULTS

Pixel-wise Labeling Accuracies	
Methods	***Labeling Accuracy***
J. Verbeek and Triggs B. [4]	67.1%
Schroff et al. [5]	72.2 %
Our method with CH	68.2%
Our method with CH+GLCM	72.4%

The above table shows labeling accuracy of our approach using color features alone is 68.2% and the combination of color and texture features is 72.4%. We compared our method with state-of-the-art techniques in [4] and [5] on the 9-class MSRC dataset. Our labeling approach is still competitive with the state-of-the-art techniques.

V. CONCLUSION

The proposed method generates segmented regions with nine meaningful labels: aero-plane, cow, bicycle, building, face, car, sky, vegetation and unknown. This paper shows the proposed method of labeling accuracies 68.2% for color features alone and combines with texture features is 72.4%. The proposed method with the combination of color and texture features observes most of the vegetation classes. The method has difficulty in capturing the entire body of aero-plane and face, as some parts are mis-categorized as vegetation and building.

REFERENCES

[1] B. Fulkerson, A. Vedaldi, and S. Soatto, “Class Segmentation and Object Localization with Super pixel Neighborhoods”, In Proceedings of the International Conference on Computer Vision (ICCV), 2009.

[2] Vogel, J. and Schiele, B. “Natural scene retrieval based on a semantic modeling step” in Proceedings of International Conference on Image and Video Retrieval, Dublin.

[3] A. Criminisi. Microsoft research cambridge object recognition image database. Version 1.0, 2004 - http://research.microsoft.com/vision/cambridge/recognition/

[4] J. Verbeek and Triggs B. “Scene Segmentation with CRFs Learned from Partially Labeled Images”. In NIPS, 2008.

[5] Schro, F., Criminisi, A., Zisserman, A. “Object class segmentation using random forests”, In Proceedings of the British Machine Vision Conference, 2008.