

Semantic Concepts Classification on Outdoor Scene Images Based on Region-Based Approach

Kyawt Kyawt Htay and Nyein Aye

Abstract—Outdoor scene analysis is a complex problem for both image processing and pattern recognition domains. There are two methods of segmenting images to look for objects in an image, block-based and region-based. Region-based method can provide some useful information about objects even though segmentation may not be perfect. There are three phases in this system: segmentation, features extraction and classification. The basic idea of this system is to classify local image regions into semantic concept classes such as tree, sky and road etc. In this paper, modified Marker-Controlled Watershed (MCWS) algorithm is proposed. Firstly, the modified (MCWS) algorithm is used to segment input image. And then, texture feature vectors are extracted from segmented regions by Gray-Level Co-occurrence Matrix (GLCM). Finally, classification is performed by 3-layer Artificial Neural Network (ANN). This system is applied on real scene images dataset.

Index Terms—Marker-controlled watershed, outdoor scene analysis, texture.

I. INTRODUCTION

The semantic information of an image carries the meaning of that image. It is trivial for the human eye to extract semantic information from photos. However, for a computer, it is difficult to identify the semantic features of high-level images in photos. Therefore, if a computer can be made to correctly identify the semantic features of objects in photos, it will enhance the image identification rate. Towards this goal, segmentation of an image into regions has been used in recent years. Some researchers believe that a segmentation of images into regions can provide more semantic information than the usual global image features. Scene classification has become a popular research topic in recent years. However, even though much research has been done before, classifying photos into semantic types of scene (e.g., portrait, landscape) is still a difficult problem. The popular method for classifying scenes is to use low-level features (e.g., texture, color). Therefore, image classification can be achieved either by only using low-level features, or by integrating the low-level and high-level features [1]. Popular low-level feature is texture. Texture is the structure arrangement of the surface of an object, which assists users in recognizing objects or regions. Textures have been expressed using several methods [2], such as Gabor filters, wavelets, local binary patterns

(LBP), and so on. Vogel and Schiele [3] modeled the semantic content of an image and used this model to classify local image regions into semantic concept classes. Mode-based systems rely upon the configuration of the scene components. Luo and Stephen [4] proposed a model-based approach to detecting sky. This approach consists of color classification, region extraction and physics motivated sky signature validation.

II. RELATED WORKS

Image segmentation is a preliminary and critical step in segment-based image analysis. The best segmentation result is used in image classification. Scene-classification system needs to be sure which objects are in the image before classifying the scene. The common method of doing this is to segment the objects in the image and then identify the low-level features and semantic concepts. There are two methods to segment images, block-based and region-based [5], [6]. The block-based method simply segments the image into several rectangles. The region-based method also segments the image, but the objects are more meaningful to the human eye. The objective of a Scene Understanding System consists of recognizing and localizing the significant imaged objects in the scene and identifying the relevant object relation-ships. Consequently, a system must perform segmentation, region characterization and labeling processes. As Haralick and Shapiro suggested [7], the way to carry out these three tasks depends on the strategy used: bottom-up, top-down, or hybrid strategies. Gao-Hua, Iaeng and Zhao Chun-Xia, Zhang Hao-Feng [8], proposed a scene classification method. Firstly, texture features are extracted in gray channel. Then, color moments are computed in all three color channels of RGB color space. The parameters of GMM are estimated by training these labeled samples. Finally, new scene images are classified by the trained GMM model. In the system of Town and Sinclair [9] neural networks are trained to classify previously segmented image regions into one of eleven semantic classes such as brick, cloud, fur or sand. Serrano *et al.*, [10] proposed a set of low-dimensional, computationally efficient low-level features that are extracted from LST color space and wavelet texture features.

In this paper, region-based image segmentation method is used for semantic concepts classification on outdoor scene images. As a first step, modified Marker-Controlled Watershed algorithm (MCWS) is used to solve the problem of segmented regions generation. Therefore, we will focus our work in the problem of image regions labeling to classify

Manuscript received March 27, 2014; revised June 28, 2014.

Kyawt Kyawt Htay is with University of Computer Studies, Mandalay, Myanmar (e-mail: kyawtkyawhtay@gmail.com).

Nyein Aye is with the head of the Department of Hardware, University of Computer Studies Mandalay, Myanmar (e-mail:nyeinaye@gmail.com).

every region of a given image into one of several predefined classes.

III. PROPOSED SYSTEM

A. Overview of Proposed System

There are three mainly phases in this proposed system: segmentation, feature extraction and classification.

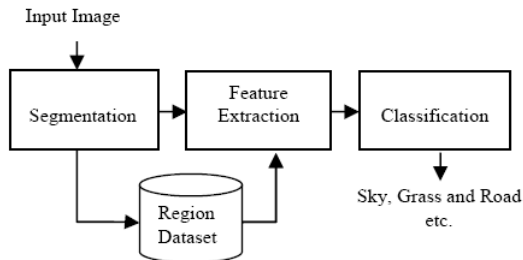


Fig. 1. Overview of proposed system.

According to Fig. 1, our proposed system used Modified Marker Controlled Watershed (MCWS) algorithm in the segmentation phase, in the feature extraction phase texture features are extracted from the segmented regions, and then classification is performed by 3-layer ANN by using these extracted features.

B. Modified Marker-Controlled Watershed Segmentation

There are three steps in the image segmentation phase: pre-processing, watershed transform and post-processing.

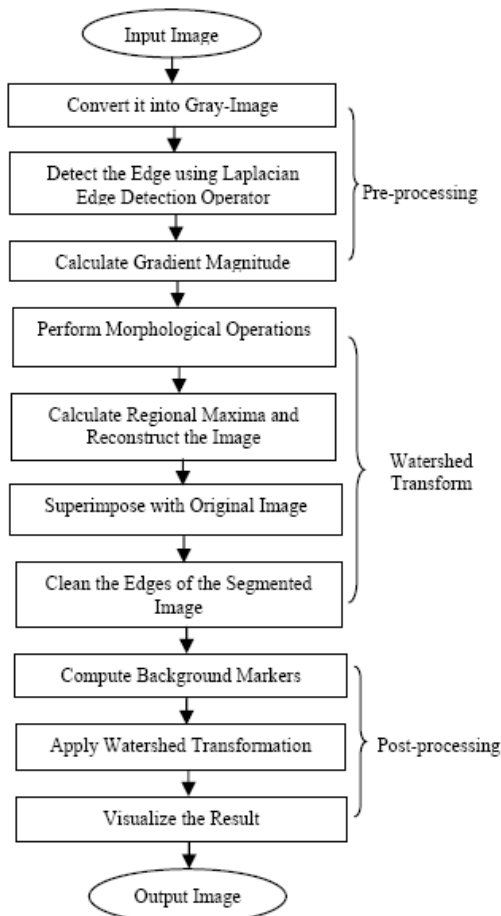


Fig. 2. Modified marker-controlled watershed (MCWS) algorithm.

According to Fig. 2, in the pre-processing step, scene

image is given as input and filter the image by using Laplacian operator instead of Sobel. Marker controlled watershed segmentation: Now mark the background objects with the help of marker. A variety of procedures could be applied here to find the foreground markers which must be connected blobs of pixels inside each of the foreground objects. Here the method that is used is morphological techniques called opening by reconstruction and closing by reconstruction to clean up the image. These operations will create flat maxima inside each object that can be located using `imregionalmax`. Now compute background markers. Now you need to mark the background. In the cleaned-up image, the dark pixels belong to the background, so you could use a thresholding operation to mark the background objects. Now compute the watershed transform of the segmentation function. The function `imimposemin` can be used to modify an image so that it has regional minima only in certain desired locations. Here you can use `imimposemin` to modify the gradient magnitude image so that its only regional minima occur at foreground and background marker pixels. Finally compute the watershed transform of the modified segmentation function.

IV. SEGMENTATION AND CLASSIFICATION

With the segmentation, objects of interest from image are extracted. Various techniques discovered till now for segmentation, here watershed algorithm is used. Watershed is also based on morphology. It is a region based algorithm having low computational complexity and high efficiency. It provides complete division of the image. Besides, all these advantages, it has a major drawback; it suffers from over-segmentation. Due to this, image content is distorted completely. So, some modifications are required to remove the problem of over-segmentation. In this paper, propose a pre-processing step in Marker-Control Watershed (MCWS) algorithm which actually reduces the number of segments produced by watershed algorithm.

A. Marker Control Watershed Transformation

The advantages of watershed transformation are that it is simple, instinctive knowledge, and can be parallelized. The main drawback of this method is the over-segmentation due to the presence of many local minima. To decrease the effect of severe over-segmentation, marker-controlled watershed transformation is proposed. Separating touching objects in an image is one of the more difficult image processing operations. The watershed transform is often applied to this problem. The watershed transform finds "catchment basins" and "watershed ridge lines" in an image by treating it as a surface where light pixels are high and dark pixels are low. Segmentation using the watershed transforms works well if you can identify, or "mark," foreground objects and background locations. Marker-controlled watershed segmentation follows this basic procedure:

- Compute a segmentation function. This is an image whose dark regions are the objects you are trying to segment.
- Compute foreground markers. These are connected blobs of pixels within each of the objects.

- Compute background markers. These are pixels that are not part of any object.
- Modify the segmentation function so that it only has minima at the foreground and background marker locations.
- Compute the watershed transform of the modified segmentation function.

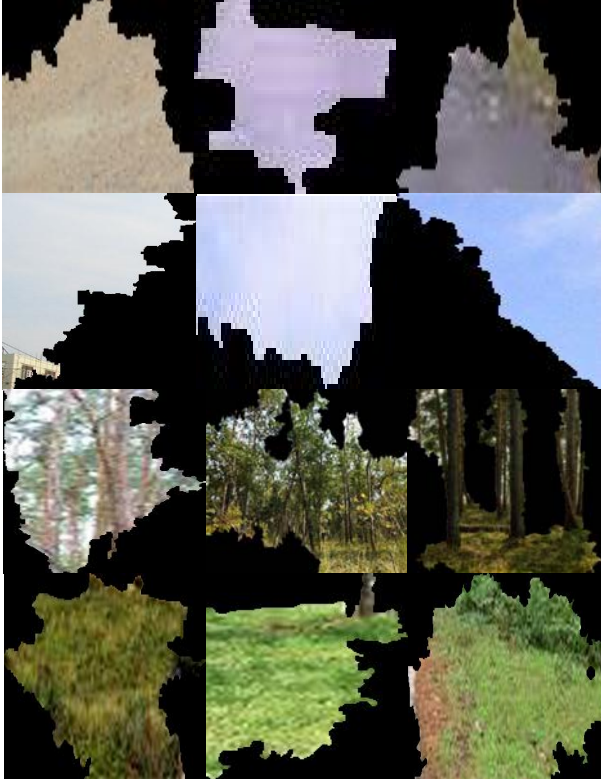


Fig. 3. Segmented regions dataset for road, sky, tree and grass.

The segmented regions dataset is obtained by using our proposed modified Marker Controlled Watershed (MCWS) algorithm in segmentation phase (see Fig. 3).

B. Classification

This system uses 3-layer Artificial Neural Network (ANN) for training and classification. In this architecture, consists of 81 input nodes, 100 neurons and 5 output nodes. It tends to classify previously segmented image regions into one of five predefined semantic classes such as sky, grass, road, tree and unknown etc.

V. FEATURES EXTRACTION

The type of features to be extracted from an image depends on the nature of the scene classification task. Semantic concepts classification on outdoor scene images based on color and texture features have been addressed by many researchers. In the present work, we deal with the scene images primarily containing natural regions. Although not sufficient, low-level features such as color and texture contain good representation power for the region classification of natural scenes.

It is well-known that color-based terrain classification of outdoor scene images is difficult when environmental effects, such as those due to weather, illumination, and atmospheric conditions, are present. In order to overcome the

shortcomings of color-based classification methods, other visual features can be used.

A. Texture

Texture can provide good discrimination between natural classes similar in the color domain e.g. water and sky. However, in our case, texture should be dealt with more care, as any single class does not have a unique texture. Within a semantically coherent region, there might be areas of high or low textures, with different scales and directional uniformity. In the classification of such regions, a very strong texture measure can sometimes undo the good work done by the color features. Several techniques have been reported in the literature to compute the texture in a pixel neighborhood.

B. Gray-Level Co-occurrence Matrix (GLCM)

Texture is a commonly used feature in the analysis and interpretation of images. Texture is characterized by a set of local statistical properties of pixel intensities. The GLCM method considers the spatial relationship between pixels of different gray levels. The method calculates a GLCM by calculating how often a pixel with a certain intensity i occurs in relation with another pixel j at a certain distance d and orientation θ . Each element (i, j) in the GLCM is the sum of the number of times that the pixel with value i occurred in the specified relationship to a pixel with value j in the raw image [11]. GLCM is calculated four directions: 0° , 45° , 90° and 135° degrees. The eight texture descriptors are extracted from each co-occurrence matrices which are computed in each of four angles. Thus 32 features are used to represent an image.

TABLE I: COMPUTATION OF TEXTURE FEATURES

FEATURES	FORMULA
Energy	$\sum_{i,j=0}^{G-1} p_{d,\theta}(i, j)^2$
Inertia	$\sum_{i,j=0}^{G-1} (i-j)^2 p_{d,\theta}(i, j)$
Entropy	$\sum_{i,j=0}^{G-1} p(i, j) \log_2 p_{d,\theta}(i, j)$
Homogeneity	$\sum_{i,j=0}^{G-1} \frac{p_{d,\theta}(i, j)}{1 + (i-j)^2}$
Maximum Probability	$\max_{i,j} p_{d,\theta}(i, j)$
Contrast	$\sum_{i,j=0}^{G-1} (i-j)^2 \sum_{i,j=0}^{G-1} p_{d,\theta}(i, j)$
Inverse	$\sum_{i,j=0}^{G-1} \frac{p(i, j)}{(i-j)^2}$
Correlation	$\frac{\sum_{i,j=0}^{G-1} ij p_{d,\theta}(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}$

The table shows eight equations that are used to extract texture feature vectors. And then, these extracted texture feature vectors are used as input to 3-layer ANN to classify

segmented regions such as Road, Sky, Grass and Tree etc.

VI. RESULTS AND DISCUSSION

The proposed system is applied on our real scene images dataset by various edge detection filters (Disk, Average, Motion, Gaussian, Laplacian, LOG, Prewitt, and Sobel). Laplacian operator is used for our proposed Modified Marker-Control Watershed (MCWS) Algorithm. For each filter: three different parameters are calculated i.e. ENTROPY, MSE, and PSNR. On the basis of these values, final result and conclusion has been drawn.

Entropy:

$$\text{Entropy} = \sum_i p_j \log_2 p_j \quad (1)$$

Mean Square Error (MSE):

It considers the quantity of the removed noise. The mean square error (MSE) is defined as:

$$\text{MSE} = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} (f(m, n) - f'(m, n))^2 \quad (2)$$

Value of MSE should be low for an efficient filter [12].

A. Peak Signal to Noise Ratio (PSNR)

PSNR is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. A higher PSNR generally indicates that the reconstruction is of higher quality. PSNR is most easily defined by the Mean Squared Error (MSE). PSNR can be defined as:

$$\text{PSNR} = 10 \times \log_{10} \left[\frac{255^2 \times M \times N}{\sum \sum (x(i, j) - y(i, j))^2} \right] \quad (3)$$

Value of PSNR should be high for an efficient filter [12].



Fig. 4. (a) Original image (b) Gradient image by standard watershed algorithm.

According to Fig. 4, the main drawback of traditional watershed algorithm is over-segmentation problem that is shown in (b) and (a) is original scene image. Therefore, to reduce over-segmentation problem of traditional watershed algorithm Fig. 5 (a) and (b) are shown using Marker-Controlled Watershed Segmentation (MCWS) algorithm and our proposed Modified Marker-Controlled Watershed Segmentation (MCWS) algorithm. Our proposed

Modified Marker-Controlled Watershed (MCWS) algorithm is reduced over-segmentation than traditional Marker-Controlled Watershed (MCWS) algorithm.

TABLE II: COMPARISON OF VARIOUS FILTER METHODS ON INPUT IMAGE

Filter	ENTROPY	PSNR	MSE
Disk	1.3339	21.1108	503.5052
Average	1.5666	20.2588	612.6373
Motion	1.4368	21.6421	445.5202
Gaussian	2.2609	19.7207	662.2395
LOG	2.2596	18.6852	880.1554
Laplacian	2.3615	19.8868	682.1544
Prewitt	2.2993	19.1281	794.8312
Sobel	2.2655	19.7871	682.9267

On the basis of Table II, it is observed that Laplacian filter has highest in Entropy value. Moreover, it has higher in PSNR value and lowers in MSE value than Sobel filter in existing marker-controlled watershed segmentation algorithm. Therefore, Laplacian filter is selected as an edge detection filter for our proposed Modified Marker-Controlled Watershed (MCWS) algorithm.



Fig. 5. (a) Marker-controlled watershed algorithm, (b) Modified marker-controlled watershed algorithm.

VII. CONCLUSION

Semantically meaningful labels to local image regions that can be used for subsequent retrieval step. In this paper, propose a method for semantic concepts classification based on region-based approach. Firstly, input image is segmented by Modified Marker-Control Watershed algorithm (MCWS) in order to get more accurate compact regions. And then, texture feature is extracted by GLCM. Finally, classification is performed by 3-layer ANN to label the regions produced by the segmentation process. For future work, the proposed method can be tested with various segmentation algorithms and low-level features. The classification results vary depending on the segmentation algorithms and features used.

ACKNOWLEDGMENT

We would like to thank my supervisor, Dr. Nyein Aye, at University of Computer Studies, Mandalay for providing

excellent supporting and the feedback from the discussions of our research topic.

REFERENCES

- [1] J. Luo and A. Savakis, "Indoor vs outdoor classification of consumer photographs using low-level and semantic features," in *Proc. the 2001 International Conference On Image Processing (ICIP 01)*, Thessaloniki, Greece, 2001, vol. 2, pp. 745-748.
- [2] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 837-842, Aug. 1996.
- [3] J. Vogel and B. Schiele, "Semantic modeling of natural scenes for content-based image retrieval," *International Journal of Computer Vision*, vol. 72, no. 2, 2004.
- [4] J. Luo and S. P. Etz, "A physical model-based approach to detecting sky in photographic images," *IEEE Transactions of Image Processing*, vol. 11, no. 3, pp. 201-212, Mar. 2002.
- [5] C. Ko, H. S. Lee, and H. Byun, "Image retrieval using flexible image subblocks," in *Proc. the 2000 ACM symposium on Applied computing 2000*, pp. 574-578, March 2000.
- [6] M. Szummer and R.W. Picard, "Indoor-outdoor image classification," in *Proc. IEEE International Workshop on Content-based Access of Image and Video Databases, Bombay, India, 1998*, pp. 42-51.
- [7] B. L. Saux and G. Amato, "Image recognition for digital libraries," in *Proc. the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval*, pp. 91-98, 2004.
- [8] G. Hua, C. X. Zhao, and Z. H. Feng "Visual features fusion for scene images classification," in *Proc. the International Multi-Conference of Engineers and Computer Scientists*, pp. 14-16, 2012.
- [9] B. L. Saux and G. Amato, "Image classifiers for scene analysis," presented at the International Conference on Computer Vision and Graphics 2004.
- [10] N. Serrano, A. Savakis, and J. Luo, "Improved scene classification using efficient low-level features and semantic cues," *Pattern Recognition*, vol. 37, no. 9, pp. 1773-1784, Sep. 2004.
- [11] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, no. 6, 1973, pp. 610-621.
- [12] H. P. Ng, "Medical image segmentation using Watershed segmentation with texture-based region merging," in *Proc. the 30th Annual International IEEE EMBS Conference, 2008*, pp. 4039-4042. Vancouver, Canada.



Kyawt Kyawt Htay received the B.C.Sc. and M.C.Sc. degrees in computer university (Mandalay) and Computer University (Loikaw), Myanmar in 2006 and 2010, respectively. She is also a tutor and a Ph.D candidate of University of Computer Studies, Mandalay, Myanmar. My research fields are image processing and computer vision.



Nyein Aye received the Ph.D degree from Engineering Physics Institute, Moscow. He currently serves as an associate professor of Computer University, Mandalay, Myanmar. His current research interests include pattern recognition, image processing and information security.