Volume 3, Issue 12, December 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Study of Few Approaches and Novel Descriptors used for Scene Image Classification

**Mr. Aniket D. Pathak*, Prof. Dinesh D. Patil**
*Department of CSE, SSGBCOET*
*Bhusawal, India*

***Abstract— Among the challenging topics in the area of computer vision scene classification is one. There are various approaches present through which this can be achieved, this paper mainly emphasizes on the two approaches one is holistic approach and other is Object-Based approach. This paper describes in detail the two approaches and also shows effect on combining both approaches. Also we explain the different novel image descriptors with their role. It depicts the fact that how performance gets increases when any of these are fused or combined to form a new one.***

***Keywords— scene classification, holistic approach, objectbased approach, CENTRIST, HaarHOG descriptor, 3D-LBP descriptor, 3DLH descriptor***

## I. INTRODUCTION

Scene classification is among the challenging topic of computer vision. It has attracted significant attention in the past decade, and many approaches have been proposed to resolve this problem. Basically, out of these we study two categories: holistic and object-based. Holistic approach represents scene which had content in simple form. However, since it didn't consider the internal object relationship, hence this approach didn't well characterize complex scenes with multiple objects. In contrasting manner, object-based approach estimated the scene class by analyzing the object co-occurrence information, as a result of which it is proved better in characterizing scenes with complex content. But object-based approach is not that better at classifying simple scenes. Instead the new approach which combines both approaches performs better. This paper surveys three novel descriptors for scene image classification which are useful as applications to image search and retrieval. Among them first is 3-Dimensional Local Binary Pattern (3DLBP) descriptor is for color image local feature extraction. Second, a new shape descriptor (HaarHOG) which combines Haar wavelet transformation and Histogram of Oriented Gradients (HOG). Third, descriptors are fused using an optimal feature representation technique to generate a robust 3-Dimensional LBP-HaarHOG (3DLH) descriptor that can perform well on different scene image categories.

## II. APPROACHES FOR SCENE IMAGE CLASSIFICATION

In this section we study three approaches used for scene image classification namely, Holistic based approach, Object based approach and the combinational approach which combines both holistic and object based approaches.

### *A. Holistic Approach*

Holistic strategy characterizes a scene image without considering the regional semantics. A popular holistic approach is bag-of-words [2], which characterizes the images using an intermediate representation. In bag-of words approach a dictionary of visual words is created for the scene images first, and then bag-of-words models are constructed to represent scenes. When the scenes have very few objects or simple global visual features, then this approach achieves superior's performance. But, for scenes which contained multiple objects and the objects play important roles in the scene discrimination, the holistic approach doesn't work well, as it didn't take into consideration the object semantic information. A number of holistic image representations have been proposed [1], [2], [3]. In this paper, we define two approaches namely: Spatial Pyramid Matching (SPM) and CENTRIST.

SPM characterizes image visual features using a number of SIFT descriptors. Then a vocabulary of visual words is builded based on the SIFT descriptors using k-means clustering. After that, each SIFT descriptor can be represented by the visual words. Finally, the visual words are concatenated through spatial pyramid to represent the images. Fig.1 shows the spatial pyramid split of an image.

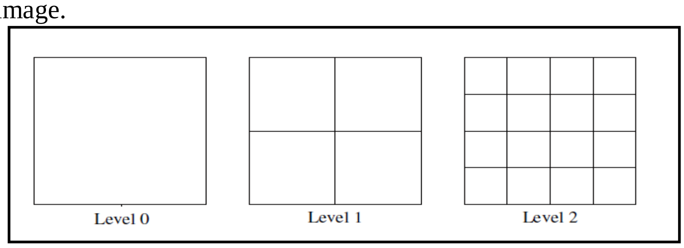



Fig.1 Spatial Pyramid Split of an Image

There are three levels: level 0, level 1, and level 2, respectively splits the image into 1 block, 4 blocks, and 16 blocks. Hence, in total 21 blocks are obtained. If the vocabulary contains v visual words, then the image is finally represented by a $v \times 21$ dimension feature vector.

CENTRIST [4] represents an image using the histogram of Census Transform (CT). CT compares the intensity value of a pixel with its eight neighboring pixels. If a pixel is larger than or equal to one of its neighbors; a bit 1 is set in the corresponding neighbor location. Otherwise, a bit 0 is set. The eight bits are then collected from left to right, top to bottom to form a binary number, which is consequently converted to a decimal number in [0- 255]. The decimal number is the CT value for the center pixel. Fig. 2 shows an example to calculate the CT value for the center pixel. CENTRIST is the histogram of the CT values for all the pixels.

$$\begin{matrix} 32 & 64 & 96 \\ 32 & 64 & 96 \\ 32 & 32 & 96 \end{matrix} \Rightarrow \begin{matrix} 1 & 1 & 0 \\ 1 & & 0 \\ 1 & 1 & 0 \end{matrix} \Rightarrow (110110)_2 \Rightarrow CT \Rightarrow 214$$

Fig.2 Calculation of CT value

### *B. Object based Approach*

It represents a scene image by modeling the object semantics. This approach usually segments the images into regions, and then recognizes semantic objects for each region. Finally, the images get classified as per the object co-occurrence relationship. The object based approach is most useful and advantageous when the scenes consist of different objects. However, if the scenes are simple that contain very few objects and the objects are useless to distinguish the scenes, the object-based approach may be inefficient. Furthermore it must be noted that the accuracy of regional object recognition significantly influences the final classification performance.

Object-based approach first recognizes objects appearing in the image, and then classifies the image based on the object co-occurrence distribution. Spatial information and object correlation can be explored to refine the classification accuracy. The object-based approach is implemented as follows. First, partition an image into $10 \times 10$ regular regions. For each region, recognize the semantic objects. Second, consider three parts of the image: top, middle, and bottom. Calculate the area ratio dominated by each object for each image part. Finally, concatenate the area ratio features of three parts for the image representation. Use this representation for scene classification. Suppose m semantic objects are used. Then for each image part, $m$-dimensional feature vector is attained, and totally the image is represented by a $3m$-dimensional feature vector.

### *C. Combinational Approach*

As holistic and object-based strategies represent images in different ways, they complement each other. The combinatory approach is achieves better performance by taking advantages of these strategies. The combination scheme is carried out as follows. If the decisions of holistic and object-based approaches are the same, the scene class agreed by them is selected as the final decision. Otherwise, majority voting will be employed to make the final decision based on the results of all the classifiers of both holistic and object-based approaches. The combination approach is shown in Fig. 3.

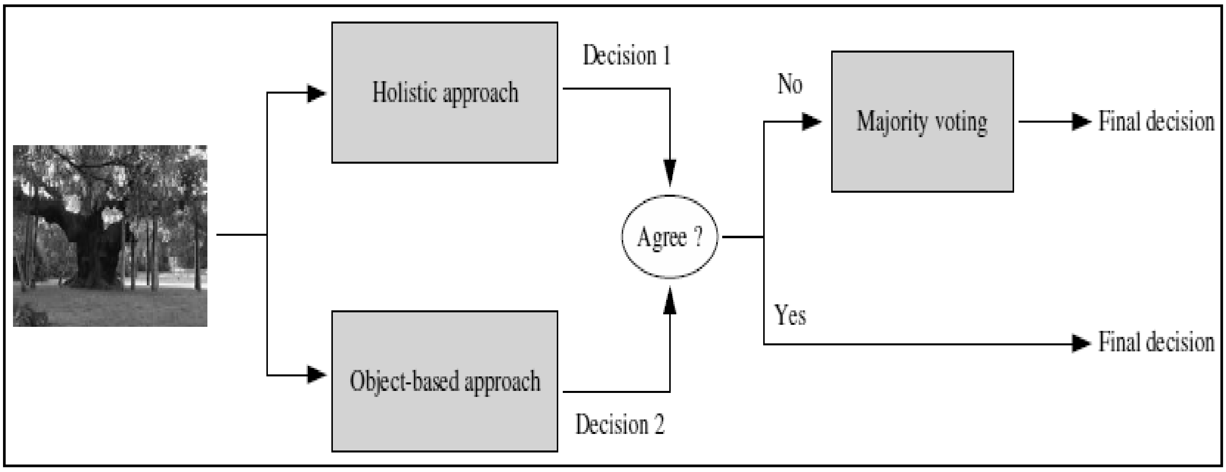



Fig.3 Combinatory approach

First of all, the input image is classified by the holistic and object based approaches separately. The decisions of them are then compared. If the classification results of two approaches are the same, the scene class agreed by them is selected as the final result. Otherwise, all the classifiers of the holistic and object-based approaches are used to re-vote for the final decision through majority voting principle. That is, the class that gains the maximum number of votes is chosen as the scene type of the input image. Using the classifier SVM and one versus one scheme we had, suppose there are in total n scene classes. We have $n(n-1)/2$ SVM classifiers for either holistic or object-based approaches. In the re-voting phase, $n(n-1)/2 \times 2 = n(n-1)$ classifiers involved, and the final scene class is

$$a_{final} = \arg\max_{a_j} \sum_{i=1}^{n(n-1)} \delta(y_i a_j)$$

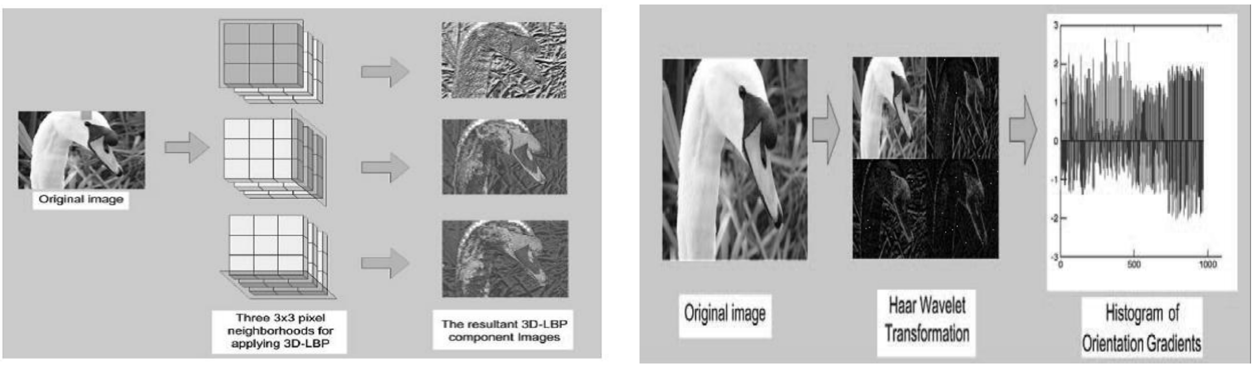
Where $1 \leq j \leq n$, $y_i$ is the decision of classifier i. $\delta(y_i a_j) = 1$, if $y_i = a_j$; and $\delta(y_i a_j) = 0$ if $y_i \neq a_j$.

## III. DESCRIPTORS USED FOR SCENE IMAGE CLASSIFICATION

In this section we study the few descriptors used for the scene image classification and which are discussed below.

### *A. 3-Dimensional Local Binary Pattern(DLBP) descriptor*

The LBP descriptor assigns an intensity value to each pixel of an image based on the intensity values of the eight neighboring pixels using a 3×3 mask. Since a color image is represented by a three dimensional matrix, this concept is extended to assign an intensity value to each pixel based on its neighboring pixels not only on the same color plane but on other planes as well. This method as explained in [5] is shown in Fig. 4(a). For doing this operation, replicate the first and third image planes on opposite sides of the three existing planes to create a five-plane matrix. After the 3D-LBP operation, only the three middle planes are retained. The 3D-LBP method produces three images. Concatenate the dense histograms of these three images to obtain the 3D-LBP feature vector.




(a) (b)

Fig. 4 (a) A 3×3×3 pixel region of the original image is magnified to show the 3D-LBP neighborhoods and the resulting LBP images (b) The original image undergoes Haar Wavelet Transformation and then HOG is generated for each component of the resulting image and concatenated

### *B. HaarHOG descriptor*

To form the HaarHOG feature vector, apply the Haar wavelet transformation to each component of the original image to divide each component image into four distinct regions that separate the local features of each image. Then generate the HOG descriptor for each of these regions and then concatenate them for all the components to get our final HaarHOG feature vector. This process as illustrated in [6] is shown in Fig. 4(b).

### *C. 3-Dimensional LBP-HaarHOG (3DLH) descriptor*

It should be noted that the generation time for both the 3DLBP and HaarHOG descriptors is linear with respect to the number of pixels. Finally, extract the most expressive features from both these vectors and fuse them in the PCA space to form the 3DLH feature vector which outperforms both 3D-LBP and HaarHOG based classification.

### *D. CENTRIST*

CENsus TRansform hISTogram (CENTRIST), is a visual descriptor for recognizing topological places or scene categories [4], It is seen that in place and scene recognition, especially for certain environments, it requires visual descriptor to possess properties that are different from other vision domains (e.g., object recognition). It also satisfies properties and is perfect for task of place and scene recognition. It's holistic representation and has strong generalizability for category recognition. It mainly encodes the structural properties within an image and suppresses detailed textural information.

## IV. CONCLUSIONS

This paper mainly focused on scene image classification which is one of the challenging topics in computer vision and which is also important in terms of image search and image retrieval applications. In this the holistic and object based approaches as well as combined approach is studied and results carried out shows that the combination approach showed a better classification rate that is performance than the individual. The descriptors used in these approaches used spatial pyramid matching and CENTRIST[4] which is a visual descriptors also Bag of Words[1] descriptors can also be used. The dataset tested consists of scene of different categories. In addition to this some novel descriptors were also studied such as the 3D-LBP, HaarHOG and combination of these two using the different color spaces available. Also the visual descriptor CENTRIST has an advantage in scene image classification done on basis of location or place but still it has limitation.

**REFERENCES**

[1] A. Bosch, A. Zisserman and X. Munoz, *"Scene classification using a hybrid generative/discriminative approach,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 4, pp. 712-727, 2008.

[2] L. Fei-Fei and P. Perona, *"A Bayesian hierarchical model for learning natural scene categories,"* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 524-531, 2005.

[3] S. Lazebnik, C. Schmid and J. Ponce, *"Beyond bags of features: spatial pyramid matching for recognizing natural scene categories,"* Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2169-2178, 2006.

[4] J. Wu and J. M. Rehg, "*CENTRIST: A Visual Descriptor for Scene Categorization,"* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1489-1501, 2011.

[5] Zenghai Chen; Zheru Chi; Hong Fu; Dagan Feng, *"Combining Holistic and Object-Based Approaches for Scene Classification,"* Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on , vol.1, no., pp.65,68, 28-29 Oct.2012.

[6] Banerji, S.; Sinha, A.; Chengjun Liu, *"Scene image classification: Some novel descriptors,"* Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on , vol., no., pp.2294,2299, 14-17 Oct. 2012.

[7] A. Oliva and A. Torralba, *"Modeling the shape of the scene: A holistic representation of the spatial envelope,"* International Journal of Computer Vision, vol. 42, no. 3, pp. 145-175, 2001.

[8] D. G. Lowe, *"Distinctive image features from scale-invariant key points,"* International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[9] C. Liu and J. Yang, *"ICA color space for pattern recognition,"* IEEE Transactions on Neural Networks, vol. 20, no. 2, pp. 248-257, 2009.

[10] S. Banerji, A. Verma, and C. Liu, *"Novel color LBP descriptors for scene and image texture classification,"* in 15th International Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada, July 18-21 2011.

[11] A. Verma, S. Banerji, and C. Liu, *"A new color SIFT descriptor and methods for image category classification,"* in International Congress on Computer Applications and Computational Science, Singapore, December 4-6 2010, pp. 819-822.

[12] C. Zhu, C. Bichot, and L. Chen, *"Multi-scale color local binary patterns for visual object classes recognition,"* in International Conference on Pattern Recognition, Istanbul, Turkey, August 23-26 2010, pp. 3065-3068.