A Novel Approach to Scene Classification using K-Means Clustering

Padmavati Shrivastava
Research Scholar,
Dr. C.V. Raman University,
Bilaspur, India

K.K. Bhoyar, PhD
Professor (Dept. of IT),
Y.C.C.E., Nagpur
Maharashtra, India

A.S. Zadgaonkar, PhD
Advisor,
Dr. C.V. Raman University,
Bilaspur, India

ABSTRACT

A challenging problem of computer vision is scene classification. An efficient method for classifying natural scenes from the Oliva – Torralba dataset is proposed. The method is based on K-Means clustering algorithm followed by a novel two phase voting method for classification which is the main contribution of this paper. Two distinct feature sets have been used. The first feature set is used for grouping perceptually similar images into two clusters based on K-Means algorithm. The second feature set is selected based on observed visual attributes of images in these two clusters. Classification is achieved by a novel voting method which firstly assigns test image to the most similar cluster. Each cluster contains images from four categories. Therefore to assign test image to correct category within an assigned cluster, candidate voters from the assigned cluster are selected. The category of majority candidate voters decides the class of test image. The efficiency of the proposed voting scheme is that 83.4% test images are correctly classified. Silhouette index, purity, variance, F-measure and Rand's metric are used for cluster validation.

Keywords

Scene classification, K-means clustering, holistic features, image mining, cluster mapping, semantic labelling, purity, silhouette index, F-measure

1. INTRODUCTION

Due to recent advancements in multimedia technology, large numbers of images are being generated and stored. These images can be used to extract useful patterns which can be classification patterns [1, 2], association rules [3, 4], clustering characterizations [5, 6] or summarizations. Image classification (supervised categorization of images) and image clustering (forming homogeneous groups of images in an unsupervised way) are the two major techniques of image mining [7]. Automatic scene recognition and classification is an important area in the field of computer vision. Owing to ambiguity and variations in the appearance of scenes of various categories, scene classification becomes a challenging task. Many pioneering works in scene classification have used low-level information such as color and texture to classify scenes. Some complex applications recognize objects in images which serve as cues to scene understanding. Few works such as the method of semantic modelling presented in [8] are based on identifying the presence of various semantic concepts such as sky, grass, water etc. which help to identify the category of a scene. Some scene classification systems such as the work in [9] extract global semantic properties from scenes. Different clustering techniques have been extensively used in pattern recognition [10, 11], machine learning [12] and various tasks related to analyzing and understanding images. Authors in [13] perform K-Means clustering of only the HSV values of pixels to initially label the images, which are then used to train a multi-class nearest mean classifier. The clustering algorithm is used to generate codebooks which are further utilized for annotation and retrieval. In [14] authors have performed K-Means clustering based on color and texture features of images. The cluster centroids obtained are used to generate a dictionary of representative values which is then used in classification phase to respond to queries by returning most similar entries.

Natural scene images may contain inherent patterns which may seem similar to human eyes but may have different intrinsic details. The main aim of this work is to analyze these patterns in natural scenes using clustering and utilising these observed patterns to extract appropriate features which successfully categorize these images. Unlike other methods the proposed method does not rely on recognition of individual objects for final classification.

The rest of the paper is organized as follows. Section 2 discusses the description of the dataset used and the extraction of various low level features. Section 3 presents the novel voting scheme for scene classification using K-means algorithm. In section 4 the experimental results of the proposed method along with various evaluation measures are presented. Section 5 presents conclusions and future research directions.

2. IMAGE DATASET AND IMAGE FEATURES

2.1 Dataset Description

The Oliva-Torralba dataset (OT) [15] which is a subset of the Corel database has been used to evaluate the proposed system. The dataset contains 2688 outdoor images broadly classified into two categories of scenes: natural scenes and urban scenes, with each category containing four classes. All the images are in color (RGB) and of size 256 x 256 pixels in JPEG format. The sources of the images vary (from commercial databases, websites to digital cameras). In this work experiments have been performed only on the natural scenes consisting of total 1472 images which are diverse due to interclass similarity and intra class differences. Table 1 presents details of only those categories of the dataset which have been used in this setup. Figures 1a-1h show example images from each category. The dataset is split into training data (used for clustering) and test data (for classification).

Table 1: Description of the Dataset used

Broad category	Internal category	Number of images
Natural	Coast	360
	Forest	328
	Mountain	374
	Open Country	410
Total		1472 images

(a)   **(b)**   **(c)**   **(d)**   **(e)**
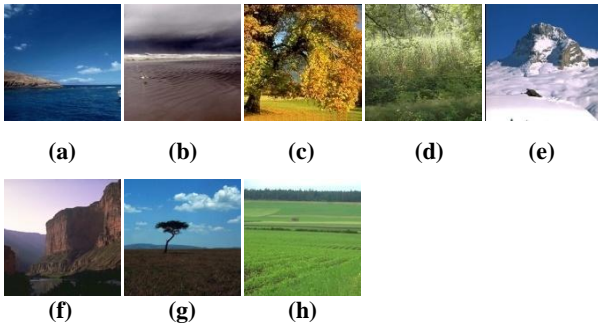

(f)   **(g)**   **(h)**

Figure 1: Example images from each category

Images shown in 1(a) and 1(b) are from Coast category, 1(c) and 1(d) are from Forest category, 1(e) and 1(f) are from Mountain category, 1(g) and 1(h) are from Open Country category.

2.2 Feature Description

Feature extraction from images and their selection is the key to the success of any image mining task [16]. The proposed work uses low level color, texture and edge descriptors since they are closely related to the underlying image semantics of natural scenes and help in discriminating one category from another. Two distinct feature sets has been used in the system which are obtained as follows:

1. The first feature set comprises of feature vectors extracted using Color Moments, Gabor Filters and Edge Orientation Histogram. K-means algorithm is applied on individual features of this set to form clusters of images which have similarity of certain patterns. Careful analysis of these patterns indicates certain discriminative features which can be further extracted for efficient classification. For example when k-means clustering of training images is performed based only on color moments feature two clusters are obtained based on highest silhouette value. Each cluster contains images from all four categories. It can thus be said that each cluster has four partitions, out of which two partitions containing maximum number of images from a single category are called dominant partitions. However coast images and mountain images (in both these categories, images have 'blue' color regions occupying a major area) form dominant partitions of one cluster while forest and open country images (in both these categories images have 'green color regions occupying a major area) form the dominant partitions of another cluster as shown in first three columns of the first row in Table 2. The dominant partitions of these clusters indicate that images in which major portions have similar hue are grouped together. Similarly clustering using Gabor filters (Refer first three columns of the second row in Table 2) indicates that images are grouped on the base of granularity and orientation of texture. Clusters obtained using Edge direction histogram features (Refer first three columns of the third row in Table 2) indicate that images with prominent horizontal directions (Coast and Open Country images) are put in one cluster and those having prominent vertical and diagonal edges (Forest and Mountain images) are grouped in other cluster. The clusters obtained using k-means algorithm on individual features are not used for classification but only meant for efficient selection of features for the second feature set.

2. The second feature set (as shown in last column of Table 2) comprises of Chromatic features, Tamura features and Mean of edges along horizontal, vertical and two

principle diagonal directions calculated using edge direction histogram. These features are selected based on the observed patterns in clusters formed using first feature set.

Table 2: Observed patterns from clusters obtained by K - Means algorithm on individual features of first feature set

First Feature Set	Cluster 1	Cluster 2	Second Feature Set
	Dominant Partitions		
Color Moments	Coast , Mountain	Forest, Open Country	Chromatic features (Dominant Hue*)
Gabor filters	Coast, Open Country	Forest, Mountain	Tamura's Coarseness (Texture Granularity*) Tamura's Contrast (Presence of orientation in texture*) Tamura's Directionality (Sharpness of edges*)
Edge Direction Histogram	Coast, Open Country	Forest, Mountain	Mean of directional edges (Edge Direction*)

* represents the patterns observed from clusters based on first feature set

The details of different features already mentioned in Table 2 are:

Color Moments: They are scaling and rotation invariant. Most of the color distribution information is contained in the low-order moments. Therefore in this work three moments Mean, Standard deviation and Skewness for each channel in HSV space are extracted resulting in a nine-dimensional feature vector.

Chromatic features: Mean and standard deviation of hue and saturation channels are extracted. These features signify the color purity. A four dimensional feature vector is this obtained.

Gabor features: They are invariant with respect to scale, rotation and displacement and are useful in analyzing textured patterns. A Gabor filter bank for 4 scales and 6 orientations (window size 39x39) resulting in a total of 24 filters is created. For each scale and orientation, mean and standard deviation of magnitudes of transformed coefficients are calculated. A forty-eight dimensional feature vector is this obtained.

Tamura features: The three most important Tamura features coarseness, contrast and directionality have been used resulting in a three-dimensional feature vector. These features are selected because they are rotation invariant and have a strong correlation to human perception.

Edge Direction Histogram: Orientation of edges is evaluated by searching the maximum response over a set of edge filter kernels. Firstly canny edge detector is used for edge detection. The edge pixels in vertical, horizontal and two diagonal directions are then counted. The remaining edges are non-

directional and are also taken into consideration as a feature. This results in a five-dimensional feature vector.

3. NOVEL VOTING SCHEME FOR SCENE CLASSIFICATION USING K-MEANS CLUSTERING

The main aim of this paper is to analyse semantics and global patterns relevant to the domain of natural scenes. A scene classification system based on unsupervised learning (clustering) and a novel voting method has been presented in this work. The details of the two main phases of the proposed method are given below:

1.   **Phase I (Clustering):** Unsupervised learning using K-means clustering

2.   P**hase II (Novel Voting Method):** This phase is used for classification and has two steps: Cluster Assignment and Semantic Labelling.

Figure 2 shows the workflow for Phase I (Clustering) phase along with the cluster assignment step of the novel voting method for classification.
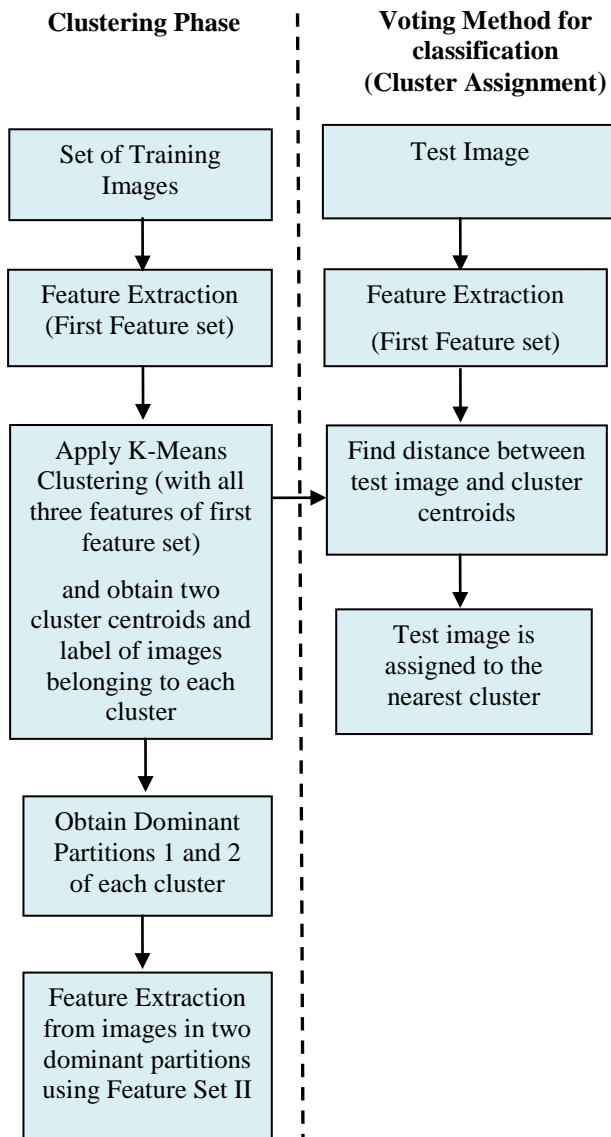
Figure 2: Clustering phase and cluster assignment step of novel voting method for the proposed scene classification system

Figure 3 shows the workflow for semantic labelling phase of the proposed novel voting method for classification.
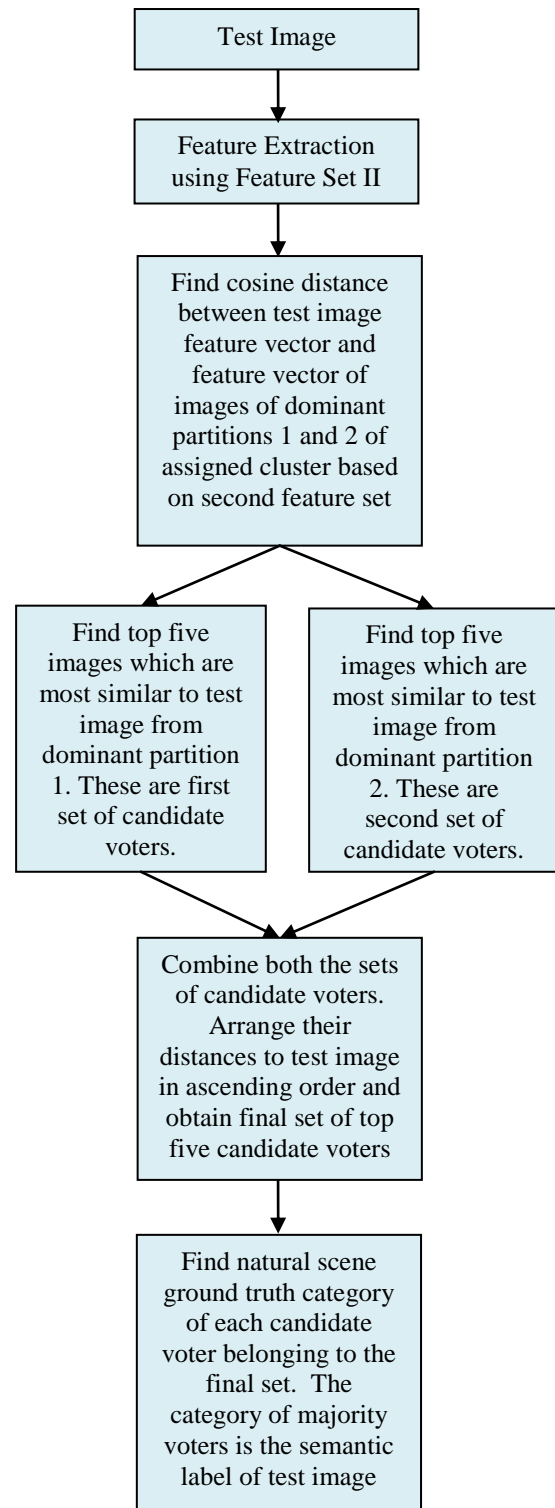
Figure 3: Semantic labelling step of novel voting method for the proposed scene classification system

The different steps of the algorithm for scene classification based on K-Means clustering and novel voting method are elaborated below:

Steps of Clustering Phase

1. Extract color moments (9-dimensional), Gabor features (48-dimensional) and Edge direction histogram (5-dimensional) from training images (for clustering) and test images (used for classification).
2. Concatenate the features so that each vector is 62-dimensional.
3. Store the feature vectors as training features and test features (using Feature Set I)
4. Apply k-means algorithm on training database such that $k_1 \leq k \leq k_2$ and plot silhouette index for each output. The values of k1 and k_2 are to be determined experimentally.
5. The value of k for which silhouette index is highest is optimal_k.
6. Obtain cluster centroids and labels of images belonging to each cluster for k= optimal_k.
7. Each cluster contains images from all the four categories. A partition is a set of images belonging to a particular scene category within a cluster. Find the size of each partition within each cluster. Identify two dominant partitions of each cluster as those which contain maximum number of images of a single scene category. Call these dominant partition 1 and dominant partition 2.
8. Extract chromatic features (4-dimensional), Mean of Directional Edges (using Edge Direction Histogram (1-dimensional) and Tamura features (3-dimensional) from the set of all images in dominant partitions 1 and 2 of each cluster. Also extract these features from test images.

Since there are four scene categories ideally four clusters should be formed using k-means algorithm. However the value of 'k' is obtained by comparing the average silhouette coefficient values for varying is number of clusters and selected the one with highest silhouette value. As shown in Figure 4 the highest mean silhouette value using k-means algorithm based on feature set I is obtained for two clusters. This natural grouping of scenes of different categories into two clusters helps to obtain collection of images which are similar in semantic content. This is useful in understanding relevant patterns and extract appropriate features.
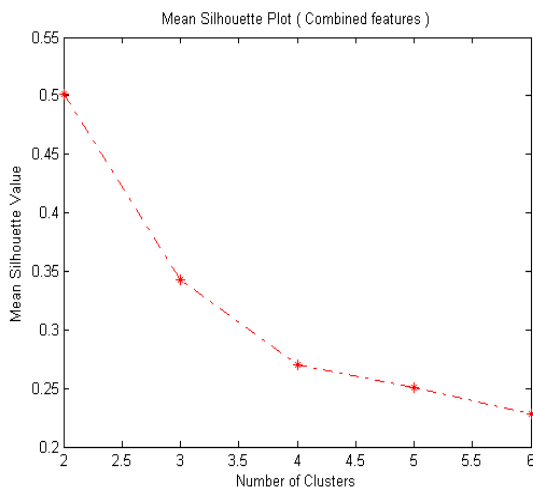
Figure 4: Mean silhouette values for varying number of clusters

Voting Algorithm for Scene Classification

The two major phases involved in the proposed voting algorithm for classification are:

Steps of Cluster Assignment Phase

1. Calculate distance between test image and cluster centroid obtained in Clustering phase with Euclidean distance measure using equation (i).

$$d(A,B) = \sum_{i=1}^{n}(A_i - B_i)^2$$

....equation (i)

where A_i, B_i is the ith element of test image

feature vector A and cluster centroid feature vector B based on features of first feature set.

2. Repeat step 1 for all cluster centroids.
3. Find the cluster for which Euclidean distance between test image feature vector and cluster centroid feature vector has minimum value. The cluster with mimimum distance is the nearest cluster and the test image is assigned to it.
4. Repeat steps 1-3 for all test images.

Steps of Semantic Labeling Phase

1. Extract chromatic features (4-dimensional), Mean of Directional Edges (using Edge Direction Histogram (1-dimensional) and Tamura features (3-dimensional) from test image and concatenate the features to form a 8-dimensional feature vector based on Feature Set II.
2. Retrieve the feature vectors of a dominant partition of the cluster to which the given test image was assigned in the Cluster Assignment phase.
3. Calculate cosine similarity between feature vector test image and feature vector of image from a dominant partition of the cluster to which the given test image was assigned using equation (ii).

$$similarity(A,B) = \cos(\theta) = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$$

.... equation (ii)

where A_i, B_i is the ith element of feature vector for test image and feature vector of an image from dominant partition based on second feature set. The angle between the two feature vectors is represented by θ and $\cos(\theta)$ is the measure of similarity between two images based on the content.

4. Repeat step 2 for all images of dominant partition.
5. Repeat steps 2-3 for each dominant partition of the cluster to which the test image is assigned.
6. Using the cosine similarity measure of step 2 find

top five images which are most similar to test image from dominant partition 1. These are first set of candidate voters. Also find top five images which are most similar to test image from dominant partition 2. These are the second set of candidate voters.

7. Combine the first and second set of candidate voters. Arrange their distances to test image in ascending order and obtain final set of top five candidate voters.

8. Find the ground truth label of each final candidate voter. The category of majority voters is the semantic label of test image.

9. Repeat steps 1-8 for each test image.

4. EXPERIMENTAL RESULTS

In this section the results of the proposed algorithm are presented. Out of 1472 natural scenes from Oliva-Torralba dataset 1357 images have been chosen for clustering and the remaining 115 images are used as test images using stratified cross validation technique. The contingency table is used to explore and document the relationship between images of different categories within a cluster. The contingency table for two clusters and four categories obtained by k-means algorithm on the first feature set is shown in Table 3.

Table 3: Contingency Table for two clusters obtained using K-Means algorithm

Cluster Details	Partition 1 (Coast)	Partition 2 (Forest)	Partition 3 (Mount.)	Partition 4 (Open Country)	Total
Cluster 1	**311**	55	201	**276**	843
Cluster 2	16	**248**	**149**	101	514
Total	327	303	350	377	1357

It can be seen from the contingency table that the dominant partitions (shown in bold face) for cluster 1 are those containing images from coast category and open country category whereas in cluster 2 the dominant partitions (in bold face) contain forest and mountain images.

4.1 Cluster Validation

Different validation measures have been used to evaluate the goodness of the clusters obtained using k-means algorithm such that each cluster should have maximum cohesion and one cluster is well separated from other clusters. The measures used are purity, variance, F-measure and Rand's metric as shown in Table 4 for individual clusters. In this work images are clustered based on holistic color, texture and edge values therefore a single cluster contains images from more than one category. As the number of categories within a cluster increases the value of purity decreases. High purity values can be achieved by increasing the number of clusters. Variance denotes the spread of points around the cluster centroid. In Table 4 lower variance values indicate that the obtained clusters are compact and high F-measure values denote good quality of clustering. Rand statistic measures the fraction of object pairs where clusters C and ground truth class labels T agree that they belong together or do not belong together. Table 5 shows values of cluster validation measures for overall clustering. Table 6 shows the classification results by voting method.

Table 4: Cluster Validation Measures (Individual Clusters)

Cluster Description	Purity	F-measure	Variance
Cluster 1	.3689	.5316	.3488
Cluster 2	.4825	.6071	.1649

Table 5: Cluster Validation Measures for overall clusters

| Overall Purity	Rand's Metric
.4213	.5559

Table 6: Classification results by voting method

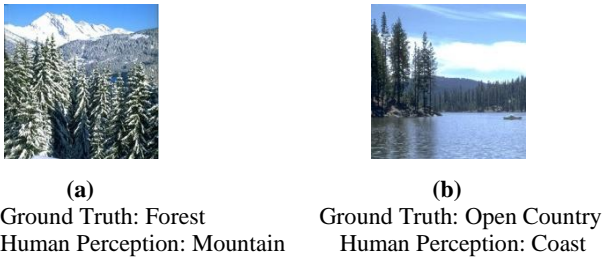| Ground Truth Category of Test Images	Number of test images	Classification by voting method	Classification accuracy (Percent)
Coast	33	28	84.84
Forest	25	23	92
Mountain	24	18	75
Open Country	33	27	81.81
Total	115	96	83.4(average)

Once the test image is assigned to a cluster the successful classification depends on the candidate voters selected by the voting algorithm. Figure 5 and Figure 6 show the candidates selected by voting algorithm for an example test image of coast and forest categories respectively. Both the figures reveal that the voting algorithm successfully selected all five voters from the same category as that of test image. This results in correct classification.

Figure 5: Correctly classified test image along with candidate voters (Coast category)
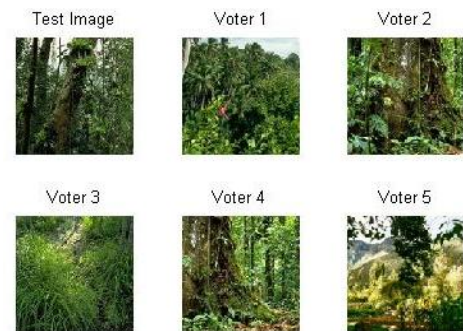
Figure 6: Correctly classified test image along with candidate voters (Forest category)

(a)
Ground Truth: Forest
Human Perception: Mountain

(b)
Ground Truth: Open Country
Human Perception: Coast

Figure 7: Confused examples from the original database

Figure 7 shows that some images are confusing in terms of the category to which they are actually assigned (ground truth) and their category in terms of human perception which can be one of the reasons for misclassifications. Figures 8 and 9 show misclassified test images.

Figure 8: Misclassified test image along with candidate voters (Image of Coast category misclassified as Open Country category)

Figure 9: Misclassified test image along with candidate voters (Image of Mountain category classified as Forest)

5. CONCLUSIONS AND FUTURE WORK

The main aim of this paper is to implement two major image mining tasks: clustering and classification. The main contribution is that natural scenes are classified without the need of segmentation or object recognition. The natural scene categories of the dataset used were not clearly separable. Therefore the clusters obtained after k-means clustering contained images from different categories. However the amount of variation within clusters is low as indicated by the variance values. Patterns from individual feature based clusters have been analysed. The features indicated by these patterns were discriminatory as the overall classification results are good. In the proposed method a test image is first assigned to the most similar cluster and then a novel voting mechanism is used for final classification of the test image to a partition (category) within the assigned cluster. Classification accuracy of 83.4% shows the effectiveness of the proposed approach. Overlapping categories such as coast and open country have similar object regions such as sky. Novelty and effectiveness are the advantages of the proposed method. In future work the aim is at performing hierarchical clustering of these images which shall result in compact and cohesive clusters. A modified version of this work for image retrieval can also be implemented.

6. REFERENCES

[1] Gandhe, S. T., Talele K. T. and Keskar, A.G. (2007),

[2] "Image Mining Using Wavelet Transform", Knowledge-Based Intelligent Information and Engineering Systems, pp. 797-803.Uma Shankar, B., Meher, Saroj K. and Ghosh, Ashish (2011), "Wavelet-fuzzy hybridization: Feature-extraction and land-cover classification of remote sensing images", Applied Soft Computing (Elsevier), Volume 11, Issue 3, pp 2999–3011.

[3] Lee , A.J.T. , Hong , R.W. , Ko , W.M. , Tsao, W.K. and Lin , H.H. (2007), "Mining spatial association rules in image databases", Information Science 177(7), pp. 1593-1608.

[4] Ordonez, C. and Omiecinski, E. (1999), "Discovering association rules based on image content", Proceedings of the IEEE Advances in Digital Libraries Conference , pp. 38-49

[5] Singh, Divakar and Jain, R.C. (2012), "Image Clustering by Neural Networks (SOM)", International Journal of Electronics Communication and Computer Engineering, Volume 3, Issue 2, pp. 257-263.

[6] Wong, M. T., He, X. and Yeh W. (2011),"Image clustering using Particle Swarm Optimization", Evolutionary Computation (CEC), IEEE Congress on, pp. 262-268.

[7] Zhang, J., Hsu, W., and Lee, M. L., (2001), "Image Mining: Issues, Frameworks and Techniques", Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), in conjunction with ACM SIGKDD conference, San Francisco, USA

[8] Serrano N., Savakis A., Luo J. (2004), "Improved scene classification using efficient low-level features and semantic cues", Pattern Recognition 37, pp. 1773–1784.

[9] Oliva A. and Torralba A.,(2002), "Scene-centered description from spatial envelope properties", in: International Workshop on Biologically Motivated Computer Vision, LNCS, Vol. 2525, Tuebingen, Germany, pp. 263–272.

[10] Duda, R. O. and Hart, P. E. Pattern Classification and Scene Analysis, John Wiley and Sons, 1973.

[11] Fukunaga, K. Introduction to Statistical Pattern Recognition. Academic Press, 1990.

[12] Michalski, R. S. and Stepp, R. E. Learning from observation: Conceptual clustering. In R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, editors, Machine Learning: An Artificial Intelligence Approach, Volume I, pp. 331-363. Morgan Kaufmann, 1983.

[13] Cavus, O. and Aksoy, S. (2008), "Semantic Scene Classification for Image Annotation and Retrieval", in IAPR International Workshop on Structural and Syntactic Pattern Recognition, (also in Lecture Notes in

Computer Science, volume 5342), pp: 402-410, Orlando, Florida.

[14] Kobyli´nski,,L. and Walczak, ,K. (2006) "Image Classification based on Customized Associative Classifiers", Proceedings of the International Multi conference on Computer Science and Information Technology, pp. 85–91.

.

[15] Torralba A., Oliva A., "Semantic organization of scenes using discriminant structural templates", in: International Conference on Computer Vision, Korfu, Greece, 1999, pp. 1253–1258.

[16] Foschi, P. G. , Kolippakkam, D. , Liu H. and Mandvikar, A. (2002), "Feature Extraction for Image Mining", Proceedings in International workshop on Multimedia Information System, pp. 103-109.