Automatic Multi-Label Color Image Annotation for Natural Images

A. Kalaivani ,
Dept. of Computer Applications,
Easwari Engineering College,
Chennai, Tamilnadu,India.

Dr. S. Chitrakala,
Dept. of Computer Science and Engineering,
Anna University Chennai,
Tamilnadu, India.

***Abstract*— Digital images are available at ever growing rate. However, to find a required image for an ordinary user is a challenging task. In content based image retrieval , images are retrieved based on the image features. There exist a semantic gap between the image features and image semantics understood by the humans. To overcome semantic gap , automatic image annotation is a solution. Automatic Image annotation is to automatically learn semantic concept model from large number of image samples and use the model to label new images. Once the test images are annotated with the semantic labels, images can be retrieved by keywords which is similar to text document retrieval. In this paper , a system is proposed for annotating image of multiple regions. Images are initially segmented into multiple regions , image region color and texture features are extracted. A model is generated based on the image features using supervised classifiers. Model is also tested with sample images and the performance classifiers are analyzed.**

Keywords— Fast K-Means Clustering Segmentation, Color and Texture Features, Supervised Classifier, Multi-Label Image Annotation.

I. INTRODUCTION

Due to the growth of digital technologies a large volume of digital data are created, stored and processed. A demand rised for effective and efficient tool to find visual information.. A large amount of research has been carried on image retrieval (IR) for the last two decades. Image Retrieval is broadly classified into two approaches as text based Image Retrieval and Content Based Image Retrieval. In the case of text based image retrieval images(TBIR), images are annotated by humans and images are then retrieved in the same way as text documents. However, it is impractical to annotate a huge amount of images manually. Images which are annotated are too subjective and ambiguous. In the case of content based image retrieval(CBIR) , images are automatically indexed based on image features. A semantic gap exist between the low level image content features and semantic concept used by humans to understand the images. To overcome semantic gap automatic image annotation(AIA) is a solution. The main idea of AIA is to automatically learn semantic concept model from large number of image samples and use the model to label new images. Once the test images are annotated with the semantic labels, images can be retrieved by keywords which is similar to text document retrieal. The key characteristic of AIA is that image retrieval is done on keyword search which is generated based on low level image content.

The focus of the paper is to propose a system for annotating images of multiple regions using various supervised classifiers. The classifiers performance are analyzed to identify the better classifier for annotating images. To implement the above said system, first the input images are initially segmented into multiple regions. Image segmentation is done automatically by identifying the number of regions based on local maxima of gray level co-occurrence matrix. The auto generated region count K is then passed into fast K-Means Clustering algorithm for segmenting the images into multiple regions. Secondly, image regions color and texture features are extracted. Color features include color moments based on HSV color model chosen to map human interpretation. Texture features are GLCM features and Harlick features are chosen for four direction such as 0degree, 45 degree, 90 degree and 135 degree . Thirdly, input images are classified into66% of training set and 34% of test for model building and effective testing. Several supervised classifier are chosen such as J48 decision tree, decision table, Naïve bayes classifier, SMO, Attribute selected classifier. Model is constructed on the training set and tested on the test set. Performance are analyzed at multiple dimensions which leads to produce better classifier to be used for the proposed system.

The paper is organized as follows: Section 2 discuss about the related works carried out in the field of clustering based region segmentation. Section 3 highlights the process modules of the proposed system. Section 4 focus on the discussion of the experimental results. Section 5 finally concludes the paper with future research direction.

II. RELATED WORK

Many approaches to color image segmentation have been proposed over the years. Image Clustering is the simplest and widely used method to segment gray or color images into regions. Researchers contributed on the improvements of K-Means Clustering algorithm. The papers surveyed on the area of color image segmentation using Clustering segmentation are discussed below.

D T Pham [12] reviews the current methods on selecting number of clusters and also discuss the factors which affects user selection with its pros and cons.

Kitti Koonsanit [3] proposed a method to identify the number of clusters for satellite image based on co-occurrence matrix and then the images are segmented through K-means clustering algorithm.

Ms.Chinki Chandhok [2] applied K-Means clustering algorithm where the pixels are clustered based on their color and spatial features to form image regions.

Monika Xess [4] analyze two unsupervised clustering based color image segmentation such as K-means clustering and Fuzzy C-means clustering.

In automatic image annotation, images are automatically classified into a set of pre-defined concepts without human intervention. For a given input image, a single concept or multiple concepts, can be assigned to the image based on feature descriptors. Based on the survey paper [1], multi-label image annotation is classified into two approaches : statistical and machine learning approach .

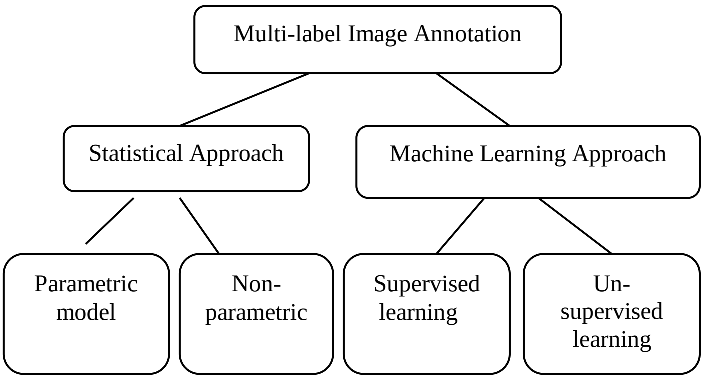


Fig.1. Multi-label image annotation approaches[1]

Based on statistical approach the various models identified by various researchers are co-occurrence model[18],translation model[17], cross media relevance model[15], continous space relevance model[16], multiple bernouli relevance model[13],dual cross media relevance model[10], Gaussian mixture model and its variants[14,8]. Based on machine learning approach the various supervised classifier used for automatic image annotation are SVM[11] , Neural Network[10], Decision Tree [9,5,6] and unsupervised clustering [7].

From the papers surveyed, for segmenting images, popular algorithm used are K-Means or Fuzzy C-Means. The major two issues are identified: first user has to specify K - number of clusters and second one is that sometimes K – Means clustering produces zero clustering. To overcome the above said issue our proposed system devise a technique to automatically identify the number of regions- K and based on K, the images are segmented using fast K-Means clustering which yield better performance than K-Means Clustering. In the case of supervised classifier for automatic image annotation yield different performance , sometimes involved high computation and also dependency exist on the training set. To overcome the above said issue in the proposed system multiple supervised classifiers are trained to build the model. An optimal classifier is finalized based on the comparative analysis of multiple classifier for the same training set.

III. PROPOSED SYSTEM

Proposed system is to annotate the images based on supervised classifiers. The overall framework for annotating the image is shown in figure 2.

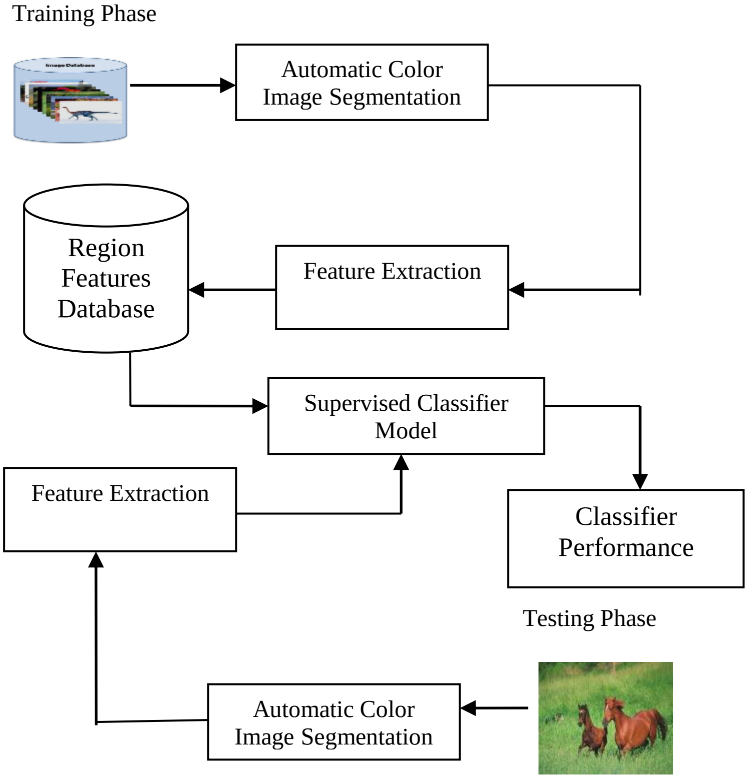


Fig.2. Proposed System Architecture

A.Automatic Color Image Segmentation

Automatic color image segmentation are done by determination of number of regions in the input image based on local maximum. The value of K is then passed to Fast K-Means Color Image Segmentation algorithm. An overview of the algorithm proposed for determination of region count – K is placed in table 1.

Table-1. Determination of Regions Count - K.

Determination of Regions Count - K
Input : RGB Image
Output : Regions count K
1: Read input RGB Image
2. Image is transformed from rgb color space into gray color space
3. Apply Image preprocessing to remove noise
4. Compute Gray Level Co-occurrence matrix
5. Extract GLCM Diagonal Elements
6. Find Peaks for GLCM diagonal elements
7. Calculate local maxima on diagonal elements for varying gray levels

K-means depends mainly on distance calculation between all data points and the centers. The time cost is high when the size of the dataset is large. A two stage refinement algorithm is already proposed to reduce the time cost of distance calculation for huge datasets. The first stage is fast distance calculation use small portion of data to produce the best centre location. The second stage is the slow calculation in which the whole data set is used to get the exact location of the centers.

The time cost of the distance calculation for the fast stage is very low due to the small size of the training data chosen. The time cost of the distance calculation for the slow stage is also minimized due to small number of iterations. Algorithm for fast K-Means clustering is shown in table 2.

Table-2. Fast K Means Clustering Algorithm

Segmenting Color Images to Multiple Regions
Input : RGB Image & Number of clusters
Output : Segmented Image
1: Read input RGB image
2. Image Transformation – RGB to Lab Color space
3. Initialize centroids of k clusters randomly
4. Assign each sample to the nearest centroid
5. Calculate centroids of k clusters
6. If centroids are unchanged, process completed else go to step 2.

The complexity of the k-means clustering algorithm is $O(KQN)$where K is the number of clusters, Q is the number of iteration required to get to the stopping criteria and N is the input patterns. The time complexity of the fast stage of clustering: $O(KQfNf)$ where Nf is the number of data for the fast stage and Qf is the iterations during the fast stage only.

B. Feature Extraction

Feature Extraction is a method of capturing visual content of images for indexing & retrieval. Primitive or low level image features can be either general features, such as extraction of color, texture and shape or domain specific features. The various features extracted for the proposed system are color features – color moments and texture features – GLCM texture features and Harlick Texture features.

Color Moments Features

Color moments are the measurements which is used to differentiate images based on the image color . There are three primary color moments. Mean, Standard Deviation and Skewness. The segmented dominant region is converted into HSV color space and these color moments are extracted for Hue, Saturation and Value in HSV color space. A nine dimensional color feature vector is extracted for the input image in the HSV colour space.

Mean is defined as the average image color. Image Mean for the region is calculated by using equation 1.

$$E_i = \sum_{N}^{j=1} \tfrac{1}{N} p_{ij} \qquad (1)$$

Standard Deviation is the square root of the variance of the distribution. Image Standard deviation for the region is calculated by using equation 2.

$$\sigma_i = \sqrt{\left(\tfrac{1}{N} \sum_{N}^{j=1} \left(p_{ij} - E_i \right)^2 \right)} \qquad (2)$$

Skewness is the measure of the degree of asymmetry in the distribution. Image skewness for the region is calculated by using equation 3.

$$S_i = \sqrt[3]{\left(\tfrac{1}{N} \sum_{N}^{j=1} \left(p_{ij} - E_i \right)^3 \right)} \qquad (3)$$

Where p_{ij} is the color value of i th color component and N is the total number of pixels in the image.

Texture Features

Image Texture features describes the visual patterns having the property of homogeneity. The texture features are extracted from the regions using gray level co-occurrence matrix. A gray level co-occurrence is defined as the joint probability density of the gray levels of the two pixels separated by a given displacement d and angle θ. The extraction of GLCM features are divided into two processes: formation of co-occurrence matrix and extraction of GLCM descriptors against the co-occurrence matrix. Co-occurrence matrices are constructed for four orientations (horizontal, vertical and two diagonals). The GLCM descriptors includes energy, entropy, contrast, homogeneity, correlation.

Angular Second Moment or Energy shows the texture uniformity or homogeneity. Image energy for the region is calculated by using equation 4.

$$\text{Energy} = \sum_i \sum_j p^2(i,j) \qquad (4)$$

Entropy shows the degree of randomness. Homogeneous scenes will have high entropy. Image entropy for the region is calculated by using equation 5.

$$\text{Entropy} = -\sum_i \sum_j p(i,j) \log p(i,j) \qquad (5)$$

Contrast or Second order Element Difference Moment shows the contrast texture value. Image contrast for the region is calculated by using equation 6.

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 p(i,j) \qquad (6)$$

Homogeneity shows the first order inverse element difference moment. Image homogeneity for the region is calculated by using equation 7.

$$\text{Homogeneity} = \sum_i \sum_j \frac{p(i,j)}{1+|i-j|} \qquad (7)$$

Correlation ia a measure of gray level linear dependence between the pixels. Image correlation for the region is calculated by using equation 8.

$$\text{Correlation} = \sum_i \sum_j \frac{(i-\mu i)(j-\mu j)\bar{p}(i,j)}{\sigma_i \sigma_j} \quad (8)$$

where p(i,j) is an element of matrix co-occurrence, μ is the the mean value of matrix co-occurrence.

Haralick's texture features were calculated using the haralick function. The basis for these features is the gray-level co-occurrence matrix in equation 9. This matrix is square with dimension *Ng*, where *Ng* is the number of gray levels in the image. Element [*i,j*] of the matrix is generated by counting the number of times a pixel with value *i* is adjacent to a pixel with value *j* and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered to be the probability that a pixel with value *i* will be found adjacent to a pixel of value *j*.

$$G = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix} \quad (9)$$

Since adjacency can be defined to occur in each of four directions in a 2D, square pixel image (horizontal, vertical, left and right diagonals-Fig:4), four such matrices can be calculated. Fig 3: Four directions of adjacency are defined for calculation of the Haralick texture features.

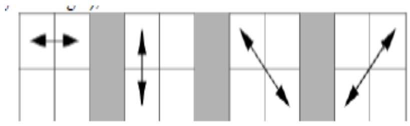


Figure 3: Four Directions of adjacency

The Haralick statistics are calculated for co-occurrence matrices generated using each of these directions of adjacency. Haralick Texture Features which can be extracted from each of the gray-tone spatial-dependence matrices. The various Haralick texture features calculated are angular second moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, information of correlation, maximal correlation coefficient.

C.Supervised Classifier

Classification is a data mining algorithm to determine the output of a new data instance. It classifies each item in a set of data into one of the predefined set of classes. There are two types of machine learning classification: supervised classification and unsupervised classification. In supervised classification, classifier knows the class labels. In unsupervised classification, classifier cluster the data into several group based on interclass and intra-class similarity. Several supervised classifier are found in the literature.

J48 Classifier

J4.8 classifier is an open source java implementation of C4.5 algorithm uses a simple procedure. To classify a new item a decision tree is created based on the attributes of the training data. Whenever it encounters the data in training set, it identifies the attributes that differentiates various instances clearly. When the data instances fall within a category that has the same value for the target variable then the branch is terminated and assigned to the target value obtained. For the other data instances this classifier searches for another attribute that can create another category. This process is continued until we get a clear decision of what combination of attributes gives a particular target value or if we run out of attributes. In case of inadequate attributes the branch is assigned a target value which is possessed by the majority of item under the particular branch.

*A. **Bayes Network***

Bayes network or Belief network is one of the probabilistic graphical models which is used to represent knowledge about uncertain domain. Each node in a graph is a random variable and the edges between them represent the probabilistic dependencies among the random variables. Bayesian networks are directed acyclic graph which is used to represent joint probabilistic distribution over a set of random variables. A Bayesian network **B** is an annotated acyclic graph that represents a joint probability distribution over a set of random variables **V**. The network is defined by a pair **B** = (G,Θ) where G is the directed acyclic graph whose nodes $X_1 \ldots X_n$ represent the random variables and their edges represent the direct dependencies between these variables

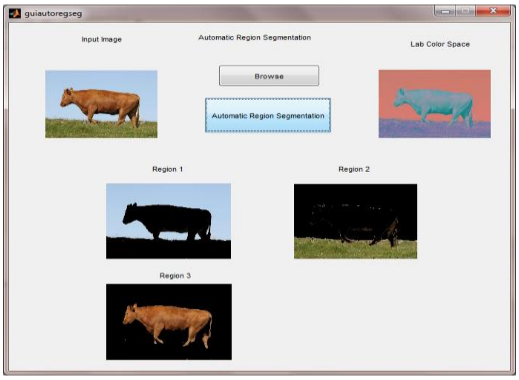
*B. **Sequential minimal optimization***

SMO is an algorithm used for solving optimization problems during the training of support vector machines. It breaks the problem into sub-problems and they are solved. It is developed in a decomposition method to solve the dual problems arising during SVM formulation. In each iteration it reflects two coefficients $ß_i$, $ß_j$, the other coefficients keep their current values. This selection of coefficients can be used to solve the optimization sub-problems analytically. The iteration is continued until the conditions are achieved.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Proposed System is experimented on the natural images taken from COREL , MSCROID data sets and Web images. Total images taken for the study is 352 of which 11 images belongs to elephant, 39 images belongs to goat, 67 images belong to horse, 27 images belongs to furnitures, 121 images belongs to cow and 87 images belongs to birds. Out of all these category 177 grass images and 59 sky regions are formed. 66% of the feature data sets category are grouped as training set and remaining 34% are grouped as testing set. The various classifier under which the training instance are tested are J48 classifier, Decision Table , Naive Bayes classifier,multi-layer perceptron, attribute selected classifier.

The Graphical user interface designed for segmenting color image into multiple regions is shown below in figure 4.



Feature vector for sample image of color moments and GLCM texture features are listed in table 3 and table 4.

Table-3. Color Momemts Features

MeanH	StdH	SkewH
0.1923	0.4071	0.6298

MeanS	StdS	SkewS
0.0912	0.1319	0.2922

MeanV	StdV	SkewV
5.0638	-0.6367	-1.0628

Table-4. GLCM Texture Features

0 degrees	0.5065	0.9297	0.1544	0.8476
45 degrees	0.6294	0.9129	0.1426	0.8216
90 degrees	0.4384	0.9392	0.1549	0.8490
135 degrees	0.6264	0.9133	0.1417	0.8188

In most of the literature the performance of annotation system is calculated by using prescision and recall. Precision and Recall values in annotation system are evaluated for each word and the mean of all words are consider as the performance of the system. Accordingly,

$$\text{Precision} = \frac{\text{Number of correct annotated label}}{\text{Total annotated label}} \quad (9)$$

$$\text{Recall} = = \frac{\text{Number of correct annotated label}}{\text{Total label in test set}} \quad (10)$$

For comparing system using only single value F-score is good choice.

$$\text{F-score} = \frac{2*\text{precision}*\text{recall}}{(\text{precision} + \text{recall})} \quad (11)$$

Precision, Recall, F-score, True positive rate and false positive rate for different classifiers are shown in table -5.

Table-5.Classifier Performance

Classifier	Precision	Recall	F-Score	TP rate	FP rate
J48 DT	0.775	0.768	0.769	0.768	0.033
Decision Table	0.715	0.687	0.668	0.687	0.063
Naïve Bayes	0.674	0.652	0.624	0.652	0.055
SMO	0.736	0.727	0.724	0.727	0.056
MLP	0.741	0.737	0.735	0.737	0.04
ASC	0.797	0.788	0.784	0.788	0.041

The error results of the supervised classifier are mean absolute error (MSE), root mean squared error (RMSE), relative absolute error(RAE), root relative squared error(RRSE) are shown in table 6.

Table-6.Classifier Error Performance

Classifier	Precision	Recall	F-Score	TP rate
J48 DT	0.0614	0.23	30.29%	72.68 %
Decision Table	0.1548	0.2588	76.41%	81.78%
Naïve Bayes	0.0869	0.2913	42.90%	92.04%
SMO	0.0709	0.2535	34.98%	80.10%
MLP	0.709	0.2144	35.02%	76.56%
ASC	0.0605	0.2423	29.84%	67.74%

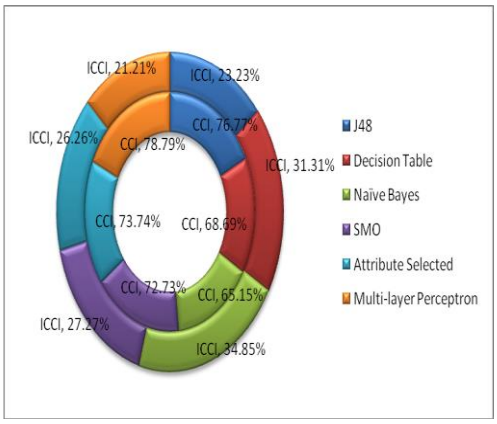


Figure 5:Classifier Performance

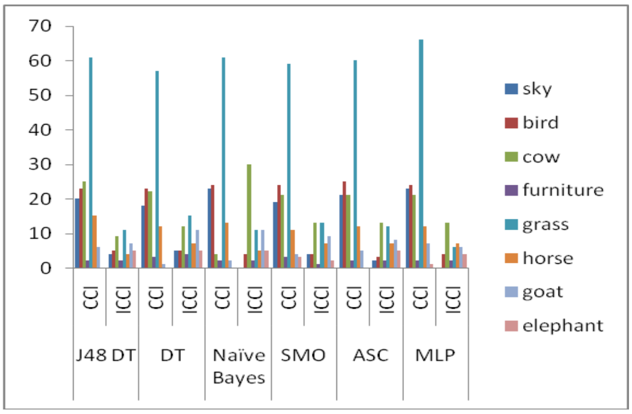


Figure 6:Multi-Classifier Multiple Class Performance

From the above performance analysis for multiple classifier, the performance is higher 78.79 % for correctly classified instance in the case of multi-layer perceptron when compared with decision tree of 76.77% . But the time taken to build the model for the given instances is 0.28 seconds for decision tree and 66.21 seconds for multi-layer perceptron. Kappa statistic is used to assess the accuracy of the classifier. The average kappa score from the selected algorithm is from 0.57-0.73. When compared with other classifier algorithms Multi-layer perceptron yields better classification of the class labels and also yields higher classifier performance . But J48 Decision tree yields next higher level performance which can also be selected as the model for further processing.

IV. CONCLUSION

Image classification classify images into several classes based on the supervised classifier. A system is proposed to identify efficient classifier to annotate the multiple region based on multiple supervised classifier. From the experiment results, the performance of the multiple classifier are analyzed and is concluded that J48 decision tree classifier yield better performance than other classifier even with respect to time to build the model. Multilayer perceptron also yields better accuracy than J48 decision tree but takes more time to build the model for this data sets. Proposed system can be further expanded to include more number of class labels and still some more image features can be extracted for improving the annotation performance. Further a semantic based image retrieval system can be proposed to retrieve the images based on the class labels generated.

REFERENCES

[1] A.Kalaivani, ,Dr.S.Chitrakala,"Challenges and Approaches in Multi- Label Image Annotation", 4th IEEE International Conference on Computing , Communication and Network Technologies, pp.1-8, July 2013.

[2] Ms.Chinki Chandhok, Mrs.Soni Chaturvedi, Dr.A.A Khurshid, "An Approach to Image Segmentation using K-means ClusteringAlgorithm", Int. Journal of Information Technology (IJIT), Vol.1, Issue – 1, pp. 11-17, August 2012.

[3] Kitti Koonsanit, Chuleerat Jaruskulchai, and Apisit Eiumnoh, "Determination of the Initialization Number of Clusters in K-means clustering Application Using Co-Occurrence Statistics Techniques for Multispectral Satellite Imagery", International Journal of Information and Electronics Engineering, Vol. 2, No. 5,pp.785-789, September 2012.

[4] Monika Xess, S. Akila Agnes, "Survey on Clustering based Color Image Segmentation and novel approach to FCM Algorithm", International Journal of Research in Engineering and Technology, Vol. 1, Issue. 1, pp. 346-349, August 2012.

[5] Shaohua Wan, "Image Annotation using the Simple Decision Tree", International Conference on Management on e-Commerce and e-Government, IEEE,, pp.141 - 146 , 2011.

[6] Ali Fakahari, Amir-Masoud Eftekhari-Moghaddam,"Hierarchical Decision Tree (HDT) Approach for Image Annotation, IEEE, 2011.

[7] Gulisong Nasierding, Grigorios Tsoumakas, Abbas Z.Kouzani, "Clustering Based Multi-Label Classification for Image Annotation and Retrieval, International Conference on Systems, Man and Cybernetics, IEEE, pp. 4514 - 4519 , 2009.

[8] Changhu Wang, Shuicheng Yan, Lei Zhang, Hong-Jiang Zhang, " Multi Label Sparse Coding for Automatic Image Annotation", IEEE, pp.1643-1650, 2009.

[9] Y.Liu, D.Zhang, G.Lu, "Region-based image retrieval with high level semantics using decision tree learning", Pattern Recognition, vol.41, issue no.8, pp.2554-2570, 2008.

[10] J.Liu, B.Wang, M.Li, Z.Li. W.Y.Ma, H.Liu and S.Ma, "Dual Cross Media Relevance Model for Image Annotation", 15th International Conference on Multimedia, pp.605-614, 2007.

[11] Changbo Yang, Ming Dong, Jing Hua, "Region- based Image Annotation using Asymmetrical Support Vector Machine-based Multiple Instance Learning", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-5,2006.

[12] D T Pham_, S S Dimov, and C D Nguyen, "Selection of K in K-means clustering", J. Mechanical Engineering Science, Vol. 219 ,pp.103-119, 2005.

[13] S.Feng, R.Mammatha and V.Lavrenko, "Multiple Bernoullli Relevance Model for Image and Video Annotation", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.1002-1009, 2004.

[14] G.Carneiro, A.B. Chan, P.J.Moreno, N.Vasconelos, "Supervised learning of semantic classes for image and video annotation", Proceedings of CVPR04, pp.1002-1009,2004.

[15] J.Jeon, V.Lavrenko and R.Mammatha, "Automatic Image Annotation and Retrieval using Cross Media Relevance Model", Proceedings of the 26th Annual International ACM, 2003.

[16] V.Lavrenko, R.Mammatha and J.Jeon, "A Model for Learning the Semantics of Pictures", Proceedings of Advance in Neural Information Processing, 2003.

[17] P.Duygulu, K.Barnard, J.Freitas and D. Forsyth, "Object Recognition as Machine Translation Learning a Lexicon for a fixed image vocabulary", Proceedings of the 7th European Conference on Computer Vision, pp.97-112, 2002.

[18] Y.Mori, H.Takahashi and R.Oka, "Image-to-word Transformation Based on Dividing and Vector Quanitizing Images with words", MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.