Multi-relation Image Retrieval and Annotation based on Holistic Approach and Decision Tree

Sarla More
School of IT, RGPV
Bhopal, India

Nishchol Mishra
Asst. Professor
School of IT, RGPV
Bhopal, India

ABSTRACT
This work of multi-relation image retrieval and annotation is based on the holistic approach and the decision tree. In this we have proposed that for the retrieval of similar images as that of query image Dominant color descriptor (DCD) is used, this descriptor uses the color feature for the retrieval of images. This creates the feature vector index. Test keywords are correlated with the feature vector index, the correlation is performed by multi class association to get the classes for processing on them. Classes which are not necessary are discarded using cross validation in the decision tree process Decision tree used to take the relevant classes and finally we calculate the Gain of feature vector and we get the retrieval of the images based on the query image with the associated keywords or annotation. This work has been implemented on MatLab 7.5 simulator.

Keywords
DCD, Multirelational association, automatic image annotation

1. INTRODUCTION
Automatic image annotation is the process in which we get the keywords that are human visual in the image. Image retrieval is the task in which we get the images according to some features extracted such as color, texture and shape. The retrieval of similar images of the query images is done using the Dominant color descriptor in the proposed work which is able to retrieve the images along with the associated keywords. The holistic approach used in this work is a hierarchical approach which reduces the semantic gap problem between the low level and the high level features. Proper indexing of the features set is done and the feature vectors are correlated to the test keywords to extract the desired information. Finally the decision tree is used for the multiple classes which make the distinction between the relevant and irrelevant classes which are used for further processing and other are discarded which are not used. The gain is a useful property which tells that the weightage of the vector value. If the vector value is greater and equal to the gain the output is generated otherwise process the next index.

1.1 Content-Based Image Retrieval
An application of computer vision technique i.e. Content-based image retrieval (CBIR) [10] is the process of searching for digital images in large databases. The goal of CBIR systems is to operate on collections of images and, in response to visual queries, extract relevant image that are used for many applications.

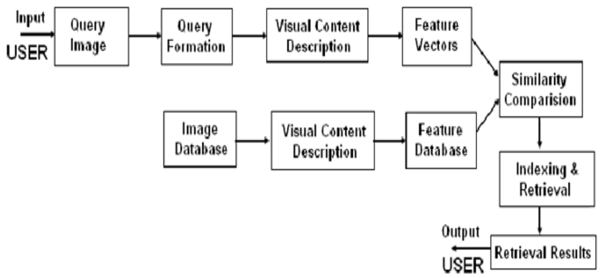

Figure 1: content based image retrieval

1.2 Features Extraction
Feature extraction [3] is the core of the content based image retrieval. Two reasons behind raw image data that cannot use directly is the high dimensionality of the image which makes it difficult to use the whole image and a lot of the information embedded in the image is redundant and un-useful. The process of finding the expressive representation is called feature extraction and the resulting representation is called the feature vector. Feature extraction can be defined as the act of mapping the image from image space to the feature space. Image content can differentiate between visual and semantic content. Features usually represent the visual content, which can be further divided into general or domain specific. For example the features that can use for searching would be representing the general visual content like color, texture, and shape and the features that are used for searching human faces are domain-specific and may include domain knowledge. Annotation based on the visual content help to obtain semantic content. Many concerns are associated for choosing the features to be extracted:

- There should be enough information about the image and any domain specific knowledge is not required
- For large image collection and retrieval the computation must be easy.
- Human perceptuation characteristic is an important think which must be related well. Finally the user will determine the retrieved images suitability.

1.3 Automated Image Annotation
Automatic image annotation [4] also known as automatic image tagging is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database. This method can be regarded as a type of multi-

class image classification [5] with a very large number of classes - as large as the vocabulary size. It consists of a number of techniques that aim to find the correlation between low level visual features and high level semantics. The main challenge in automated image annotation [6] is to create a model able to assign visual terms to an image in order to successfully describe it. Image analysis techniques are used to extract features from the images such as color, texture and shape in order to model the distribution of a term being present in the image. Features can be obtained from the whole image (global approach), or from blobs, which are segmented parts of the image (segmented approach) or from tiles which are rectangular partitions of the image. The next step is to extract the same feature information from an unseen image in order to compare it with all the previously created models. The result of this comparison yields a probability value of each keyword being present in an image. The block diagram of typical image annotation framework [7] is shown in following figure.

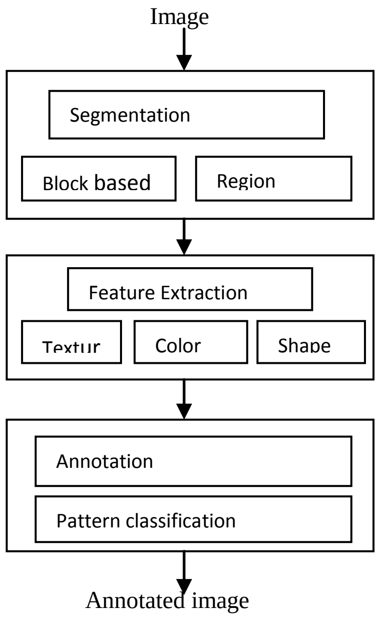

Figure 2: A typical image annotation frame work

There are three types of image annotation approaches available: manual, automatic and semi automatic. Manual annotation needs users to enter some descriptive keywords when perform image browsing. While automatic annotation detects and labels semantic content of images with a set of keywords automatically. In semi-automatic annotation, it needs user's interaction to provide an initial query and feedback for image annotation while browsing. The comparison of the three annotation techniques and advantages and disadvantages of them are shown in Table 1 and Table 2.

The Manual image annotation is considered expensive and time consuming. While semi annotation is very efficient compared to manual annotation and more accurate than automatic annotation based on experiment evaluation by. Automatic image annotation is the best in term of efficiency but less accuracy. Independently of the method used to define the annotation, automated image annotation system generate a set of keywords that help to understand the scene represented in the image. Many experiments show that current image annotation techniques present poor performance in the context of image retrieval. Image annotation surveys have been reviewed by many researchers according to the demanding the needs for annotating images.

TABLE 1: Annotation techniques comparison

Annotation techniques	**Manual**	**Semi Automatic**	**Automatic**
Initial Human Interaction	Enter some descriptive keyword	Provide initial query at the beginning	No interaction
Machine task	Provide storage for annotation to be saved such as disk space or database	Parse Human's query and extract semantic information to perform annotation	Detect labels semantic keywords automatically using recognition technology
Human effort	Perform sufficient semantic information for retrieval purposes	Perform some annotation and work with machine's output	Verify and correct machine's output finally for annotation accuracy

TABLE 2: Advantages and Disadvantages on Annotation

Annotation techniques	**Manual**	**Semi Automatic**	**Automatic**
Advantages	The most accurate annotation	Quality of the annotation improves in the interactive manner after correction	The most efficient, the least time consuming
Disadvantages	Time consuming, expensive, difficult, subjective, inconsistency	Less time than automatic greater time than manual annotation	

1.4 Annotation by Association

This methodology [8] [9] is a direct extension of the traditional association rule mining that was developed to mine patterns of associations in transaction databases. Each transaction involves certain items from a set of possible items. Given N transaction and d as the size of the set of possible items, collection of transactions can be represented as a size $N \times d$ matrix. Each transaction involves only very few items, the transaction matrix is very sparse. Association mining [11][12] tries to discover frequent item sets, the items that appear together frequently in the transaction matrix, in the form of rules wherein the presence of a particular item in a transaction predicts the likely presence of some other items. A rule has the form $X \Rightarrow Y$ with support s and confidence c, implying that s% of the transactions contain both X and Y and c% of the transactions that support X also support Y . Methods for annotation by association mining differ in terms of how items and transactions are defined to take advantage of existing association rule mining algorithms, for example the well-known Apriori algorithm.

1.5 Dominant Color Descriptor

The Dominant Color Descriptor (DCD) [13] [14] provides a compact description of the representative colors in an image or image region. Its main target applications are similarity retrieval in image databases and browsing of image databases based on single or several color values. Unlike the traditional histogram based descriptors, the representative colors are computed from each image instead of being fixed in the color space, thus allowing the color representation to be accurate and compact.. The DCD allows for efficient indexing 2of large databases as presented in [14]. The DCD is defined to be

$$F= \{\{Ci, Pi, Vi\}, s\}, (i=1, 2, 3\ldots N)$$

Where N is the number of dominant colors, each dominant color value *ci* is a vector of corresponding color space component values (for example, a 3-D vector in the RGB color space). The percentage *pi* (normalized to a value between 0 and 1) is the fraction of pixels in the image or image region corresponding to color *ci*. The optional color variance *vi* describes the variation of the color values of the pixels in a cluster around the corresponding representative color. The spatial coherency *s* is a single number that represents the overall spatial homogeneity of the dominant colors in the image. The number of dominant colors N can vary from image to image and a maximum of eight dominant colors were found to be sufficient to represent an image or an image region. The color space quantization depends on the color space specifications defined for the entire database and need not be specified with each descriptor. The binary syntax of the dominant color descriptor specifies 3 bits to represent the number of dominant colors and 5 bits for each of the percentage values (uniform quantization of [0,1]). The Color Space and Color Quantization descriptors are referred to by this descriptor and RGB is the default color space. The optional color variances are encoded at 3 bits per color with non-uniform quantization.

1.5.1 DCD Extraction

The extraction procedure [15] described in dominant color uses the Generalized Lloyd Algorithm to cluster the pixel color values. It is recommended that the clustering be performed in a perceptually uniform color space such as the CIE LUV. The distortion *i D* in the i-th cluster is given

$$Di=\sum h(n)||x(n)-Ci||^2, x(n)€Ci$$

Where *ci* the centroids of cluster Ci, x (n) is the color vector at pixel n, and h(n) is the perceptual weight for pixel n. The perceptual weights are calculated from local pixel statistics to account for the fact that human visual perception is more sensitive to changes in smooth regions than in textured regions. The update rule for the above distortion metric can be derived to be:

$$Ci= \sum h (n)\, x (n)/\sum h (n), x (n) €C$$

The procedure is initialized with one cluster consisting of all pixels and one representative color computed as the centroids (center of mass) of the cluster. The algorithm then follows a sequence of centroids calculation and clustering steps until a stopping criterion (minimum distortion or maximum number of iterations) is met. The clusters with highest distortion are divided by adding perturbation vectors to the centroids until the maximum distortion falls below a predefined threshold or the maximum number of clusters is generated. The percentage or fraction of pixel in the image belonging to each of the quantized colors is then calculated and these resulting percentages are uniformly quantized to 5 bits. The color values are quantized according to the specifications of the color space and the associated color quantization descriptors. A simple connected component analysis is performed to identify groups of pixels of the same dominant color that are spatially connected. The normalized average number of connecting pixels of each dominant color is then computed. A 3x3 masking window is used for this purpose. This is used as a measure of spatial coherency for that dominant color. The overall spatial coherence is then a linear combination of the individual spatial coherence values with the corresponding percentages *pi* being the weights. The spatial coherence value is then nonuniformly quantized to 5 bits, where 31 means highest confidence and 1 means no confidence. The value 0 is used for cases where it is not computed. Finally, the color variances are computed as variances of the pixel values within each cluster and nonuniformly quantized to 1 bit per color component.

1.6 Decision Tree

Decision trees are hierarchical, sequential classification structures that recursively partition a set of data, thus representing the rules underlying the data. Decision tree may be used for exploring data in any following ways:

- To uncover mappings from independent to dependent variables that can be used to predict the value of the dependent variable in the future.
- To reduce the volume of data to a more compact form, this provides an accurate summary.
- To discover if the data contains well-separated clusters of objects, so they could be interpreted meaningfully in the context of a substantiated theory.

Consider a set of objects, each of them completely described by a set of attributes and a class label. A tree is a rooted connected acyclic graph, consisting of a set of internal nodes (denoted by ovals in a graphic representation) and a set of leaf nodes (denoted by rectangles). A decision tree is a tree induced on a training set, which consists of objects. In a decision tree, a test (split) is associated to every internal node. Such tests are logical expressions involving the object's attributes. Each edge from an internal node T to its children is labeled with a distinct outcome of the test at node T. A class label is associated to each leaf node. When classifying an example, the role of an internal node is to test the value of the expression based object's attributes, and to send it "down" the corresponding edge. We call a leaf node pure if all the training examples at that node belong to the same class.

1.7 Multirelational Association Rule

Consider a database *D* that contains a set of transactions *T.* If *X*, *Y* are items in T then an association rule is an implication of the form $X \rightarrow Y$ and has a support *s* in the database *D* if *s*% of the transactions in *D* contains both *X,Y.* Similarly rule $X \rightarrow Y$ has a confidence *c* if *c*% of the transactions in *D* that supports *X* also supports *Y.* The task of association rule mining is to thus generate a set of rules that have a minimum support and minimum confidence above certain user specified thresholds. Association rules over multiple tables can be described similarly. Association rules over multiple tables *X* and *Y* are atom sets of the form $p(t1, t2, \ldots, tn)$ where *ti* is a variable or a function of the form $f(t1, t2, \ldots, tn)$. The confidence is defined as follows: if *c*% of transaction in *D* are covered by *X* then they are also covered by $X \cup Y$. Similarly the support is defined as follows: *S* is the support of rule if *s*% of all transactions in *D* are covered by $X \cup Y$

2. RELATED WORK

In the Comma Framework [1] the auto annotation problem is posed as a multi relational association rule mining for making the relation between the images based feature and textual annotation. Using the multi relational association rule mining across multiple tables for generating association rules we combine the low level image features (color, orientation, intensity) and the corresponding text annotation. These rules are used to auto annotate test images. For performing the association rule mining task effectively an FP-tree algorithm is done. In multimedia data mining the multi-relational association rule mining is used because of multiple description of the same image such as multiple people labeling the same image differently. Multi-relational association rule mining is also advantageous as it provides auto annotation by pruning the number of trivial associations that are generated if text and image features were combined in a single table through a join. On the different test sets the auto-annotation is performed.

Another approach which describes the automatic image annotation is the HANOLISTIC approach [2], which is called Hierarchical image annotation system using Holistic Approach. The process of Automatic image annotation is assigning keywords to digital images based on the content information and it is a mapping from visual content information to semantic context information. There is a mapping of whole image which are set of independent descriptor spaces to class labels of different set of words. In this approach there are two layers. Former layer is level-0 that contains annotators of a set of distinct descriptors that extracted from whole image. **S**upervised learning paradigm is used for training of each annotator. Where class label defined by each word. In this approach each image has one or more annotating words, and assumed that an image belongs to more than one class. The output is the membership value obtained from level-0 annotators that shows the words in the vocabulary, which belongs to an image. To obtain meta-level annotator at the second layer the membership values of former layer are then aggregated from each annotator. Finally based on the ranking of the output of meta-level a set of words from the vocabulary is selected.

3. PROPOSED WORK

The aim of the proposed work is to retrieve an image and assign keywords or caption which is called annotation to the image based on the MPEG-7 Dominant Color Descriptor (DCD) and decision Tree. There are many algorithms for image retrieval and automatic image annotation on the basis of color but have not very much efficient. In this algorithm we apply the dominant color descriptor on the database and query image that generate vector value for image which can be correlated with the test keywords by their index values, in the multiclass association we generate the support and confidence values to check the frequency of the feature vector so that we are able to proceed further. By applying Decision Tree we calculate the gain of feature vector whether they satisfy the condition or not, if satisfy we are able to get the output i.e images that are more closed to the query image and their annotation are generated. If the condition is wrong process the next index. The Decision Tree divided the classes into leveled and unleveled classes for validation purpose. The leveled classes are used for further processing and unleveled classes are discarded. Using the cross fold validation function we check ratio between leveled and unleveled data that is 3:1

4. ALGORITHM OF PROPOSED WORK

Step 1: Firstly load RGB image from the database
Step 2: Convert this RGB image into a HSV image that gives gray value of the image. HSVmap = rgb2hsv (RGBmap)
Step 3: HSV space is generated
Step 4: By the HSV space calculate the Histogram of the image which is represented into blocks of bins.
Step 5: Apply the Dominant color descriptor on the Histogram of the Image.
Step 6: Generate Feature Vector
Step 7: Generate classes of feature vector.
Step 8: Go to the next level of Database
Step 9: Apply the DCD extraction on the image
Step10: We get Class of features of DCD
Step 11: Create the feature vector index
Step 12: The feature vector index is sorted according to Feature frequency.
Step 13: Index of feature vector is generated.
Step 14: Index of feature vector and test keyword are correlated using index values and applied on multi classification association to get the frequency of feature by their support and count values.
Step 15: After the multi-classification association we get the generated index of keywords
Step 16: We get assigned class of vector
Step 17: Apply the Decision tree where we are able to get leveled and unleveled data.
Step 18: On this leveled and unleveled data we apply the cross validation
Step 19: at the time of applying Decision tree we get calculated gain at input image.
Step 20: calculate the condition of gain. I.e Gain >=Vi if Yes then generate the result go to its next level and finally retrieve the output. If No then go to the processed next index.

Firstly from the database we read the input image to convert it from RGB to HSV color space for the better operations of annotation and retrieval then calculate the histogram of the image. A histogram is a standard statistical description of a distribution in terms of occurrence frequencies of different event classes for color, the event classes are regions in color space. On the histogram of the image we apply Dominant Color Descriptor to find the dominate value of the data. This will create the feature vector of the different features according to these features different vector classes are generated. Now at this stage we are having the various classes of feature vectors. On each image of the database the DCD extraction will create the class of feature of DCD that create the feature vector index which is sorted according to the features and get an index of feature vector. This indexed feature vector and the test keyword are correlated using multiclass association for giving the value of support and confidence. Now the image has index with the keyword to which we assign class of vector. Now we apply decision tree for class validation purpose and then calculate the gain for the final result. The decision tree will level or unlevel the classes, which can be validated for actual processing. Leveled classes are used and unleveled are discarded. The gain of input image is compared with the vector value. If gain>= vector value we obtain the result as retrieved images with their annotation of Query image and we go to the next level where again the image retrieval with annotation is performed, and if gain>=vector value is not satisfied we process the next index.

4.1 Flow Graph for proposed work
User Input:

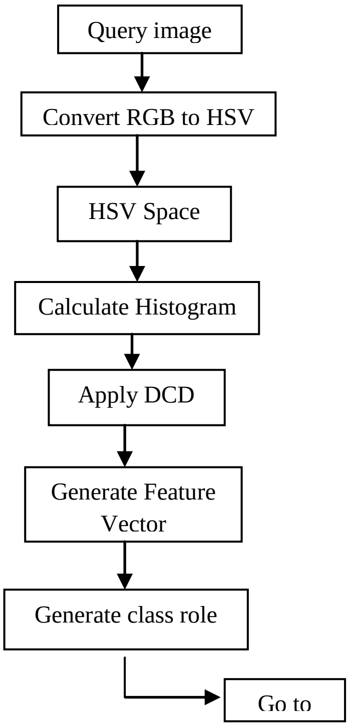

Figure 3: User input for proposed work

Database:

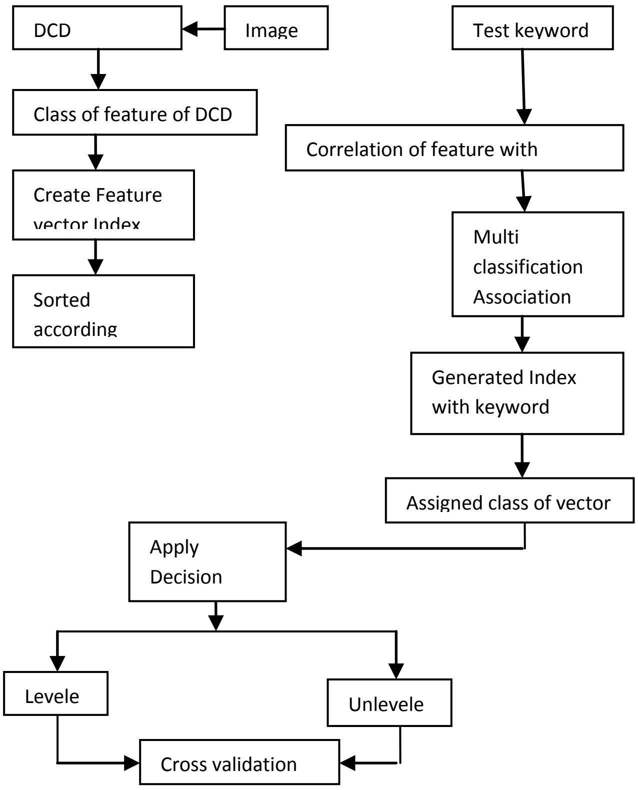

Figure 4: Database for proposed work

4.2 Results and Analysis
Here we read an input image from Database

Here we read an input image from the Database:

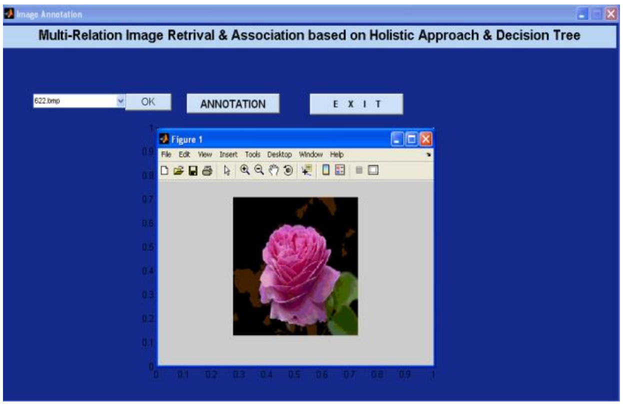

Figure 5: Read an image

Input image becomes Query Image after processing:

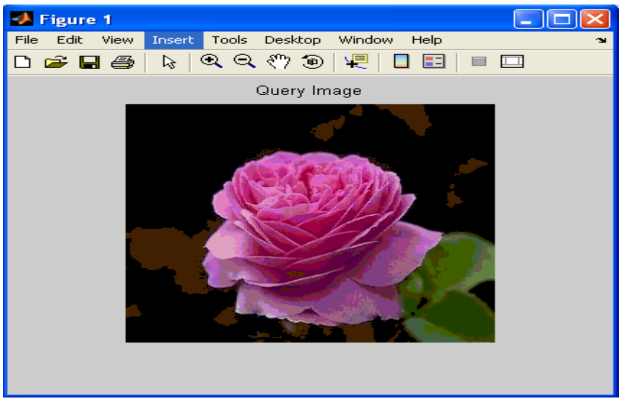

Figure 6: Query image of input image

Retrieved image based on the DCD generates automatic image annotation

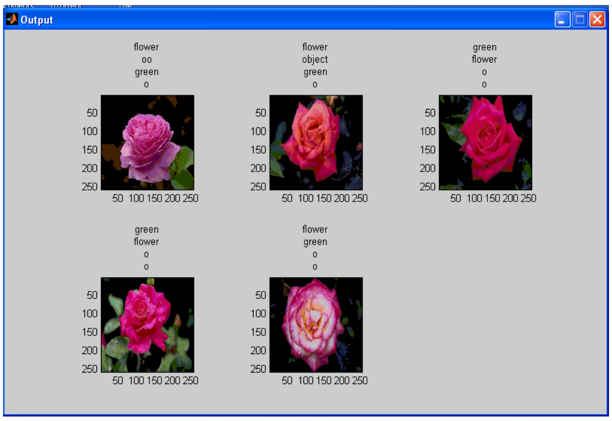

Figure 7: Image retrieval and Annotation

5. CONCLUSION

In the field of content based image retrieval color is very powerful feature representing image content, in this thesis we use very well known standard MPEG-7 Dominant Color Descriptor for color feature extraction for image retrieval and using the correlation between the test keywords and the class of feature vector we generate the auto annotation. This thesis proposes a new technique using Dominant Color Descriptor and Decision tree for the efficient retrieval and proper annotation for the query image. The Holistic approach used here for multiclass association provides the desired result. The results show the accuracy of the algorithm for better retrieval and annotation.

6. FUTURE WORK

The simulation result shows that this approach of image retrieval and annotation has better performance than the traditional approaches varying other parameter and performance metrics. We can improve the rate of annotation keyword by other methods. The work in the future is to improve the algorithm, and make it be suitable for complex environment. To improve the performance of retrieval we can use more combination of descriptor in MPEG-7 like color, shape, etc. the rate of retrieval annotation can also be improved using other methods.

7. ACKNOWLEDGMENTS

I am deeply thankful to my Guide Prof. Nishchol Mishra, Head of the Department Dr. Sanjeev Sharma, School of Information Technology, my parents and my best friend Mandakini whose guideline, suggestions and encouragement helped me in all the time for my paper.

8. REFERENCES

[1] Ankur M. Teredesai, Muhammad A. Ahmad, Juveria Kanodia, Roger S. Gaborski, “CoMMA: a framework for integrated multimedia mining using multi-relational associations” Springer-Verlag London Ltd. 2005.

[2] O zge, O ztimur Karadag, Fatos¸ T. Yarman Vural, Member IEEE “HANOLISTIC: A Hierarchical Automatic Image Annotation System Using Holistic Approach”, 2009 IEEE

[3] Manjunath T.N, Ravindra S Hegadi, Ravikumar GK, “A Survey on Multimedia Data Mining and Its Relevance Today”, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.11, November 2010 K. Elissa

[4] T. Sumathi, C.Lakshmi Devasena, and M.Hemalatha “An Overview of Automated Image Annotation Approaches” International Journal of Research and Reviews in Information Sciences Vol. 1, No. 1, March 2011

[5] Fuming Sun, Yong Ge, Dongxia Wang and Xueming Wang “A Collaborative Approach for Image Annotation” 2010 IEEE

[6] Dimitris K. Iakovidis, and Christos V. Smailis, “Efficient Semantically-Aware Annotation of Images”, 2011 IEEE

[7] Ja-Hwung Su, Chien-Li Chou, Ching-Yung Lin, and Vincent S. Tseng, Member, IEEE *“*Effective Semantic Annotation by Image-to-Concept Distribution Model” IEEE Transactions on multimedia, vol. 13, no. 3, june 2011

[8] Bo Li, Hong Li, Min Wu, Ping Li “Multi-label Classification based on Association Rules with Application to Scene Classification” 2008 IEEE

[9] Ran Li, YaFei Zhang, Zining Lu, Jianjiang Lu, Yulong Tian, “Technique of Image Retrieval based on Multi-label Image Annotation” 2010 IEEE

[10] Ritendra datta, dhiraj joshi, jia li, and james z. Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age”, ACM Computing Surveys, Vol. 40, No. 2, Article 5, Publication date: April 2008

[11] Ameesh Makadia, Vladimir Pavlovic, and Sanjiv Kumar, “A New Baseline for Image Annotation” Google Research, New York, NY, Rutgers University, Piscataway, NJ

[12] Peng Huang, Jiajun Bu, Chun Chen, Kangmiao Liu, Guang Qiu “Improve Image Annotation by Combining Multiple Models”, 2008 IEEE

[13] Jens-Rainer Ohm, Leszek Cieplinski, Heon Jun Kim, Santhana Krishnamachari, B. S. Manjunath, Dean S. Messing, Akio Yamada “The MPEG-7 Color Descriptors”

[14] Hong Shao, Yueshu Wu, Wencheng Cui, Jinxia Zhang “Image Retrieval Based on MPEG-7 Dominant Color Descriptor“, The 9th International Conference for Young Computer Scientists, 2008 IEEE

[15] Zhiping Shi & Xi Liu & Qingyong Li & Qing He & Zhongzhi Shi, “Extracting discriminative features for CBIR”, Springer Science Business Media, LLC 2011