

# MULTI MODAL ONTOLOGY SEARCH FOR SEMANTIC IMAGE RETRIEVAL

R.I. Minu<sup>1</sup> and K.K. Thyagarajan<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Anna University, Tiruchirappalli, India

E-mail: r\_i\_minu@yahoo.co.in

<sup>2</sup>RMD Engineering College, India

E-mail: kkthyagarajan@yahoo.com

## Abstract

*In this world of fast computing, automation plays an important role. In image retrieval technique automation is a great quest. Giving an image as a query and retrieving relevant images is a challenging research area. In this paper we are proposing a research of using Multi-Modality Ontology integration for image retrieval concept. The core strategy in multimodal information retrieval is the combination or fusion of different data modalities to expand and complement information. Here we use both visual and textual ontology contents to provide search functionalities. Both images and texts are complimentary information units as the human perspective will be different. So, the computational linguistic of images will lead to disambiguate text meaning when it is not quite clear in right sense of several words. That's why the Multi-Modal information retrieval may lead to an improved operation of information retrieval system.*

*If we go for automation we are in need of a fuzzy technique to predicate the result. So in this paper we using Support Vector Machine classifier to classify the image automatically by using the general feature such as color, texture and texton of an image, then by using this result we can create both feature and domain ontology for an particular image. Using this Multi-Modality Ontology we can refine our image searching system.*

## Keywords:

*Image Retrieval, Multi-Modality Ontology, SVM Classifier, Domain Ontology, Feature Ontology*

## 1. INTRODUCTION

An image is equivalent to thousand words that is the reason why we have enormous number of image database. Lots of research for retrieving is been proposed yet we haven't [1] reach the effective way of retrieving images. The main reason is that lack of unique features, for identifying a human can be done by his biometric feature like eyes and thumb impression which provide the unique feature, for general images such unique feature is inadequate.

In this paper we are proposing a research work of using Multi Modal Ontology integration for image retrieval concept. The need for such a work is that like our literal data we are in need of effective search retrieval technique for images [2][3].

The main reason for the lack of imperfection in image retrieval, is that like our literal data the image input cannot be made generic collection as the image is of different format and the low level features of same image with different format is completely different. So there is a need to fill the semantic gap between low-level and high-level feature. For this we are proposing a concept of integrating the Ontology with the image feature. As simply integrating a single ontology with a single feature won't provide generalized hypothesis, here we are using multiple feature and ontology.

We sub divided the work into four different phases: 1) Image Classification Phase 2) Multi Modal Ontology Creation Phase 3) User's Keyword or Image Query processing and 4) Retrieving and Relevant Ranking Phase, with this heading the following paper is been organized.

## 2. PROPOSED FRAMEWORK

In [7] Huan Wang 's proposal they have taken an Animal Ontology of Canine family domain with that they integrate the visual feature of the fox family with its domain ontology. In [4] they concentrate more on how to integrate the ontology of visual feature and domain ontology, so specific way to represent the visual feature is not given.

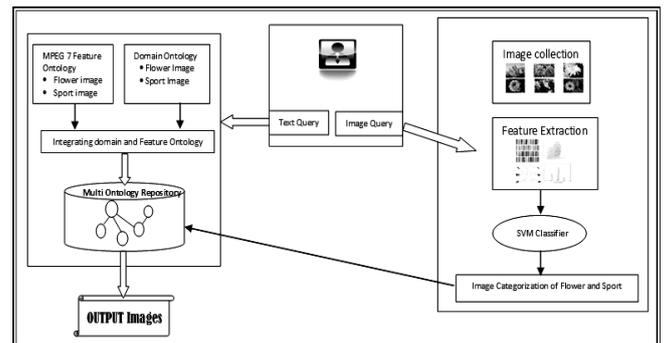


Fig.1. Proposed Framework

So all those proposal has technique to integrate the multimodal ontology concept but has not a way to get a generalized visual and domain feature Ontology. In our proposal we concentrate on how to get different visual feature and how to classify different image. Here we have taken two different domain set of images color flower images and sports images. Figure 1 shows the general idea behind the multimodal fusion image retrieval.

As shown in Fig.1 a user can provide either text or image as a query. The framework has an Image Classification Phase where is classifies whether it is a flower image or a sports images for this classification we use Semi-supervised SVM technique. Then it has an Multi Modal Ontology Creation Phase where we create an Multi Ontology such as feature and domain for Flower and Sports images and then an User Interface phase through which an user can give an query either as Keyword or as image.

## 3. IMAGE CLASSIFICATION PHASE

Human perceptive of analyzing and classifying the images is not equivalent to that of a machine. So as we are in need of a

generalized hypothesis concept for classification of images here we are considering only two domains of image collection. Namely Sports image collection and Flower image collection.

In related work [8] the author would consider either one of domain major work has been done in sports image classification.

As shown in Fig.1 the main component of this phase is the MPEG7 feature extractor and the SVM classifier.

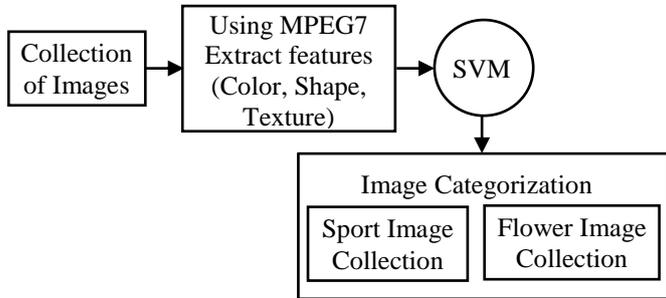


Fig.2. Image Classification Phase

**3.1 MPEG 7 FEATURE EXTRACTOR**

MPEG 7 is an ISO/IEC [6] standard for describing the multimedia content using different standard audio/video descriptor.

Table.1 shows the list visual Descriptor specified by the MPEG 7 standards. As shown in Fig.1 from the collection of images or from the given query image instead of extracting general features like the one used for Content Based Image Retrieval here we are using the standard MPEG 7's Color, Texture and Shape descriptor.

In most of the related work [5] they use any one of the feature here we use the combined features which provide some relevant images.

In this [17] paper we are proposing an idea of extracting the Dominant Color Descriptor (DCD), Color Layout Descriptor (CLD) and Edge Histogram Descriptor (EHD) for classification of images.

Table.1. MPEG 7 Visual Descriptor

VISUAL DESCRIPTOR			
COLOR	TEXTURE	SHAPE	MOTION
Dominant color	Texture Browsing	Contour Shape	Camera Motion
Color Layout	Homogenous Texture	Region Shape	Motion Trajectory
Scalable Color	Edge Histogram		Parametric Motion
Color Structure			Motion Activity
GOF/GOP			

DCD D <sub>1</sub> D <sub>2</sub> .....D <sub>8</sub> 24 byte	CLD Y <sub>dc</sub> Cr <sub>dc</sub> Cb <sub>dc</sub> 2 byte	EHD B1 <sub>ehd</sub> B2 <sub>ehd</sub> ---- B16 <sub>ehd</sub> 10 byte
---	--	---

Fig.3. DCE Chuck

As shown in Fig.3 the first 24 byte of the chunk is the Dominant color of the image. In MPEG 7 total of 8 dominant color would be extracted, each dominant color D<sub>n</sub> has the value from (0 0 0 to 255 255 255) so it contribute total of 24 bytes. In Color Layout Descriptor we are more concentrated only on the DC coefficient block's value so it is of totally 2 byte and the last few bits gives the detail of Edge Histogram value where each block value has 16 bit data. Table.2 shows the DCE Chuck values for different images.

**3.2 SVM CLASSIFIER**

Support Vector Machine classifier [9, 10, 11] is a kind of unsupervised learning algorithm. It is a kind of learning technique where we have a set of input but not have a deterministic idea about the output. In the proposal need an unsupervised learning algorithm to determine whether the given input image is a flower or sports image from its feature vector as a percepts for this we are using the SVM classifier.

Table.2. DCE Chuck values

IMAGES	DCD								CLD			EHD																
	D1	D2	D3	D4	D5	D6	D7	D8	Y	CR	CB	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	E16	
	64	192	0	255	255	255																						
	64	192	0	175	0	255																						
	64	192	0	175	0	255			16	22	28	23230	23454	44565	33566	42652	13667	42357	43664	34673	32657	32567	42466	24556	33526	24665	33565	
	192	576	0	605	255	765																						
	64	192	0	255	255																							
	64	192	0	175	255				18	34	50	23632	35572	24742	43231	42672	42554	24464	51474	42371	34373	25762	61652	21230	13171	33762	33652	
	192	576	0	605	765																							
	255	255	0	255	255	255	64	192																				
	0	255	0	200	175	255	64	192	48	19	34	54663	35672	36564	34772	42556	33764	44664	446633	32755	34132	32664	45542	26211	04110	03111	24122	
	0	0	0	0	175	255	64	192																				
	255	510	0	455	605	765	192	576																				
	192	64	255	0	255	255																						
	192	64	255	0	175	200																						
	192	64	255	0	175	0			40	18	29	03200	03122	23211	33524	15235	13656	52443	23555	06322	43744	61323	23344	04313	54434	26534	07222	
	576	192	765	0	605	455																						

A learning algorithm study its environment from its past history of input output data sets. Such kind of data sets is said to be a training data sets. So here the prediction is depends upon the hypothesis space or the function used. Here in SVM more generally called as a Kernel Machine uses a complex, nonlinear function for effective hypothesis function [12, 13].

Let  $(X_i, Y_i)$  be the attributes in our training example. Assume  $X_i$  belongs to a hypothesis space  $R^D$ , Then  $Y_i$  is -1 if it not belong to  $R^D$  or +1 if it belong. So,

$$Y_i(WX_i, b) \geq 0 \tag{1}$$

which determines the hyperspace plane [11].

From the above statement the main issues in SVM is to find an effective kernel function for separating different attributes in the training sets as per their feature. The optimal solution has been derived in SVM using quadratic programming concept.

Optimization [13] can be done by linear programming and by Quadratic Programming. If we have only Linear Constrain we will go for Linear Programming concept.

**Example:**  $2X_1 + 3X_2 + X_3$ , in this simple equation by substituting the appropriate value for the constraint  $X_1, X_2$  and  $X_3$  we can get an optimal solution where in Quadratic Programming the equation would be Quadratic in nature.

$$F = 2X_1^2 + 3X_2^2 + X_3 \tag{2}$$

To determine the value of  $(X_1, X_2, X_3)$  we require a kernel function say  $K$ ,

$$K = H(X_1, X_2, X_3) \tag{3}$$

where,  $H$  is a hypothesis function. So now to determine these values in quadratic programming,

$$F_{min} = \frac{1}{2} X^T H X C^T X + \alpha \tag{4}$$

The SVM is mostly stated in matrix or in vector form. Here the value of  $H$  is a symmetric matrix of the given quadratic equation and  $C$  is the vector of the attribute used and  $\alpha$  is a constant.

So with reference to Quadratic Programming the in Eq.(1) for optimal solution the  $W$  should be minimizes as given by,

$$\phi(W) = \frac{1}{2} W^T W \tag{5}$$

As in Eq.(4) here the  $W$  is minimized.

Fig.4 [10] shows a SVM classification of two set of data class in Magenta and Blue color. Like the color dots the MPEG 7 feature would be densely distributed and can be classified using the SVM algorithm.

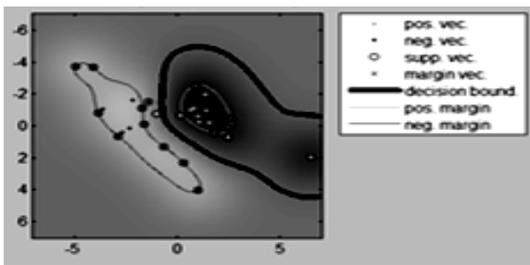


Fig.4. SVM Demo output

We can use the MPEG 7 Chunk [17] value as describe in Fig.3 as data set through which we can generate a complex quadratic equation as shown in Eq.(6) through which by implementing the SVM quadratic programming we can optimize and classify the image,

$$255D_1^3 + 175D_2^2 + 255D_2 + \dots + Y14 + Cr17 + Cb7 + 3E_1^2 + E_1 \dots \tag{6}$$

We just try this concept with simply giving a query flower image and show the relevant flower image is shown in Fig.5 which is an half way implementation of said proposal as some irrelevant image is also classified as flower image.

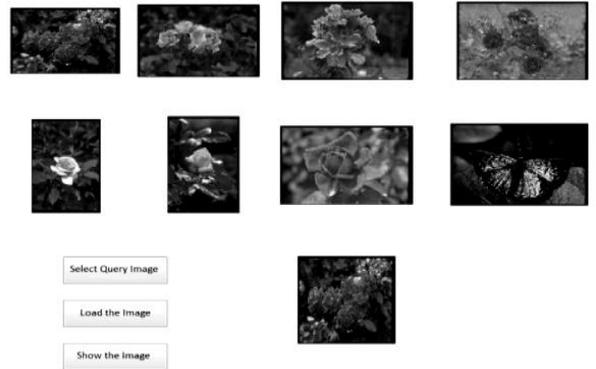


Fig.5. Simple implementation

#### 4. MULTI MODAL ONTOLOGY CREATION PHASE

Once we have classified the images as sport or flower images, we can integrate those results with this phase.

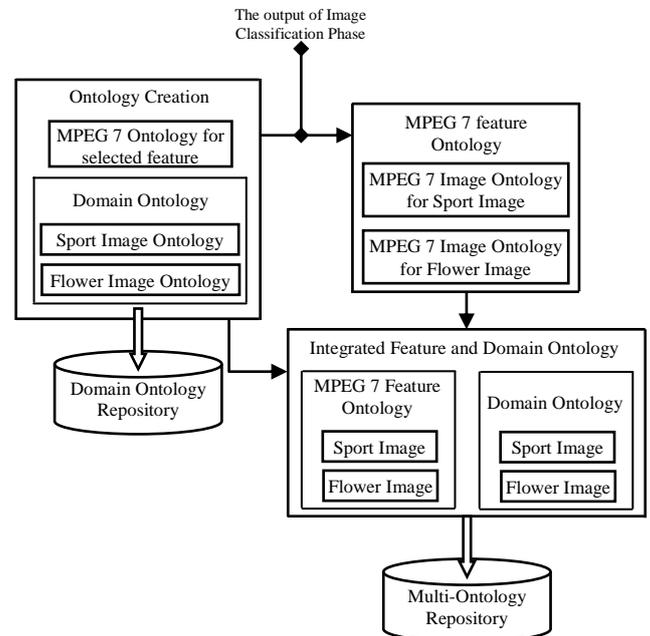


Fig.6. Multi Modal Ontology Creation

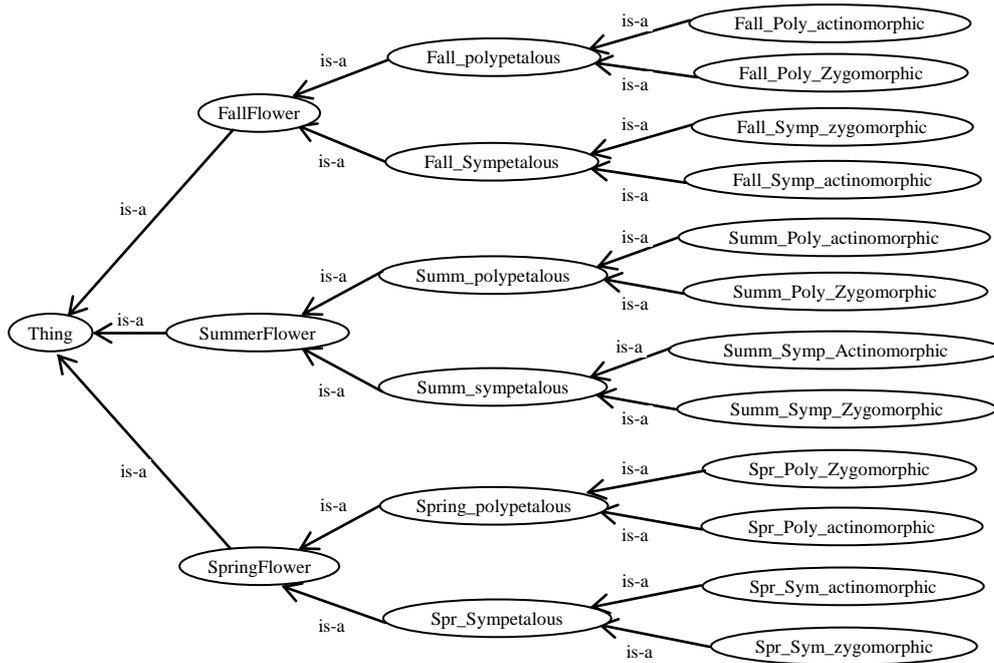


Fig.7. Flower Ontology

From the block diagram of Fig.6, here we are fusing multiple ontologies so it called as [16] Multi-Modal Ontology. First we have to create some general domain ontology for Sports and Flower as in this work we are concentrated only on those two domain. Then for image feature extraction as we are using MPEG 7 descriptors we are also in need of its domain ontology so creation of all is what given in Ontology Creation blocks. Then the feature value derived from the previous phase is matched with the domain ontologies though integrating all ontologies [14].

As shown in Fig.6, here we have to concentrate on two main things one on the Domain ontology of the sport and Flower image. One can create domain ontology if they have well rich information about that domain. As for sports [8] provided a complete ontology for sports. For flower ontology can be of like one shown in Fig.7.

As MPEG 7 is a standard way of representing multimedia details the complete ontology of MPEG 7 is shown in Fig.8.

As these ontology is of in OWL language which as some similarity of XML language the feature vector value such of DCD, CLD, EHD etc can to substituted to the concern OWL subclass as the output of a mere MPEG 7 for an given image is an XML file with all element in it so we can merge the XML element of MPEG 7 with the OWL classes thus can create a repository of OWL image files for classified image. So for a given flower images we have an Domain ontology to explain the domain of the concern image and also have a feature ontology through which for each and every images we have provided the semantic to it.

## 5. USER'S KEYWORD OR IMAGE QUERY PROCESSING PHASE

In this Ontology based image retrieval system we have tried to provide semantic to all the images. As the fact if we provide some meaning to the images stored in the repository to the system, a computer based intelligent agent can provide you with relevant images. In this semantic based system we can provide either image or query as the input as shown in Fig.9.

If a user given text as an query, the query is been analyzed and can be compare with all the class and sub class of the domain ontology. Once the domain is identified then the concern image from the repository can be displayed as per ranking [15].

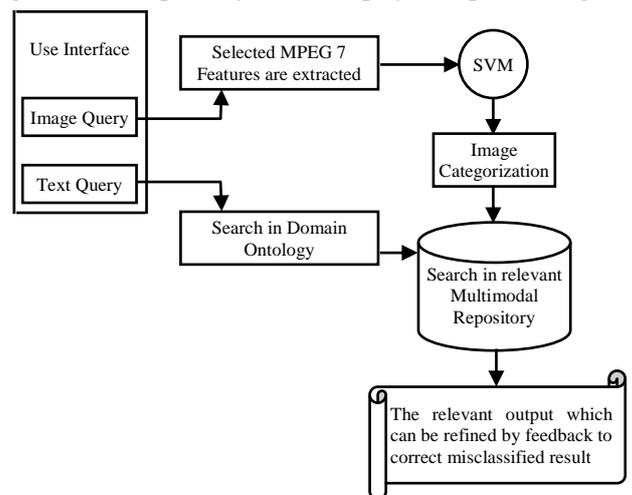


Fig.9. User Interface

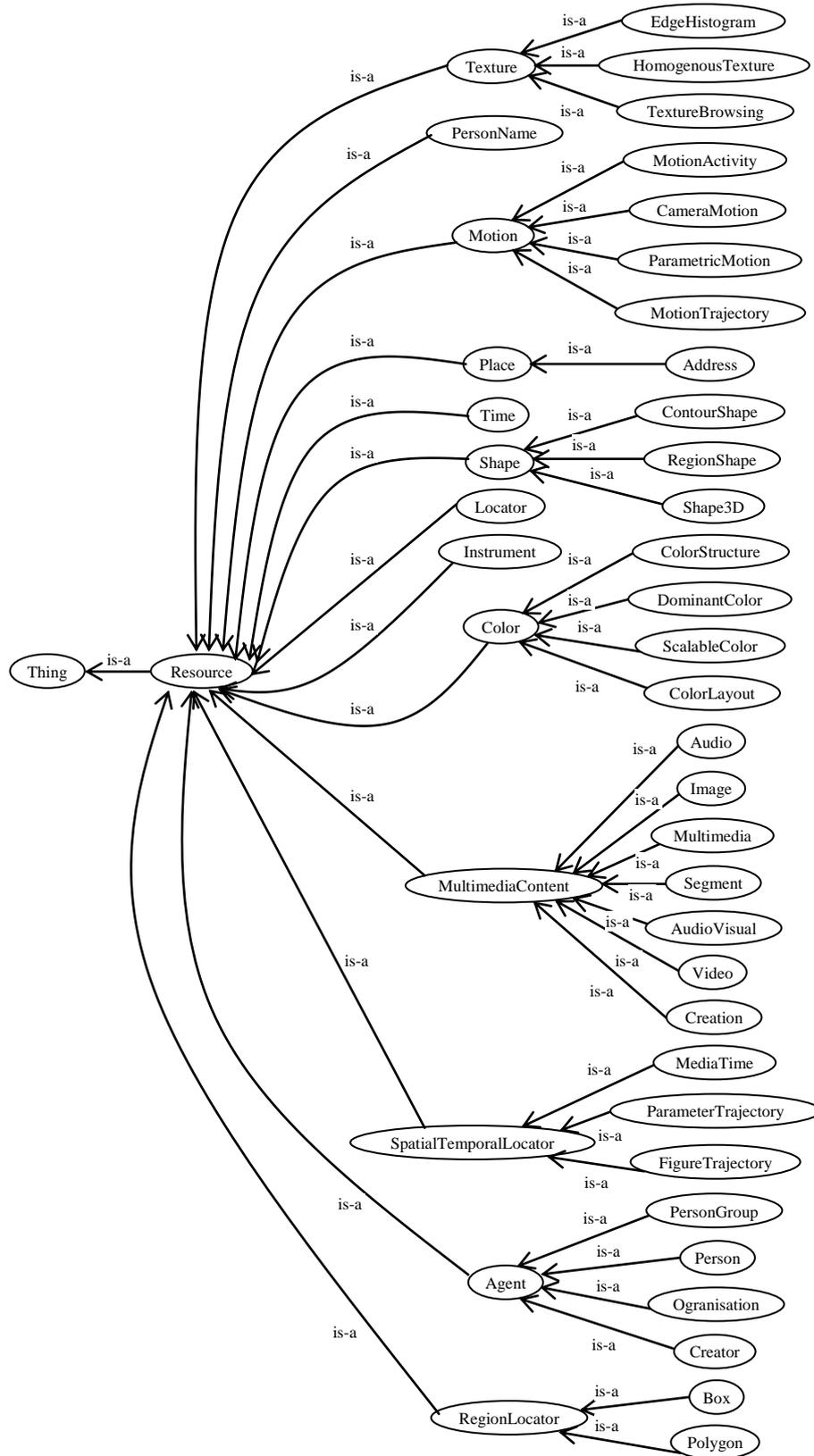


Fig.8. MPEG 7 Multimedia content descriptor ontology

If a user provide image as a query because as for flower kingdom the user will unaware of more than 70,000 different flower names. The feature of the given image is extracted then the DCE chunk is determined through our SVM classifier we can

classify whether it is a flower or sport image. Once the flower is been classified form the flower repository a mere search is made and the relevant result is listed out with ranking strategy.

## 6. EXPERIMENTAL RESULT

The MPEG 7 features mainly dominant color descriptor, scalable color descriptor and edge histogram descriptor of different flower family is extracted and we tried to classify as per the DCE chunk, manually in that few flower image got classified in class of 10 categories. Then we tried on different types of images such as flower image, cup image and motorbike images. In that classification the images doesn't clustered as per there class group. The problem in this classification is that, these features alone cannot give the efficient classification. So, in our future work we try to incorporate few more semantic feature such as texton [21] and shape features.

## 7. CONCLUSION

Relevancy based image retrieval is a wide exploring research region. As for today's technology researchers are trying to put Robotics to Cloud computing network, so to learn a Robotic agent it can take a image of current issues and search for relevant information from Cloud to perform further action. So the importance of relevancy and recall factor for image retrieval system should be very accurate, this kind of accuracy can be provide only if we teach our system the meaning alias semantic of the images. That's why in our work we tried an approach of merging the domain knowledge to feature of the images.

## REFERENCES

- [1] Ying Liu, Dengsheng Zhang, Guojun Lu and Wei-Ying Ma, "A survey of content based image retrieval with high-level semantics", *Pattern Recognition*, Vol. 40, No. 1, pp. 262-282, 2007.
- [2] Nidhi Singhai and Shishri K.Shandilya, "A Survey on: Content Based Image Retrieval Systems", *International Journal of Computer Applications*, Vol. 4, No. 2, pp. 22-26, 2010.
- [3] Yihun Alemu, Jong bin Koh, Muhammed Ikram and Dong Kyoo Kim, "Image Retrieval in Multimedia database: A Survey", *Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 681-689, 2009.
- [4] Ja-Hwung Su et. al., "Multi-Modal Image Retrieval by integrating web image annotation, concept matching and fuzzy RANKING technique", *International Journal of Fuzzy Systems*, Vol. 12, No. 2, pp. 136-149, 2010.
- [5] B.S. Manjunath and W.Y. Ma, "Texture features for browsing and retrieval of image data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, pp. 837-842, 1996.
- [6] J. Martinez, "MPEG-7 Overview (version 10)", *Palma de Mallorca*, 2005  
<<http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm>>.
- [7] Huan Wang, Song Liu and Liang-Tien chia, "Does ontology help in image retrieval?: A comparison between keyword, text ontology and multi modality ontology approaches", *Proceedings of the 14<sup>th</sup> annual ACM International Conference on Multimedia*, pp. 109-112, 2006.
- [8] Kraisak Kesorn, "Multi modal multi semantic image retrieval", Ph.D. diss., School of Electronics Engineering and Computer Science, University of London, 2010.
- [9] Khalid Sayood, "*Introduction to Data Compression*", Second Edition, Morgan Kaufmann Publishers, 2000.
- [10] Lei Zhang, Fuzong Lin and Bo Zhang, "Support Vector Machine Learning for Image Retrieval", *IEEE International Conference on Image Processing*, Vol. 2, pp. 721-724, 2001.
- [11] Haiying Guna, Sameer Antani, L Rodney Long and George R Thoma, "Bridging the semantic gap using ranking SVM for image retrieval", *IEEE International Symposium on Biomedical Imaging*, pp. 354-357, 2009.
- [12] Patrik Ndjiki-Nya, Oleg Novychny and Thomas Wiegand, "Merging MPEG 7 descriptors for image content analysis", *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 453-456, 2004.
- [13] Chengcui Zhang et. al., "A Multiple instance learning approach for content based image retrieval using One class support vector machine", *IEEE International Conference on Multimedia and Expo*, pp. 1142-1145, 2005.
- [14] Ahmad Adel Abu Shareha, Madava Rajeswari and Dhanesh Ramchandram, "Multimodal Integration(image and text) using ontology alignment", *American Journal of Applied Sciences*, Vol. 6, No. 6, pp. 1217-1224, 2009
- [15] Apostol Natsev, Alexander Haubold, Lexing Xie and Rong Yan, "Semantic concept based query expansion and re ranking for multimedia retrieval", *Proceedings of the 15<sup>th</sup> International Conference on Multimedia*, pp. 991-1000, 2007.
- [16] Huan Wang, Liang-Tien Chia and Song Liu, "Semantic retrieval with enhanced matchmaking and multi-modality ontology", *IEEE International Conference on Multimedia and Expo*, pp. 516-519, 2007.
- [17] R.I. Minu and K.K. Thyagarajan, "Automatic image classification using SVM Classifier", *CiiT International Journal of Data Mining Knowledge Engineering*, 2011.
- [18] G. Nagarajan and K.K. Thyagarajan, "A Survey on the Ethical Implications of Semantic Web Technology", *Journal of Advanced Research in Computer Engineering*, Vol. 4, No. 1, pp. 123-133, 2010.
- [19] R.I. Minu and K.K. Thyagarajan "Scrutinizing the Video and Video Retrieval Concept", *International Journal of Soft computing and Engineering*, Vol. 1, No. 5, pp. 270 – 275, 2011.
- [20] R.I. Minu and K.K. Thyagarajan, "A Novel Approach to Build Image Ontology Using Texton", *Intelligent Informatics: Advances in Intelligent Systems and Computing*, Vol. 182, pp. 333-339, 2013.