Hybrid based Semantic Image Annotation using SVM and DT

Manochitra.S
PG Scholar, CSE Department
SREC, Coimbatore, India

Janice E J
PG Scholar, CSE Department
SREC, Coimbatore, India

Suganya S
PG Scholar, CSE Department
SREC, Coimbatore, India

ABSTRACT

This system proposes an ontology based framework for the semantic search in the image annotation process. The main objective of this approach is to use ontology for the semantic search in the image retrieval process. The ontology based framework is developed to define the image space. This system proposes a construction of semantic based approach for image representation using SVM and decision tree classifiers for learning and retrieval of relevant images. So the performance is significantly enhanced by using the SVM and decision tree as a classifier for retrieving the similar images.

Keywords

Automatic image annotation, support vector machine, Decision Tree learner, ontology based retrieval, Content Based Image Retrieval.

1. INTRODUCTION

Content-based Image Retrieval (CBIR) means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and the descriptions associated with the image. The term 'content' in this context might refer to colors, textures, and shapes or any other information that can be derived from the image itself. CBIR is desirable because image search engines rely purely on metadata and this produces a lot of garbage in the results also having humans manually enter keywords for images in a large database can be inefficient and expensive and may not be capture for every keyword that describes the image. Thus a system that can search the images based on their content would provide better indexing and return more accurate results. It generally deals with the image visual features such as color, shape and texture. The problem associated with the CBIR is a large semantic gap between the low level visual features and the high level textual features and the textual information about the images in the database are described by the human.

Automatic image annotation (AIA) 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. In a typical AIA method, the model corresponding to each keyword is built from training samples using machine learning. Several typical machine learning tools have been used extensively in AIA, such as Support Vector Machine (SVM),[3] neural network Minimum Reference Set based Multiple-Instance Learning (MRS-MIL)[15] and group sparsity method proposed a Supervised Multiclass Labelling (SML) method.

The ontology based image retrieval is to represent the image in a semantic manner and machine understandable format. The semantic concept is included in order to combine the low level features and high level textual features.

Through semantic annotation, both images and retrieval queries can be formalized as XML files. In semantic annotation, the semantic meanings of images and queries are described based on a combination of concepts defined in ontology. In image retrieval, the goal is to determine the similarity between images and a retrieval query. To achieve this objective in ontology-based image retrieval, we implement similarity comparison in two steps: extraction of combined concept entities and similarity comparison between images and a retrieval query.

2. RELATED WORK

Different methods of image annotations are proposed by various researchers.

Ruhan He , Naixue Xiong , Laurence Yang , Jong , 2011, [12] proposed an approach based on Multi-Modal Semantic Association Rule (MMSAR).This method aims to combine the keywords and visual features automatically for Web image retrieval. A MMSAR contains a single keyword and several visual feature clusters, which crosses and associates the two modalities of Web images.

Anne-Marie Tousch, 2012,[1] conducted a survey on image annotation uses and user needs, and the need for automatic annotation. The survey exposes the difficulties posed to machines for this task and how it relates to controlled vocabularies.

Yakup Yildirim, Adnan Yazici, and Turgay Yilmaz, 2011,[13], has proposed a semantic content extraction system. It allows the user to query and retrieve the objects, events and concepts that are extracted automatically. A novel ontology-based fuzzy video semantic content model approach is introduced in this method. It uses spatial/temporal relations in event and concept definitions. This meta-ontology definition provides a wide domain applicable rule construction standard that allows the user to construct ontology for a given domain.

Dmitri V. Kalashnikov,Sharad Mehrotra,JieXu, and Nalini Venkatasubramanian ,2011,[4],has proposed how semantic knowledge in the form of co-occurrence between image tags can be exploited to boost the quality of speech recognition.

Ning Yu,Kien [9] introduced a novel Multi-Directional Search framework for semi-automatic annotation propagation. Here the user interacts with the system to provide example images and the corresponding annotations during the annotation propagation process. Example images are clustered each iteration and the corresponding annotations are propagated separately to each cluster. Then some of the images are returned to the user for further annotation. As the user marks more images, the annotation process goes into multiple directions in the feature space. Multi path navigation is used to treat query movements. User input based path splitting mechanism is evolved. The system provides accurate annotation assistance to the user

images with the same semantic meaning but different visual characteristics can be handled effectively.

3. PROPOSED SYSTEM

In the proposed system initially the image visual features are extracted and from the visual features select some of the regions as the training dataset for the semantic learning. Then manually annotate the images in the training dataset. Next is to apply the machine learning algorithms such as support vector machine and decision tree in order to get the annotated rules. Simultaneously ontology structure is constructed and mapped to the indexing to retrieve the category images.

4. ARCHITECTURE

The main purpose of this work is to develop a machine learning approach and ontology hierarchy to translate an unstructured image into a structured textual document, that is, to translate regions in an image into textual keywords. Fig 1 represents the block diagram of the proposed work and the steps are described as follows:

4.1 Feature Extraction

In feature extraction phase the three visual features are retrieved, they are color, shape and texture feature.

4.1.1 Color feature : The color feature value which is extracted using the HSV (Hue Saturation Value) color histogram. Here the extracted color feature value is dominant color values which are 3-dimentional vectors [6].

4.1.2 Texture feature: The wavelet HAAR transform is used for the texture feature extraction. [3] The wavelet transform includes the average values and standard deviation.

4.1.3 Shape feature: Edge orientation histogram is utilized for the shape feature extraction. In edge orientation histogram the JPEG images are divided into 4 segments, each segment is further divided into 4 in order to find the edge for each region.

The idea is that of local processing. The image is divided into 4×4 sub-images. Each sub-image is further divided into smaller image blocks (typically 4×4 pixels). The standard allows for the having vertical, horizontal, diagonal (45 and 135 degrees) and non-directional edges. No edge is counted for monotone image block. Simple filtering of the image blocks allows obtaining the most prominent edge in the block.

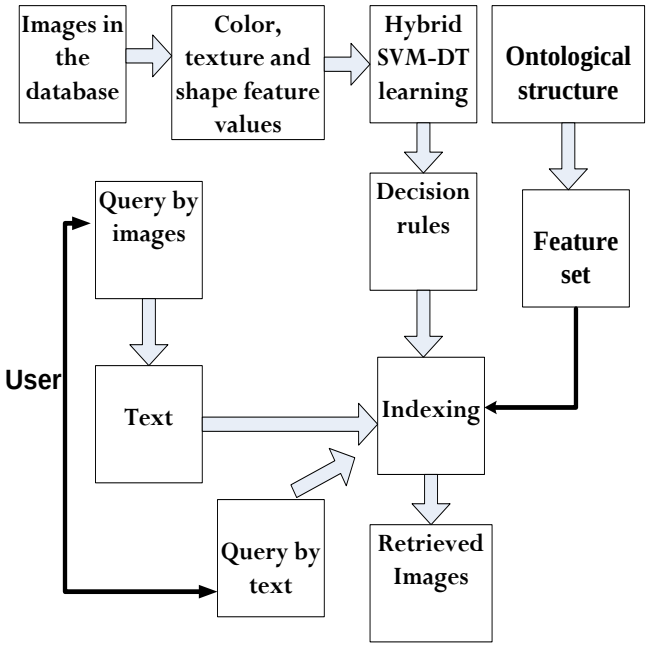

Fig.1 Block Diagram of Proposed System

A histogram of 5 bins (1 for each edge type) is computed over all the image blocks in the sub-image. This procedure is repeated for all the 16 sub-images and hence we obtain 80 histogram coefficients. The standard proposes non-linear quantization for the sake of storage.

4.2 Machine Learning Algorithms

Once the values are established as the training regions, they are used to learn semantic concepts and annotate unknown regions by a machine learning mechanism. Different from other classification models, the input-output relationship in DT can be represented using human comprehensible rules, i.e., ‘‘if-then’’ rules. DT induction algorithms ID3 [9] and C4.5 [10] are widely used. DT [11], proposed a hybrid tree simplification method using both pre-pruning and post-pruning. Fig 3 depicts the HSVM-DT process. The steps are as follows:

Step 1: Given the training set X = {[C1, T1, S1], . . . , [Cm, Tm, Sm]} and the labels Y = {y1, . . . , yn}, where Ci, Ti and Si (i = 1, . . . , m)are the inputs of color, texture and shape feature values respectively, m = 1200 and n = 40. Each xie X is associated with a yj.

Step 2: For the training set of {X, Y} above, train the SVM by using N-fold cross validation. According to the distribution characteristic then divide the training set into 10 equal size subsets, i.e., SS1, . . . , SS10. Each subset includes 120 training samples. Then test each of the 10 subsets in 10 iterations. At each iteration, select a different subset as a test set (SVM_test) for the SVM classifier. Then use the union set of the remaining nine subsets as a training set (SVM_train) for the SVM classifier.

Step 3: For each SVM_test, choose those that are correctly classified by the SVM as a new dataset (SVM_new) from the SVM_test.

Step 4: After the 10 iterations, use the union set of all SVM_new as the new input training set (SVMDT_train) instead of the original training set X for DT learning.

Step 5: Generate the decision rules in ''if-then'' format by growing a DT for the training set SVMDT_train.

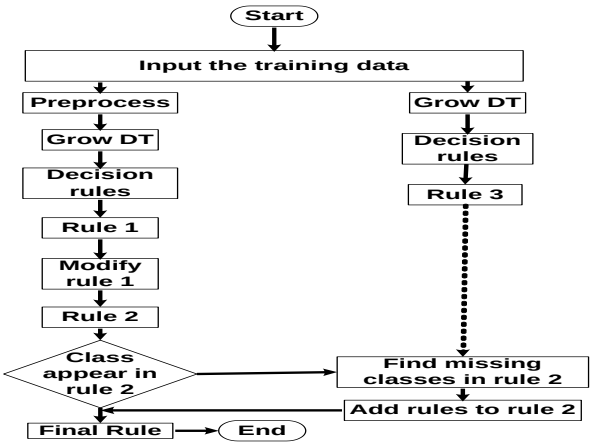

Fig. 2 Flow Chart of HSVM-DT

The first stage of the hybrid algorithm is N-FOLD cross validation. . For each fold, randomly assign data points to two sets a_0 and a_1, so that both sets are equal size (this is usually implemented as shuffling the data array and then splitting in two).Then train on a_0 and test on a_1, followed by training on a_1 and testing. This has the advantage of both training and test sets are large, and each data point is used for training and validation on each fold on a_0.

N-fold cross-validation, shown in fig.2 the original sample is randomly partitioned into N subsamples. Of the N subsamples, a single subsample is retained as the validation data for testing the model, and the remaining N − 1 sub - samples are used as training data. The cross-validation process is then repeated N times (the folds), with each of the N sub samples used exactly once as the validation data. The N results from the folds then can be averaged (or otherwise combined) to produce a single estimation. The advantage of this method over repeated random sub-sampling is that all observations are used for both training and validation, and each observation is used for validation exactly once. 10-fold cross-validation is commonly used, but in general k remains an unfixed parameter

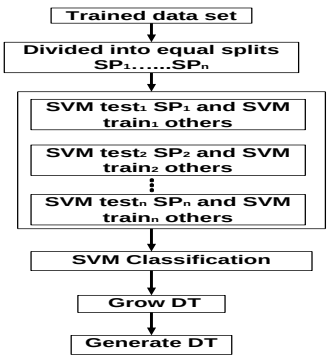

Fig. 3 N-FOLD Cross Validation

4.3 Ontological Structure

This is the phase where the ontology structure is constructed and used for image annotation. Finally the query image is mapped to the rules refined by the HSVM-DT machine learning algorithms.

4.3.1 Class hierarchy: Ontology expresses the elements of the domain using classes, properties and individuals, and the intended meaning of the elements are given by hierarchical relationships that may have cardinality, restrictions, combinations, union, intersection etc between the concepts [7] and [8]. The images are mapped to existing class hierarchy as instance with complete relationship (property) and content description. The lower class inherits the properties of upper classes. The fig. 4 represents the general ontology hierarchy for the image annotation.

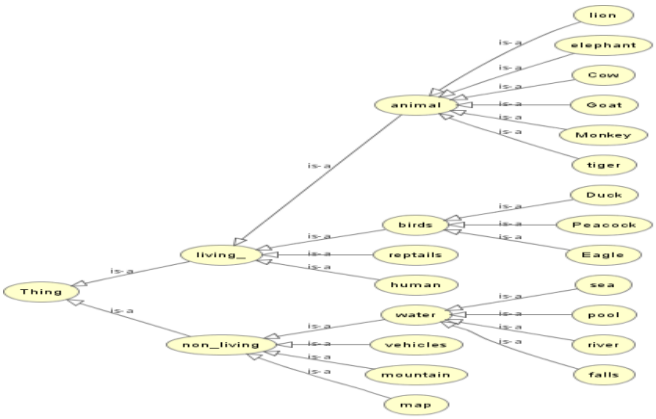

Fig. 4 General Ontology Hierarchy

4.3.2 Image Annotation: Image annotation is process of relating unknown image to the named class. That is mapping the unknown image to one of a number of known classes. This approach is based on the idea of image representation using ontology and retrieval using decision tree learning method. [14]High-level concepts are efficiently stored as meta data and automatically mapped to objects.[5] Low level features of images are extracted by various image analysis techniques such as edge histogram, color histogram, Zernike moments and co-occurrence matrix.

4.3.3 Image Retrieval: When query image is given, the color, edge and texture feature of that image is identified and which is compared with the features of the images in the database [2]. Similarly, one after one best matching images is found by measuring the one after one shortest distance in order to retrieve the similar images from the database. In order to get all the similar images, generate ontology and classify the initial retrieved images by employing ID3 algorithms. The query image is match up to the images in the database for image retrieval.

4.3.4 Image ontology: Image Ontology is constructed using Class, Properties and instances. A nouns class hierarchy of image which is also instances of leaf classes representing images. Common descriptions arranged for class hierarchies (data type properties).Object properties are used to connect instances of semantic classes with instances from classes containing description. The Image is annotated to more than one class by using assertion

property. Other general properties are added to classes as necessary (Such as bird has two legs).

Let Q be the query image. Perform the shape extraction process for query image Q. Then calculate the similarity measure between shape feature of query image Q and the features of each image in the database using Euclidean distance. The feature of query image is represented as Q_f and F_i represents the feature vector for each image in the database.

$$d = \sqrt{\Sigma(Q_f - F_i)^2} \quad (1)$$

4.4 Indexing

Then finally the indexing is to be done for the purpose of efficient image retrieval. The indexing is done based on the mapping of ontology with machine learning algorithm. Create an inverted file to index textual documents and rank the annotated images in databases, then it is easy to retrieve the images as same as the text retrieval. For query by images, the system annotates each of the segmented regions of the input image in the same way as above. Then use the keywords from the annotation as queries.Indexing must be done with images and its associated keywords [14].

5. RESULTS AND DISCUSSION

Precision is the percentage of retrieved imagssses that are relevant to the query. It measures the accuracy of retrieval and is computed as

$$R = \frac{number\ of\ relevant\ images\ retrieved}{total\ number\ of\ relevant\ images\ in\ the\ database} * 100 \quad (2)$$

Recall is the percentage of all relevant images that are retrieved. It measures the robustness of the system and is calculated as

$$P = \frac{number\ of\ relevant\ images\ retrieved}{total\ number\ of\ images\ retrieved} * 100 \quad (3)$$

The experimentation is done on the collection of corel image database. Approximately there are about 35 categories of images. For example grass, plant, bus, rose, horse, tree, flowers, elephant, plane, mountain, tiger, wave, ground, river, pond, leaf, stone, and rock.

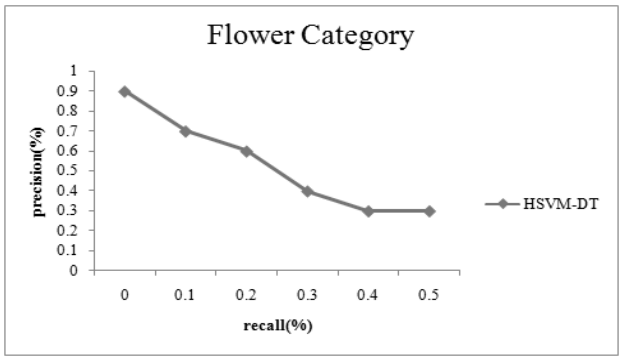

Fig. 5 Precision and Recall Rate for Flower Category

The corel database consists of 1000 images and each image is classified on its associated category. For example in flower category there are approximately 100 images. In the image retrieval, there is an inverse relationship between precision and recall. Precision falls as recall increases. The query image is a flower and the resultant will be a collection of flower category images.

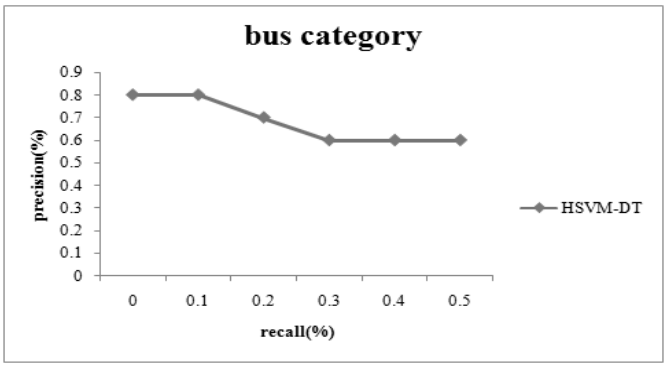

Fig. 6 Precision and Recall Rate for Bus Category

The precision and recall rate for the flower category is depicted in fig. 5. Here the flower category consists of 237 images. The precision and recall rate for the bus category is depicted in fig. 6.

6. CONCLUSION

In this paper, a hybrid methodology of SVM and DT is combined with ontology to complete the task of image annotation and retrieval as semantic process. By taking the SVM classifier as pre-processing of DT, plus SVM predicting, then obtain the robust DT rules. Thus the creation of ontology for collection of images with complete relationship will enhances the retrieval process at a faster rate when compared to ordinary retrieval. Ontology-enriched knowledge base of image metadata can be applied to constructing more meaningful answers to queries than just hit-lists.

7. REFERENCES

[1] Anne-Marie Tousch, Stephane Herbin, and Jean-YvesAudibert ," Semantic hierarchies for image annotation: A survey", Pattern Recognition Vol 45,pp.333–345 2011.

[2] Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.,, "Supervised learning of semantic classes for image annotation and retrieval", IEEE Trans. Pattern Anal. Mach. Intell. , 2007 Vol.29 No. 3, 394–410

[3] Deng, Y., Manjunath, B.S., " Unsupervised segmentation of color-texture regions in images and video" IEEE Trans. Pattern Anal. Mach. Intell. Vol. 23 No. 8,pp. 800–810, 2001.

[4] Dmitri V. Kalashnikov,Sharad Mehrotra,JieXu, and Nalini Venkatasubramanian,"A Semantics-Based Approach for SpeechAnnotation of Images" IEEE transactions on knowledge and data engineering, 2011 Vol. 23, No. 9, September 2011,pp.1373-1387.

[5] Han, Y., Qi, X.,"A complementary SVMs-based image annotation system" In: Proceedings of the International Conference on Image Processing (ICIP'05), Genoa, Italy, pp. 1185–1188 2005.

[6] Mylonas, P., Spyrou, E., Avrithis, Y., Kollias, S., 2009. Using visual context and region semantics for high-level concept detection. IEEE Trans. Multimedia 11 (2), 229–sss243.

[7] N.Magesh, P.Thangaraj, 2010, "Semantic Image Retrieval Based on Ontology and SPARQL Query, IJCA, PP 12-16.

[8] N.Magesh, P.Thangaraj,(2011) An Image Retrieval System Based on Extensive Feature Set using ID3 Decision Tree Algorithm, (2012), European journal of scientific research Vol 89 No. 1,121-135.

[9] Ning Yu , Kien A. Hua, Hao Cheng,2012, "A Multi-Directional Search technique for image annotation propagation",Visual Communication, Vol. 33 pp.237-244.

[10] Quinlan, J.R., 1986. Induction of decision trees. Springer Machine Learning 1, 81–106.

[11] Quinlan, J.R., 1996. Improved use of continuous attributes in C4.5. J. Artificial Intelligence Res. 4, 77–90.

[12] Ruhan He, Naixue Xiong, Laurence T. Yang and Jong Hyuk Park ,(2011)' Using Multi-Modal Semantic Association Rules to fuse keywords and visual features automatically for Web image retrieval' Elsevier science, Information Fusion Vol. 12 , pp.223-230.

[13] Yakup Yildirim, Adnan Yazici, and Turgay Yilmaz(2011)' Automatic Semantic Content Extractionin Videos using a Fuzzy Ontology and Rule-based Model' IEEE Transactions On Knowledge And Data Engineering,pp.1-15.

[14] Zeng Chen, Jin Hou , Dengsheng Zhang , Xue Qin .,2012.An annotation extraction algorithm for image retrieval. Elsevier Pattern recognition (33),pp. 1257-1268.

[15] Zhao, Y., Zhao, Y., Zhu, Z., Pan, J.S., 2008. MRS-MIL: minimum reference set based multiple instance learning for automatic image annotation. In: Proceedings of the International Conference on Image Processing (ICIP'08), San Diego, California, USA, pp. 2160–2163.