Semantic Scene Classification for Image Annotation

Vaibhav Vatsal[1], Shreya Srivastava[2]

[1,2]student of G.C.E.T B Tech(IT),Galgotias College of Engineering And Technology, Greater Noida, U.P

Abstract - *Nowadays The problem of annotating an image falls under pattern classification and computer vision methods. Various systems have been proposed for the purpose of image annotation, content based retrieval, and image classification and indexing. Most of these systems are based on low level features such as color and texture statistics. We believe that the detection of semantically meaningfully regions in digital photographs could be productively employed in classification and retrieval systems for image databases. The image retrieval problem has not been successfully solved in spite of decades of work, dozens of research prototypes and commercial products that have been developed lately. In addition to its primary goals, consistent, cost-effective, fast, intelligent annotation of visual data, the method proposed can also be used to improve the performance of subsequent image query and retrieval operations on the annotated image repository. Pure content based image retrieval (CBIR) also falls short of providing an adequate solution to the problem, mostly due to the limitations of current computer vision techniques in bridging the gap between visual features and their semantic meaning. There is a clear need to cleverly combine the two in order to yield a better solution.*

Keywords: Preprocessing, Image Segmentation Feature Extraction, semantic scene classification.

1. INTRODUCTION

Semantic scene classification corresponds to the task of assigning a textual meaning to the image. This kind of operation is useful in large databases where images need to be categorized under different labels such as a forest, city etc. In order to perform annotation it is important that the objects of the image need to be identified thus object recognition is achieved through image segmentation techniques. Further attaching labels to these objects helps us to identify the semantic regions of the image. When the classification is done using the meaningful components (semantic), not only the object but also its surrounding play a important role in deciding the final labeling to the image. In our approach an image segmentation methodology is employed to identify the semantic regions of the image. This is followed by feature extraction of images which involves the analysis of color feature to train a non-linear multi-classifier to assign a semantic label to each segment. Multiple annotations thus obtained serve as semantic information corresponding to regions of the image. A second level classifier is then trained to annotate the image based on the semantic content of the image obtained from the first level. Since images are an essential component of the Web, this work focuses on an intelligent approach to semantic annotation of images. Our key contribution is the use of machine learning approach to perform image annotation using color at the semantic label. Further the images are provided with labels using the knowledge of the meaningful components present in the image.

2. PREPROCESSING:

Preprocessing involves performing initial operations on any form of data to make it more understandable and clear. Similarly an image at hand might not be a perfect distribution of pixels that are similar in intensity values and is highly likely that there would be pixels not belonging to that group. These are referred to as noise pixels and affect the image in its appearance which may also hinder in the annotation process. Noise may correspond to reflections, blurs, dark spots etc. Therefore prior to computational processes it is always useful to preprocess the image i.e. smoothing the image portions such as edges by certain masking techniques.

2.1 Image Segmentation:

An image consists of coherent portions which could be in terms of color, texture or any intensity values. Dividing images in terms of these semantic portions would make classification much easier hence image segmentation techniques need to be deployed. Image segmentation is the process of dividing an image into coherent components for the pur- pose of study. The goal of segmentation is to represent an image into something that is more meaningful and simpler to analyze. Basically image segmentation is the process of assigning to every pixel, a tag such that pixels with the same tag share certain characteristics. Post the image segmentation the image is partitioned into a set of regions where in each region, every pixel is related with another pixel by some characteristic or computed properties like color, texture, intensity etc. various approaches are used such as

- Thresholding
- Edge Based Segmentation
- Region Growing

2.2 Feature Extraction:

Color Color image segmentation is a process of extracting connected regions which satisfy some similarity criteria based on the feature derived from the spectral components. The color information is used to create histograms, information about the pixels that denote the edges or boundaries of the image. This feature is used in many applications, for example in effective scanning large amount of images and videos, they need to be compiled, sorted and stored, and color.is a very efficient method for retrieving the images and videos on the basis of their content so the color segmentation is broadly used in indexing and management of data.

Fig -1 : Image of a bird in different color spaces

3 . CLASSIFICATION

Classification methodology encompasses the techniques of supervised and unsupervised learning techniques. Unsupervised learning are not efficient for high-dimensional data for classification purposes Supervised techniques on the other hand use the concept of training the classifier using training set which contains n-dimensional input along with its class label. Once the trained classifier is obtained, class labels for test examples are predicted. In the problem of Semantic scene classification, there are two stages where we need to perform supervised learning classification.

There are various methods which are already used in the semantic scene classification. Some of them are listed below

- Bayesian classification
- Markov Model
- SVM(Support Vector Machines)

3.1 Three Layered Architecture

The first step toward building a classifier is to identify meaningful image categories which can be automatically identified by simple and efficient pattern recognition techniques. These classes match human perception, they allow organizing the database for effective browsing and retrieval. Natural scenes and sunset images are further grouped into the landscape class. City shots, monuments can be grouped into the city class. Finally, the miscellaneous, face, landscape, and city classes can be grouped into the top level class of vacation scenes. Given an input image, the classifier computes the class-conditional probabilities for each of the features. The three layered architecture is the one in which the inter planar relationships are used as a basis of making inferences about the semantic content of not yet annotated images. The visual plane organizes the visual information extracted from the raw image contents. The information in the visual plane is mapped in semantically meaningful keywords in the semantic plane. The keywords in the semantic plane, in turn map to specific slots in the domain specific schema(s), contained in the ontological plane. The three layers are as follows:

- The bottom layer is a K Dimensional space containing the visual feature vectors corresponding to each image in the annotated database. The size and nature of these feature vectors is purely dependent on the visual feature extraction algorithm that one wishes to implement.

- The middle layer contains all the keywords that have been used to annotate the images

- The top layer contains the schemas and ontology to which the keyword belong. It contains information about the relationships between the various tags in the RDF schema. This information in separate plane allows us to easily extend the RDF schema and automatically re-annotate the images in the annotated image database with the new keyword tag combinations used in the extended schema.

- RDF schema: It is a set of classes with certain properties using the RDF extensible knowledge representation language, providing basic elements for the description of ontology

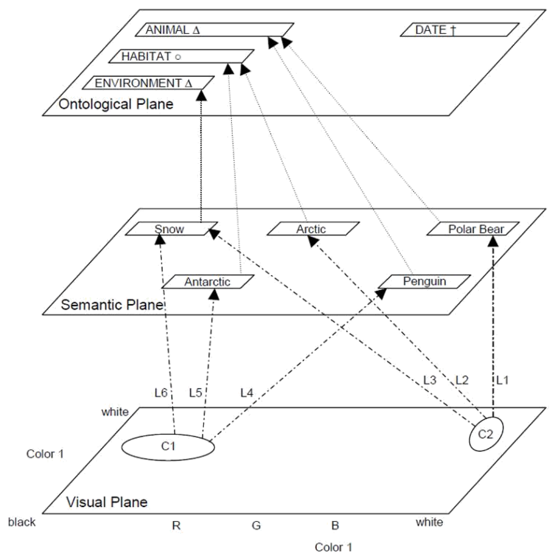

Fig-2 – Three Layered Architecture

4 . SEMANTIC SCENE CLASSIFICATION

We need images with detailed annotation. Detailed annotation is necessary for several reasons. First, labeled data is necessary to quantitatively measure performance of different methods. Second, labeled data is necessary because current segmentation and interest point techniques are not capable of discovering the outlines/shapes of many object categories, which are often small and indistinct in natural images. Some kind of focus of attention is necessary. The final reason to believe that large databases will be useful is by analogy with the speech and language communities, where history has shown that performance increases drastically when more labeled training data is made available.

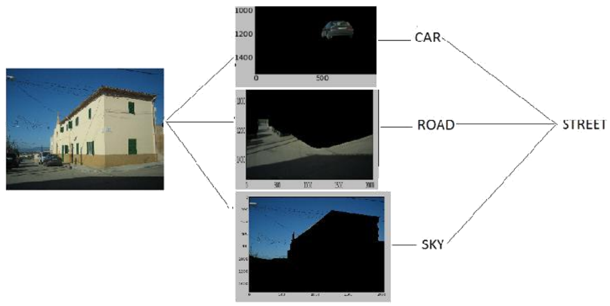

Fig-3 A Segmented Image with its label

4.1 Classification Of Segments Based On Color:

Once the segmentation is performed, an SVM classifier is trained. A predefined set of classes is defined such as street, car, building, people, sky, water, grass, animals etc. Hence at the end of the second step we have multiple annotations for an image.

4.2 Classification Of Images Based On Semantic Regions:

This more particularly means to establish a relationship among the semantic regions of the image.

For example,

CAR + STREET + BUILDING = CIT Y

This step is very essential since classification based on visual parameters can not be fully trusted. There are chances that the classification at this level may be faulty. Hence an approach of classification to also look into the surrounding objects of a segment plays an important role.

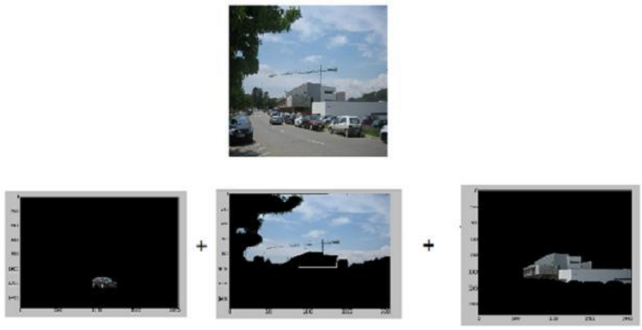

Fig -4 original image

Now a combination of car, sky and building can serve as a possible annotation of forest. Whereas if sky is visually misclassified as snow, then such an error can be avoided.

5 . RESULTS

We have used Scikit and OpenCV open source image processing tools for our implementation purpose.Some of the results of the segmentation of images with their labels. We took 5 major classes to categorize the images. Every class was trained using training set of 30 images.

	City	Marketplace	Scenery	Grassland	Village
City	75%	25%	0%	0%	0%
Marketplace	30%	70%	0%	0%	0%
Scenery	0%	0%	50%	30%	20%
Grassland	0%	0%	15%	60%	25%
Village	0%	0%	30%	30%	40%

Fig-5 Confusion matrix

Figure 5 is the confusion matrix obtained for the training examples

The accuracy of the classifier is calculated by adding the percentage of images which have been classified correctly to their respective labels

Accuracy (75 + 70 + 50 + 60 + 40)/5 = 59%

7. CONCLUSIONS

Semantic Image Annotation classifies the image segments obtained to finally obtain a label for the image. It is important to consider the surroundings of an object in order to firmly provide a label to the image. The visual parameter analysis of segments helps in judging the insights of an image better than performing low level feature extraction globally. Although presently we have just computed the color parameter for the segments, other important visual parameters such as texture, shape can be incorporated in order to obtain better representation of the image. Although incorporating more features in the feature vector increases its dimension and interfacing problems with the classifier, techniques of normalization can be successfully employed in order to deal with such problems.

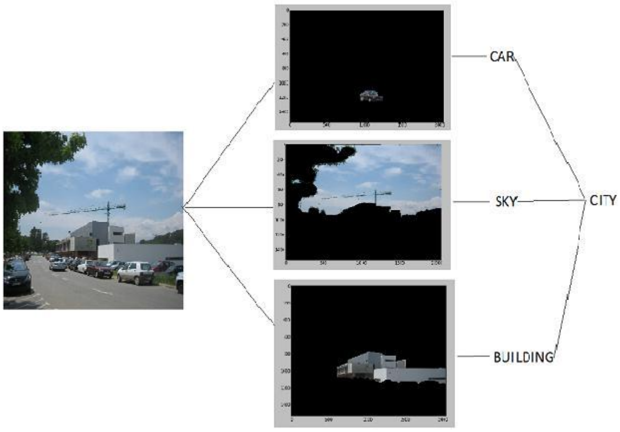

Fig -6 Segmentation and Labeling of City Image.

REFERENCES

[1] V. L. Jeon, Jiwoon and R. Manmatha, "Automatic image annotation and retrieval using crossmedia relevance models," 2003.

[2] e. a. Bloehdorn, Stephan, "Semantic annotation of images and videos for multime-dia analysis," The semantic web: research and applications. Springer Berlin Hei-delberge, 2005.

[3] T. S. H. Rui, Yong and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," Journal of visual communication and image represen-tation, 1999

[4] U. H.-G. Kressel, "Advances in kernel methods: Support vector learning," MIT Press, 1999.

[5]] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," . Data Mining and Knowledge Discovery, vol. 2, no. 2,, 1998.

[6] K. M. W. T. F. B. Russell, A. Torralba, "Labelme: a database and web-based tool for image.