Asian Journal of Information Technology 5 (4) : 408-412, 2006
© Medwell Online, 2006

Automatic Image Annotation Using Binary Decision SVM-AN Integration Framework

G. Suresh Kumar, R. Baskaran and A. Kannan
Department of Computer Science and Engineering,
Anna University, Chennai-600025, Tamil Nadu, India

Abstract: Automatic image annotation is a process of assigning semantic keywords to images and these annotations are used to retrieve the unlabeled images from large image collections by using semantic query texts. We are proposing the AIAS (Automatic Image Annotation System), which provides a effective mechanism for Annotating images using an active learning framework. The visual features like color and shape gives a great evidence for representing image blobs and its usage for image annotation has been explored in this study. The extracted image feature vectors (color, shape) and training keywords are used by machine learning techniques to automatically apply annotations to new images. During training phase the SVM (Support Vector machine) generation process learns the correlations between image features and training keywords. The trained model provides the mapping between training image data set and semantic keywords and then the trained decision model can be used for the automatic image annotation process.

Key words: Image features, color, shape, annotataion, support vector machines

INTRODUCTION

The choice of image processing and learning techniques determines the accuracy and efficiency of the image annotation system. Earlier, lots of proposals were made for annotating images with various combinations of segmentation and statistical modeling techniques. HuaMin Feng *et al.*,[1] proposed an approach for annotating online image collections. They proposed a novel approach to annotate images by disambiguating the region overlapping regions obtained by using two segmentation methods (JSEG, Blobworld)[2] and they used soft decision binary SVM model for active learning[1,3].

This framework proposes a SVM based learning technique which is designed to associate low level image features like shape and color with conceptual categories which can be used as query key words or annotations. Supervised machine learning and classification techniques are used for predicting labels for newly observed image data. For retrieving images from the large image collection based on their syntactical features, the Annotation Based Image Retrieval (ABIR) mechanism must provide strong association between the low level structural features like color or shape and their semantic descriptions. This generalized framework is trying to reduce this semantic gap.

The visual feature shape is less utilized in image retrieval system because the complexity in representing and processing shape data. Comparatively the shape feature provides effective retrieval mechanism in both CBIR (Content Based Image Retrieval) and ABIR[4]. The image pre-processing section of this AIAS consists of region based segmentation module and feature extraction module. The region labeling and region merging process are discussed in section 2. The active learning part of AIAS consists of the SVM classifier generation module and a shape thesaurus which is discussed in section 3.

IMAGE PRE-PROCESSING

This image pre-processing module performs image segmentation and feature extraction.

Segmentation: Is to split an image into disjoint regions while each region represents a meaningful object. The definition of object applied here is a group of image pixels that are specially connected (4-Neighborhood) and belong to the same region. Any object segmentation method like Edge-flow or Region-Grow can be used for segmentation. The region labeling and merging algorithm is used to segment the input image into meaningful regions.

Feature extraction: Is to extract the color and shape visual feature data, which is used to represent the objects by associating with known classes. The color feature extraction is done by color clustering algorithm. The section 4 describes the algorithms in detail.

Corresponding Author: Suresh Kumar, Department of Computer Science and Engineering, Anna University, Chennai-600025, Tamil Nadu, India

Shape: Identification, Extraction, Representation are done by the region labeling and merging object segmentation algorithm. There are two main steps involved in shape feature extraction; object segmentation and shape representation. Object segmentation is possibly the most challenging part of shape feature extraction. Once objects are segmented, their shape features can be represented and indexed. The object shapes are represented as regions.

Color: Is the most used visual feature for image retrieval due to the simplicity in similarity measurement and less complexity of its extraction. And the shape and color of the object can best describe it. One of the desirable characteristics of an appropriate color space for image retrieval is its uniformity. In this AIAS system the RGB (Red, Green and Blue) color model is used. Color Histogram is a most general method to represent the proportion of number of pixels of each color for the given image. To minimize the number of color bins the image is pre-quantized.

SVM CLASSIFIER FOR ACTIVE LEARNING

The SVM model is designed for binary classification which separates a set of training vectors which belong to two different classes, (x1, y1), (x2, y2),…, (xm, ym), where $x_i \in R^d$ denotes vectors in a d-dimensional feature space and $y_i \in \{1, +1\}$ is a class label. During the SVM model generation, the low-level input feature vectors like color and shape for image retrieval, are mapped into a new higher dimensional feature space denoted as $F: R^d \rightarrow H^f$ where $d<f$. Then, an optimal separating hyper-plane in the new feature space is constructed by a kernel function, $K(x_i, x_j)$. The most widely used kernel functions is the Gaussian Radial Basis Function (RBF) kernel functions which has the form,

$$K_{gaussian}(x_i - x_j) = e^{|x_i - x_j|^2/2\sigma}$$

Where σ is Gaussian sigma.

$$\sigma = 1/(1+e)^{ac+\beta}$$

Where each concept corresponds to one group of α and β

The mapping function separates the object vectors present in the input space by a hyper-plane. All vectors lying on one side of the hyper-plane are labeled as 1 and all vectors lying on another side are labeled as +1. The training instances that lie closest to the hyper-plane in the transformed space are called support vectors. The number of these support vectors is usually small compared to the size of the training set and they determine the margin of the hyper-plane and thus the decision surface. In order to produce good generalization, the SVM maximizes the margin of the hyper-plane and diminishes the number of support vectors[3].

IMPLEMENTATION

This section describes a prototype development of the proposed system AIAS and the following subsection describes the system architecture and implementation techniques.

The proposed frame work AIAS consists of four compartments and two phases as shown in Fig. 1.

- The input section provides the unlabeled image collections as the input for training and testing modules.
- The image pre-processing module segment the image into regions by using region labeling and merging algorithm and the color clustering algorithm is used to extract the color feature.
- The active learning module perform training data extraction from the feature vectors and the SVM classifier is used to build the association between the training set and concepts (keywords). The concepts are obtained from the lexicon shape thesaurus.
- The output section represents the set of keywords which semantically related to the input image, used as annotations.

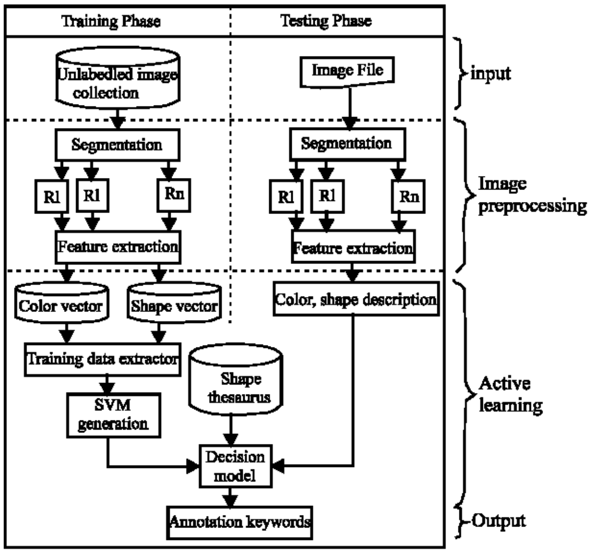


Fig.1: System architecture

The training and testing modules shares a common shape thesaurus which includes a set of collected concepts or lexicons.

Segmentation: This AIAS system uses object segmentation using the region split and merge algorithm, which is divided into two sub algorithms:

- Region labeling algorithm
- Region merging algorithm

Region labeling algorithm: *num_reg;* denote the number of total regions.
T(i,j) represents the region type of the pixel (i,j).
L(i, j) denotes the region label of the pixel

- let *num_reg* = 0;
- scan the image from left to right, top to bottom:

previous region labels should be modified by executing subroutine adjust(i,j).

- Calculate the information for each region(region area).
- Call the subroutine adjust(i,j)

Region merging based on edge merit: The algorithm begins with the smallest regions and terminates when there are only big regions and those that cannot be merged remaining.
When there are tiny regions left in the image do the following:-

- Get the smallest region, calculate the length of its boundaries, find out the longest boundary.
- Merge the two regions separated by that boundary, update the region labels and region information.

When there are small and medium size regions not processed left in the image merge the regions.
Merging terminates.

Color feature extraction
Color clustering algorithm

- Compute the RGB color histogram (number of pixels having the same color)
- Find all the color peaks from the histogram
- A peak corresponds to a color cluster, for each cluster note the RGB values and population from the histogram.
- Sort the peaks in descending order based on the cluster population
- Determine the number of peaks which do not have a very small population.

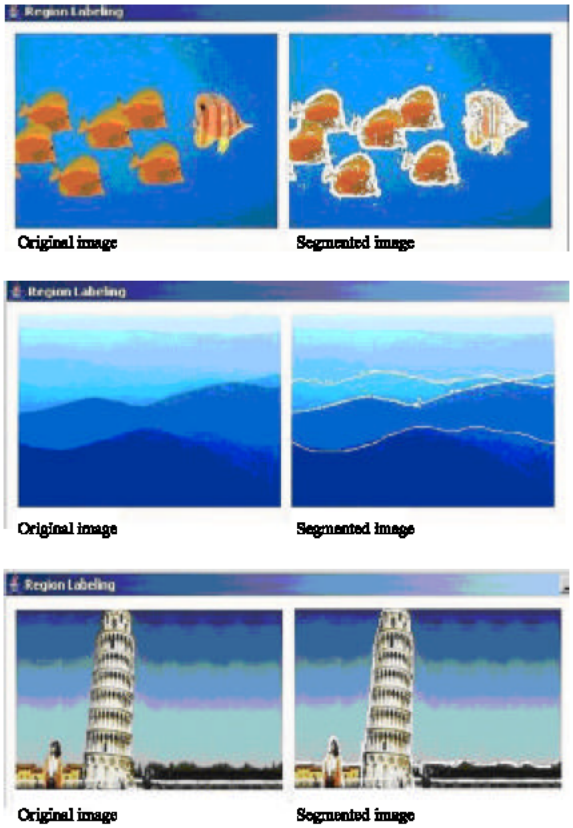


Fig. 2: Sample segmented images

- If the number of peaks found in the step 5 is less then in step 3, merge the very small clusters to their nearest color clusters. The nearest cluster is computed based on the color distance metric in (L*u*v*) space.
- The representative color for this merged cluster is the weighted mean of the two original clusters.
- For each pixel compute the color distance to the different clusters. Assign the pixel to the cluster for which color distance is maximum. Thus every pixel gets assigned to one of the clusters.

The Fig. 2 shows the regions generated by the above region labeling and merging algorithm.

Active Learning using SVM Classifiers

Training data extraction: The training data set is formed by a data extraction module as the combination of both color and shape. The Fig. 3 shows the data extraction process.

The Fig. 4 present the learning procedure to generate SVM classifiers and the automatic annotation of unlabeled images. The training set contains low-level feature vectors (color and shape) and their associated class labels are given.

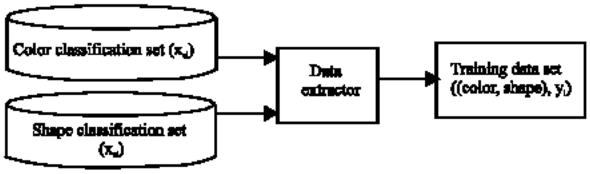

Color classification set (x_c)
Shape classification set (x_s)
Data extractor
Training data set ((color, shape), y_i)

Fig. 3: Training data extraction

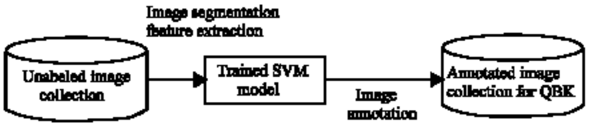
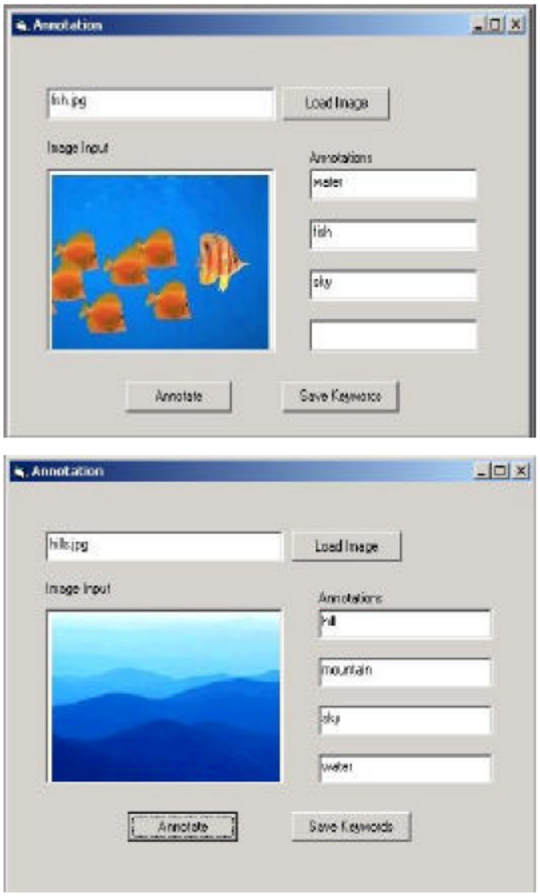


Image segmentation feature extraction
Unabeled image collection
Trained SVM model
Image annotation
Annotated image collection for QBK

Fig. 4: Test or annotation phase



fish.jpg

water

fish

sky

hills.jpg

hill

mountain

sky

water

Fig. 5: Sample annotation results

Table 1: Number of images tested in generic and specific category

Number of images	25	50	75	100	200	500
Generic class	25	45	50	50	100	300
Specific class	0	5	25	50	100	200

The color training set, $P_c = \{(x_{c1}, y_{c1}), (x_{c2}, y_{c2}), \ldots, (x_{cm}, y_{cm})\}$ where each y_{ci} is a label of the color names associated with each x_{ci} and shape training set, $P_s = \{(x_{s1}, y_{s1}), (x_{s2}, y_{s2}), \ldots, (x_{sm}, y_{sm})\}$ where each y_{si} is a label of the shape names associated with each x_{si} are used to train the color and shape classifiers. The shape descriptions are obtained from the shape thesaurus.

RESULTS

This work is tested with the dataset consists of 2,000 images from NOAA Photo Library and Corel image CDs. The data set includes both generic and specific scenes and objects, 750 images from each category was taken for training the classifier and the trained system was tested with 500 images. The Table 1 shows the sample data of the number of images considered in each category with different training data sizes. The lexicon consists of 50 set of keywords in both generic and specific category.

The following Fig. 5 shows the sample annotation results for specific objects like fish, pisa tower and generic objects like hills.

CONCLUSIONS

The current automatic annotation systems are statistical model based and the complexity is more. This AIAS proposes a binary decision SVM based framework for effectively annotating image collections. And the proposed system was developed and tested with sufficient data. The use of Region labeling and merging algorithm simplifies complexities involved in object based segmentations, further the low level visual features like color and shape provides good results. The 2 sigma RBF kernel and *m* class SVM provides better classification accuracy. The use of shape thesaurus is well explored for bridging the semantic gap between the structural features and syntactical features. This AIAS is a general framework in which different segmentation methods can be applied to improve the accuracy. In order to improve the annotation accuracy some future enhancements can be done in this work by alternating the segmentation methods and by utilizing more SVM models for feature extraction.

REFERENCES

1. HuaMin Feng and Tat-Seng Chua, 2004. A Learning-based Approach for Annotating Large On-Line Image Collection. Proceedings of the 10th International Multimedia Modeling Conference (MMM'04) IEEE.
2. Carson, C., *et al*., 1999. BlobWorld: A system for region-based image indexing and retrieval. Intl. Conf. Visual Inform. Sys.

3. Jeon, J., V. Lavrenko and R. Manmatha, 2003. Automatic image annotation and retrieval using cross-media relevance models. Proceedings of the ACM. SIGIR. Conf. Res. Develop. Inform. Retrieval, pp: 119-126.
4. Metzler, D. and R. Manmatha, 2004. An inference network approach to image retrieval. Proc. Intl. Conf. Image and Video Retrieval, pp: 42-50.
3. Barnard, K. and D.A. Forsyth, 2001. Learning the semantics of words and pictures. Proc. Intl. Conf. Comput. Vision, pp: 408-415.
4. Wang, J.Z. and J. Li, 2002. Learning-based linguistic indexing of pictures with 2-D MHMMs. Proc. ACM Multimedia, pp: 436-445.
5. Li, J. and J.Z. Wang, 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. IEEE Trans. Pattern Analysis and Machine Intelligence, pp: 1075-1088.
6. Cusano, C., G. Ciocca and R. Scettini, 2004. Image annotation using SVM. Proceedings of Internet Imaging IV.
7. Jeon, J. and R. Manmatha, 2004. Using maximum entropy for automatic image annotation. Intl. Conf. on Image and Video Retrieval (CIVR 2004), pp: 24-32.
9. Lavrenko, V., R. Manmatha and J. Jeon, 2003. A model for learning the semantics of pictures. Proceedings of the 16th Conference on Advances in Neural Information Processing Systems NIPS.
10. Jin, R., J.Y. Chai and L. Si, 2004. Effective automatic image annotation via A coherent language model and active learning. Proceedings of MM'04.
12. Feng, S., R. Manmatha and V. Lavrenko, 2004. Multiple bernoulli relevance models for image and video annotation. IEEE Conference on Computer Vision and Pattern Recognition, pp: 1002-1009.
13. Fan, J., Y. Gao, H. Luo and G. Xu, 2004. Automatic image annotation by using concept-sensitive salient objects for image content representation. Proc. 27th Ann. Intl. Conf. Res. Develop. Inform. retrieval, pp: 361-368.
14. Yavlinsky, A., E. Schofield and S. Rüger, 2005. Automated image annotation using global features and robust nonparametric density estimation. Intl. Conf. Image and Video Retrieval (CIVR, Singapore).