*International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*
*Volume 3 Issue 6, June 2014*

# Automatic Annotation of Images using Label Transfer

Ebsil Sherly G.T, Mrs.S.Vanitha Sivagami

***Abstract*— Automatically assigning relevant keywords for an image provides easier retrieval of images and understanding of large collection of image data. In this paper, a new image annotation algorithm was introduced which is used for image retrieval process. The proposed system is based on the similarity measure between images which is referred as compactness. The compactness based matching model shows how close the test image features lie to the training image cluster centers. The usage of low level features maintains low running time and gives more accurate detail information. At annotation time similar images are retrieved from the database using similarity measure and the labels from these images are used to form the annotation. This method makes use of different label transfer techniques to transfer the labels from the retrieved images. This annotation process requires only simple training algorithm and can efficiently work with different low level image features.**

***Index Terms* – automatic image annotation, compactness, image retrieval, object recognition, scene analysis**

## I. INTRODUCTION

Automated image annotation refers to the process of assigning few relevant keywords for an image from the dictionary and this annotation represents the visual content of the image. Nowadays a large amount of images are stored in database either locally in the computer or widely in internet. As the storage content increases it is difficult for one to retrieve the necessary image. Thus saving image in annotation form provide efficient access and more useful to organize them.

Image annotation task mainly suffers from to different problems: *semantic gap and weak labeling.* The first problems deals with the difficulties in extraction of semantic information only by usage of low level visual features such as color, edges and texture that can be easily retrieved from the image but difficult to organize them. The second problem arises due to the lack of correspondence between the labels and image regions in the training data. In this process, for some available objects in the image, the labels are not present in the database and another effect is that the labels are not linked to specific regions instead they are given for the entire image. Due to this it is infeasible to determine exact region to which the label refers.

The simplest technique is to treat the annotation scheme as the problem of image retrieval. In this scenario, for the given test image find the nearest neighbor from the training set and assign all the keywords of the nearest image to test image. For extracting the features, the local features are more efficient than global features because while using the global features the fine details may lose, also low level feature reduce the computation time.

The main contribution of this system lies definition of compactness as a similarity measure between two images, which is used to retrieve the related images and also provides formalism for defining label transfer techniques based on weight function.

## II. RELATED WORK

Recently a large number of methods were proposed by the researchers in various domains such as object recognition and scene recognition for automatically annotation the objects in the scene image. Basically these methods are categorized as keyword based annotation and retrieval based annotation.

In the Keyword Based Annotation Method, the models which provides the representation for each keyword (label) was created and at annotation time the closely related labels are obtained from the models. The limitation of this approach is that the annotation with a keyword is based on only single decision and not based on multiple possibilities.

G. Carneiro. [2] proposed the Gaussian Mixture Models which is used to characterize the label as a multimodal Gaussian distribution defined on the feature space. Another model defined by F.F.Li [5] was Bayesian Hierarchical Model, the labels are inferred by patch based representation of the input image using Bayesian inference rule. The Markov Model [6] was created to learn spatial and multi-resolution relationship between image features, which increases the performance of annotation. All these works are based on the classification of global image features which are in the form of bag-of-words features for scene recognition. A histogram is

***Manuscript received May, 2014*.**
***Ebsil Sherly G.T***, *PG Scholar,Computer Science and Engineering, Mepco Schlenk Engineering College, Sivakasi, Tamilnadu, India.*
*S.Vanitha Sivagami,Assistant Professor, Computer Science and Engineering, Mepco Schlenk Engineering College.Sivakasi, Tamilnadu, India.*


ISSN: 2278 – 1323 

constructed and SVM classifier is used for training the features. The main drawback of keyword based annotation was some images are not labeled with the corresponding keywords.

In Retrieval Based Annotation method, the similarity measure between the two images is determined to retrieve the similar images. Also the labels from these images are transferred to produce annotated image. The global image features such as color descriptors and texture descriptors are used to represent the image. The similarity is determined by taking the inverse of the distance between the extracted global image features. The raw pixels from the global image feature vector and a distance function is used as a similarity measure. Finally the Labels are transferred from the best matching images.

## III. SIMILARITY MEASURE

Commonly the annotation methods are constructed using the global features, because it enables a fast computation but the information about the image may lose. To efficiently perform retrieval of similar images, compactness is used in which the local features are used for similarity matching. From the extracted training image features, a set of relevant descriptor centers are obtained using K-means clustering algorithm.

### A. Compactness Matching

Compactness defines the measure which shows how close the test image features lies to the training image feature centers. It is defined by using

$$c(X,C) = \frac{1}{|X|}\sum_{i=1}^{|X|}\min_{j} d(x_i, c_j) \qquad (1)$$

Where X denotes the set of features extracted from the test image, C represents the set of centers extracted from the training image and $j \in 1,K$. The distance of each point from the X to the closest point from C can measured by $L^p$ norm instead of distance function.

$$c(X,C) = \frac{1}{|X|}\sum_{i=1}^{|X|}\min\left\|x_i - c_j\right\|_p^p \qquad (2)$$

To measure the similarity between two images by using compactness, first the features are extracted which are represented as X and Y respectively. Then, from one of image consider the second image Y, is characterized by the cluster centers of the feature and denoted as C. finally the similarity is calculated as the compactness of the features from the first image to the cluster centers of the second image c(X, C).

## IV. LABEL TRANSFER

The Keywords (Labels) from the database of the retrieved similar images are transferred to form annotation. This is done by the construction of histogram h, in which each bins corresponds to a concept from L, where L= $\{l_1,l_2,\ldots,l_n\}$ the vocabulary containing the semantic labels. Using compactness the best N matching images are retrieved. After construction of histogram, the labels with the highest corresponding bin value will be taken for annotation. Different label transferring schemes are used based on the value incremented in histogram bins. The matching images are represented a $\mu$, which returns the ordered list based on the measured compactness. The labels from the trained image increments the histogram bin by the weight function defined as:

$$h = \sum_{n=1}^{N}\sum_{l \in L_n}\omega(n)\delta_l$$

Where $L_n$ represents the labels from the $n^{th}$ match of the training image set and $\delta$ is a vector in which the position of currently associated label is one and all other positions are zero. Various types of label transfer techniques can be achieved by changing the value of weight function. Some of the label transferring techniques used is as follows:

### A. Equal Contribution Transfer

This is the simplest label transfer weighing method, in which all the match image from the training set contributes evenly to the histogram. It is denoted by,

$$\omega_0(n) = 1 \qquad \forall n=1,N \qquad (4)$$

The benefit is that it can eliminate the labels which are appear only in few matches, thus it doesn't give important to best matches.

### B. Rank based Transfer

Here the weighing the best matching image is done based on the rank associated. The highly matching image has more weight and it is decreases exponentially based on the rank.

The weight function defined for rank based method is as follows:

$$\omega_a(n) = 2^{a\left(1-\frac{k-1}{N-1}\right)}, \qquad \forall n=1,N \qquad (5)$$

In this method, the rank is allotted based on the position not based on the distance of the test image. If



ISSN: 2278 – 1323 

all the matches are very close, then this rank method is not relevant.

*C. JEC Type Transfer*

The JEC type transfer is used only for the best matching labels and the remaining labels are transferred based on their appearance frequencies in the training set. The weight function corresponds to this technique is defined as:

$$\omega_J(1) = 10, \omega_J(n) = 1, \quad \forall n=1,N \tag{6}$$

In the final step the histogram is updated in order to take the label frequencies.

*D. Distance based Transfer*

The distance based transfer method is based on the compactness value calculated for each of the corresponding matching image. The weight function is defined as,

$$\omega_d(n) = 2^{b\left(1-c_n/c_1\right)}, \ \forall n=1,N \tag{7}$$

This method is used only when the compactness measured values are relevant and the first match will always get the same weight.

*Label Transfer Techniques*

Using the calculated weight function, the histogram was constructed and the labels from the best match images are transferred to the test image to form annotation. The labels are transferred based on the single feature, multiple features and also appearance frequency. Based on the single features the histogram is constructed using the eqn (3).

*A. Based on Multiple Features*

The construction of histogram using weight function can be done by training multiple features. Save the histogram and repeat the process for all the feature types by iteration. The histogram was constructed using the formula defined as,

$$h = \sum_m \eta_m \sum_n \sum_{l \in L_{m,n}} \omega(n)\delta_l \tag{8}$$

Where m denotes the feature type, $\eta_m$ refers the weight assigned for the corresponding feature and $L_{m,n}$ is the set of labels from the n-th match for the feature type m.

*B. Based on Appearance Frequency*

In this method, the frequency of each label from the training set is used to update the histogram. Here the histogram was constructed using one of the weight function calculated described in the above techniques.

The histogram based on frequency was constructed using the formula:

$$h = \sum_m \eta_m \sum_n \sum_{l \in L_{m,n}} \omega(n)\delta_l + \varphi \sum_{l \in \bigcup L_{m,n}} f_l \delta_l \tag{9}$$

This type of label transferring method is used in rare cases such as when the histogram value for two concepts are same.

## V. ALGORITHM USED

**Algorithm 1:** Training Images

```
Algorithm TrainingImage(Input images)
input : set of input images I_n
output : cluster centers C_n
{
   for all training images I_n do
   {
      X_n =extractLocalFeatures(I_n )
      C_n = K-means(X_n )
   }
}
```

The algorithm describes the steps involved for the training process. The local features are extracted from the images and are clustered using K-means clustering. The clusters centers are saved and are further used for similarity measurement. The number of clusters used for the k-means clustering was 20.

**Algorithm 2**: Testing Process

```
Algorithm TestingImage(Image centers, Test Images)
Input : Training image centers C_n , Test Images T_k
Output : Annotations for test images L_k
{
   for all test images T_k do
   {
      Y_k =extractLocalFeatures(T_k )
      B_k = selectSamples(Y_k)
      for all training image In do
      {
```

$D_{k,n}$= findCompactness ($B_k$, $C_n$ )

}

$M_k$ = findMatches($D_k$)

$BM_k$= selectNBestMatches($M_k$)

$H_k$= constructHistogram($BM_k$)

$L_k$=assignAnnotations($H_k$)

}

}

In the testing process the features are extracted from the test image and calculate the compactness using these features and the centers saved during training process. The best matching images are retrieved by similarity measure, and then select the related matching images. Construct the histogram and perform annotation by transferring the labels from the trained image to test image.

## VI. IMPLEMENTATION

The images from the *IAPR-TC12* dataset, which consists of 20,000 still natural images; the database comprise of 17825 training images and 1980 test images with 291 labels. The color and texture features are extracted from these images.

### A. Color Feature

The simple color feature was obtained by extracting the features from three color spaces such as RGB, Lab, and HSV. Initially the image dimension is reduced to 64, which reduces the execution time. The extracted feature vector dimension is 9. This color descriptor is used for their simplicity.

### B. Texture Feature

The Texture features are obtained by using Gabor filter and Haar wavelet. In this experiment the Histogram of Oriented Gradients (HOG) features are extracted from the input image. The number of angle bins used here is 12, and thereby the feature dimension is 12 for gray image and 36 for color image.

More number of features can be extracted to increase the accuracy and efficiency. Some of the other features such as WLD descriptor which gives the excitation and gradient orientation values at each pixel and others like DCT features, SIFT features, Law texture descriptors are also used.

TABLE I

LOCAL FEATURE DESCRIPTOR

| Descriptor | Type | Dimension |
|---|---|---|
| RGB+Lab+HSV | Color | 9 |
| HOG | Texture | 12 |

### C. Clustering the Feature Vectors

Determining the similarity between two images based on local descriptors is infeasible. For each local descriptor, a set of relevant descriptor centers are obtained by using k-means clustering algorithm. These algorithms provide better performance and are used for its simplicity.

### D. *Retrieval of Similar Images*

After extracting the color and texture features from the input image, find the closest data points using compactness measure mentioned in section B. The distance from the trained feature cluster centers and the test features are calculated by Lp norm distance function. The images with minimum compactness are close to the test image.

### E. Label Transfer

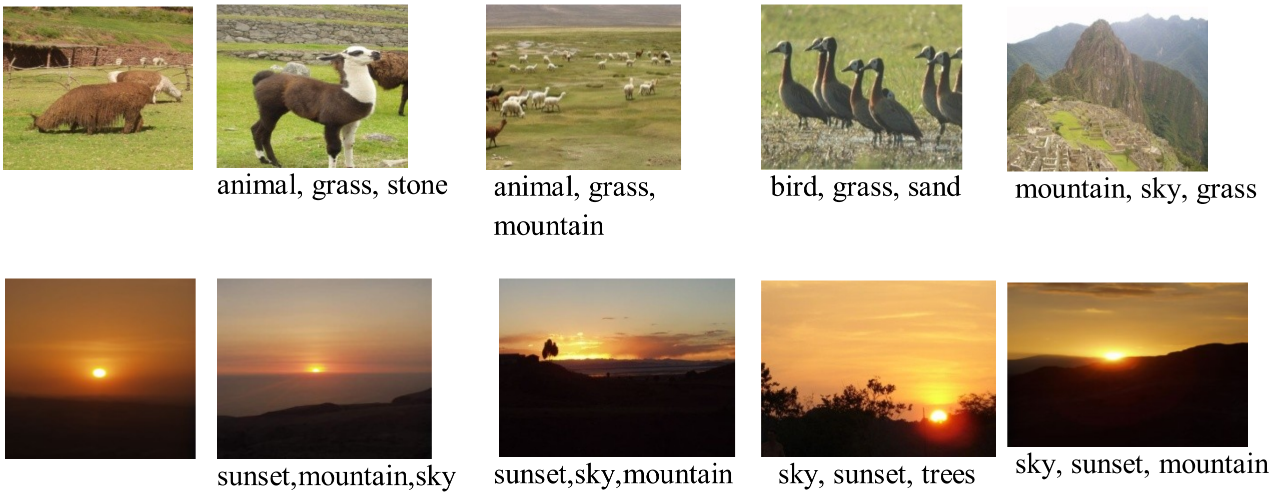
The Labels from the similar images are retrieved and a histogram for these labels is constructed. The labels with highest corresponding bin value in the constructed histogram will be chosen to form the final annotation. The labels are transferred based on the above mentioned techniques in section IV. The *JEC* type transfer gives better result.

## VII. CONCLUSION AND FUTURE WORK

In this paper, usage of compactness for image matching makes the annotation process efficient. The advantages of this approach are simple and fast; efficiently combine different feature types and also does not need any segmentation process. This system can improve by making use of different combination of feature types to obtain optimal results. The labels from the best matching images are to be transferred in the next process. The future work lies in this paper, comparing the system with various label transfer techniques for finding the accuracy and performance.



ISSN: 2278 – 1323 




(a) (b) (c) (d) (e)

Fig 1: (a) Test image and (b), (c), (d), (e) retrieved best matches

## ACKNOWLEDGMENT

The authors wish to express their heartfelt thanks and gratitude to the Department of Computer Science and Engineering of Mepco Schlenk Engineering College, Sivakasi for providing good support and encouragement for this work. The authors also thank their principal and management for providing the necessary facilities to carry out this work.

## REFERENCES

[1] Robert Varga and Sergiu Nedevschi, *Member, IEEE,* "Label Transfer by Measuring Compactness" in *Ieee Transactions On Image Processing*, Vol. 22, No. 12, December 2013 4711.

[2] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "An Efficient Mining Model For Enhancing Text Classification Using k-NN" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3, Issue 2, February 2014.

[3] J. Li and J. Z. Wang, "Automatic inguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 25, no. 9, pp. 1075–1088, Sep. 2003.

[4] A. Llorente, R. Manmatha, and S. M. Rüger, "Image retrieval using Markov random fields and global image features," in *Proc. CIVR, 2010,* pp. 243–250.

[5] F. F. Li and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. CVPR*, 2005, pp. 524–531.

[6] J. Verbeek, M. Guillaumin, T. Mensink, and C. Schmid, "Image annotation with tagprop on the MIRFLICKR set," in *Proc. 11th ACM Int.Conf. Multimedia Inf. Retr., 2010, pp. 537–546.*

[7] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen,X. Chen, and W. Gao, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal.Mach. Intell., vol. 32, no. 9, pp. 1705–1720, Sep. 2010.*

[8] J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 6, pp. 985–1002, Jun. 2008.

[9] J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 9, pp. 1075–1088, Sep. 2003.

Ebsil Sherly G.T received her B.E degree in Computer Science and Engineering from Noorul Islam College of Engineering, Tamilnadu, India in 2011. She is currently pursuing M.E in Computer Science and Engineering in Mepco Schlenk Engineering College, Sivakasi, India. Her research interests are in Image Processing and Image annotation.