

# ACIR with Image Annotation by Label Correlation and Interactive Boosting Algorithm to Improve the Retrieval Accuracy

Miss. Pallavi H. Dhole

Department of Computer Science and Engg.  
P.R.Patil College of Engg. & Tech. Amravati  
Amravati University, India

Mr. Ajay B. Gadicha

Department of Computer Science & Engg.  
P.R.Patil College of Engg. & Tech. Amravati  
Amravati University, India

**Abstract**— Now a days the popularity of digital images increases due to the improving digital imaging technologies and convenient availability facilitated by the internet. So to find user intended images from large image dataset is very difficult. It is an important issue that to categorized and retrieve the image accurately. As a way to facilitate image categorization and retrieval, automatic image annotation has received much research attention. Considering that there are a great number of unlabelled images available, it is beneficial to develop an effective mechanism to leverage unlabelled images for large-scale image annotation. Meanwhile, a single image is usually associated with multiple labels, which are inherently correlated to each other improves the performance significantly. A straightforward method of image annotation is to decompose the problem into multiple independent single-label problems, but this ignores the underlying correlations among different labels. In this topic, to bridge the gap between high level semantic concept and low level image features well, we propose a new annotation & content based image retrieval (ACIR) system with image annotation by integrating label correlation and visual similarity mining into a joint framework to annotate the unlabelled images and interactive boosting algorithm to enhance the retrieval accuracy.

**Keywords**— ACIR, Image Annotation, label correlation, visual features, interactive boosting algorithm.

## I. INTRODUCTION

With the development of computer network and storage technologies, we have witnessed explosive growth of web images. There are large amounts of digital images generated, shared, and accessed on different websites, e.g., Flickr. With the popularity of digital cameras, we are able to create personal photos easily. To categorize these images manually is now become the most tedious and totally not human friendly. It becomes an important challenge to organize these resources effectively, categorized and retrieves accurately. The growing number of web image requires an effective retrieval and browsing mechanism in either a content- or keyword-based manner [8], [10].

Due to the proliferation of image data in digital form, CBIR has become a prominent research topic. From historical perspective, one shall notice that the earlier image retrieval systems are rather text-based search since the images are required to be annotated and indexed accordingly. However, with the substantial increase of the size of images as well as the size of image database, the task of user-based annotation

becomes very cumbersome, and, at some extent, subjective and, thereby, incomplete as the text often fails to convey the rich structure of the images. The extracted visual information is natural and objective, but completely ignores the role of human knowledge in the interpretation process. The bottleneck to the efficiency of Content-based approaches is the semantic gap between the high-level image interpretations of the users and the low-level image features stored in the database for indexing and querying. Among others, automatic image annotation technology, which associates images with labels or tags, has received much research interest [10]. Automatic image annotation enables conversion of image retrieval into text matching. Indexing and retrieval of text documents are faster and usually more accurate than that of raw multimedia data. Image annotation thus brings several benefits in image retrieval, such as high efficiency and accuracy.

Image annotation is essentially a classification problem. In field of multimedia and computer vision, many researchers have proposed a variety of machine learning and data mining algorithms for automatic image annotation recently. These works have shown promising achievements in overcoming the well-known semantic gap by applying machine learning algorithms to image annotation.

Usually, a single image may be associated with multiple labels, and the image annotation is a typical multi-label classification problem. A straightforward way to deal with this problem is to decompose it into several binary classification problems, with one for each label. However, the limitation is that this type of approach does not consider correlations among different class labels. Thus, another method to reduce the required labor in image labelling is to utilize the label correlation for image annotation.

In this topic, we propose a new ACIR system with Image annotation for automatic web and personal image labelling by integrating shared structure label correlation and visual features into a joint framework. We annotate the unlabelled images with the help of label images by label correlation and by using its visual features. Compared with other existing system, our system simultaneously utilizes the information in the unlabelled data and the label correlation information for improving the retrieval efficiency and accuracy.

## II. RELATED WORK

In [2], proposed a semantically shared annotation technique to enhance retrieval accuracy that combines semantic annotation, textual features and visual features which are collected from different web pages that share visually similar images which help to widely increase the amount of annotation not from single web page as traditional approaches. This system use visual ontology, which is a concept hierarchy, is built according to the set of annotations.

J. Jeon, V. Lavrenko and R. Manmatha, proposed relevance model for automatic approach to annotating and retrieving images based on a training set of images. We assume that regions in an image can be described using a small vocabulary of blobs. Blobs are generated from image features using clustering. Given a training set of images with annotations, they show that probabilistic models allow us to predict the probability of generating a word given the blobs in an image. This may be used to automatically annotate and retrieve images given a word as a query [3]. Wu et al. proposed a probabilistic distance metric learning scheme for retrieval-based image annotation [4]. Because web images with user-generated tags are comparatively easy to obtain, image tagging has the advantage that less human labor is required.

As pointed out in [5], it is helpful to utilize unlabeled data in many applications. To relieve the required human labor in image labeling, an effective way is to leverage the unlabeled data to learn an accurate classifier for image annotation and some other applications, such as multimedia retrieval [7]. For example, Cai et al. proposed semi-supervised discriminant analysis (SDA) [9] for image annotation and retrieval. Some other related work can also be found in [5]. These algorithms achieve promising performance in overcoming semantic gap by exploiting both labeled and unlabeled data during the training stage, particularly when there are only a small amount of labeled images.

In [6], an efficient hashing scheme is proposed for image tagging. The system in [6] first searches for semantically and visually similar images from the web and then annotates images by mining the search results. In [11], proposed a semi-supervised learning framework to solve the automatic image annotation problem. A new multi-label correlated Green's function approach was proposed to propagate image labels over a graph with considering the correlations among labels. An adaptive decision method is proposed to deal with the unbalanced distribution of the training data.

In the field of multimedia, Z. Zha, T. Mei, J. Wang, Z. Wang, and X.-S. Hua, proposed "the graph-based transductive algorithms have been applied to many applications, such as content-based multimedia retrieval, image annotation". These research efforts have shown that transductive algorithms are effective in overcoming the well-known semantic gap [13]. Fergus et al. pointed out that the predicted label matrix can be specified by the eigenvectors of  $L$ , and they proposed a fast algorithm to compute the approximated result. Although the linear system can be solved in linear time approximately, a limitation of the existing transductive classification algorithms is that they cannot annotate the images that are unseen during the training phase. Each time we annotate one image outside

the training set, we must rerun the whole training procedure. In other words, the transductive classification algorithms are more suitable for static image databases. In contrast, both web and personal image databases are dynamic, i.e., the number of web and personal images keeps increasing. It is infeasible to rerun the training algorithm to annotate each new image added into the database. Therefore, transductive classification algorithms are not applicable to web and personal image annotation.

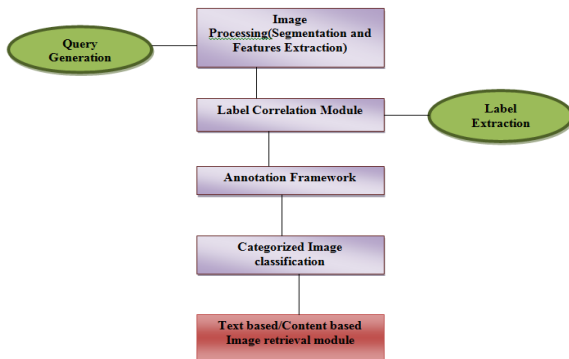
**Semisupervised Inductive Learning:** As discussed previously, although transductive classification is comparatively effective for image annotation, it is not suitable for large-scale image databases whose size grows dynamically. On the one hand, manually annotating many training data is expensive and time consuming. On the other hand, insufficient labeled training data may induce overfitting. To relieve the tedious work in supervised learning, some researchers suggest improving the learning performance by leveraging unlabeled data, e.g., [14]. Compared with traditional supervised learning algorithms, such as linear discriminant analysis (LDA), this type of algorithm is able to reduce the required number of labeled data during the training stage. Compared with transductive learning, the inductive algorithm is able to predict the labels of unseen data, which are outside the training set. It is therefore more suitable to apply the algorithm to dynamic image database annotation. However, in most of existing semisupervised learning algorithms such as [14], a linear constraint is imposed on the image labels, whereas data distribution of multimedia data is demonstrated to be more of a nonlinear manifold. It is beneficial to make the classifier more flexible.

Yi Yang, Fei Wu, Feiping Nie, Heng Tao Shen, Yueting Zhuang, and Alexander G. Hauptmann, proposed a framework for web and personal image annotation. They proposed to simultaneously mine label correlations and visual similarities by integrating SSL and relaxed visual graph embedding into a joint framework, which usually learn the classifiers by minimizing the regularized empirical error with respect to a graph embedded label prediction matrix. They shows problem is non-convex, a global optimal solution can be obtained by performing generalized eigen-decomposition [1].

## III. PROPOSED WORK & ARCHITECTURE

Currently, most image retrieval systems use either purely visual features or textual metadata associated with images. Recent research in web image retrieval suggested a combine use existing textual context and visual features can provide better web image retrieval result. To overcome drawbacks and improve the performance without sacrificing the efficiency, the new web image retrieval systems should pay a great attention for these features. This paper introduces a ACIR system with a joint framework for image annotation by integrating label correlation and visual features mining which leads to free of the previously mentioned limitations. In this proposed system a multi-label classifier is trained by simultaneously uncovering the shared structure common to different labels for image annotation. The proposed system is capable to utilized both labelled and unlabelled data for image annotation and categorized images accurately.

In this proposed ACIR system consists mainly two phases which are very important components. Image Pre-processing phase and image retrieval phase. The first phase is responsible for data, image collection, image annotation and image categorization. In second retrieval phase we need to retrieve the user intended images by using text based query or by using content based query. The next following sections explain in details each phase and its important components. Figure 1 shows basic architecture of ACIR system.



“Figure 1. Architecture of ACIR System”

#### A. Image Pre-processing phase

The image pre-processing phase has the following main modules as shown in figure 1. These modules are: (1) Image Processing module, (2) Label Correlation module, (3) Annotation Framework module, (4) Categorized Image Classification module. Figure 1 shows these modules. Each module from these modules will be composed to a set of functions in terms of system functionality. The following section of the research contains the description of each module and its functions in details.

1. Image Processing module:-This module is responsible for performing the function that is related to the image. As the image passes through the segmentation and classification process the system automatically identifies regions, scenes, objects, facial aspects and spatial positions of those regions, objects and faces within the image. As part of this process the attributes within the image are given statistical relevancy based on how they typify the concept.

Features extraction: The Features extraction is the very important and critical step in processing phase. To extract patterns and derive knowledge from large collections of images, deals mainly with identifications and extraction of unique features for a particular domain. Due to the limitations of space and time, most of the data are represented in compressed forms. As a result, techniques used for editing, segmenting, and indexing images directly in the compressed domain have become one of the most important topics in digital libraries. So in this paper we introduce the use of DCT which are the important components of image in CBIR. In general, CBIR emphasizes rough image matching rather than exact matching. The DCT domain, to a certain extent, has unique scale invariance and zooming characteristics which can provide insight into object and texture identification, therefore it is naturally considered to be a potential domain in mining visual features. Because of this the amount of data used for processing and analysis is significantly reduced. This can lead

to simple yet efficient ways of indexing and retrieval in a large scale image datasets.

2. Label Correlation:- Usually, a single image may be associated with multiple labels, and the image annotation is a typical multi-label classification problem. A straightforward way to deal with this problem is to decompose it into several binary classification problems, with one for each label. However, the limitation is that this type of approach does not consider correlations among different class labels. Intuitively, such information is helpful for us to better understand image content. For example, the keyword “sea” may often be accompanied with the keyword “beach.” Such information is quite helpful to better understand the multimedia semantics. Thus, another method to reduce the required labor in image labeling is to utilize the label correlation for image annotation. In the field of machine learning and data mining, some researchers have also suggested that incorporating the information of label correlation into multi-label learning is beneficial for a reliable classification result. These research efforts have shown that utilizing class correlation information can improve the performance of multi-label classification in many domains.

Let us consider the scenario shown in Fig. 2. In part A of the figure, many training images are labeled as “beach” and “sea.” Ideally, the system should learn a pattern that there is a strong relationship between “beach” and “sea.” Then, for the training image in part B, which is labeled as “beach” and “sunset,” the system additionally labels it as “sea.” The unlabeled training image in part C, which is visually similar to the image in part B, is then labeled as “beach,” “sunset,” and “sea.” In that way, both the label correlation and visual information are considered for image annotation during training.

To exploit label correlations for image annotation, it is reasonable to assume that different image labels are related and built on some underlying common structures. For example, different photos taken at the beach share common characteristics, including sea, sky, and sand. We assume that there is a common subspace shared by multiple image labels. The final label of each image is predicted by its vector representation in the original feature space, together with the embedding in the shared subspace.

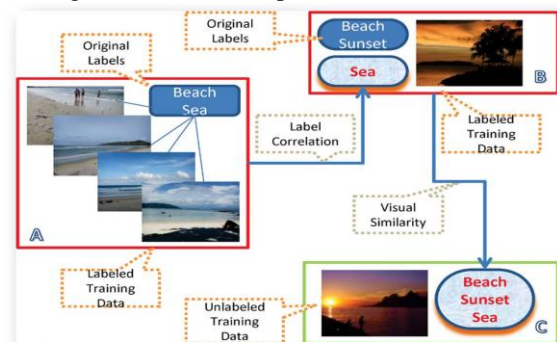


Figure 2. Illustration of Label Correlation for Image Annotation

3. Annotation Framework:-Image annotation, whose goal is to automatically assign relevant text keywords to any given image reflecting its content, is significant to the management of large scale image data. This application of computer vision



techniques is used in image retrieval system to organize and locate images of interest from a database. This method can be regarded as a type of multiclass image classification with very large number classes as large as the vocabulary size. It consists of a number of techniques that aim to find the correlation between low level visual features and high level semantics. In this paper the main challenge in automated image annotation is to create a model able to assign visual terms to an image in order to successfully describe it by realizing annotation by classification solution. In this framework learn a classifier for each multiple label and used them to predict whether the test image belong to the class defined by a particular label.

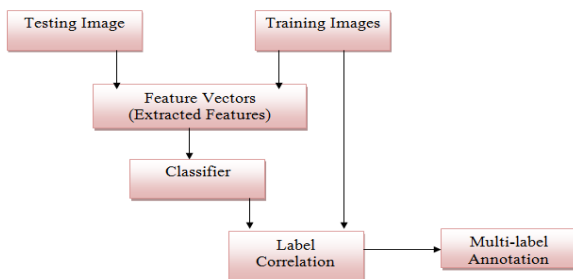


Figure3. Annotation Framework with Label Correlation

In this annotation framework we will utilize both labelled and unlabelled data for the annotation. Figure 3 shows the overview of system from training images the features are extracted to represent image in learning phase of classifier when the model is learnt. Information about correlation between labels is also obtained from training images. In testing phase classifier assigns probabilities of labels to the test image represented with visual features. The multiple labels annotate to the unlabelled image are adjusted by the label correlation module.

Following are the steps used to implement the annotation framework for unlabelled images.

- We randomly collect some personal images as training data, in which some images are labelled and remaining images are used as testing images which are not labelled (unlabelled data).
- Next step is to decide manually the set of queries that is categorized keywords like sky, water, buildings, Everest etc
- Find the visual features of training images with one for each class.
- To learn a better classifier we will use K-N-N clustering algorithm

Consider that there are  $N$  training data from  $C$  classes  $X = \{X_1, X_2, \dots, X_n\}$  where  $X_i$  is the vector representation of the  $i$ th instance in the training set and that there are  $m$  labelled data pairs  $(X_1, Y_1), \dots, (X_m, Y_m)$  where  $Y_i$  is the label of  $X_i$ . If  $X_i$  is in  $J$ th class and  $X_j$  is not labelled then the similarity between the  $X_i$  and  $X_j$  is calculated with the help of following equation,

$$\exp(-\|X_i - X_j\|^2)$$

By mapping the multi-class K-N-N algorithm outputs into probabilities we get the posterior

probabilities of each image belonging to each category. In other words,  $i$ th image characterized with  $J$ th class and this image can be annotated by word  $W_i$  with probabilities  $P(w_i|x_i)$ .

- Next step is to collect the labels from that particular Labeled image.
  - To exploit the label correlation, the first step is to extract the visual information from training images and put them in prediction matrix. These training images have label associates with them. For each image in dataset we need to extract the visual graph from each dataset image and put them in prediction matrix.
  - So we need to compare each matrix with training prediction matrix and based on the closest eigen matrix we need to label the image. We also consider the principal component analysis so that we can handle the case of relationship as explained in scenario which is shown in fig.2 to minimize our multiple comparisons.
- With this we introduce the predicted label matrix  $F$  for all training images  $X$ , where  $F_i$  is the predicted label of  $x_i \in X$  which is consistent with both visual features and labels [1].
- Our framework of annotation propagates the Labels of the nearest neighbors to the test image by considering its visual features and its label correlation. Given the test image  $x$ , compute its features vectors and find its  $K$  nearest neighbors in the training set, and annotate the test image with predicted labels.
  - Last categorized the annotated images with the labels and sort for retrieval purpose.

### B. Image Retrieval phase

Content-based image retrieval applications often suffer from small sample set and high dimensionality problem. Relevance feedback and boosting have been widely used to alleviate those problems. We will use interactive boosting framework to integrate relevance feedback into boosting scheme for content-based image retrieval. Compared to the traditional boosting scheme, the proposed method obtains more performance improvement from the relevance feedback by putting human in the loop to facilitate learning process. It has obvious advantage over the classic relevance feedback method in that the classifiers are trained to pay more attention to wrongfully predicted samples in user feedback through a reinforcement training process. It is clear that the framework can bridge the gap between high-level semantic concept and low-level image features better.

The process can be described in the following steps:

Step 1: Train weak classifiers on the original labelled data set and assign weights to classifiers based on their performance.

Step 2: Predict the labels of unlabeled data and present a subset of unlabeled data with their predicted labels to the user.

Step 3: User gives feedback on the retrieved data.

Step 4: Data obtained from user relevance feedback is added to construct a new labelled data set and removed from unlabeled data set.

Step 5: The labelled data are weighted according to their predicted label correctness.

Step 6: Go back to Step 1.

#### IV. CONCLUSION

After a review of existing techniques related to automatic annotation and image retrieval, we note that these methods are not powerful enough to efficiently retrieve relevant images including the concept of semantic web.

We propose a new annotation & content based image retrieval (ACIR) system with image annotation by integrating label correlation and visual similarity mining into a joint framework to annotate the unlabelled images and to enhance the retrieval accuracy. We annotate the unlabelled images with the help of label images by label correlation and by using its visual features. We also use interactive boosting algorithm to bridge the gap between high level semantic concepts and low level image features better. Compared with other existing system, our system simultaneously utilizes the information in the unlabelled data and the label correlation information for improving the retrieval efficiency and accuracy.

#### V. REFERENCES

- [1]. Yi Yang, Fei Wu, Feiping Nie, Heng Tao Shen, Yueting Zhuang, and Alexander G. Hauptmann, "Web and Personal Image Annotation by Mining Label Correlation With Relaxed Visual Graph Embedding", *IEEE Transactions On Image Processing*, Vol. 21, No. 3, March 2012.
- [2]. Riad, A., Elaminir, H. and Elghany, S.A. "Web Image Retrieval Search engine based on Semantically Shared Annotation", *IJCSI International Journal of Computer Science Issues*, 2012, pp. 1694-0814.
- [3]. J. Jeon, V. Lavrenko and R. Manmatha, "Automatic Image Annotation and Retrieval using Cross-Media Relevance Models," *SIGIR'03*, July 28–August 1, 2003
- [4]. L. Wu, S. Hoi, R. Jin, J. Zhu, and N. Yu, "Distance metric learning from uncertain side information with application to automated photo tagging," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 2, pp. 13:1–13:28, Feb. 2011.
- [5]. Y. Yang, F. Nie, D. Xu, J. Luo, Y. Zhuang, and Y. Pan, "A multimedia retrieval framework based on semi-supervised ranking and relevance feedback," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2011, to be published.
- [6]. X.-J. Wang, L. Zhang, X. Li, and W.-Y. Ma, "Annotating images by mining image search results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1919–1932, Nov. 2008.
- [7]. Y. Zhuang, Y. Yang, and F. Wu, "Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval," *IEEE Trans. Multimedia*, vol. 10, no. 2, pp. 221–229, Feb. 2008.
- [8]. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, 2008.
- [9]. D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. ICCV*, 2007, pp. 1–7.
- [10]. M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 2, no. 1, pp. 1–19, Feb. 2006.
- [11]. H. Wang, H. Huang, and C. Ding, "Image Annotation Using Multi-label Correlated Green's Function", in 2009.
- [12]. A. Makadia, V. Pavlovic, S. Kumar, "Baselines for Image Annotation.
- [13]. Z. Zha, T. Mei, J. Wang, Z. Wang, and X.-S. Hua, "Graph-based semisupervised learning with multiple labels," *J. Vis. Commun. Image Represent.*, vol. 20, no. 2, pp. 97–103, 2009.
- [14]. X. Zhu and A. Goldberg, *Introduction to Semi-Supervised Learning*. San Rafael, CA: Morgan & Claypool, 2009
- [15]. Kinh Tieu & Paul Viola, "Boosting Image Retrieval", *IEEE conference on computer vision and pattern recognition* 2002.