Automatic Image Annotation: A Review

Shalini K.Kharkate  
Rajiv Gandhi college of Engg , Research and Technology,Chandrapur ,Maharashtra,India  
shalin_k1@rediffmail.com

Nitin J.Janwe  
Rajiv Gandhi college of Engg , Research and Technology  
Chandrapur ,Maharashtra,India

Abstract-  
Automatic image annotation is the process of assigning keywords to digital images depending on the content information. Automatically assigning keywords to images is of great interest as it allows one to index, retrieve, and understand large collections of image data. Many techniques have been proposed for image annotation in the last decade that gives reasonable performance on standard datasets. This paper presents detailed literature about the Automatic Image Annotation. Literature gap is presented here which helps anyone to find different problem definitions. Based on the survey and literature gap the approach to deal with image annotation is presented.

Keywords: Automatic image annotation, Literature, feature vector, concept modeling.

I. INTRODUCTION

Due to easy availability and low cost of high resolution digital cameras digital image archives increasing rapidly. For effective search of image there is urgent need of successful indexing and searching image retrieval system. In the text based image retrieval the images are annotated and the data base management system retrieves them same way as text documents. However it is very difficult and time consuming task due to need of human intervention. To address these drawbacks, content based image retrieval using low level image feature such as color, texture, shape was proposed. Some representative system CBIR system such as QBIC, visual SEEK, photo book support content based retrieval by color, texture and shape [15].

To search for images relevant to a query some image processing technique are used for features extraction. Unfortunately the naive user may not be familiar with low level visual features and it is difficult to specify there query concepts by using the low level visual features directly, that is users are familiar with natural language like queries such as text and typically query images by semantics. Semantic gap between low level image content and high level content is main drawback of CBIR system eg consider a image of a mountain in low level terms it is a composition of colors, lines of different length and different shapes, in high level terms it is a mountain. If user wants to search for mountain image they need to specify the low level features such as green texture or they should the key word mountain.
The next improvement in process of image retrieval is IA approach which associates one or more multiple concepts with objects and images. Image annotation is the professional way for content based image retrieval. Here images are tagged with meaningful keywords or captions semantically. It is used in many applications domains such as entertainment, commerce education, biomedicine, military, and web image classification. Image annotation and retrieval has been very challenging research topic.

Image annotation techniques are classified in various categories as follows.

1. Making use of textual information: The huge numbers of images are available on the World Wide Web. In order to categorize and competently retrieve them, background information of the images such as surrounding content and associations can be used for image annotation. Automatically we can obtain semantic knowledge for Web images. Similarly context can be assigned to web images by using page layout analysis method. At the same time, the accuracy of image retrieval will be low as it can retrieve many irrelevant images. There are three reasons for low accuracy. Firstly, Web images can be used by anyone in the Web pages and there is no standard exists for the relationships between the texts and inserted images in the Web pages. Secondly, Web images are fairly wide-ranging in meaning, because they are created by different group for different reasons. Thirdly, the qualities of the Web images vary significantly. The users require passing through the whole list of retrieved images to search the preferred ones.

2. Manual Annotation: In manual annotation users have to enter some descriptive keywords when the images are browsed. In Manual annotation accuracy is high, since keywords are selected based on human determination of the semantic content of images. But at the same time, it is an effort intensive and monotonous process. But users can forget the annotations they have used after a long period of time.

3. Ontology driven approach: Semantic hierarchies and ontologies are also used for image annotation. Here conceptual knowledge i.e topological and spatial information are used to describe image.[24]explores the role of ontologies and knowledge-based approaches in the modeling and understanding of images semantics. use of explicit representations of background knowledge, i.e. Ontology driven approaches is powerful as they provide a formal framework that may contain explicit semantic definitions, which can be directly processed by a machine, and allow at the same time to derive implicit knowledge by automatic inference. The keyword-based approach is user friendly and it can be easily applied with satisfactory retrieval accuracy, while semantically rich ontology concentrate on the need for complete descriptions of image retrieval and advances the accuracy of retrieval. Ontology performs well with the combination of low level image features with high level textual information due to effectiveness of visual information to sort out the most of imprecise results.

4. Semi automatic annotation: In semi automatic image annotation, it requires user participation in the image annotation process[18].Due to human correction quality of annotation is improved as compared to manual annotation.

5. Automatic image annotation: Main idea of AIA is to capture semantic content of images to provide better means to organize and search on image database. It is mapping from the visual content information to the semantic context information automatically. It is best case in terms of efficiency and time but more error prone.
But AIA is difficult task because visual content to be analyzed depend on many factors such as shooting conditions, instances of objects, lighting condition, resolution of camera, and the background clutters. State of the art automatic image annotation systems can be analyzed and grouped from various point of views. The remainder of paper is organized as follows. Review of the existing approaches is explain in section II. The literature gap is described in section III. Proposed method is explain in IV and we conclude our work in section 5.

II. LITERATURE REVIEW

Probabilistic model biased towards the words that occur more frequently. Tianxia Gong, Shimiao Li, Chew Lim Tan[1] proposed a framework of using language models to represent the word-to-word relation utilizing probabilistic models. Also Propose - the semantic similarity language model to estimate the semantic similarity among the annotation words so that annotations that are more semantically coherent will have higher probability to be chosen to annotate the image. Probabilistic model TM and CMRM use word probability conditioned on image but language model use image probability conditioned on annotation word to reduce bias.

Lei Ye, Philip Ogunbona and Jianqiang Wang[6] recognized that some concepts can be characterized by structural visual features of images, some concepts are not directly characterized by structural visual features of images but can be described by certain concepts that can be characterized by structural visual features of images, and some concepts cannot even be derived from image visual features without further information or domain knowledge. Hence propose two categories of concepts, atomic concepts and collective concepts, and formulate the annotation problems as feature classification and concept inference problems. Images are first annotated with atomic concepts that can be characterized by visual features and the collective concepts that cannot be directly characterized by visual features but can be characterized by atomic concepts are then inferred from the atomic concepts. Features are first mapped to atomic concepts and atomic concepts are then mapped to collective concept. Yunhee Shin, Youngrae Kim, Eun Yi Kim[4] suggest a method to annotate textiles using emotional concepts such as romantic, classic, cute to reduce the semantic gap between low-level features and the high-level perception of users, and develop a method that automatically predicts the emotional concepts associated with images using machine learning algorithms. The performance of the proposed method was tested with 3600 textile images, and when the results were compared with those for other methods, the proposed prediction method achieved the best annotation accuracy of above 92% thus can be used for image retrieval.

Rami AlbatalL, Philippe mulhem Yves chiaramella propose model [2] where the regions of Interest (ROI) are successfully used in automatic image annotation through Bag of Visual Words (BoVW)
models. The obtained results indicate that this method outperforms (+6.2% of MAP) the CBOBW method. These results encourage us to do further analysis on the topological Visual Phrases in order to find interesting patterns for object classes.

Ran Li, YaFei Zhang, Zining Lu, Yu [12] propose a novel approach of multilabel image annotation for image retrieval based on annotated keywords. A novel annotation refinement approach based on PageRank is also proposed to further improve retrieval performance. Multi-label annotation contains two main stages: training and annotation. At training stage, considering different importance of features and removing redundancy, a bi-coded genetic algorithm is employed to select optimal feature subsets and corresponding optimal weights for every pair of classes in training set. At annotation stage, after unlabeled image is segmented, a set of pre-trained classifiers are used to vote and annotate each region, the final label of image is merged through all the region labels. Annotation refinement based on PageRank is employed to rank the candidate annotations and to deselect irrelevant labels which lowers down there recall. Therefore require more accurate refinement algorithm.

Dongjian He, Yu heng, Shirui Pan, Jinglei Tang[7] in order to boost annotation performance and to show one to one correspondence between image region and keyword proposed a novel algorithm, EMDAIA for automatic image annotation based on ensemble of descriptors . EMDAIA regards the annotation process as a multi-class image classification .First, each image is segmented into a collection of image regions. For each region, a variety of low-level visual descriptors are extracted. All regions are then clustered into k categories with each cluster associated with an annotation keyword. Moreover, for an unlabeled instance, distance between this instance and each cluster center is measured and the nearest category’s keyword is chosen to annotate it. But accuracy dependant on selection of segmentation algorithm.

S. Hamid Amiri, Mansour Jamzad[8] propose an annotation approach which follows the ALIPR structure. To describe the image contents, authors proposed an approach which extracts two discrete distributions as signatures for color and texture. These signatures are determined by applying clustering algorithms to the color and texture content of images. Major advantage of the signature extraction is that the number of segments for color and texture contents is determined automatically. The similarity of two nodes is defined based on the Mallows distance which provides more robust clusters. The required time for training the model of one concept was reduced substantially. Golnaz Abdollahian, Murat Birinci [9] low level features and local descriptors used combinely to obtain similarity between images.To identify the large homogeneous areas in the images as “texture areas” and exclude them from the keypoint detection and matching process . The texture areas are matched between images based on their overall color and texture properties. The non texture areas are handled by local descriptor matching The matching scores from different areas of the images are then combined to obtain an overall matching score for an image.

In automatic image annotation, systems that rely on a single classifier cannot achieve satisfactory results. Therefore, the combination of multiple classifiers for annotation has been an active research area. This kind of combination, which is also known as ensemble classifiers. Yinjie Lei Wilson Wong[10] simply combine the results of individual classifiers to obtain the final annotation without consideration for the potential errors that may be introduced by each individual classifiers. To address this problem, a novel AIA System based on a Two-Stage Feature Mapping (TSFM) model and a Training Set Construction (TSC) module using term extraction and image localization techniques is proposed. Hua Wang, Heng Huang and Chris Ding[11] propose “Image Annotation Using Bi-Relational Graph of Images and Semantic Labels”, Bi-relational Graph (BG) approach performs random walks on the BG that comprises both image vertices and class vertices, the resulted equilibrium distributions measure the relevances not only between class and image but also between class and class. They applied the proposed approaches in automatic image annotation and semantic image retrieval tasks. Encouraging results in extensive experiments demonstrated their effectiveness. A graph learning framework[3] for image annotation, which contained the image-based
graph learning and the word-based graph learning. The image-based graph learning generated the candidate annotations, and the word-based graph learning refined the candidate annotations to output the final results. To better capture the complex distribution of the image data, the NSC-based technique was proposed to construct the image-based graph. Graphs edge-weights are derived from the chain-wise statistical information instead of the traditional pairwise similarities. The word-based graph learning was performed by exploring three kinds of word correlations. One is the word co-occurrence in the training set, and the other two are derived from the web context. Extensive experiments on the Corel dataset and the web image dataset demonstrated the effectiveness of the proposed method.

In some literature, image annotation is formulated as a multi-class classification problem [23], which deals with the weak annotation problem and works with image-level ground truth training data. The relationship between low-level visual features and semantic concepts is found by supervised Bayesian learning. For each region in the test image, a posterior probability for each concept is calculated from class densities estimated from the training set and then the probability is modified using relevance with the other regions in the image. The image-level posterior probabilities are obtained by combining the regional posterior probabilities and keywords are selected according to their ranks. In [18], a novel Multi-Directional Search framework for semi-automatic annotation propagation. In this system, the user interacts with the system to provide example images and the corresponding annotations during the annotation propagation process. In each iteration, the example images are clustered and the corresponding annotations are propagated separately to each cluster: images in the local neighborhood are annotated. Furthermore, some of those 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. The query movements can be treated as multiple path navigation. Each path could be further split based on the user’s input.

III. LITERATURE GAP

Images are annotated to simply access them by using metadata that being added to images in order to allow more effective searches. If the images are described by textual information, then text search technique can be used to perform images searches [11]. Atomic and collective concepts can be used but Concept space construction is domain dependent [6]. However, there is a need to improve generation of automated metadata for images called AIA. Many researchers have proposed various techniques in attempting to bridge the well-known semantic gap. In [7,12], segmental approaches images are segmented into region and relation is find out between image region and word. Segmentation process is fragile and erroneous which make annotation process unreliable. Holistic approach [27] is to estimate the probabilities of images queries that then will be ranked according to their probabilities. No segmentation in holistic approach makes fast feature extraction but no direct correspondence between image region and word. Many of them realize another problem which is dependency on the training dataset to learn the models [12]. Image annotation surveys have been reviewed by many researchers according to the demanding the needs for annotating images. The graph model based image annotation methods’ time complexity and space complexity are always high, and it is difficult to apply it directly in real world image annotation [3]. In [4] shape feature in not consider which may improve accuracy of emotion prediction. Jiayu [14] has classified image annotation approaches into statistical approaches, vector-space related approaches and classification approaches. Probabilistic approaches have computational overhead. Classification model performance is superior to probabilistic. However, classification approach cannot be extended to unsupervised learning which is inherently supervised. Each model has its own advantages and disadvantages.

IV. PROPOSED WORK
A training set $S$ consisting of $N$ images with $n$ feature vectors. $n$ feature vectors forms a feature matrix and A pair of similar and dissimilar images($L$). The main purpose of this paper is to investigate the feature selection properties in the image annotation task. This image pair setting helps us to create a feature matrix that contains the same groups of features. Thus, we can directly do feature analysis on this matrix within the same framework.

Calculate the weight assign to each feature vector by using feature matrix and $L$, which is final step of training stage. Weight vector is used to find relevancy of keyword to the image. Sufficient training is necessary to have correct annotations to the image in testing stage.

Feature vector of input image is compared with feature matrix of the training images. Based on the weights calculated, most similar images are find out from $L$. Keywords from $L$ are getting assigned to test image which are annotations.

![Proposed Architecture](image)

**Fig 2.** Proposed Architecture

### V. CONCLUSION

In this paper, detailed literature about various image annotation algorithms is mentioned. These various image annotation techniques have its own advantages and disadvantages. A detailed literature gap is presented based on the literature which helps to find a new method of image annotation. The image annotation framework is proposed here. Future work is to implement this algorithm and to compare with various algorithms available.

### VI. REFERENCES

1. Tianxia Gong, Shimiao Li, Chew Lim Tan,"A Semantic Similarity Language Model to Improve Automatic image annotation" 22nd International Conference on Tools with Artificial Intelligence, 2010
2. “A new ROI grouping schema for automatic image annotation” by Rami Albatal, Philippe Mulhem, Yues Chiaramella


5. Y. Alp Aslandogan and Clement T. Yu, Senior Member, IEEE, Techniques and Systems for Image and Video Retrieval VOL. 11, NO. 1, JANUARY/FEBRUARY 1999


9. Golnaz Abdollahian, Murat Birinci †, Fernando Diaz-de-Maria ‡, Moncef Gabbouj †, Edward J. Delp “A region-dependent image matching method for image and video annotation” 2011 IEEE


11. “Image Annotation Using Bi-Relational Graph of Images and Semantic Labels” by Hua Wang, Heng Huang and Chris Ding, pages 126–139, 2011

12. “Technique of Image Retrieval based on Multi-label Image Annotation” by Ran Li, YaFei Zhang, Zining Lu, Jianjiang Lu, Yulong Tian-2010 Second International Conference on MultiMedia and Information Technology


23. Fei Shi, Fangfang Yang, Jiajun Wang, ”Supervised Semantic Image Annotation Using Region Relevance 2012 International Conference on Medical Physics and Biomedical Engineering


25. Feichao Wang “A survey on automatic image annotation and trends of the new age. 2011,Elsevier

26. Hao Ma, Jianke Zhu, Member, IEEE, Michael Rung-Tsong Lyu, Fellow, IEEE, and Irwin King, Senior Member, IEEE “Briding the semantic gap between Iage content and bags”