www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-7, Issue-4, July-August 2017

International Journal of Engineering and Management Research

Page Number: 54-58

Image Annotation and Retrieval- An Overview

Sayantani Ghosh[1], Prof. Samir Kumar Bandyopadhyay[2]
[1,2]Department of Computer Science and Engineering, University of Calcutta, INDIA

ABSTRACT

Structured knowledge models, such as semantic hierarchies and ontologies, appear to be a way to improve the accuracy of automatic image annotation. It allows modeling many valuable semantic relations between concepts based on image annotation using contextual and spatial relationships. Indeed, these relationships have been proved to be of prime importance for the understanding of image semantics. Moreover, such structured knowledge models about high-level concepts enable to reduce the complexity of the large-scale image annotation problem. The proposed paper discusses the different methods for the task of image annotation and retrieval.

***Keywords*--** Image annotation, Image Classification, image retrieval

I. INTRODUCTION

Efficient access to multimedia information requires the ability to search and organize the information. While, the technology to search text has been available for some time and in the form of web search engines is familiar to many people for using the technology to search images and videos, is much more challenging.

Several researchers have investigated techniques to retrieve images based on their content but many of these approaches require the user to query based on image concepts like color or texture which most people are not familiar with. In general, people would like to pose semantic queries using textual descriptions and find images relevant to those semantic queries. For example, one should be able to pose a query like "find me all images of tigers in grass". This is difficult if not impossible with many of these image retrieval systems and hence has not led to widespread adoption of these systems.

The traditional solution to this problem, used by libraries and other organizations is to annotate such images manually and then search those annotations. Although this allows semantic image retrieval manual annotations are expensive and do not always capture the content of images and videos well. The manual image annotation is an expensive and labor intensive procedure and hence there has been great interest in coming up with automatic ways to retrieve images based on content.

Initially, images were manually annotated by text descriptors (or tags) which are then used by an image retrieval system. This process is called 'iconography'. 'Iconography' means the description and interpretation of works in visual arts. With respect to image retrieval field, 'iconography' means the process of human annotation of images. The iconography may be of prime importance for image retrieval systems since it provides valuable information about image content. It allows to dispose of the visual content description of an image, and very often of its subjective, spatial, temporal and social dimensions which are priceless for understanding image semantics. Online photo sharing systems, such as Flickr, Picasa and Getty Images, provide a valuable source of human-annotated photos. However, the iconography requires a considerable level of human labor, and it cannot be considered for large image databases.

Automated image annotation can be defined as the process of modeling the work of a human annotator when assigning words to images based on their visual properties .up to now most of the image annotation systems are based on the combination of Image analysis and statistical machine learning techniques. To improve retrieval accuracy, the research focus has been shifted from designing sophisticated low level feature extraction algorithm to reducing the semantic gap between the visual features and the richness of human semantics. Given an input image, the goal of automatic image annotation is to assign a few relevant text keywords to the image that reflects its visual content. Utilizing image content to assign a richer, more relevant set of keywords would allow one to further exploit the fast indexing and retrieval architecture of Web image search engines for improved image search. This makes the problem of annotating images with relevant text keywords of enormous practical interest.

Image annotation is a difficult task for two main reasons: First is the well-known pixel-to-predicate or semantic gap problem, which points to the fact that it is hard to extract semantically meaningful entities using just low level image features, e.g. color and texture. Doing unambiguous recognition of thousands of objects or classes reliably is currently an unsolved problem. The second difficulty arises due to the lack of correspondence between the keywords and image regions in the training data. For each image, one has access to keywords assigned to the entire image and it is not known which regions of the image correspond to these keywords. This makes difficult the direct learning of classifiers by assuming each keyword to be a separate class.

Image indexing and retrieval has been a very active research domain since two decades, hence many approaches have been proposed to solve this problem. These different approaches can be classified in several different ways and from different points of views, as for example, the application domain, the indexing technique, the used content (modality) for image description, and so on.

II. REVIEW WORKS

Automatic Image Annotation has been active research topic in the last few years due its high impact on the Web search. To simplify the image retrieval metadata is added to images by an automatic image annotation. Recently, techniques have emerged to circumvent the correspondence problem under a discriminative multiple instance learning paradigm [1] or a generative paradigm [2].

One approach to automatically annotating images is to look at the probability of associating words with image regions. Some researchers used a Co-occurrence Model in which they looked at the co-occurrence of words with image regions created using a regular grid [3]. More recently, a few other researchers have also examined the problem using machine learning approaches [4]. Some other proposed to describe images using a vocabulary of blobs [5]. Each image is generated by using a certain number of these blobs. Their Translation Model - a substantial improvement on the Co-occurrence Model - assumes that image annotation can be viewed as the task of translating from a vocabulary of blobs to a vocabulary of words. Given a set of annotated training images, they show how one can use one of the classical machine translation models to annotate a test set of images [6]. Isolated pixels or even regions in an image are often hard to interpret. It is the context in which an image region is placed that gives it meaning. Query expansion is a standard technique for reducing ambiguity in information retrieval. One approach to doing this is to perform an initial query and then expand queries using terms from the top relevant documents (often approximated by the top documents). This expanded query when used for retrieval increases the performance substantially. In the image context, tigers are more often associated with grass, water, trees or sky and less often with objects like cars or computers.

Relevance-based language models [7-9] were introduced to allow query expansion to be performed in a more formal manner. These models have been successfully used for both ad-hoc retrieval and cross-language retrieval. Here, we investigate the problem of automatically annotating images as well as the ranked retrieval of images using a modification of the relevance model.

One of the significant ideas to bridge the semantic gap is to annotate the image which simplifies the image retrieval. In general, research efforts on image retrieval can be distributed into three types of approaches [1].

The first approach is the traditional text based annotation. In this approach, images are annotated manually by humans and images are then retrieved in the same way as text documents. It is impractical to annotate large amounts of image manually as it is time consuming and expensive. Also human annotations are too subjective and vague.

The second type of approach centers on content based image retrieval (CBIR), where images are automatically indexed and retrieved with low level content characteristics like color, shape and texture. However, recent research has shown that there is a significant gap between the low level content features and semantic concepts used by humans to interpret images. In addition, it is impractical for general users to use a content based image retrieval (CBIR) system because it furnishes users to provide query images.

The third approach of image retrieval is the Automatic Image Annotation so that images can be retrieved same as the text documents and extracts semantic features using machine learning techniques.

In general, the image annotation refers to process of assigning relevant keywords to the image to bridge the semantic gap between low level content features and semantic concepts understand by the humans. The basic purpose of automatic image annotation is to improve image retrieval accuracy which will reduce the irrelevant images in image retrieval system.

III. IMAGE ANNOTATION AND RETRIEVAL

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. The starting point for most of these algorithm is a training set of images that have been already been annotated by humans. These metadata made up of simple keywords that describe the content of the image. Image analysis techniques are used to extract features from the images such as color, texture and shape in order to model the

distribution of a term being present in the image. Features can be obtained from the whole image (global approach), or from blobs, which are segmented parts of the image (segmented approach) or from tiles which are rectangular partitions of the image. The next step is to extract the same feature information from an unseen image in order to compare it with all the previously created models. The result of this comparison yields a probability value of each keyword being present in an image.

The general steps for the automatic image annotation is classified into following steps:

1. Segmentation
2. Feature Extraction
3. Clustering/Classification
4. Annotation Model

Images are partitioned into group of pixels which are homogenous in nature. Image segmentation identifies visual features of image which can be merged or split in order to build objects of interest on which image analysis and interpretation can be performed. It is used in segmentation phase.

In the next phase low level visual information from segmented image is extracted using various feature descriptors like color, texture, shape. Color and texture are the most expressive to extract visual features of image [12].

It refers to the reducing high dimensional feature space to low dimensional feature space by using statistical techniques such as Principal Component Analysis and Particle Swarm Optimization Algorithm [10-11]. Feature selection is the basis of this process.

The group of feature vectors is formed depending on efficient clustering techniques such as k means, fuzzy clustering [16] clustering partitions the group of feature vector based on some specified common features and various similarity measures for image retrieval. The classification techniques such as k nearest neighbor, SVM, Decision Tree can also be used for grouping of feature vectors based on some predefined class label.

In the annotation model phase the testing image is done on the annotation model chosen such that labels are transferred from training to test images based on the annotation model specified used for the annotation of the image. Annotate the image based on annotation model such as probabilistic model, classification model, nonparametric approach or graph based approach.

Automatic image annotation can be defined as the process of automatically assigning a text description (often reduced to a set of semantic keywords) to a digital image through a computational model (or a computer system). Automatic image annotation is usually used in image retrieval systems in order to index and retrieve images of interest from a large database. Very often, this task is regarded as an image classification problem (usually a multiclass classification problem), and is involved by the following steps:

1. Gather a training image dataset consisting of a set of images with their textual annotations. These textual annotations generally consist of a set of high-level concepts (semantic concepts) depicting image content, and are usually called the ground truth. The set of all concepts composed the annotation vocabulary.
2. Build a computational model enabling to find a correspondence model between the low-level or mid-level representations of images and the concepts of the annotation vocabulary.
3. Test the system and adjusting the parameters of the computational model.

Figure 1 illustrates the aim of automatic image annotation by an example.

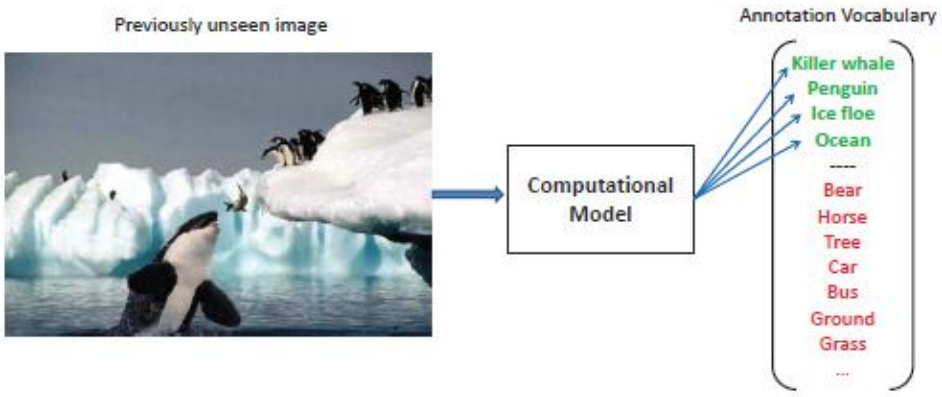

Figure 1 Example of automatic image annotation

The image retrieval field is carrying on the set of techniques/systems for browsing, searching and retrieving images from a large collection of digital images. Usually, such systems operate in two steps: i) image indexing: which could be defined as the process of extracting, modeling and storing the content of the image, the image data relationships, or other patterns not explicitly stored, and ii) image search: which consists in executing a

matching model to evaluate the relevance of previously indexed images with the user query. Figure 2 illustrates the workflow of image retrieval systems. It is common sense that image search in these system is based on the same features (or modalities) than the ones used for image indexing.

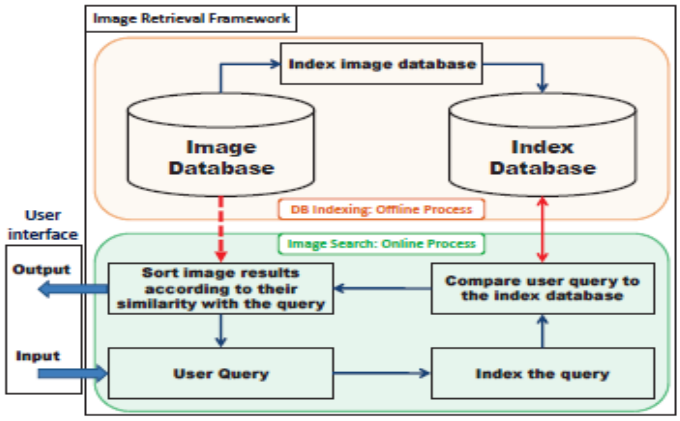

Figure 2 Workflow of image retrieval systems

Current approaches for automatic image indexing and retrieval can be classified into text-based approaches or content-based approaches, according to the used content (or modality) to index images. In the text-based approaches, images are indexed by a set of text descriptors which are extracted from the surrounding context. Nevertheless, although this paradigm is adopted by many current image search engines.

Sometimes the surrounding context is not always relevant with the image content, or sometimes only a small part of it describes its content. Consequently, the surrounding context is not always relevant for indexing images. Moreover, these methods do not consider the image content during the indexing process, and therefore there is no guarantee that the provided annotation is relevant with respect to image content. To overcome these problems, Content-Based Image Retrieval is used. Automatic Image annotation can be of four classes: Probabilistic Modeling, Classification, Graph Based, Parametric approach. The various methods for image annotation will be presented in the following paragraphs.

In Probabilistic modelling approach annotation of image is done by estimating the joint probability of an image with a set of words. It utilizes the probability table to estimated correspondences and using it to refine the estimate of the probability table. It annotates the image by partitioning the segments into blobs and finding the association of words and blobs by selecting the words with highest probability [13].

In Classification Based Approach low level features are extracted from image content, and the features are fed directly into a conventional binary classifier which gives a yes or no vote. The common machine learning tools include Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT) [14].

The automatic image annotation is highly affected by the segmentation results so to avoid prior segmentation. Researchers proposed an approach called, latent semantic analysis (LSA) based neural network (NN) annotation scheme. The annotation scheme is comprises of three parts. First, LSA is introduced to reveal the latent contextual correlation among the keywords. Second, with the labelled training images, Neural Network is obtained for characterizing the hidden association between the visual content of the image and the textual keyword. Third, given a test image, the learnt Neural Network is able to effectively provide the keywords to be annotated [15]. It is also classification based approach.

In Parametric Approach the feature space is assumed to follow a certain type of known continuous distribution. The conditional probability p(x|c) is modelled using multivariate Gaussian distribution where x and c are mean and concept label associated with feature vector. Some researchers use the conditional probability models concept by concept and then use the models to annotate unknown images [14].

Graph Based Approach for annotation using image based graph learning and word based graph learning. For a given the annotated training set and the visual features of all the images, the image-based graph learning aims to propagate labels from the annotated images to the unannotated images by their visual similarities. The labelling matrix is another necessary component during the graph learning. The image-based graph learning only focuses on the visual similarities among images, while the word correlations are not analyzed.

Two words with high co-occurrence in the training set will lead to high probability to annotate certain image jointly, such as `cloud' and `sky', `water', and `fish'. Therefore, the word co-occurrence becomes an informative

representation of the word correlation. To better capture the complex distribution of the image data, the Nearest Spanning Chain-based technique was proposed to construct the image-based graph. 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.

IV. CONCLUSIONS

The automatic image annotation is a very promising and important issue to improve image search and retrieval. Nevertheless, most of the current approaches are still insufficient to satiate the user need due to the following problems:

• Uncertainty introduced by machine learning algorithms.
• Scalability problems in terms of the image number, the dimension of the representation space of image visual features, and the concept number (dimension of the semantic space).
• Semantic gap problem.
• Subjectivity of image semantics (these approaches do not adapt to the user background).
• Sensitivity to the accuracy of the ground truth of the learning dataset.
• Polysemy problem.

Therefore, automatic image annotation is still a challenging problem and current approaches need to be improved.

REFERENCES

[1] Yang, C., Dong, M., Hua, J.: Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2006.
[2] Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, and N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007.
[3] Y. Mori, H. Takahashi, and R. Oka. Image-to-word transformation based on dividing and vector quantizing images with words. In MISRM'99 First International Workshop on Multimedia Intelligent Storage and Retrieval Management, 1999.
[4] K. Barnard and D. Forsyth. Learning the semantics of words and pictures. In International Conference on Computer Vision, 2001.
[5] P. Duygulu, K. Barnard, N. de Freitas, and D. Forsyth. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In Seventh European Conference on Computer Vision, 2002.
[6] P. Brown, S. D. Pietra, V. D. Pietra, and R. Mercer. The mathematics of statistical machine translation: Parameter estimation. In Computational Linguistics, 1993.
[7] V. Lavrenko and W. Croft. Relevance-based language models. Proceedings of the 24th annual international ACM SIGIR conference, 2001.
[8] V. Lavrenko, M. Choquette, and W. Croft. Cross-lingual relevance models. Proceedings of the 25th annual international ACM SIGIR conference, 2002.
[9] Fengxi Song; Zhongwei Guo; Dayong Mei, 2010. "Feature Selection Using Principal Component Analysis," System Science, Engineering Design and Manufacturing Informatization (ICSEM), International Conference, 2010.
[10] Jafar, O.A.M., Sivakumar R., "A study on fuzzy and particle swarm optimization algorithms and their applications to clustering problems," IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2012.
[11] D. Zhang, Md. M. Islam, G. Lu, 2012. "A review on automatic image annotation techniques", Pattern Recognition, vol. 45, no. 1, pp. 346–362.
[12] Manjunath, B.S.,Ohm, J.-R., Vasudevan, V.V.; Yamada, A., "Color and texture descriptors," Circuits and Systems for Video Technology, IEEE Transactions on , 2011.
[13] P. Duygulu, K. Barnard, N. de Freitas, D. Forsyth, "Object recognition as machine translation: learning a lexicon for a fixed mage vocabulary", In Seventh European Conference on Computer Vision (ECCV), 2002.
[14] Chapelle, O, Heffner, P., Vapnik, V.N., "Support vector machines for histogram-based image classification," IEEE Transactions on Neural Networks, vol.10, no.5, pp.1055, 1999.
[15] Yufeng Zhao,Yao Zhao, Zhenfeng Zhu; Jeng-Shyang Pan, "A Novel Image Annotation Scheme Based on Neural Network," Intelligent Systems Design and Applications, ISDA '08. Eighth International Conference, 2008.
[16] J. Liu, M. Li, Q. Liu, H. Lu, and S. Ma, "Image annotation via graph learning," Pattern Recognition, 2009.