Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 9, Issue, 07, pp.54622-54625, July, 2017

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

APPROACHES FOR AUTOMATIC IMAGE ANNOTATION

Sayantani Ghosh and *Samir Kumar Bandyopadhyay

Department of Computer Science and Engineering, University of Calcutta, India

ARTICLE INFO

Article History:
Received 28th April, 2017
Received in revised form
11th May, 2017
Accepted 17th June, 2017
Published online 31st July, 2017

Key words:

Automatic image annotation,
Ontology Engineering,
Image Retrieval.

ABSTRACT

The usual reason to annotate data (i.e. add metadata to it) is to simplify access to it. This is one of the key ideas behind the semantic web. The metadata added to documents or images allow for more effective searches. In the case of images, if they are completely described by a textual annotation, then many image searches can be done effectively by text search techniques.In this paper the different approaches of annotation of images are discussed.

Citation: Sayantani Ghosh and Samir Kumar Bandyopadhyay, 2017. "Approaches for Automatic Image Annotation", *International Journal of Current Research*, 9, (07), xxxxxxxxxx.

INTRODUCTION

The problem with adding metadata manually is that it is an extremely labor intensive and time-consuming task. In order to maximize the benefit of the colossal repository of digital images available both publicly and in private collections, intelligent matchmaking tools are required. Many World Wide Web image search engines attempt to automate this task by using text from the image filename and text near the image on a webpage. However, search results using this method usually contain many irrelevant images. With the aim of improving the automated metadata generation for images, automated image annotation and object recognition are currently important research topics in the field of computer vision (Gabrilovich and Markovitch, 2005). 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. Medical images play a central role in patient diagnosis, therapy, surgical planning, medical reference, and medical training. With the advent of digital imaging modalities, as well as images digitized from conventional devices, collections of medical images are increasingly being held in digital form. It becomes increasingly expensive to manually annotate medical images. Consequently, automatic medical image annotation becomes important. In medical image retrieval problems, the problem addressed has the following properties:

1. The images in the retrieval database can be annotated into one of the pre-defined labels, which are denoted as the ground truth labels of the images. Due to the ground truth complexity, only a small portion of the whole image collections have their ground truth labels available.
2. Given a specific query, the correctly retrieved images should have the same ground truth label, which may not necessarily equal to the ground truth label of the query image provided that the query image and the retrieved images share a sufficient semantic similarity. This means that a user may query the database with an image that is close to but not exactly what he/she expects.

The automatic generation of image metadata should allow image searches and Content-Based Image Retrieval (CBIR) is more effective. In image retrieval scenario an image database could be annotated offline by running a keyword annotation algorithm. Suppose images contain a particular keyword and user wishes to find the specific keyword in the database, e.g. for an on-line shopping task, he/she would select a region containing the target keyword from an image. An object recognition algorithm could then categorize the selected region

****Corresponding author: Samir Kumar Bandyopadhyay***
Department of Computer Science and Engineering, University of Calcutta, India.

of the keyword and a text search could be carried out to find all images in the database with the associated keyword. This would significantly reduce the number of images in which it would be necessary to attempt to recognize the specific keyword selected by the user. In recent years, Content-Based Image Retrieval System is mostly used in all sort of domain. CBIR is a method of retrieving the image based on the input image. In this CBIR system, the content of the image is being analyzed attributes such as color, shapes and texture of an image. This technique is done with various algorithms and methods for matching the images from database. Nowadays, CBIR system is a powerful resource used widely in Medical. CBIR has the possibility to give medical doctors with accurate result in diagnoses. The conventional goal of CBIR for medical is to retrieve the closest images to the query-image from the dataset (Grubinger *et al*., 2006). Automatic Image Annotation is a process assigning the information or data to an image. It is done with the caption or the keyword of the image. Machine learning techniques are commonly used for the image classification and image feature analysis such as segmentation and so on. In this image annotation method, a general term is usually quoted called 'blob'. A blob is a part of an image with a vocabulary meaning. The image is separated into blobs based on the region or cluster and the corresponding blobs are labeled with a vocabulary. In the medical imaging, image annotation helps the doctor to find out the description of an image.

Affordable access to digital technology and advances in Internet communications have contributed to the unprecedented growth of digital media repositories (audio, images, and video) over the past few years. Retrieving relevant media from these seemingly ever-increasing repositories is an impossible task for the user without the aid of search tools. Efficient content-based retrieval of image and video databases is an important application due to rapid proliferation of digital video data on the Internet and corporate intranets. Text either embedded or superimposed within video frames is very useful for describing the contents of the frames, as it enables both keyword and free-text based search, automatic video logging, and video cataloging. To measure progress towards successfully carrying out the task, evaluation of algorithms which automatically extract the metadata is required. Some algorithms providing global annotations, such as distinguishing between city and landscape images or between images acquired indoors and outdoors, have a higher success rate than algorithms attempting to detect specific objects, such as cars, cows and sunglasses. Automatic recognition of activities, events and abstract or emotive qualities in images currently performs rather poorly.

Many of the existing approaches to extracting text from images and video suffer from one or more limitations such as locating only the bounding blocks of the text (therefore requiring human involvement to recognize the characters), sensitivity to font sizes and styles, restrictions on the appearance characteristics of text that can be handled, restrictions on the type of text that can be extracted (e.g. captions only), and inability to handle normal and inverse video modes of text. In this paper we review the different approaches of annotation of images. Our aims are to discuss three types of annotation such as free-text annotations, keyword annotations and annotations based on ontologies. The particular attention is given for the creation of vocabularies for image annotation and to methods which have been applied for reducing the amount of effort required for image annotation.

Different Approaches Image Annotation

One of the significant ideas to annotate the image for easy retrieval. The traditional approach is text based annotation. Here images are annotated manually by humans and images are then retrieved in the same way as text documents. It is time consuming and expensive. Human annotations are normally too subjective and vague. Different types of information can be associated with images or videos. They are:

Content-independent metadata is related to image or video content, but does not describe it directly. For examples author's name, date of publication, book title, cost, etc.

Data directly refers to the visual content of images. It can be classified further as Content-dependent metadata and Content-descriptive metadata. The first one refers to low/intermediate-level features (colour, texture, shape, motion, etc.). The second one describes the relationships of image entities with real-world entities or temporal events, emotions and meaning associated with visual signs and scenes. Content-dependent metadata is easy to extract since one can extract huge feature vectors containing colour histogram features, texture features calculated by different algorithms, etc. Content-descriptive metadata can have No pre-defined structure for the annotation as well as arbitrarily chosen keywords or keywords chosen from controlled vocabularies. Classifications based on ontologies are similar to classification by keywords, but the fact that the keywords belong to a hierarchy enriches the annotations. For example, it can easily be found out that a "Cow" is a subclass of the class "animal". In Annotation using keywords each image is annotated by having a list of keywords associated with it. There are two possibilities for choosing the keywords:

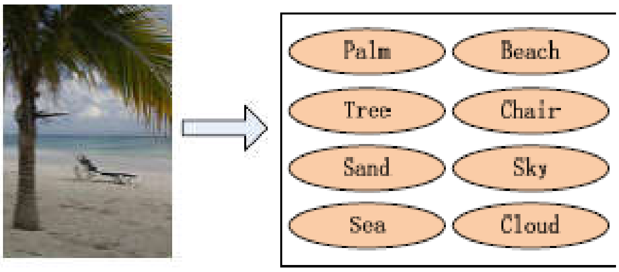
(1) The annotator can use arbitrary keywords as required.
(2) The annotator is restricted to using a pre-defined list of keywords (a controlled vocabulary).

This information can be provided at two levels of specificity:

(1) A list of keywords associated with the complete image, listing what is in the image.
(2) A segmentation of the image along with keywords associated with each region of the segmentation. In addition, keywords describing the whole image can be provided. Often the segmentation is much simpler than that shown, consisting simply of a rectangular region drawn around the region of interest or a division of the image into foreground and background pixels.

If one is searching within a single image database that has been annotated carefully using a keyword vocabulary, then one's task is simplified. In practice Different image collections are annotated using different keyword vocabularies and differing annotation standards. A naive user does not necessarily know the vocabulary which has been used to annotate an image collection. This makes searching by text input more difficult. The following figure 1 shows the basic task of automatic image annotation. If the user want to choose from an on-screen list of keywords then it is difficult for large number of keywords. A more sophisticated approach is to extend the annotation of a document by using ontologies and other information available on the World Wide Web. This has been done in the text retrieval domain in the biomedical abstract retrieval domain

(Hardoon *et al*., 2006; Datta *et al*., 2008). An ontology is a specification of a conceptualization (Jing Liu, 2007). It basically contains concepts (entities) and their relationships and rules. Adding a hierarchical structure to a collection of keywords produces a taxonomy, which is an ontologyas it encodes the relationship "is a" (a dog is an animal). An ontologycan solve the problem that some keywords are ambiguous. Ontologies are important for the Semantic Web, and hence a number of language sexist for their formalization, such as OWL and RDF.



Unlabelled Image **Annotation Result**

Work on the development of ontologies which aim to arrange all the concepts in the world into a hierarchical structure is not new. One of the first comprehensive attempts was made by researcher in 1668. One of the main problems is that there are many possible logical ways of classifying concepts, which also depend for example on the influence of culture (Jing Liu, 2007). Developing ontologies to describe even very limited image domains is a complicated process (Hardoon *et al*., 2006; Datta *et al*., 2008). In the domain of image description, Icon class is a very detailed ontology for iconographic research and the documentation of images, used to index or catalogue the iconographic contents of works of art, reproductions, literature, etc. It contains over 28 000 definitions organised in a hierarchical structure. Each definition is described by an alphanumeric code accompanied by a textual description (textual correlate). For example, the code 47D31 refers to "windmill" and translates into the following hierarchy:

- 4 Society, Civilization, Culture
- 47 crafts and industries
- 47D machines; parts of machines; tools and appliances
- 47D3 machine driven by wind
- 47D31 windmill
- Note that this is distinct from the concept of "windmill in landscape" which,falls into a completely different category. It has the code 25I41, which translatesinto:
- 2 Nature
- 25 earth, world as celestial body
- 25I city-view, and landscape with man-made constructions
- 25I4 factories and mills in landscape

In free text annotation the user can annotate using any combination of words or sentences. This makes it easy to annotate, but more difficult to use the annotation later for image retrieval. Often this option is used in addition to the choice of keywords or ontology. Creation of keyword vocabularies and methods for making manual annotation is more efficient. There are two approaches to associating textual information with images described in the computer vision literature: annotation and categorization. In annotation, keywords or detailed text descriptions are associated with an image, whereas in categorization, each image is assigned to one of a number of predefined categories (Datta *et al*., 2008).Categorization can be used as an initial step in image understanding in order to guide further processing of the image. For example, in (Jing Liu, 2007) a categorization into textured/non-textured and graph/photograph classes is done as a pre-processing step. Recognition is concerned with the identification of particular object instances. Object recognition would distinguish between images of two structurally distinct cups (Datta *et al*., 2008), while category-level object recognition (Jing Liu, 2007) would place them in the same class. Recognition also has its uses in annotation, for example in the recognition of family members in the automatic annotation of family photos. Category-level object recognition canat present be seen as annotation using a small keyword vocabulary. This is because current category-level object recognition algorithms tend to be capable of recognizing only a few objects. As object recognition algorithms improve, it is to be expected that the vocabulary sizes will increase.

Categorization can be used as an initial step in image understanding in order to guide further processing of the image. For example, a categorization into textured/non-textured and graph/photograph classes is done as a pre-processing step. Recognition is concerned with the identification of particular object instances. Object recognition would distinguish between images of two structurally distinct cups (Datta *et al*., 2008), while category-level object recognition would place them in the same class. Recognition also has its uses in annotation, for example in the recognition of family members in the automatic annotation of family photos. Category-level object recognition cant present be seen as annotation using a small keyword vocabulary. This is because current category-level object recognition algorithms tend to be capable of recognizing only a few objects. As object recognition algorithms improve, it is to be expected that the vocabulary sizes will increase. While a number of ontologies and vocabularies are available but they are used for commercial purposes and also for specific areas of application.

There are a number of criteria that affect the construction and usefulness of avocabulary. One is the range of terms to be included (Jing Liu, 2007; Hare *et al*., 2006). This is tied closely to the planned use of the vocabulary and the specification of which information should be included in an image annotation. A vocabulary including a wide range of terms, ranging from names of objects to emotions provoked by an image is applicable in a wide range of situations. However, annotating an image with all the expressive capability of such a vocabulary will most likely be time-consuming. If the annotated images are to be used to evaluate object recognition algorithms, then some of the annotation will exceed the requirements of the task. Solutions are to use an extensive vocabulary with additional annotation guidelines which restrict the parts of the vocabulary to be used, or to create a restricted vocabulary containing only keywords suitable to the task at hand. A further design criterion to be considered is how to impose a suitable hierarchical (or other) structure on the vocabulary. Word Net is an on-line lexical reference system which organizes English nouns, verbs and adjectives into synonym sets, each representing one underlying lexical concept. For example, some researchers gave the full Word Net vocabulary along with a set of annotation guidelines to people producing the ground truth for their recognition evaluation dataset. Word Net has also been used as the basis for creating a more restricted vocabulary. Another researcher constructs an ontology of portray able objects by pruning the Word Net tree. They began with the subclass “object” of the class “entity” and extracted a tree with 102 nodes in the level below “object” and 24 000words describing portray able objects in the leaf nodes of the tree. An effort was begun to create a vocabulary of 12 000 to15 000 terms for general collections of images. This was done in a first stage by gathering a large number of terms from existing vocabularies for image classification followed by the merging of vocabulary lists created by a number of participants. The expansion of the vocabulary in the second stage was done by examining sources of images such as multi-language visual dictionaries and specialized reference works. Unfortunately the work on this vocabulary seems to have been abandoned. In probability and statistics, a generative model is a model for randomly generating observable data, typically given some hidden parameters. It specifies a joint probability distribution over observation and label sequences. Generative models are used in machine learning for either modeling data directly (i.e. modeling observed draws from a probability density function), or as an intermediate step to forming conditional probability density function. A conditional distribution can be formed from a generative model through the use of Bayes’ rule.

A dual cross-media relevance model (DCMRM) for automatic image annotation estimates the joint probability by the expectation over words in a pre-defined lexicon. DCMRM involves two kinds of critical relations in image annotation. One is the word-to-image relation and the other is the word-to-word relation. Both relations can be estimated by using search techniques on the web data as well as available training data (Putthividhy *et al*., 2010). Discriminative models are a class of models used in machine learning for modeling the dependence ofan unobserved variable y on an observed variable x . Within a statistical framework, this is done by modeling the conditional probability distribution P(y| x), which can be used for predicting y from x. Discriminative models differ from generative models in that they do not allow one to generate samples from the joint distribution of x and y. However, for tasks such as classification and regression that don’t require the joint distribution, discriminative models generally yield superior performance.

On the other hand, generative models are typically more flexible than discriminative models in expressing dependencies in complex learning tasks. In addition, most discriminative models are inherently supervised and cannot easily be extended to unsupervised learning. Graph model has successfully resolved many machine learning problems in recent years, there have been some graphical model-based image annotation methods and by which image annotation performance could be promote obviously. Someone discussed the annotation process theoretically by reviewing some related work, and proposes unified annotation framework via graph learning. The framework includes two sub-processes, i.e., basic image annotation and annotation refinement. In the basic annotation process, the image-based graph learning is utilized to obtain the candidate annotations.

In the annotation refinement process, the word based graph learning is used to refine those candidate annotations from the prior process. However, 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.

Conclusion

The approach for the annotation process was user-focused and took into account the dynamics of the retrieval process. It was based on a sentence-based template suitable for modeling user-queries that can involve complex relationships between the query keywords. The retrieval algorithm is based on a variation of the nearest-neighbor search technique for traversing the ontology tree and can accommodate complex, relationship-driven user queries. The algorithm also provides for using pre-defined weightings to qualify the search result in accordance to user preferences. Due to the rapid advancement of digital technology in the last few years, there has been an increasingly large amount of images available on the Web. Therefore, it is of great importance to automatically annotate images. In this survey, the existing approaches for automatic image annotation are summarized.

REFERENCES

Datta, R., Joshi, D., Li, J. and Wang, J.Z. 2008. Image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys. 2008.

Gabrilovich, E. and Markovitch, S. 2005. Feature generation for text categorizationusing world knowledge, in: Proceedings of The Nineteenth International JointConference for Artificial Intelligence, Edinburgh, Scotland, 2005.

Grubinger, M., Clough, P., Muller, H. and Deselaers, T. 2006. The IAPR TC-12 benchmark- a new evaluation resource for visual information systems, in: Proceedings ofthe International Workshop OntoImage’2006.

Hardoon, D. R., Saunders, C., Szedmak, S. and Shawe-Taylor, J. 2006. A correlation approach for automatic image annotation, in: Proc. 2nd Int. Conf. Advanced Data Mining and Applications, 2006.

Hare, J.S. *et al*. 2006. Mind the gap: another look at the problem of the semantic gap in image retrieval. Multimedia Content Analysis, Management, and Retrieval 2006.

Jing Liu, Bin Wang, Mingjing Li, *et al*., 2007. Dual cross-media relevance model for image annotation, In Proceedings of the 15thinternational conference on Multimedia, 2007.

Kutics, A., Nakagawa, S., Arai, H. Tanaka, S. and Ohtsuka, 2004. Relating words and image segments on multiple layers for effective browsing and retrieval, in:Proceedings of the International Conference on Image Processing, IEEE, Vol. 2, PP.2203-2206.

Putthividhy, D., Attias, H.T. and Nagarajan, S.S. 2010. Topic regression multi-modal latent dirichlet allocation for image annotation, In Proceedings of IEEE Computer Vision and Pattern Recognition, 2010.

****