International Journal of Computer science engineering Techniques-– Volume 2 Issue 4, May - June 2017

RESEARCH ARTICLE OPEN ACCESS

Comprehensive Overview of Existing Semantic Annotation Approaches

Mrs. Sayantani Ghosh , Prof.Samir Kumar Bandyopadhyay

Department of Computer Science and Engineering,University of Calcutta

Abstract:

The semantic features that based on keywords or annotationsmaybe very subjective and time consuming. Whereas,the semantic features that based on visual contentis complex becauseof the inference procedures. Automatic image annotation is good approach toreduce the semantic gap. This paper provides an overview of the most common techniques of different types of annotated image retrieval systems along with classification methods.

Keywords — Contextual Information,Automatic Image Annotation,Content-based Image Retrieval,Text-Based Image Retrieval, Ontology, Semantic Annotation

Introduction

In real world objects are seen embedded in a specific context and its representationis essential for the analysis and the understanding ofimages. Contextual knowledge may stem from multiple sources of information, includingknowledge about the expected identity, size, position and relative depth of anobject within a scene [1-2]. For example, topological knowledge canprovide information about objects that are most likely to appear within a specificvisual setting, for example an office typically contains a desk, a phone, and a computer. Spatial information can also provide informationabout which locations within a visual setting are most likely to contain objects, e.g.in a beach scene, the sky is usually placed at the top, while the sea is below. Given aspecific context, this kind of knowledge can help reasoning on data to improve imageannotation.

Contextual information means the collection of relevant conditions and surrounding influences that make asituation unique and comprehensible. While contextual knowledge is the information,and/or skills that have particular meaning because of the conditions that form part oftheir description.It is of prime interest to make efficient use of contextual knowledgein order to narrow the semantic gap, and to improve the accuracy of image annotation.

Images get their semantic meaning for image interpretation or understanding, and it is consequently difficult for an imageretrieval system to discern the meaning sought by a user when he is searching for aparticular image.Image semantics seems to be important forimage retrieval related tasks.

Image semantics is that it is not fully, norexplicitly stored in the image pixels, and it is usually hard for a machineto access the image semantics using only image features.We can therefore conclude that the imageinterpretation process requires often a reasoning mechanism over the detected objectsin the image, which is usually based on cognition and on the past experiences.

Knowledge modelsshould go further than the simple description of specific objects that may appear inimages, and rather model the image context through the description of conceptsand the semantic relationships between them.Image semantics is a multi-level paradigm, i.e.there are several levels of semantics (or interpretation) for a given image and themajor challenge of image retrieval systems is then to be able to extract such semanticsfrom images and to adapt to the user background in order to be efficient and useful.

If we look at image at Figure 1, the semantics at theobject level could be {"Bear", "Iceberg"}, the semantics at the partial level could be"Polar Bear standing on a small iceberg" and the semantics at the full level could be"global warming threatens the survival of the polar bears". Therefore, we can noticethat the difficulty of processing and extracting the semantics from images increasessignificantly according to the sought level of abstraction. Currently, most approachesfor image retrieval deals with the first level of semantic content. These approachestarget to provide efficient methods to learn semantics classes from visual image features.

Figure 1 Global warming threatens the survival of polar bears.

Today, to use automatic image annotation in order to fill the semantic gap between low level features of images and understanding their information in retrieving process has become popular. Since automatic image annotation is crucial in understanding digital images several methods have been proposed to automatically annotate an image. This paper reviews current methods for visualizing semantic effects on the annotated images.

Different Methods for Semantic Image Annotation

The World Wide Web has become one of the most important sources of information due to the fast development of internet technology. Search engines are the most powerful resources for finding visual content (e.g., images, videos, etc.) from World-Wide Web, These search engines use the surrounding text near the image for describing the content of an image and rely on text retrieval techniques for searching particular images [1].However, there are two significant drawbacks of such engines; (a) when the surrounding words are ambiguous or even irrelevant to the image; search results using this method usually contain many irrelevant images. (b)The retrieval of images will be ineffective when different languages are used in the description of the images if this image collection is to be shared globally around the world. It is difficult to map semantically equivalent words across different languages [2-3].

The rapid growth of multimedia content comes with the need to effectively managethis content by providing mechanisms for image indexing and retrieval that canmeet user expectations. Towards this goal, semantic image analysis and interpretationhas been one of the most interesting challenges during this last decade, andseveral attempts have addressed the, previously introduced, semantic gap problem.In particular, a typical method for narrowing the semantic gap is to perform automaticimage annotation.

Automatic image annotation was introduced in the early2000s, and first efforts focused on statistical learning approaches as they providepowerful and effective tools to establish associations between the visual featuresof images and the semantic concepts.A recent review on automaticimage annotation techniques was proposed [4-5].

Early efforts aim to narrow the semantic gap for mapping low-level features (such as color, texture, shape andsalient points) directly to some specific semantic concepts such as indoor/outdoor, nature, animal, food, and pedestrian.These approaches havequickly become cumbersome and impractical following the normal request of a largerannotation vocabulary. Indeed, it would be impossible to build a detector for eachpotential concept, as they are too many [5].

Content-based image retrieval (CBIR) is used to solve text based image retrieval [4]. In this technique different low-level visual features are extracted from each image in the image database and then image retrieval is to search for the best match to the features that are extracted from the query image.CBIR based approaches shows good accuracy for detecting specificobjects/concepts, such as faces, pedestrians, cars, etc. These approaches selectthe parameters of the model so as to minimize the detection error on a set oftraining images by machine learning.In the text-based approaches, images are indexed by a set of text descriptors which are extracted from the surrounding context.

Content-based means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results. Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results. The basic CBIR will look as shown in the figure 2.

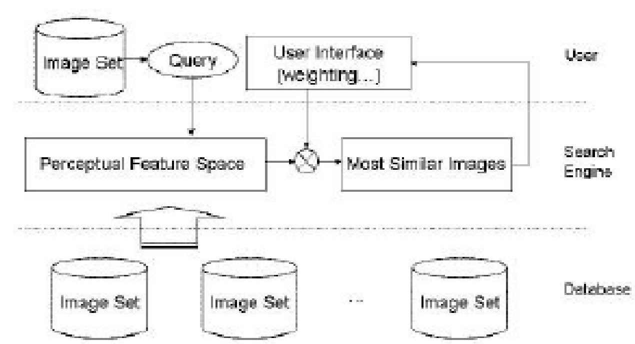

Figure 2 Basic System of CBIR

CBIR systems is classified into two categories: text query or pictorial query. In text query based systems, images are characterized by text information such as keywords and captions. Text features are powerful as a query, if appropriate text descriptions are given for images in an image database. However, giving appropriate descriptions must be done manually in general and it is time consuming. There are many ways one can pose a visual query. A good query method will be natural to the user as well as capturing enough information from the user to extract meaningful results. In pictorial query based systems, an example of the desired image is used as a query. To retrieve similar images with the example, image features such as colours and textures, most of which can be extracted automatically, are used.

The typical CBIR system performs two major tasks. The first one is feature extraction, where a set of features, called image signature or feature vector, is generated to accurately represent the content of each image in the database. A feature vector is much smaller in size than the original image, typically of the order of hundreds of elements (rather than millions). The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their signatures is computed so that the top "closest" images can be retrieved. Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image.

Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Different similarity/distance measures will affect retrieval performances of an image retrieval system significantly. For content-based image retrieval, user interaction with the retrieval system is crucial since flexible formation and modification of queries can only be obtained by involving the user in the retrieval procedure. User interfaces in image retrieval systems typically consist of a query formulation part and a result presentation part. There are various techniques have been proposed to retrieve the image effectively and efficiently from the large set of image data. These are as follows:

- Gaussian Mixture Models
- Semantic template
- Wavelet Transform
- Gabor filter
- Support Vector Machine
- Color Histogram
- 2D Dual-Tree Discrete Wavelet Transform
- There are three fundamental bases for content based image retrieval, i.e. visual feature extraction, multidimensional indexing, and retrieval system design.
- Feature extraction and indexing of image database according to the chosen visual features, which from the perceptual feature space, for example color, shape, texture or any combination of above.
- Feature extraction of query image.
- Matching the query image to the most similar images in the database according to some image-image similarity measure. This forms the search part of CBIR systems.
- User interface and feedback which governs the display of the outcomes, their ranking, the type of user interaction with possibility of refining the search through some automatic or manual preferences scheme etc.

Ontology is a specification of a conceptualization. Ontology defines a set of representational terms called concepts; each concept has three basic components: terms, attributes and relations. Terms are the names used to refer to a specific concept, and can include a set of synonyms that specify the same concepts. Attributes are features of a concept that describe the concept in more detail. Finally relations are used to represent relationships among different concepts and to provide a general structure to the ontology.The main parts of image annotation are shown in figure 3.

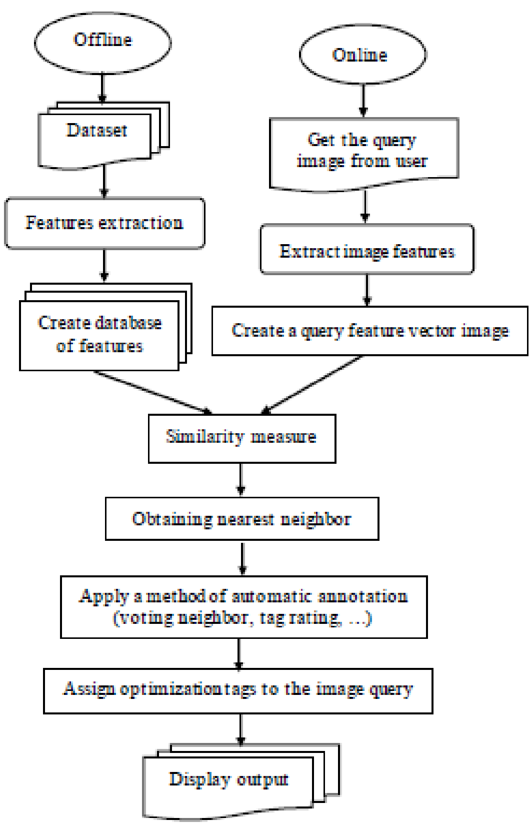

Figure 3 Main Parts of CBIR

CBIR has many applications in real world such as:

i) The advantages of such systems range from simple users searching a particular image on the web.

ii) Various types of professionals like police force for picture recognition in crime prevention.

iii) Medicine diagnosis

iv) Architectural and engineering design

v) Fashion and publishing vi) Geographical information and remote sensing systems

Text-Based Image Retrieval (TBIR) is currently used in almost all general-purpose web image retrieval systems today. This approach uses the text associated with an image to determine what the image contains. This text can be text surrounding the image, the image's filename, a hyperlink leading to the image, an annotation to the image, or any other piece of text that can be associated with the image [6].

In image mining, meaningful information can automatically extract meaningfulinformation from a huge of image data are increasingly indemand. It is an interdisciplinary venture that essentiallydraws upon expertise in artificial intelligence, computervision, content based image retrieval, database, data mining,digital image processing and machine learning.

Image mining frameworks [7] are grouped into two broadcategories: function-driven and information-driven. Theproblem of image mining combines the areas of content-basedimage retrieval, data mining, image understanding anddatabases. Image mining techniques include image retrieval,image classification, image clustering, image segmentation,object recognition and association rule mining.Image Retrieval is performed by matching the features of aquery image with those in the image database. The collectionof images in the web are growing larger and becoming morediverse. Retrieving images from such large collections is achallenging problem. The research communities study aboutimage retrieval from various angles are text based and contentbased. The text based Image retrieval is used for traditionaltext retrieval techniques to image annotations.

Digital images are currently widely used in medicine,fashion, architecture, face recognition, finger print recognitionand bio-metrics etc. Recently, Digital image collections arerapidly increased very huge level. That image contains a hugeamount of information. Conversely, we cannot make sure thatinformation is useful unless it is implemented so we needsufficient browsing, searching, and retrieving the images.Retrieving image has become a very dynamic researcharea.

Two major research communities such as databasemanagement and computer vision have study image retrievalfrom various ways such as text based and content based. Late1970s, the text-based image retrieval had been traced back. Avery popular framework of image retrieval was to annotate theimages by keyword and they used text based databasemanagement system for operating image retrieval. Emergence of large-scaleimage collections in the early 1990s, the major difficulties aremanual image annotation is also accurate.

To avoid thissituation, content-based image retrieval was improved. Itmeans, instead of using text based key words, images shouldbe defined by their visual contents as colour and texture. Manytechniques in this research area have been developed for manyimage retrieving systems as research and commercial, havebeen built. It has established a general framework of imageretrieval. In this paper we will focus our effort mainly to thecontent-based image retrieval.Text-based image retrieval [7-8] can be based onannotations that were manually added for disclosing theimages (keywords, descriptions), or on collateral text that isavailable with an image (captions, subtitles,nearby text). It applies traditional text retrieval techniques toimage annotations or descriptions. Most of the image retrievalsystems are text-based, but images frequently have little or noaccompanying textual information.

Text data present in multimedia viz. video and images contain useful information for automaticannotation, indexing. The Process of Extraction of information is detection, localization, tracking,extraction, enhancement, and recognition of the text from a given image [9]. However, there aredifferences in text in style, orientation, size, and alignment, as well as low contrast image andcomplex background make the automatic text extraction problem more difficult and timeconsuming. While critical surveys of related problems such as document analysis face detectionand image & video indexing and retrieval can be found, the problem of text extraction isnotsurveyed well.

A variety of approaches to text extraction from images and video have beenpresented for many applications like address block location [14], content-based image/videoindexing [10, 16], page segmentation [12-13], and license plate location [11-15]. In spite of such incritical studies, it is still not easy to design a general-purpose Text Extraction system. This isoften a result of so many possible sources of variation once extracting text from complex images,or from images having difference in style, color, orientation, font size and alignment.Although images non-inheritable by scanning book covers, CD covers, or different multi-coloreddocuments have almost similar characteristics as the document images.

Text in video images can classify intocaption text and scene text. The caption text is artificially overlaid on the image and scene textexists naturally in the images. Some researchers prefer to use the term 'graphics text' for scenetext, and 'superimposed text' or 'artificial text' for caption text [17-18]. It is documented that scenetext is harder to detect.The text of input images need to be identified as the input image contains any text, the existenceor non-existence of text among the image. Several approaches assuming certain types of videoframe or image contain text (e.g., recording cases or book covers). However, in the case of video,the amount of frames containing text is far smaller than the amount of frames while not text. Thetext detection stage detects the text in image.

The unique properties of video collections (e.g., multiple sources, noisy features and temporal relations) examine the performance of these retrieval methods in such a multimodal environment, and identify the relative importance of the underlying retrieval Components. Based on query string matching videos are retrieve from database and sort it based on relevance. Video in the figure 4 the word "Sania" is extracted from the figure and is shown in figure 5.Text-based image retrieval has some limitations such as task of determiningimage content is highly perspective.

Figure 4 Result for Query video Sania

Figure 5 Query word "Sania"

Model-based approaches for automatic image annotation are based on the idea of finding a mapping between low-level image features and semantic concepts (e.g. sky, car, sea). This is achieved by analyzing a set of already labeled images, called the training set, and creating a corresponding prediction model. Model-based approaches can be classified into two categories: probabilistic modeling methods and classification-based methods. In probabilistic modeling, it aims to learn the joint probability distribution between image features and keywords. Classification-based approaches treat theproblem of automatic image annotation as a classification problem. For this purpose,each keyword is considered as an independent class and a classifier is learned to predictthe right class(s) of test images. A widely used method to construct the classifier is thetechnique of support vector machines[19-23].

Search-based Automatic Image Annotationretrieves a set of similar imagesfrom a large scale database of already labeled images, such as the web or specializedphoto sharing platforms, e.g., Flickr. Subsequently, tags/keywords of similar imagesare analyzed and propagated to the target image. More specifically, to identify similarimages, a two-phase search process is applied- Semantic/Contextual Search and Search by Image Contents.

Manual image annotation is a time consuming task and as such it is particularly difficult to be performed on large volumes of content. There are many image annotation tools available but human input is still needed to supervise the process. So, there should be a way to minimize the human input by making the annotation process fully automatic. In Automatic image annotation images are automatically classified into a set of pre-defined categories (keywords). Low-level features of the training images are extracted. Then, classifiers are constructed with low-level features to give the class decision. Lastly, the trained classifiers are used to classify new instances and annotate un-labelled images automatically. Automatic image annotation plays an important role in bridging the semantic gap between low-level features and high-level semantic contents in image access.

Photos represent one of the most common content types which are contributed and shared among the users of the Internet. This can be explained according to the availability of digital photography devices which provide an easy and a cheap medium for producing photos. At the same time, the bandwidth of the current Internet connections allows fast upload of photos. There are also several social aspects that make photos that popular. Photos are not only a documentary or reminders; they are also an emotional journal. Moreover, photos are a rich type of content that"is worth a thousand words", they capture our moods and feelings and provide a proof that we have been there. Additionally, photos represent a subtle means of social communication. People post their photos as a statement of positive affirmation regarding the way they live, what they do and what they achieved.

To address the limitations of manual tagging, research on automatic image annotation has received a considerable attention. Automatic image annotation aims at associating unlabeled images with keywords that describe their contents. Early research on automatic annotation techniques focused on using machine learning techniques. The idea is to use a dataset of already labeled images in order to train models for predicting labels for un-annotated images. However, creating good training datasets is a challenging and time consuming task. Indeed, most available datasets are limited to images corresponding to small set of predefined concepts. Therefore, the annotations generated by such approaches are also limited and they cannot meet the diverse ways in which people describe and search for images.

The aim of automatic image annotation is to generate descriptive keywords (tags) for unlabeled images without (or with only a little) human interference. Many methods have been proposed for automatic image annotation, which can be roughly categorized into two groups: keyword-based methods and ontology-based methods [19]. Keywords-based methods: Arbitrarily chosen keywords from controlled vocabularies, i.e. restricted vocabularies defined in advance, are used to describe the images. The basic goal of image annotation is presented in figure 6.

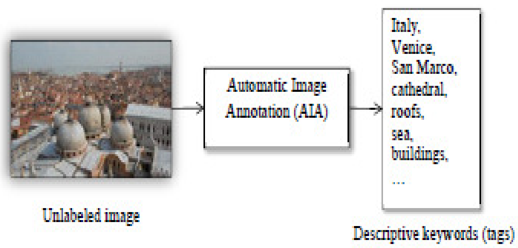

Figure6 The goal of automatic image annotation

Information Retrieval techniques are well-established, they are not effective when problems of concept ambiguity appear. On the other hand, neither search based only on semantic information may be effective, since: a) it does not take into account the actual document content, b) semantic information may not be available for all documents and c) semantic annotations may cover only a few parts of the document. Hybrid solutions that combine keyword-based with semantic-based search deal with the above problems. Developing methodologies and tools that integrate document annotation and search is of high importance. For example, researchers need to be able to organize, categorize and search scientific material (e.g., papers) in an efficient and effective way. Similarly, a press clipping department needs to track news documents, annotating specific important topics and searching for information.

Ontology based method is a way of describing concepts and their relationships into hierarchical categories [20]. This is 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 car or bus is a subclass of the class land vehicle, while car and bus have a disjoint relationship.

Ontology-based label extraction is extensively used to interpret the semantics found in image and video data. Particularly, ontology-based label extraction is one of the main steps in object class recognition, image annotation, and image disambiguation. These applications have important roles in the field of image analysis, and as such, a number of variations of the ontology-based label extraction used in these applications have been reported in the literature. These variations involve ontology development and utilization, and can affect the applicability (e.g., domain- and application-dependency) as well as the accuracy of the output. Unfortunately, the variability aspect of this variation has neither been established nor tracked. Thus, the variations were not configured.

Ontology is a conceptual knowledge source, which mainly consists of concepts and their hierarchical relationships. A concept is a tag identified by a word, phrase or label, and describes a real-world entity. Ontology may also have properties that describe the concepts and nonhierarchical relationships among the concepts of the ontology. Ontology may be used as a hierarchicallyenabled browsing mechanism and can be employed in semantics extraction, the process of accessing ontology and inferring knowledge based on its concepts and relationships. Ontology-based label extraction, a type of semantics extraction, produces true labels for an input image. It is shown in figure 7.

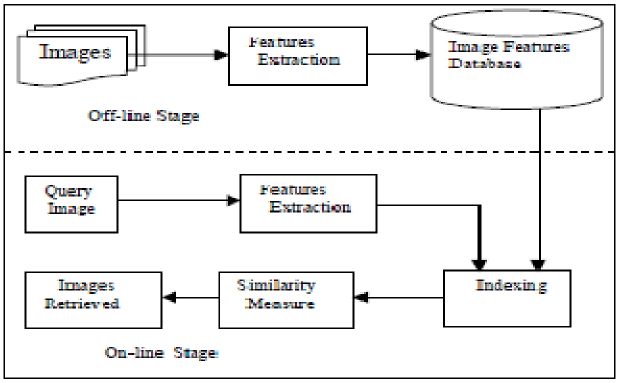

Figure 7 A typical Content-Based Image Retrieval system

Generally, given an input image, the ontology-based label extraction process has several steps. First, the input image features are projected and matched with concepts in the ontology through a process called mapping. The relationships connected to the matched concepts are then analyzed, and new concepts are identified sequentially until the final output is extracted. This process is called mining. Existing surveys mainly focus on a single application/problem (e.g., recognition, annotation, and disambiguation) and have reviewed the existing literature from several perspectives.

Generally, the existing literature focuses on comparing and analyzing methods based on the characteristics of the output, with no linkage to the technique and type of ontology used. The input for ontology-based label extraction may be image features or object labels (maps) extracted using various image annotation techniques [21-24]. The mapping procedure is constrained with the type of input and ontology characteristics. Features input, which have a wide range, require learning techniques. Meanwhile, maps can be mapped directly (e.g., using syntactic string matching). Ontology is task-independent and is developed by domain experts. Existing ontologies, such as WordNet [29] and Cyc [30], are upper-level ontologies that consist of a large number of concepts and their relationships. These ontologies may be used with various applications. However, an existing ontology may be customized depending on the task at hand and the desired output.

Ontology customization usually involves extracting a specific part of the ontology, which includes the required concepts and some of their relationships [25-26]. In addition, ontology-like knowledge may be developed if the required concepts or their relationships do not exist in the existing ontologies. The mining procedure depends greatly on the type of the output, that is, if the output is part of the input (i.e., image disambiguation), then a similarity technique is used; otherwise, a flooding procedure is implemented (i.e., image annotation). Image annotation and object recognition, as mentioned earlier, predict object(s) in a given scene based on the extracted features. Subsequently, these applications require an ontology that forms associations among features and labels for objects.

Generally, existing ontologies do not include the visual properties of the described objects [9, 27-28]. Thus, feature-based label extraction uses customized ontologies or ontology-like knowledge developed for the task at hand. The structure of these specific task ontologies depends on the task at hand and the desired output. Variations of this structure are reflected in the ways by which the required image features are represented.In the task-oriented category, ontology-like knowledge is developed to smoothly fit the task at hand. These ontologies, however, cannot be used elsewhere. Two main approaches, standard and advanced approaches, are then proposed. Their main differences are in the information conveyed by their ontologies, which require the use of different techniques. In the standard approach, the ontology conveys the following information: object labels, hierarchical relationships, and low-level features. In the advanced approach, the ontology has an additional feature, i.e., the spatial relationships among concepts. During the ontological construction, concepts are created based on labels obtained from a dataset of labeled images. Then, another set of concepts with coarse granularity is manually created to facilitate the categorization principle of the ontology. Low-level features are then assigned as properties to each concept using a supervised machine learning process.

In label extraction, features are extracted from the input image, labeled, and then mapped to properties in the ontology using a classification method. Mining is implemented as a propagation process, which transfers from one concept to another over the hierarchical relations in a topdown manner (from the concepts at the general level to the concepts at a specific level). The propagation process might be intermediate and have more classification processes, in order to filter out the concepts reached through the propagation process. Finally, the concepts obtained at the lowest level (i.e., leaf) of the propagation process are selected as the output.

For all its promising edge, search-based image annotation has to deal with several challenges. The first challenge is posed by community tags as a main resource from which annotations (for unlabeled images) are extracted. User-tags are created in an uncontrolled and free-style manner, thus, they are inherently noisy. Humans use inconsistent terms to describe the same thing or use the same term to express different meanings. In other words, polysemy and homonymy – two fundamental problems in information retrieval – are also present in user-provided tags. Second, as mentioned before, identifying images similar to the un-annotated image is a core component of the automatic annotation process. Accordingly, automatic image annotation has also to deal with two main challenges of CBIR techniques, namely the accuracy and the speed of the applied technique. Generally, the accuracy of CBIR is ruled by the low level image representation that is used, i.e., image features. In turn, the complexity of extracting image features, representing them as descriptor vectors and comparing the descriptors are major factors that influence the retrieval speed. Therefore, in order to ensure the efficiency of automatic image annotation, solutions for improving the accuracy and boosting the performance of the applied CBIR process have to be investigated. Third, automatic image annotation has to address the issue of estimating the relevance/importance between candidate annotations and the target image.

In general, the problem of CBIR is the semantic gap between the high-level image and the low-level image. In other words, there is a difference between what image features can distinguish and what people perceives from the image. As shown in Fig. 4, SBIR can be made by extraction of low-level features of images to identify meaningful and interesting regions/objects based on the similar characteristics of the visual features. Then, the object/region features will go into semantic image extraction process to get the semantics description of images to be stored in database.Image retrieval can be queried based on the high-level concept.Query may be done based on a set of textual words that will go into semantic features translator to get the semantic features from the query. The semantic mapping process is used to find the best concept to describe the segmented or clustered region/objects based on the low features. This mapping will be done through supervised or unsupervised learning tools to associate the lowlevel features with object concept and will be annotated with the textual word through image annotation process[1,13]. Semantic content obtained either by textual annotation or by complex inference procedures based on visual content[14].

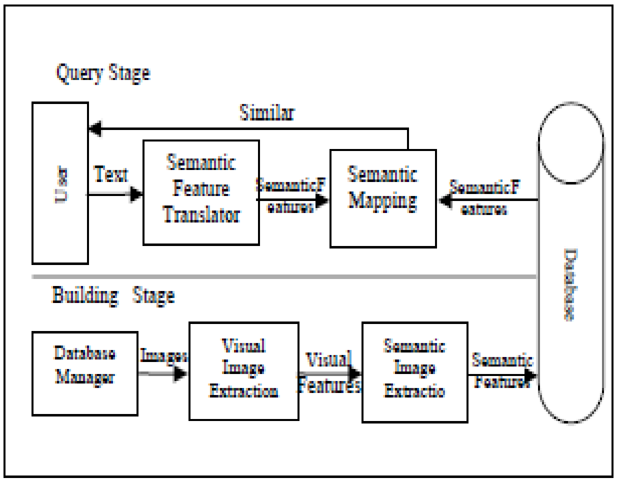

Figure 4. A typical Semantic-Based Image Retrievalsystem

The semantic annotation means to describe the semantic content in images and retrieval queries. It requires some understanding of the semantic meaning in images and retrieval query, and standardization of representation of images. Based on the semantic annotation of images and retrieval queries, semantic similarity between images and a retrieval query can be compared. At present, semantic annotation is implemented by some markup language such as XML based on a shared ontology definition.The semantic mapping process is used to find the best concept to describe the segmented or clustered region/objects based on the low features. This mapping will be done through supervised or unsupervised learning tools to associate the low-level features with object concept and will be annotated withthe textual word through image annotation process[1,13]. Semantic content obtained either by textual annotation or by complex inference procedures based on visual content[14].

Conclusions

This paper attempted to provide an overview of the most common techniques of different types of image retrieval systems. Most systems used low-level features, few systems used semantic feature. Global featuresfail to identify important visual characteristics of images but it's very efficient in computation and storage due to its compact representation. From another perspective, local features that can be extracted from images handle partial image matching or searching for images that contain the same object or same scene with different viewpoints, different scale, changes in illumination,etc. Therefore,local features can identify important visual characteristics of images but it ismore expensive computationally. The semantic features that based on keywords or annotationsmaybe very subjective and time consuming. Whereas,the semantic features that based on visual contentis complex becauseof the inference procedures. Automatic image annotation is good approach toreduce the semantic gap, butit still achallenging task due tothe different conditions of imaging, occlusionsand the complexity, and difficulty to describe objects. In future, there is a need to work more and more with available techniques to deal with the semantic gap to enhance image retrieval.

Bridging the semantic gap for image retrieval still considered a big challenge.

Even though there are a lot of efforts and works on image retrieval research, but it is not enough to provide satisfactory performance. However, there are still some spaces, which need to be improved besides the challenges that is associated with mapping low level to high-level concepts. Also overcome of the semantic gap in the broad domain database is complex because the images in broad domains can be described using various concepts.There are needto see better supportfor the image retrievalbased semantic concept with a focus on the retrieval by abstract attributes, involving a significant amount of high-levelreasoning about the meaning and purpose of the objects. In addition, the extracted semantic features should be applied for any kind of image collection. Moreover, there is need to effective ways retrieve of similar images that are conform to human perception and without human interference.

References

[1] Riad, A. M., Atwan,A., and Abd El-Ghany,S,“ Image Based Information Retrieval Using Mobile Agent”, Egyptian Informatics Journal, Vol.10, No.1, 2009.

[2] Kherfi, M.L., Ziou, D., and Bernardi,A , “Image retrieval from the World Wide Web: issues, techniques, and systems”, ACM Computing Surveys Vol. 36, No. 1, pp.35–67, 2004.

[3] Riad, A. M., Atwan,A., and Abd El-Ghany,S, “ Analysis of Performance of Mobile Agents in Distributed Content Based Image Retrieval”, In Proc. IEEE international Conference on Computer Engineering & Systems, ICCES 2008.

[4] Wang, C., Zhang, L., and Zhang, H,“Learning to Reduce the Semantic Gap in Web Image Retrieval and Annotation”, In SIGIR‟08, Singapore, 2008.

[5] P.Jayaprabhaand Rm.Somasundaram, “Content Based Image Retrieval Methods Using Graphical Image Retrieval Algorithm (GIRA)”, International Journal ofInformation and Communication Technology Research, Vol. 2 No. 1, 2012.

[6] T.Karthikeyan, P.Manikandaprabhu, "Function and Information DrivenFrameworks for Image Mining - A Review", International Journal ofAdvanced Research in Computer and Communication Engineering(IJARCCE), Vol.2, Issue 11, pp.4202-4206, Nov. 2013.

[7] S.K. Chang and A. Hsu, “Image information systems: Where do we gofrom here?” IEEE Trans. on Knowledge and Data Engineering 4(5),1992.

[8] H. Tamura and N. Yokoya, “Image database systems: A survey”, PatternRecognition 17(1), 1984.

[9] Jung, Keechul, Kwang In Kim, and Anil K Jain. "Text information extraction in images and video: asurvey." Pattern recognition 37, no. 5 (2004): 977-997.

[10] H. J. Zhang, Y. Gong, S. W. Smoliar, and S. Y. Tan, Automatic Parsing of News Video, Proc. OfIEEE Conference on Multimedia Computing and Systems, 1994, pp. 45-54.

[11] Y. Cui and Q. Huang, Character Extraction of License Plates from Video, Proc. of IEEE Conferenceon Computer Vision and Pattern Recognition, 1997, pp. 502 –507.

[12] A. K. Jain, and Y. Zhong, Page Segmentation using Texture Analysis, Pattern Recognition, 29 (5)(1996) 743-770.

[13] Y. Y. Tang, S. W. Lee, and C. Y. Suen, Automatic Document Processing: A Survey, Pattern

Recognition, 29 (12) (1996) 1931-1952.

[14] B. Yu, A. K. Jain, and M. Mohiuddin, Address Block Location on Complex Mail Pieces, Proc. OfInternational Conference on Document Analysis and Recognition, 1997, pp. 897-901.

[15] D. S. Kim and S. I. Chien, Automatic Car License Plate Extraction using Modified GeneralizedSymmetry Transform and Image Warping, Proc. of International Symposium on IndustrialElectronics, 2001, Vol. 3, pp. 2022-2027.

[16] J. C. Shim, C. Dorai, and R. Bolle, Automatic Text Extraction from Video for Content-basedAnnotation and Retrieval, Proc. of International Conference on Pattern Recognition, Vol. 1, 1998, pp.618-620.

[17] S. Antani, D. Crandall, A. Narasimhamurthy, V. Y. Mariano, and R. Kasturi, Evaluation of Methodsfor Detection and Localization of Text in Video, Proc. of the IAPR workshop on Document AnalysisSystems, Rio de Janeiro, December 2000, pp. 506-514.

[18] S. Antani, Reliable Extraction of Text from Video, PhD thesis, Pennsylvania State University,August 2001.

[19] Zhang, D., Islam, M.M., Lu, G. 2011. A review on automatic image annotation techniques. In Journal of Pattern Recognition, pp: 1-17.

[20] Feng, S., and Xu, D. 2010.Transductive Multi-Instance Multi-Label learning algorithm with application to automatic image annotation. In Expert Systems with Applications 37, pp: 661–670.

[21] A. A. Abu-Shareha and R. Mandava, Semantics Extraction in Visual Domain Based on WordNet, 5th FTRA International Conference on Multimedia and Ubiquitous Engineering (MUE), IEEE, 2011, pp. 212-219.

[22] M. Atika, A. Akbar and A. Sultan, Knowledge Discovery using Text Mining: A Programmable Implementation on Information Extraction and Categorization, International Journal of Multimedia and Ubiquitous Engineering, 4 (2009).

[23] S. Banerjee and T. Pedersen, Extended Gloss Overlaps as a Measure of Semantic Relatedness, Proceedings of the 18th International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers Inc., Acapulco, Mexico, 2003, pp. 805-810.

[24] S.-F. Chang, T. Sikora and A. Purl, Overview of the MPEG-7 standard, IEEE Transactions on Circuits and Systems for Video Technology, 11 (2001), pp. 688-695.

[25] R. Clouard, A. Renouf and M. Revenu, An Ontology-Based Model for Representing Image Processing Application Objectives., International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 24 (2010), pp. 1181-1208.

[26] T. M. Deserno, S. Antani and R. Long, Ontology of gaps in content-based image retrieval, Journal of Digital Imaging, 22 (2009), pp. 202-215.

[27] H. J. Escalante, M. Montes and E. Sucar, Multimodal indexing based on

semantic cohesion for image retrieval, Information retrieval, 15 (2012), pp. 1-32.

[28] X. Fan, Contextual disambiguation for multiclass object detection, International Conference on Image Processing, 2004, pp. 2873-2876.

[29] C. Fellbaum, WordNet: An Electronic Lexical Database, MIT Press (1998).

[30] D. Foxvog, Cyc, Theory and Applications of Ontology: Computer Applications, Springer, 2010, pp. 259-278.