

A TUTORIAL REVIEW OF AUTOMATIC IMAGE TAGGING TECHNIQUE USING TEXT MINING

Sayantani Ghosh¹, Samir Kumar Bandyopadhyay²

¹Department of Computer Science and Engineering, University College of Science, Technology and Agriculture, University of Calcutta, ²Dept. of Computer Science & Engineering, University of Calcutta, 92 A.P.C. Road, Kolkata – 700009, India, sayantani.mailme@gmail.com, skb1@vsnl.com

Abstract

With the advent of time, the number of images being captured and shared online has grown exponentially. The images which are captured are later accessed for the purpose of searching, classification and retrieval operation. Hence these images must be labelled with appropriate words, phrases or keywords so that the requisite operation can be performed efficiently. Automatic Image Tagging is such a technique which associates an appropriate keyword from a given set of words or phrases based on the relevance to the content of the image. This selection of the appropriate keyword can be performed by Text Mining which is concerned with the extraction of appropriate information from a given text.

The main objective of this paper is to depict how Text Mining technique can be implemented in the process of Automatic Image Tagging. In order to annotate an image, techniques like Content Based Image Retrieval can be used, which emphasises on the content of the image to annotate an image. However due to several constraints of the above mentioned technique, Automatic Image Tagging technique is used which chooses a tag from a given set of tags to annotate an image. The selection of the appropriate tag can be implemented using the Classification logic of the Text Mining Algorithm that assigns the given set of keywords or tags to some predefined classes. In this way the most relevant tags can be selected assigned to the given image.

Keywords: Tagging, Image Annotation, Automatic Image Tagging, Linguistic Indexing, Content Based Image Retrieval, Automatic Annotation.

1. INTRODUCTION

A picture is a resemblance of past memories which is cherished by every individual all their life. Over the years the numbers of pictures being captured and shared have grown exponentially. There are several factors responsible for this growth. Firstly, in present days the digital cameras allow people to capture, edit, store and share high quality images with great ease compared to the old film cameras. Secondly, the availability of low cost of memory and hard disk drives. Thirdly, the popularity of social networking sites like Facebook, MySpace have given the user an additional interest to share photos online with their friends across the globe.

With this rapid growth, there arises the need to perform effective manipulation (like searching, retrieval etc...) on images. Several search engines retrieve relevant images by text-based searching without using any content information. However, recent research shows that there is a semantic gap between content based image retrieval and image semantics understandable by humans. As a result, research in this area has shifted to bridge the semantic gap between low level image features and high level semantics. Thus, assigning relevant keywords is significant and can improve the image

search quality. This is known as Image Annotation. It is defined as technique of assigning semantically relevant keywords to an image.

The typical method of bridging the semantic gap is through the automatic image annotation (AIA) which extracts semantic features using machine learning techniques. **Automatic image annotation** (also known as **automatic image tagging** or **linguistic indexing**) is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database.

Apart from Automatic Image Annotation, annotation can also be done manually. However the latter technique being time consuming and involving considerable overhead motives the use of the former technique i.e Automatic Image Tagging.

1.1 Tagging - Definition

Tags are Meta data used to describe a piece of data like a web page, a digital photo, or another type of digital document. It can be defined as a non- hierarchical keyword or term assigned to a piece of information. The term "tagging" is used in context of organizing digital photos. Adobe's Photoshop Album digital photo organizer software brought the tagging concept to the mainstream for digital photography, and the popular online photo sharing service Flickr also helped to spur the trend.

1.2 Characteristics:



Figure 1 This picture can be tagged, as 'sea', 'mountain', 'nature'

A tag attain the following characteristics (from Figure 1)

- Tags are chosen by the creator and/or by the viewer of the tagged item.
- Tags are not part of a formal subject indexing term set.
- Tags are informal and personal.

1.3 Objective:

The main objective is to simplify the retrieval and search operation for an image in a photo library. Besides, in the commercial world, assigning a meaningful label or keywords to an image increases the efficiency in satisfying the needs of consumers; as an incorrectly or insufficiently labelled image is unlikely to be found, particularly within the stringent deadlines commonly experienced within the commercial world, thereby leading to a loss in operational efficiency.

1.4 Advantages of Photo Tagging:

- Tagging takes away the decision-making process of choosing the right category and risk the possibility of ending in a poorly defined sub-category.
- Tagging allows you to include associated concepts without wondering whether you have categorized the item in the correct folder or not.
- Tagging helps to create communities as users with similar interests gravitate toward similar, searchable word tags.
- Tagged content increases the amount of usable retrievals by providing more than one place you can look for information on the same topic.

1.5 Challenges of Photo Tagging:

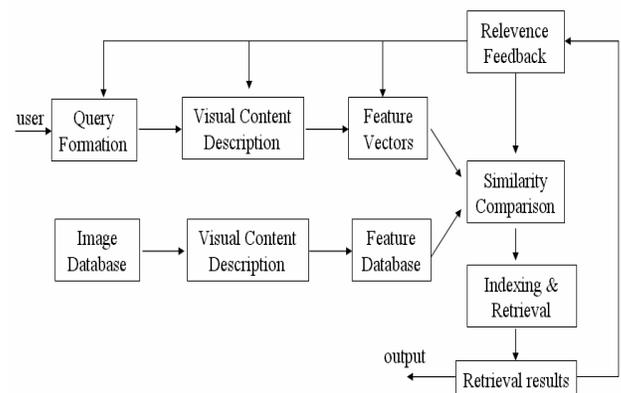
- One of the most common problems is the misspelling of tags.
- Another issue is deciding on the content being tagged. Everyone has different perceptions of what she is reading. Individual tags may tend to be disjointed, irrelevant, and often very messy. They lack precision and there is no ability to control synonyms or related terms.
- Tags with more than one meaning and can lead the searcher into an undesired area. Personalized tags are clear to a handful of people, but are not universally understood.

2. CONTENT BASED IMAGE RETRIEVAL (CBIR)

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases.

Content based means that the search will analyze the actual contents of the image.

"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 colours, shapes, textures, or any other information that can be derived from the image itself.



Content-based image retrieval uses the visual contents of an image such as colour, shape, texture, and spatial layout to represent and index the image. In typical content-based image retrieval systems (Figure), the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with example images or sketched figures.

The system then changes these examples into its internal representation of feature vectors. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. The indexing scheme provides an efficient way to search for the image database. Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results.

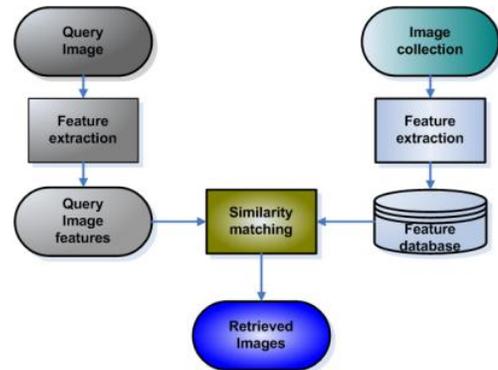
2.1 Objective of CBIR

CBIR systems aim to recover pictures from large image repositories, according to the user's interest. This technique is for retrieving images on the basis of automatically-derived features such as colour, texture and shape.

CBIR draws many of its methods from the field of image processing and computer vision, and is regarded by some as a subset of that field. It differs from these fields principally through its emphasis on the retrieval of images with desired characteristics from a collection of significant size. Image processing covers a much wider field, including image enhancement, compression, transmission, and interpretation.

2.2 Implementation of CBIR

1. Input component is the query image. This image is to be matched in the image database.
2. In Content Based Image Retrieval System, Feature Extraction module performs extraction of features of the input image. The feature extraction is normally based on colour, texture and shape of an image, where colour content of an image is the most widely used feature used by CBIR in the process of feature extraction. The three types of image features are utilized in various CBIR applications ranging from scene/object and fingerprint identification and matching to face and pattern recognition.
3. The similarity matching is performed between the extracted features of the query image with the features of the image stored in the image feature database. Similarity matching, through metrics called similarity measures, is done to determine the degree of relevance of an image in a collection to a query. Similarity matching is a key component of a content-based image retrieval (CBIR) system because finding a set of images similar to the image the user had in mind is its primary goal.
4. The feature database contains the extracted features of the images in collection.
5. Finally an image or a set of images are extracted as output if the similarity matching between the query image and the images in the image database is successful.



2.3 Advantages of CBIR

- CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results.
- 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.
- It provides better indexing and returns more accurate results.
- It does not involve additional overhead of assignment of special words or keywords to an image, hence it is less time consuming compared to automatic image annotation.

2.4 Challenges of CBIR

- The retrieval technique finds it difficult in locating a desired image in a large and varied collection.
- CBIR system can operate only at the primitive level.
- Although CBIR is based on shape, colour and texture of the image, but it does not combine global colour, colour region, colour sensation, shape and qualitative spatial relation features to query an image.
- It faces the problem of matching the human visual system that has evolved over a period of more than 100 million years as the analysis of image data by human observer is far more difficult than the interpretation of text by human observer.
- In text retrieval system users are free to compose queries using several words but this not the case with image retrieval where the user is usually asked to provide a query image or a sketch.

3. AUTOMATIC IMAGE ANNOTATION

Due to several challenges faced by the Content Based Image Retrieval System as mentioned in the previous section, an alternative image retrieval technique has been recognized and used widely in recent times. Previously image annotation for the purpose of indexing and later retrieval of the requisite image or images from the image collections was done manually. However this method is considerably laborious and expensive, hence an automatic image annotation technique is used to annotate images based on its content.

Automatic Image Tagging also known as Automatic Image Annotation is an annotation technique which assigns tags or captions to an image relevant to an image to increase the efficiency of image retrieval and classification.

The idea with automatic image tagging is that tags are automatically captioned and assigned to the digital image. These tags should describe every important part or aspect of the image and its context. Automatic image tagging can be done based on the visual content of the image, contextual information, or using a mixture of these two approaches.

Thus the main objective of Automatic Image Annotation technique is to bridge the semantic gap between the image retrieval and the image semantics understandable by humans.

3.1 Automatic Image Annotation Models

The task of developing accurate and robust automatic image tagging models entails following challenges.

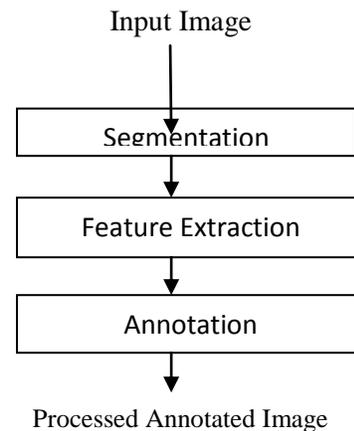
1. The availability of large and correctly annotated image databases for the training and testing of the new annotation models.
2. The extraction of useful image and text features for the construction of reliable annotation models.

Therefore the use of several natural language resources can be explored to construct image annotation models that are capable of automatically tagging images from unrestricted domains with good accuracy. Unlike traditional image annotation methodologies that generate tags using image – based features, an alternate method based on keyword extraction can be used.

Automatic Image Annotation can be defined as the probability of assigning keywords to an image on the basis of its content. The technique can be implemented using the following models.

1. **Co-occurrence Model** which concentrates at the co-occurrence of words with image regions created using a regular grid.
2. **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.

3.2 Steps involved in Image Annotation



1. Segmentation

Image segmentation is useful in many applications. It can identify the regions of interest in a scene or annotate the data. The segmentation component partitions images into local contents via either some block or region based method. The segmentation of an image includes global and local features of an image. Global features consider the visual features of the whole image; they cannot completely describe different parts of an image. On the other hand, image segmentation into local contents (i.e. different regions or areas) is able to provide more detailed information of images.

2. Feature Extraction

Feature Extraction is based on Global and Local features. Global feature includes all the pixels of an image. This extraction is used to represent the global colour of an image. There are two strategies for extracting local features. The first one is to partition a set of fixed sized blocks or tiles and the second for a number of variable shaped regions of interest. After performing block and/or region based segmentation, low-level features can be extracted from the tiles or regions for local feature representation. Low-level features such as colour, texture, shape, and spatial relationship are extracted to represent image features.

3. Annotation

Annotation is a technique of associating comments, notes, keywords, or other form of external words to a document without causing any change in the document itself. Annotations are metadata that add additional information about the existing piece of data.

There are three types of image annotation: manual, automatic and semi automatic.

Manual annotation needs users to enter some descriptive keywords when perform image browsing.

Automatic annotation detects and labels semantic content of images with a set of keywords automatically.

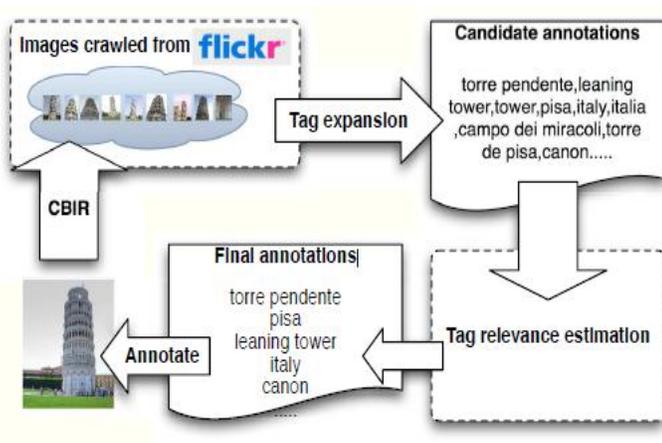
In case of Semiautomatic annotation, it needs user's interaction to provide an initial query and feedback for image annotation while browsing.

3.3 Purpose of Automatic Image Annotation

The search and the retrieval operation of requisite image by a search engine like google is usually performed based on text without much emphasis on the content of the image. Therefore assigning suitable keywords to images is a vital task. This is achieved by automatic image annotation, which assign specific words or keywords to an image relevant to the content of the image thereby bridging the semantic gap and achieving multi-level image annotation automatically.

Besides, automatic image annotation technique is used as it is less tedious and minimizes time requirement for annotation of images compared to manual annotations. This can be achieved in two steps.

1. Find an accurate keyword relevant to the query image.
2. Given one keyword, annotate the image to describe the details of this image.



Flickr is one of the most comprehensive image resources on the web. As a photo management and sharing application, it provides user with the ability to tag, organize and share their photos online. Flickr offers a public API which can be used to write applications that use Flickr in some way.

The objective of Automatic Image Annotation is to search over user- contributed photo sites like Flickr which have accumulated rich human knowledge and billions of photos, then associate surrounding tags from those visually similar Flickr photos for the unlabeled image.

For an unlabelled image, photos in the social media are extracted by the automatic image annotation technique, the annotations associated with the images are expanded, and then ranks are associated with the images, based on their relevances to the given image in terms of visual consistency and semantic consistency. Then amongst the expanded annotations, candidate annotations are retrieved and assigned to the requisite image.

3.4 Advantages of Automatic Image Annotation

1. High performance and less time- consuming.
2. It can perform effective manipulation for exponentially growing photo collection.
3. It is more flexible than the manual photo tagging.
4. Robust.

3.5 Challenges of Automatic Image Annotation

1. It is very difficult for automated algorithms to capture content and contextual information from images that do not have any associated image features.
2. Generic object limitation, which questions the use of very generic tags for the images such as "sun", "grass" and "tiger". That is most users tend to issue queries that refer to objects by proper name which usually have limited associated visual stimuli in images.
3. It is very difficult for automated algorithms to capture content and contextual information from images that do not have any associated image features.

3.5.1 Implementation – Test Case



CASE 1

Keywords: Red, Green, Flower, Yellow, Rose, Jasmine, Lotus.

Automatic Annotation: Yellow Rose Flower.

CASE 2

Keywords: Red, Green, Yellow, Jasmine, Lotus, Sunflower, Fruit, Rose

Automatic Annotation: Yellow Rose.

CASE 3

Keywords: Red, Green, Jasmine, Lotus, Sunflower, Fruit.

Automatic Annotation: Insufficient keywords for annotation.

4. TEXT MINING

Text mining can be defined as a technique that attempts to extract meaningful information from natural language text. It can be characterized as the process of analyzing text to extract information that is useful for a particular purpose. Compared with the kind of data stored in databases, text is unstructured, amorphous, and difficult to deal with algorithmically. However, text is the most common vehicle for the formal exchange of information.

The phrase “text mining” is generally used to denote any system that analyzes large quantities of natural language text and detects lexical or linguistic usage patterns in an attempt to extract meaningful information.

Text mining problem can be implemented using the following algorithms:

4.1 Classification Algorithm

The objective of Classification is to discover a model for the class in terms of the remaining attributes. The objective is to use the training data set to build a model of the class label based on the other attributes such that the model can be used to classify new data not from the training data set.

The different type of classification models are as follows:

1. Decision Tree
2. Neural Network
3. Genetic Algorithm

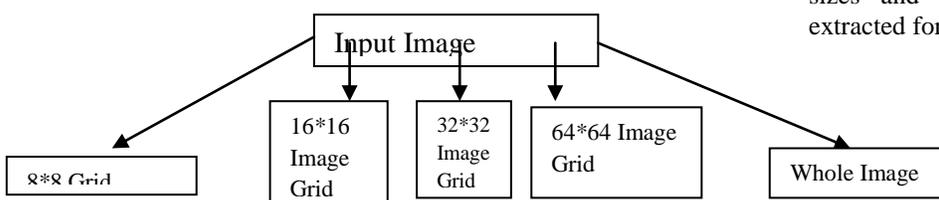
4.2 Association Algorithm

The objective of the Association Algorithm is to discover associations present in the data. The problem was formulated originally in the context of the transaction data at supermarket. This market basket data, as it is popularly known, consists of transactions made by each customer. Each transaction contains items bought by the customer. The goal is to see if occurrence of certain items in a transaction can be used to deduce occurrence of other items, or in other words, to find associative relationship between items.

Association Algorithm is of the following type:

1. Parallel Algorithm for discovering Association.
2. Sequential Algorithm for discovering Association

4.3 Clustering Algorithm



Clustering is a division of data into groups of similar objects. Each group, called cluster, consists of objects that are similar between themselves and dissimilar to objects of other groups. Representing data by fewer clusters necessarily loses certain fine details, but achieves simplification. It represents many data objects by few clusters, and hence, it models data by its clusters.

Data modelling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis.

Clustering Algorithm is of the following type:

1. Hierarchical Methods
2. Partitioning Methods.

5. ROLE OF TEXT MINING IN AUTOMATIC IMAGE TAGGING

Text mining, sometimes alternately referred to as text data mining, roughly equivalent to text analytics, refers to the process of deriving high-quality information from text. High-quality information is typically derived through the divining of patterns and trends through means such as statistical pattern learning. Text mining usually involves the process of structuring the input text (usually parsing, along with the addition of some derived linguistic features and the removal of others, and subsequent insertion into a database), deriving patterns within the structured data, and finally evaluation and interpretation of the output.

In the process of Automatic Image Tagging to increase the efficiency of image retrieval and classification, uses the technique of Text Mining to construct image annotation models that are capable of automatically tagging images from unrestricted domains with good accuracy. Given a target image and its surrounding text, we extract those words and phrases that are most likely to represent meaningful tags.

Proposal

Consider an image of nature. The input image is segmented into several components and features are extracted from the components. In order to perform the feature extraction a multi-resolution grid-based framework for image content representation and feature extraction is performed, where the images are partitioned into a set of regular grids with different sizes and multi-modal visual features are automatically extracted for each image grid.

Then A tree structure is defined that contain a predefined set of classes that includes the following classes:

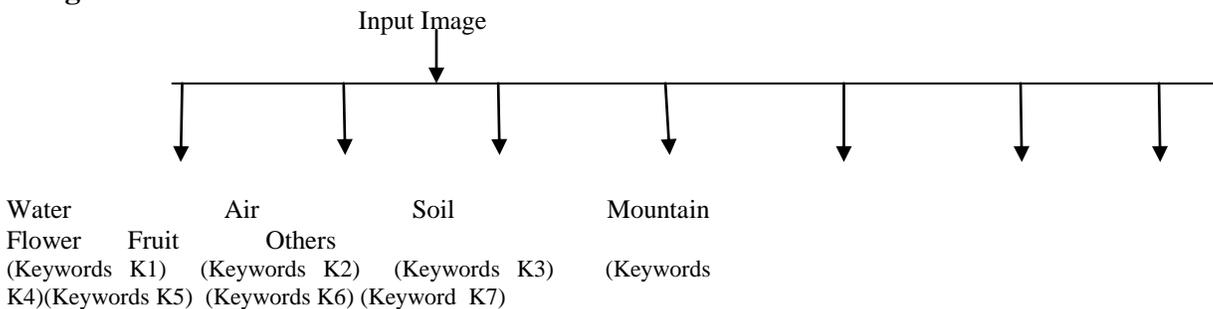
Water, Air, Soil, Mountain and Others.

When an input image is to be classified, based on its features it will be assigned to the predefined set of classes on the basis of the following assumptions:

Assumptions:

- With every class a set of keywords will be associated which is relevant to that class.
- When performing the classification, pick the first keyword from the query and associate it with the best set of match(s) that can be found from all the subclasses at level 1.
- It might be possible that one or more subclasses at level 1 has a keyword that may match the query string since

Design:



Keywords K1: The K1 set is associated with the class 'Water.' This set will contain all those keywords that is relevant to the class 'Water'.

K1={River, Sea, Pond, Lake, Blue}

Keywords K2: The K2 set is associated with the class 'Air.' This set will contain all those keywords that is relevant to the class 'Air'.

K2= {Windy, Odour}

Keywords K3: The K3 set is associated with the class 'Soil.' This set will contain all those keywords that is relevant to the class 'Soil'.

K3= {Tree, Plant, Grass, Green, Vegetables}

Keywords K4: The K4 set is associated with the class 'Mountain.' This set will contain all those keywords that is relevant to the class 'Mountain'.

K4= {Hill}

some features/functionality may be common in multiple classes.

- Form a set of all the classes that has a match with the query string (an exact match is preferred since we are assuming that the text mining algorithm will search only for exact matches and also to make the search domain smaller).
- The set formed has narrowed down our search domain to a smaller subset. The next step is to identify that in which set the query string exactly belongs to.
- To do this pick the next data from the query string is selected and search for a match in the subset formed in the previous step.
- If multiple matches are again found repeat the previous steps.

Keywords K5: The K5 set is associated with the class 'Flower.' This set will contain all those keywords that is relevant to the class 'Flower'.

K5= {Red, Yellow, Pink, Blue, Green, Rose, Jasmine, Lotus, Sunflower, Leaf}

Keywords K6: The K6 set is associated with the class 'Fruit.' This set will contain all those keywords that is relevant to the class 'Fruit'.

K6= {Red, Green, Yellow, Orange, Pink, Apple, Mango, Orange, Guava, Strawberry, Banana, Watermelon}

Keywords K7: The K7 set is associated with the class 'Others.' This set will contain all those keywords that do not belong to the other classes.

Example:

Consider the above Image. It is an image of a Yellow Rose with Green leaf. Let us assume that this is the image that is to be tagged with the appropriate keyword(s). The image is associated with a set of keywords from which the relevant keywords are to be selected and associated with the image. The keyword set includes the following.

$K = \{ \text{Red, Fruit, Name, Yellow, Rose, Green Leaf, Brown, Sunshine} \}$

In order to accomplish this task, the following steps are followed.

Step1 : Firstly we perform the segmentation and feature extraction of the input image.

Step2: The resultant image is then classified into the subclasses.

Step3: The above image is therefore classified into the class 'Flower'.

Step 4: The keywords set 'K' associated with the image is then Matched with the keyword set 'K5' to choose the appropriate keywords that is relevant to the input image.

Step 5: The input image is then annotated with the selected keywords.

Output: The output keyword which will be used to annotate the image set includes 'Yellow' 'Rose' 'Green' 'Leaf'
 $K' = \{ \text{Yellow, Rose, Green, Leaf} \}$.

CONCLUSIONS

People love to take images, but are not so willing to annotate the images afterwards with relevant tags. A requirement for effective searching and retrieval of images in rapid growing online image databases is that each image has accurate and useful annotation.

Handling large volumes of digital information becomes vital as online resources and their usage continuously grows at high speed. Online image sharing applications are getting extremely popular. Flickr1 is one of the most popular of these applications hosting over 4 billion2 images.

The idea with automatic image tagging is that tags are automatically captioned and assigned to the digital image. These tags should describe every important part or aspect of

the image and its context. Automatic image tagging can be done based on the visual content of the image, contextual information, or using a mixture of these two approaches.

By looking at the visual content of an image, it could for example be possible to predict that an image where most edges are vertical or horizontal contains a building. Another approach is to find a set of images that are visually similar to the query image in existing image databases consisting of already tagged images, and then pick the most relevant tags from the set of similar images. The context in which the image was taken can also be used to tag images.

REFERENCES

- [1]. <http://annamaebell.wordpress.com/2007/10/09/advantages-and-challenges-of-tagging-and-folksonomy/>
- [2]. <http://informatica.uv.es/~ferri/TAR3/prl08arevalillo.pdf>
- [3]. http://www.cse.iitd.ernet.in/~pkalra/siv864/Projects/ch01_Long_v40-proof.pdf
- [4]. www.jisc.ac.uk/uploaded_documents/jtap-039.doc
- [5]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.154&rep=rep1&type=pdf>
- [6]. http://www.theopavlidis.com/technology/CBIR/PaperB/ver_s3.htm
- [7]. <http://www.asis.org/Bulletin/Feb-07/rorissa.html>
- [8]. citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.103.154&rep=rep1&type=pdf - United States
- [9]. <http://www.inf.ed.ac.uk/publications/thesis/online/IM090693.pdf>
- [10]. http://www.lehigh.edu/~hol207/CVPR10_group_sparsity.pdf
- [11]. <http://www.aclweb.org/anthology-new/C/C10/C10-2074.pdf>
- [12]. <http://www.csie.ntu.edu.tw/~winston/papers/hsieh10searchbased.pdf>
- [13]. http://www.sanjivk.com/ImageAnnotation_IJCV10.pdf
- [14]. <http://www.benthamsience.com/cseng/samples/cseng1-1/Tsai.pdf>
- [15]. <http://imageclef.org/SIAPRdata>
- [16]. www.sciacademypublisher.com/journals/index.php/IJRRIS/.../13/9
- [17]. <http://disp.ee.ntu.edu.tw/meeting/%E6%98%B1%E7%BF%94/Segmentation%20tutorial.pdf>
- [18]. www.sciacademypublisher.com/journals/index.php/IJRRIS/.../13/9
- [19]. <http://coitweb.uncc.edu/~jfan/fp03c-fan.pdf>