A Survey on Image Retrieval Methods

A. Malcom Marshall[1] and **Dr.S.Gunasekaran**[2]

[1]M.E-Computer Science and Engineering
CIET-Coimbatore
malcom.research@gmail.com

[2]Prof and Head- Computer Science and Engineering
CIET-Coimbatore
gunaphd@yahoo.com

ABSTRACT

The Image retrieval plays a key role in day-to-days world. This work is a review of various references of various image retrieval methods. This paper starts with discussing the working conditions of text based image retrieval then the content-based retrieval: patterns of use, levels, the role of semantics, and the semantic gap. We briefly discuss about various techniques of content based image retrieval such as retrieval by color, shape and the texture and the various algorithms involved in content based image retrieval. Then the semantic based image retrieval aspects are discussed using local content descriptors the regions are segmented and retrieved the semantic regions of image.

Keywords: Image retrieval, color descriptor, Segmentation

1 INTRODUCTION

Image retrieval (IR) has become an important research area in computer vision where digital image collections are rapidly being created and made available to multitudes of users through the World Wide Web. Tremendous increment in the collections of images from art museums, medical institutes, and environmental agencies, to name a few. In the commercial sector, companies have been formed that are making large collections of photographic images of real-world scenes available to users who want them for illustrations in books, articles, advertisements, and other media meant for the public at large. Incredibly, the indexing of these images is all being done manually–a human indexer selects and inputs a set of keywords for each image. Each keyword can be augmented by terms from a thesaurus that supplies synonyms and other terms that previous users have tried in searches that led to related images. Keywords can also be obtained from captions, but these are less reliable. Content-based image retrieval research has produced a number of search engines. The commercial image providers, for the most part, are not using these techniques. The main reason is that most CBIR systems require an example image and then retrieve similar images from their databases. Real users do not have example images; they start with an idea, not an image. Some CBIR systems allow users to draw the sketch of the images which they wanted. Such systems require the users to have their objectives in mind first and therefore can only be applied in some specific domains, like trademark matching, and purchase of painting. Thus the recognition of generic classes of objects and concepts is essential to provide automated indexing of images for CBIR. However, the task is not easy. Computer programs can extract features from an image, but there is no simple one-to-one mapping between features and objects. Earlier CBIR systems rely on global image features, such as color histogram and texture statistics. Global features cannot capture object properties, so local features are favored for object class recognition. For the same reason, higher-level image features are preferred to lower-level ones. Similar image elements, like pixels, patches, and lines can be grouped together to form higher-level units, which are more likely to correspond to objects or object parts. Different types of features can be combined to improve the feature discriminability. For example, using color and texture to identify trees is more reliable than using color or texture alone. The context information is also helpful for detecting objects. A boat candidate region more likely corresponds to a boat if it is inside a blue region. While improving the ability of our system by designing higher-level image features and combining individual ones, to be prepared to apply more and more features since a limited number of features cannot satisfying the requirement of recognizing many different objects in ordinary photographic images.

2. IMAGE RETRIEVAL METHODS

An image retrieval system is one of the significant research area which can be used for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval system utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. In this work, the image retrieval classified mainly in three types.

a) Text Based Image Retrieval
b) Content Based Image Retrieval
c) Sketch Based Image Retrieval
d) Query based Image Retrieval
e) Semantic Based Image Retrieval
f) Annotation based image retrieval

The Various Literature regarding the above topics is represented in pictorial format Figure 2.1. This work mainly focus on Semantic based image retrieval.

3. Text Based Image Retrieval

The text based image retrieval utilizes the method of adding the metadata, such as keywords, captioning or descriptions to the images. The retrieval employed over the annotation words and it makes the annotation complex and time consuming and also requires huge labors to manually annotate the images. The semantic content is not considered in TBIR.

Dinakaran.D et al [11] proposed an effective and efficient hybrid image retrieval system by searching text with both text and image based query. The textual and visual content descriptors are generated from the text query and image query. The descriptors are converted into a vector format. Similarly textual and visual descriptors are calculated and converted into vector representation for the images stored in the database. The vector, generated by the user query is then matched with the vectors stored in the database. The text and content-based methods return two independent lists of images with different weights. These two lists must be combined in a meaningful way to give the user a combined image list.

Clough et al [8] proposed the cross language information retrieval system through which the images are captioned and the where the given textual query is preprocessed such as the normalization, removal of stop words and word stemming are used and a document ranking scheme used where captions containing all query terms are ranked and Tobbias [33] proposed the FIRE system which combines both the CBIR and Textual based retrieval. An interactive image retrieval techniques using user term feedback for a text-based approach to collect terms from all the fields which leads to more confusion and causes more possibilities to produce unrelated terms. Instead of collecting the terms from the image database, collected only from retrieved image sets for the given user query. The filename, the alternate tag and caption fields other than the user input query terms, have more probability for getting relevant terms to narrow the search.

4. Content Based Image Retrieval

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 analyzes the contents of the image rather than the metadata such as keywords, tags, 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 images are retrieved only through the Texture, Color, Shape in content based image retrieval. Smeulders et al [2] in which the collection of literatures states the earliest use of the term content-based image retrieval in the literature seems to have been described the experiments into automatic retrieval of images from a database by color and shape feature. The term has since been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as color, texture and shape) that can be automatically extracted from the images themselves and the features used for retrieval can be either primitive or semantic, but the extraction process must be predominantly automatic. Retrieval of images by manually-assigned keywords is definitely not CBIR as the term is generally understood – even if the keywords describe image content. CBIR differs from classical information retrieval in that image databases are essentially unstructured, since digitized images consist purely of arrays of pixel intensities, with no inherent meaning.

Liu et.al. [35] Stated CBIR techniques can already address many of users' requirements at level 1, and will be capable of making a significant contribution at level 2 if current research ideas can be successfully exploited. They are however most unlikely to make any impact at level 3 in the foreseeable future. Provides an in-depth review of CBIR technology, explaining the principles behind techniques for color, texture, shape and spatial indexing and retrieval in some detail and also stated the issues involved in video segmentation, motion detection and retrieval techniques for compressed images. They identify a number of key unanswered research questions, including the development of more robust and compact image content features, more accurate modeling of human perceptions of image similarity, the identification of more efficient physical storage and indexing techniques, and the development of methods of recognizing objects within images.

3.1 Semantic Gap

Smeulders et al [2] and the Ying Liu [35] and Nikhil et al [26] pointed out the fundamental difference between content-based and text-based retrieval systems are that the human interaction is an indispensable part of the latter system. Humans tend to use high-level features (concepts), such as keywords, text descriptors, to interpret images and measure their similarity.

The sensory gap is the gap between the object in the world and the information in a (computational) description derived from are recording of that scene. The semantic gap is the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.

While the features automatically extracted using computer vision techniques are mostly low-level features (color, texture, shape, spatial lawet, etc.). In general, there is no direct link between the high-level concepts and the low-level features. Though many sophisticated algorithms have been designed to describe color, shape, and texture features, these algorithms cannot adequately model image semantics and have many limitations when dealing with broad content image databases. Extensive experiments on CBIR systems show that low-level contents often fail to describe the high-level semantic concepts in user's mind. Therefore, the performance of CBIR is still far from user's expectations.

Eakins et al [18] mentioned three levels of queries in CBIR. Level1: Retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Typical query is query by example, 'find pictures like this'. Level2: Retrieval of objects of given type identified by derived features, with some degree of logical inference. For example, 'find a picture of a flower'. Level3: Retrieval by abstract attributes, involving a significant amount of high-level reasoning about the purpose of the objects or scenes depicted. This includes retrieval of named events, of pictures with emotional or religious significance, etc. Query example, 'find pictures of a joyful crowd'. Levels 2 and 3 together are referred to as semantic image retrieval, and the gap between Levels 1 and 2 as the semantic gap. More specifically, the discrepancy between the limited descriptive power of low-level image features and the richness of user semantics is referred to as the 'semantic gap'. Users in Level 1 retrieval are usually required to submit an example image or sketch as query. But what if the user does not have an example image at hand? Semantic image Retrieval is more convenient for users as it supports query by keywords or by texture. Therefore, to support query by high-level concepts, a CBIR systems should provide full support in bridging the 'semantic gap' between numerical image features and the Richness of human semantics.

4 TECHNIQUES IN CBIR

4.1 Retrieval by color

Smeulders et al [2] and Arthi et al [3] stated the main components of CBIR are the features which includes the Geometric shape, colors and the texture of the image. Schettini et al Features can be of two types like local features and global features. Object recognition can be done easily using the local features.

Chang et al [6] proposed that the image retrieval using the color distribution, mean and the standard deviation and was tested with three different databases. The other component is the relevant feedback where it helps to be more precise in searching the relevant images by taking up the feedbacks of the user.

Kekre et al [22] pointed out that Color feature is one of the most widely used features in image retrieval. Colors are defined on a selected color space. Variety of color spaces are available, they often serve for different applications. Color spaces shown to be closer to human perception and used widely in RBIR include, RGB, KVE, KEVR, KMCG, KFCG and the hue-min-max-difference.

Sun et al [20] suggested a color distribution entropy method, which takes account of the correlation of the color spatial distribution in an image. The main difference between color distribution entropy, geo stat and spatial chromatic histogram is that color distribution entropy describes how pixel patches of identical color are distributed in an image.

Xue et al [5] states separate color images and color histogram moment of extraction, and then two methods of extracting color feature vector weighted to achieve similar distance, similar to the last distance based on the size of the return search results, based on the realization of the characteristics of the color image Retrieval system.

Lin et al[7] proposed the three features of color that are co-occurrence matrix, difference between pixels of scan pattern and color histogram for k-mean.

Common color features or descriptors in CBIR systems include, color-covariance matrix, color histogram, color moments, and color coherence vector and has included dominant color, color structure, scalable color, and color lawet as color features, the authors are interested in objects taken from different point of view and illumination. As the result, a set of viewpoint invariant color features have been computed. The color invariants are constructed on the basis of hue, hue-hue pair and three color features computed from reflection model [33]. Most of those color features though efficient in describing colors, are not directly related to high-level semantics. For convenient mapping of region color to high-level semantic color names, some systems use the average color of all pixels in a region as its color feature. Although most segmentation tends to provide homogeneous color regions, due to the inaccuracy of segmentation, average color could be visually different from that of the original region. A dominant color in HSV space is defined as the perceptual color of a region. To obtain dominant color, the authors first calculate the HSV space color histogram ($10*4*4$ bins) of a region and select the bin with maximum size. Then the average HSV value of all the pixels in the selected bin is defined as the dominant color. The selection of color features depends on the segmentation results. For instance, if the segmentation provides objects which do not have homogeneous color, obviously average color is not a good choice. It is stated that for more specific applications such as human face database, domain knowledge can be explored to assign a weight to each pixel in computing the region colors. It should be noted that in most of the CBIR works, the color images are not pre-processed. Since color images are often corrupted with noise due to capturing devices or sensors, it will improve retrieval accuracy significantly if effective filter is applied to remove the color noise. The pre-process can be essential especially when the retrieval results are used for human interpretation.

4.2 Retrieval by Shape

Ying Liu et al [35] pointed out the retrieval images from the large image collections. A fast image retrieval based on object shapes extracted from objects within images. Multiple shapes at lower level can be mapped into a single shape at a higher level. Given a query shape, by searching only the relevant paths in the hierarchy, large portions of the database can thus be pruned away. An angle mapping used to transform a shape from one level to another higher level. Angle mapping replaces some edges of shape by a smaller number of edges based on the angles between the edges thus reducing the complexity of the original shapes.

Renato et al [28], as using a single color histogram for the whole image, or local color histograms for a fixed number of image cells, the one we propose (named Color Shape) uses a variable number of histograms, depending only on the actual number of colors present in the image and using a large set of heterogeneous images and pre-defined query/answer sets show that the Color Shape approach offers good retrieval quality with relatively low space overhead, outperforming previous approaches. Jagadish [15] proposed to construct an index structure on the data such that given a template shape, matching shapes can be retrieved in time that is less than linear in the size of the database, that is, by means of an indexed look-up.

4.3 Retrieval by Texture

Ying Liu et al [35] and Kekre et al [22] pointed out the texture is important component of human visual perception and can be effectively used for identifying different image regions. Compared with color and shape features, texture features indicate the shape distribution, better suits the macrostructure and microstructure of the images. Texture representation methods can be classified into three categories, namely structural, statistical and multi-resolution filtering methods. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. This two dimensional array is called as Gray level Co-occurrence Matrix (GLCM). GLCM describes the frequency of one gray tone appearing in a specified spatial linear relationship with another gray tone, within the area under investigation.

Zhang et al [10] proposed a image retrieval method based on Gabor filter. Texture features are found by calculating the mean and variation of the Gabor filtered image. Rotation normalization is realized by a circular shift of the feature elements so that all images have the same dominant direction. The image indexing and retrieval are conducted on textured images and natural images.

Yossi et al [36], proposed a retrieval method based on the Earth Movers Distance with an appropriate ground distance is shown to handle both complete and partial multi-textured queries. As an illustration, different images of the same type of animal are easily retrieved together. At the same time, animals with subtly different coats, like cheetahs and leopards, are properly distinguished.

Sandhu et al [1] pointed out the various methods of texture based retrieval such as the GLCM and histogram over shape and the texture properties.

4.4 Algorithms in CBIR

Krishna et al [25], proposed an indexing of the image using the k-means algorithm which defines first to read the image and the color separation by decorrelation stretching and the conversion RGB to L*a*b color space and the classification of color space under a*b* using k-means algorithm to separate objects. Syam et al [4] proposed a genetic algorithm method in extraction of image features which used to measure the similarity. Deepika [9], pointed algorithms such as naïve random method, the local neighbor movement method, neighboring divide and conquer method, Global divide and conquer method, Jayaprabha [16] proposed the graphical image retrieval algorithm where the keyword models built from visual feature of a set of images are labeled with keywords. It incorporates an image analysis algorithm into the text based image search engines. Sonali et al [30] proposed the SVM algorithm where it act as a classifier and includes three phases the pre-processing, feature extraction and svm classifier.

5. SKETCH BASED IMAGE RETRIEVAL

Jose M. Saavedra et al [37] pointed out the Sketch based image retrieval uses the input as sketches and based on the sketches the relevant images are retrieved and also stated that the sketch based image retrieval (SBIR) is still a young research area, there are many applications capable of using this retrieval paradigm, such as web searching and pattern detection. They also added drawing a simple sketch query turns very simple since touch screen based technology is being expanded. A novel local approach for SBIR based on detecting simple shapes which are named keyshapes, works as a local strategy, but instead of detecting keypoints, it detects keyshapes over which local descriptors are computed. It is based on keyshapes allow us to represent the structure of the objects in an image which could be used to increase the effectiveness in the retrieval task. Indeed, our results show an improvement in the retrieval effectiveness with respect to the state of the art. Furthermore, combining the keyshape approach with a Bag of Feature approach allows achieving significant improvement with respect to the effectiveness of the retrieval task.

6. SEMANTIC BASED IMAGE RETRIEVAL

Horst et al [14], proposed a method to reduce the semantic gap in the image retrieval techniques and pointed out that semantic feature layers are more than the static descriptors and classified human world features.

Athanasiadis et al [32], proposed a framework for simultaneous image segmentation and object labeling leading to automatic image annotation. Focusing on semantic analysis of images, it contributes to knowledge-assisted multimedia analysis and bridging the gap between semantics and low level visual features. The proposed framework operates at semantic level using possible semantic labels, formally represented as fuzzy sets, to make decisions on handling image regions instead of visual features used traditionally. In order to stress its independence of a specific image segmentation approach we have modified two well-known region growing algorithms, i.e., watershed and recursive shortest spanning tree, and compared them to their traditional counterparts.

Wu et al [29] proposed a framework employs ontology and MPEG-7 descriptors to deal with problems arising between syntactic representation and semantic retrieval of images. Instead of building a single ontology for a specific domain, the framework allows for the construction of incrementally multiple ontologies, and shares ontology information not only between the image seekers but also between different domains. Naïve Bayesian inference is used to estimate the similarity between query ontology and domain ontology for matching relevant images

Liu et al [35] proposed the region based retrieval using decision tree induction method.

Julia Vogel et al [19] proposed the semantic modeling of natural scenes for content based image retrieval system, discussed that the selection of scene categories and selected non-human/natural coordinates as superordinate for their experiments. A grid size of 10 x 10 placed over a selected image. In the first stage, local image regions are classified by concept classifier into semantic concept classes.

The local image regions are extracted on a regular grid of 10 x 10 regions and nine local semantic concepts s_i, i=1…M, M=9 were determined as being discriminant for the retrieval tasks [29].

These local semantic concepts are In the second stage ,the region wise information of the concept classifiers is combined to a global image representation s=[sky,water,grass,trunks,foliage,field,rocks,flowers,sand].. The concept co-occurrence vector is essentially a normalized histogram of the concept occurrences in an image [23]

Nikhil et al [26] pointed out the image similarity is measured at two levels. The first is region-level. That is to measure the distance between two regions based on their low-level features. The second is at image level. That is to measure the overall similarity of two images which might contain different number of regions. Most researchers employ the Minkowski-type metric to define region distance. Suppose we have two regions represented by two p dimensional vectors (x1, x2... xp), (y1, y2...yp), respectively. The Minkowski metric is defined as

$$\mathbf{d\ (X, Y)} = \left(\sum_{I=1}^{P} |\boldsymbol{x_i} - \boldsymbol{y_i}|\right)^{\frac{1}{r}} \qquad (1)$$

Particularly, when r =2, it is the well-known Euclidean distance (L2 distance). When r is 1, it is the Manhattan distance (L1distance). An often-used variant version is the weighted Minkowski distance function which introduces weighting to identify important features

$$\mathbf{d(X,Y)} = \left(\boldsymbol{w_i} \sum_{I=1}^{P} |\boldsymbol{x_i} - \boldsymbol{y_i}|\right)^{\frac{1}{r}} \qquad (2)$$

Where w_i=1,..., p is the weight applied to different features. Other distances are also used in image retrieval, such as the Canberra distance, angular distance, Czekanowski coefficient, inner product, dice coefficient, cosine coefficient and Jaccard coefficient. The overall similarity of two images is more difficult to measure. Basically, there are two ways. **One-One match**: This means each region in the query image is only allowed to match one region in the target image and vice versa. As each query region of the query image is associated to a single 'best match' region in the target image. Then the overall similarity is defined as the weighted sum of the similarity between each query region in the query image and its 'best match' in the target image, while the weight is related to region size. **Many-Many match**: This means each region in the query image is allowed to match more than one region in the target image and vice versa. A widely used method is the Earth Mover' Distance (EMD). EMD is a general and flexible metric. It measures the minimal cost required to transform one distribution into another based on a traditional transportation problem from linear optimization, for which efficient algorithms are available. EMD matches perceptual similarity well and can be applied to variable-length representations of distributions, hence it is suitable for image similarity measure in CBIR system.

Liu et al [35], proposed an integrated region matching scheme which allows for matching a region of one image to several regions of another image and thus decreases the impact of inaccurate segmentation . In this definition, a matching between any two regions is assigned with a significance credit. This forms a significance matrix between two sets of regions (one set is of the query image, another set is of the target image). The overall similarity of two images is defined based on the significance matrix in a way similar to EMD. Though Minkowski metric is widely used in current systems to measure region distance, intensive experiments show that it is not very effective in modeling perceptual similarity . How to measure perceptual similarity is still a largely unanswered question. There are some works done in trying to solve this problem. For example, a dynamic partial distance function is defined, which reduces the dimension of feature vectors by dynamically choosing a smaller amount of dimensions [27]

The image is segmented into local subregions. The difference is in the shape of local image subregions. While in the initial method the local image subregions are extracted on a regular grid of 10x10 regions, method tries to segment the image into arbitrary-shaped subregions, which correspond to objects boundaries. This improvement reduces the misclassification of regular subregions belonging to two or even more semantic concepts. In addition this proposed method detects presence of people in the image. This is useful because target images are typical holiday pictures from hiking outdoors. Presence of family members on this kind of images is very common, so it is important to cover also this condition into image retrieval process. So in their system it is possible to define if the retrieval pictures should contain people or not. Only local subregions that represent nature are further processed. Thus identify subregions belonging to people and separate them from others. Afterwards using low-level features classify each subregion into one of following semantic concepts: sky, water, grass, trunks, foliage, rocks, flowers and sand. The selection of these local semantic concepts was influenced by the psychophysical studies concepts used. For the classification of local image regions involved the k-Nearest Neighbor and Support Vector Machine classifiers. The last stage of this proposed method is scene categorization have six different scene categories: coasts, forests, rivers/lakes, sky/clouds, plains and mountains. To each local semantic concept, its frequency of occurrence is determined. This information enables us to make a global statement about the amount of particular concept being present in the image e.g. “There is 22% of water in the image.” Using this knowledge the most suitable category prototype is assigned to the image that gives the semantic meaning of the image.

7. ANNOTATION BASED RETRIEVAL

Hollink et al (2004) proposed and discussed a tool for semantic annotation and search in a collection of art images. Multiple existing ontologies are used to support this process, including the Art and Architecture Thesaurus, WordNet, ULAN and Iconclass and also mentioned knowledge-engineering aspect such as the annotation structure and links between the ontologies. The annotation and search process is illustrated with an application scenario.

Tatsuya et al (2010) proposed a new image annotation and retrieval method for miscellaneous weakly labeled images, by combining higher-order local auto-correlation features and a framework of probabilistic canonical correlation analysis. The distance between images can be defined in the intrinsic space for annotation using conceptual learning of images and their labels. Because this intrinsic space is highly compressed compared to the image feature space, achieves both faster and more accurate image annotation and retrieval. The higher-order local auto-correlation features are powerful global features with additive and position invariant properties. These properties work well with images, which have an arbitrary number of objects at arbitrary locations.

Gustavo et al (2007) pointed out the conventional annotation tools where the tool runs on the desktop, the online tool that can be accessed via the Internet with a web browser, and the online games. The advantage of the annotation tools that run on a desktop PC is that they can support various components that take a great deal of processor power, such as basic image processing functionality. The disadvantage is that the program has to be installed and is often platform dependent. The advantage of online tools is that they are easily accessible and that they have a larger public at their disposal to provide the annotations. However, a disadvantage is that the processor-intensive components cannot easily be incorporated. Online games have the potential to reach a far larger public than the desktop and online annotation tools. Annotation takes place quickly, but the disadvantage is that the quality is often inferior to conventional annotations. Hence, it is important to develop an image annotation game in such a manner that the game elements result in increasing the quality of the annotations to an as high a level as possible. The faster the annotations can be executed, the faster the data set will grow. With the desktop and online tools it often takes a considerable amount of time to perform an annotation. With the M-OntoMat-Annotizer tool, for example, the object must first be selected in the ontology, then the image must be opened, and next the object in the image must be outlined. It will take a user a few minutes to perform these actions, but when done properly, it yields high-quality annotations. When this process is quickened, it will decrease in annotation quality, but it will make it possible to do a large number of annotations within less time. The online games are a good example of this process. One annotation typically takes less than a minute. Besides the games, only the IBM EVA tool offers fast annotations. One of the reasons is that this tool does not offer the possibility of object-segmentation in its annotation process. In order to add semantics to annotations, ontologies such as WordNet may be used. Six of the ten annotation tools have evaluated use an ontology. This component is present in all desktop versions, the online tool IBM EVA, and LabelMe . In general, there is an ontology browser present in the desktop tools, which makes it possible to select the correct class or instance for the object of interest. When the same objects are repeatedly annotated, it will take only a short time for a user to find the object in the ontology. When another class or instance has to be found repeatedly, it can sometimes take a long time before the user finds the correct instance in the ontology was stated in Jeroen et al.,(2010)

Zhang et al proposed a probabilistic semantic model in which the visual features and the textual words are connected via a hidden layer which constitutes the semantic concepts to be discovered to explicitly exploit the synergy among the modalities. The association of visual features and textual words is determined in a Bayesian framework such that the confidence of the association can be provided. Extensive evaluation on a large-scale, visually and semantically diverse image collection crawled from Web is reported to evaluate the prototype system based on the model.

Jeroen et al (2010) stated the LabelMe the annotations are linked to WordNet. An effort is made to find the correct sense in WordNet automatically and to connect this sense to the annotation. However, many words have more than one meaning (sense), which makes it difficult to choose the correct word with the correct meaning. For this reason, the data set administrators choose the sense manually at the end of the annotation process. Unfortunately, this function has not been incorporated in the online annotation tool, nor is the WordNet ‘synsetid’ present in the data set. When annotations are too abstract, the object can often not be assigned to an image-region. This means that it is hard to allocate a word describing activities, feelings, emotions etc. to an object. For the desktop tools this is not so much of a problem, since the annotators are experts and choose their words carefully. On the other hand, in the games this is problematic. The images that are annotated in ESP Game are used as input However, the annotations that are provided by ESP Game are not filtered on abstraction. Examples of abstract words are: love, tennis, joy and nice. Although these words do have added value as an annotation, the specific location of these words cannot, or only with difficulty, be identified in the image. To prevent this from happening, ideally, these word categories must be filtered out.

Nearly all image annotation tools offer the possibility to make region-based annotations. The only exceptions are IBM EVA and the ESP Game. There are various methods to execute region-based annotations: bounding, polygonal and freehand drawing. With the bounding-box technique the object is framed by dragging a rectangle around an object. The advantage of the bounding-box selection is that it is very fast; however, the disadvantage is that the selection is inaccurate and often selects much more image data than necessary. The polygonal method offers the possibility to make a more detailed selection by drawing a polygon around the object. This method is fast, and more precise than the bounding box. Nevertheless, since it uses straight lines, it is still difficult to make a very accurate selection. With a freehand drawing tool, one can draw a free line around an object, which enables very precise selections. The obvious disadvantage is that it takes longer to draw an accurate line around an object. Two of the annotation tools, Spatial Annotation Tool and M-OntoMat-Annotizer, offer the option to use an automatic segmentation method. However, the quality of the automatic segmentations does not meet the high standard required for image region annotations. Consequently, this component has not been included in the analysis. The two methods which offer both quick and detailed image-region segmentations are the polygonal and freehand drawing methods.

Gurdeep et al (2013) listed the motivations for image annotation is many. In-depth studies on the topic, emphasizing in particular organizational and social aspects, are presented. For annotation games, also consider the scoring mechanism. This leads to three different reasons why the annotators annotate the images. These are: organizational, social, and scoring .Generally, desktop tools are used by research groups. The purpose of the annotating is to organize the annotations for research purposes. Researchers have only limited time to annotate images. They often rely on students or an external bureau, such as Mechanical Turk to provide the annotations. When researchers rely on students, the stimulus is either an interest in the research project or money. The online tools reach a larger group of annotators since the tools can be accessed by everybody via the Web. LabelMe, for example, appeals to a large group of annotators. The idea is that fellow-researchers in the object-detection field help each other to create an ever larger data set for research purposes. However, at the moment it focuses too much on the unilateral purpose of annotation, and non-researchers are unlikely to carry out large numbers of annotations. Furthermore, the annotation speed is not quick enough. Social web sites such as Flickr have been set up with the aim of sharing images with other people. The tags that have been included with the images facilitate image searches and present a potentially better representation of what the image shows, but tags are known to be ambiguous, overly personalized, and limited per item. Using games as annotation tools offers a whole new stimulus. By earning points and a high score, annotating suddenly has a different purpose to the user. This yields more annotators and is cheaper than hiring annotators. Of course, the number of annotations generated by using games is totally dependent on the popularity of the game.

7.1 FREE TEXT ANNOTATAION

For this type of 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. This is to make up for the limitation stated in [21] "There is no way the domain ontology can be complete—it will not include everything a user might want to say about a photograph". Any concepts which cannot adequately be described by choosing keywords are simply added in free form description. This is the approach used in the W3C RDFPic software [23] in which the content description keywords are limited to the following: Portrait, Group-portrait, Landscape, Baby, Architecture, Wedding, Macro, Graphic, Panorama and Animal. This is supplemented by a free text description. The IBM Video Annex software [24] also provides this option. The Image CLEF 2004 [25] bilingual ad hoc retrieval task used 25 categories of images each labeled by a semi-structured title (in 13 languages). The IAPR-TC12 dataset of 20 000 images [26] contains free text descriptions of each image in English, German and Spanish. These are divided into "title", "description" and "notes" fields. Additional content-independent metadata such as date, photographer and location are also stored.

7.2 ANNOTATIONS BASED ON ONTOLOGIES

An ontology is a specification of a conceptualization [18]. 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 ontology as it encodes the relationship "is a" (a dog is an animal). An ontology can solve the problem that some keywords are ambiguous. For example, a "leopard" could be a large cat, a tank, a gecko or a Mac operating system. Ontologies are important for the Semantic Web, and hence a number of languages exist for their formalisation, such as OWL and RDF. 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 Wilkins [19] 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 [20]. Developing ontologies to describe even very limited image domains is a complicated process, as can be seen in the work by Schreiber et al. [21], who develop an ontology for describing photographs of apes, and by Hyvonen et al. [14], who develop an ontology for describing graduation photographs at the University of Helsinki and its predecessors. In the domain of image description, Iconclass [22] 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 organized in a hierarchical structure. Each definition is described by an alphanumeric code accompanied by a textual description (textual correlate).

7.3 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 (see Figure 1a for an example).

(2) A segmentation of the image along with keywords associated with each region of the segmentation. In addition, keywords describing the whole

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. Unfortunately in practice, the following two problems arise: • 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. Forcing the user to choose from an on-screen list of keywords is a solution to the second problem, but this makes the search task more frustrating if the number of images is large pointed out by Allan Hanburry

CONCLUSION

This papers Categorizes the various concepts in image retrieval techniques and a collection of 36 papers was studied and various image retrieval techniques and their types and methods are categorized such as the text based and content based and the Semantic based image retrieval.

REFERENCES

[1] Amanbir Sandhu and Aarti Kochhar 'Content Based Image Retrieval using Texture, Color and Shape for Image Analysis' Council for Innovative Research International Journal of Computers & Technology, vol.3, no. 1, pp.2277-3061,2012

[2] Arnold W.M Smeulders, Marcel Worring and Amarnath Gupta 'Content based image retrieval at the end of the early years' IEEE Transactions on Pattern Analysis and Machine Intelligence ,vol. 22, no.12 pp. 1349-1380,2000

[3] Arthi.K and Mr. J. Vijayaraghavan 'Content Based Image Retrieval Algorithm Using Color Models', International Journal of Advanced Research in Computer and Communication Engineering,Vol. 2, Issue 3,pp.1346-1347,2013

[4]Baddeti Syam and Yarravarappu Rao 'An effective similarity measure via genetic algorithm for content based image retrieval with extensive features' International Arab journal information technology vol.10 no.2 pp.143-153,2013

[5]Bai Xue, Liu Wanjun 'Research of Image Retrieval Based on Color', International Forum on Computer Science-Technology and Applications, pp.283-286, 2009

[6]Chin-Chen Chang and Tzu-Chuen Lu 'A Color-Based Image Retrieval Method Using Color Distribution and Common Bitmap' Springer, pp. 56–71, 2005

[7]Chuen-Horng Lin, Rong-Tai Chen, Yung-Kuan Chan 'A smart content-based image retrieval system based on color and texture feature', Image and Vision Computing vol.27 pp. 658–665,2009

[8]Clough, P and Sanderson, M. 'User experiments with the Eurovision cross-language image retrieval system' Journal of the American Society of Information Science

and Technology, vol.57 no.5 pp.697 - 708,2006

[9]Deepika Nagthane 'Content Based Image Retrieval system Using K-Means Clustering Technique' International Journal of Computer Applications & Information Technology, vol. 3, no.1 pp.21-30,2013

[10] Dengsheng Zhang, Aylwin Wong, Maria Indrawan, Guojun Lu 'Content-based Image Retrieval Using Gabor Texture Features', IEEE Pacific-Rim Conference,2000

[11] Dinakaran.D, J. Annapurna, Ch. Aswani Kumar 'Interactive Image Retrieval Using Text and Image Content' Cybernetics and information technologies, vol. 10, no. 3 pp.20-30,2010

[12] Gupta A and Jain R Visual information retrieval, Communications of the ACM 40 (5), 70–79. 1997

[13]Gonzalez R. C and Richard E.W. , Digital Image Processing, Prentice Hall. 2001.

[14] Horst Eidenberger and Christian Breiteneder 'Semantic feature layers in content based image retrieval implementation of world features'.2004

[15] Jagadish.H.V 'Retrieval technique for similar Shapes' ACM,1991

[16] Jayaprabha.P and Rm.Somasundaram 'Content Based Image Retrieval Methods Using Graphical Image Retrieval Algorithm (GIRA)' International Journal of Information and Communication Technology Research,vol.2 no.1 pp.22-28,2012

[17] James. H, S. Harpreet, W. Equits, M. Flickner and W. Niblack, "Efficient Color Histogram Indexing for Quadratic Form Distance Functions", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 17, No. 7, 1995

[18] John P.Eakins , 'Towards Intelligent image retrieval' pattern recognition vol.35 pp.3-14 ,2002

[19] Julia Vogel and Bernt Scheele 'Semantic modeling of natural scenes for content based Image Retrieval' International Journal of Computer Vision,2006

[20] Junding Sun, Ximin Zhang, Jiangtao Cui, Lihua Zhou 'Image retrieval based on color distribution entropy' Pattern Recognition Letters vol.27 pp.1122–1126,2006

[21] Katerina Pastra, Horacio Saggion and Yorick Wilks 'NLP for Indexing and Retrieval of Captioned Photographs' pp.143-146,2004

[22]Kekre H.B, Sudeep D. Thepade, Tanuja K. Sarode and Shrikant P. Sanas 'Image retrieval using texture features extracted using LBG, KPE, KFCG, KMCG, KEVR with assorted Color spaces' International Journal of Advances in Engineering & Technology, Vol. 2, Issue 1, pp. 520-531,2012

[23] Kristina Lidayova and Elena sikudova 'Semantic categorization and retrieval of Natural scene Images' Proceedings of CESCG 2012: The 16th Central European Seminar on Computer Graphics,2012

[24] Lakshmana pandian.S and T.V Geetha 'Morpheme based Language Model for Tamil Part-of-speech Tagging' International journal of computer applications,2008

[25]Murali Krishna Raja.N.V, K.Shirin Bhanu 'Content Bases Image Search And Retrieval Using Indexing By K Means Clustering Technique' International Journal of Advanced Research in Computer and Communication Engineering' vol. 2, no.5 pp.2181-2189,2013

[26] Nikhil Rasiwasia, Nuno Vasconcelos and Pedro J.Moreno 'Bridging the gap: Query by Semantic Example' IEEE Transactions on Multimedia vol. 9 no.5, pp.923-938,2007

[27] Pushpa B. Patil and Manesh B. Kokare 'Interactive Semantic Image Retrieval' Journal Information Process Systems, Vol.9, No.3, pp.349-364,2013

[28] Renato O. Stehling, Mario A. Nascimento, Alexandre X. Falc , 'On

"Shapes" of Colors for Content-Based Image Retrieval' ACM, pp.171-174,2000

[29] Roung Shiunn Wu and Wen Hsien Hsu 'A Semantic Image Retrieval Framework Based on Ontology and Naïve Bayesian Inference' International Journal of Multimedia Technology, vol.2 .no2 pp..36-43,2012

[30] Sonali Jain and Satyam Shrivastav 'A novel approach for image classification in Content based image retrieval using support vector machine' International Journal of Computer Science & Engineering Technology. vol. 4 no. 03 pp.223-227,2013

[31] Tatsuya harada, Hideki nakayama, Yasuo kuniyoshi and Nobuyuki otsu 'Image Annotation and Retrieval for Weakly Labeled Images Using Conceptual Learning' New Generation Computing, vol.28 pp.277-298,2010

[32] Thanos Athanasiadis, Yannis Avrithis, Member, and Stefanos Kollias 'Semantic e Retrieval Without Segmentation',2004.

Image Segmentation and Object Labeling' IEEE transactions on circuits and systems for video technology,vol.17 no.3 pp.283-311,2007

[33] Tobias Weyand 'Combining Content-based Image Retrieval with Textual, Information Retrieval', Research project,2005

[34] Tony Rose, David Elworthy, Aaron Kotcheff and Amanda Clare 'ANVIL: a System for the Retrieval of Captioned Images using NLP Techniques' Challenge of Image Retrieval',2000

[35] Ying Liu, Dengsheng Zhang, Guojun Lu, Wei-Ying Ma 'A survey of content based image retrieval with high level semantics' Elsevier pattern Recognition vol.40 pp.262-282,2007

[36] Yossi Rubner and Carlo Tomasi 'Texture-Based Imag