Volume 3, Issue 4, April 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Design of Image Retrieval Efficacy System Based on CBIR

S.Meenachi Sundaresan
Assistant Professor/Department of CSE,
PSNITS, Tamil Nadu, India

Dr. K.G.Srinivasagan
Professor/Department of CSE-PG
National Engineering College, Tamil Nadu, India

Abstract— Content-based image retrieval (CBIR) is the mainstay of image retrieval systems. To be more profitable, relevance feedback techniques were incorporated into CBIR such that more precise results can be obtained by taking user's feedbacks into account. However, existing relevance feedback-based CBIR methods usually request a number of iterative feedbacks to produce refined search results, especially in a large-scale image database. This is impractical and inefficient in real applications. In this paper, we propose a novel method, Navigation-Pattern-based Relevance Feedback (NPRF), to achieve the high efficiency and effectiveness of CBIR in coping with the large-scale image data. In terms of efficiency, the iterations of feedback are reduced substantially by using the navigation patterns discovered from the user query log. In terms of effectiveness, our proposed search algorithm NPRF Search makes use of the discovered navigation patterns and three kinds of query refinement strategies, Query Point Movement (QPM), Query Reweighting (QR), and Query Expansion (QEX), to converge the search space toward the user's intention effectively. By using NPRF method, high quality of image retrieval on RF can be achieved in a small number of feedbacks. The experimental results reveal that NPRF outperforms other existing methods significantly in terms of precision, coverage, and number of feedbacks.

Index Terms—Texture, Colour, DWT, Euclidean, NPRF, QPM.

I. INTRODUCTION

An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval 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. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image Annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools. Image search is a specialized data search used to find images. To search for images, a user may provide query terms such as keyword, image file/link, or click on some image, and the system will return images "similar" to the query. The similarity used for search criteria could be meta tags, color distribution in images, region/shape attributes, etc [1]. Image meta search - search of images based on associated metadata such as keywords, text, etc. Content-based image retrieval (CBIR) – the application of computer vision to the image retrieval. CBIR aims at avoiding the use of textual descriptions and instead retrieves images based on similarities in their contents (textures, colors, shapes etc.) to a user-supplied query image or user-specified image features. List of CBIR Engines - list of engines which search for images based image visual content such as color, texture, shape/object, etc [2].

A. Data Scope

It is crucial to understand the scope and nature of image data in order to determine the complexity of image search system design. The design is also largely influenced by factors such as the diversity of user-base and expected user traffic for a search system. Along this dimension, search data can be classified into the following categories:

1) *Archives* - usually contain large volumes of structured or semi-structured homogeneous data pertaining to specific topics.
2) *Domain-Specific Collection* - this is a homogeneous collection providing access to controlled users with very specific objectives. Examples of such a collection are biomedical and satellite image databases.
3) *Enterprise Collection* - a heterogeneous collection of images that is accessible to users within an organization's intranet. Pictures may be stored in many different locations [3].
4) *Personal Collection* - usually consists of a largely homogeneous collection and is generally small in size, accessible primarily to its owner, and usually stored on a local storage media.
5) *Web* - World Wide Web images are accessible to everyone with an Internet connection. These image collections are semi-structured, non-homogeneous and massive in volume, and are usually stored in large disk arrays [4].

In this Approach we convert image feature vectors to weighted-term vectors and extract the frequently occurred feature vectors by implementing relevance feedback technique in content-based image retrieval to demonstrate the efficiency of this conversion. According to obtain effectiveness we extract colour and texture features because of using a single feature extraction does not give high accuracy and effectiveness. High Dimensional feature reduce the query efficiency and low level feature reduce the query accuracy. So we have to integrate those above two visual features in our approach to obtain high effectiveness [5]-[6].

II. PROBLEM STATEMENT

A web navigation behavior is helpful in understanding what information of online users demand. Following that, the analyzed results can be seen as knowledge to be used in intelligent online applications, refining web site maps, web based personalization system and improving searching accuracy when seeking information. Nevertheless, an online navigation behavior grows each passing day, and thus extracting information intelligently from it is a difficult issue. Web Usage Mining (WUM) is the process of extracting knowledge from Web user's access data by exploiting Data Mining technologies. It can be used for different purposes such as personalization, system improvement and site modification. In our system, user navigation patterns are described as the common browsing behaviors among a group of users. Since many users may have common interests up to a point during their Navigation, navigation patterns should capture the overlapping interests or the information needs of these users. In addition, navigation patterns should also be capable to distinguish among web pages based on their different significance to each pattern.

III. PROPOSED APPROACH

"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. There is a growing interest in CBIR because of the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically, e.g. from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling issues.

A. CBIR techniques

Many CBIR systems have been developed, but the problem of retrieving images on the basis of their pixel content remains largely unsolved.

1) Query techniques

Different implementations of CBIR make use of different types of user queries. Query by example is a query technique that involves providing the CBIR system with an example image that it will then base its search upon. The underlying search algorithms may vary depending on the application, but result images should all share common elements with the provided example.

Options for providing example images to the system include:

- A preexisting image may be supplied by the user or chosen from a random set.
- The user draws a rough approximation of the image they are looking for, for example with blobs of color or general shapes.

This query technique removes the difficulties that can arise when trying to describe images with words.

2) Semantic retrieval

The ideal CBIR system from a user perspective would involve what is referred to as *semantic* retrieval. Current CBIR systems therefore generally make use of lower-level features like texture, color, and shape; although some systems take advantage of very common higher-level features like faces (see facial recognition system). Not every CBIR system is generic. Some systems are designed for a specific domain, e.g. shape matching can be used for finding parts inside a CAD-CAM database.

3) Other query methods

Other query methods include browsing for example images, navigating customized/hierarchical categories, querying by image region (rather than the entire image), querying by multiple example images, querying by visual sketch, querying by direct specification of image features, and multimodal queries (e.g. combining touch, voice, etc.) CBIR systems can also make use of relevance feedback, where the user progressively refines the search results by marking images in the results as "relevant", "not relevant", or "neutral" to the search query, then repeating the search with the new information.

4) Content comparison using image distance measures

The most common method for comparing two images in content based image retrieval (typically an example image and an image from the database) is using an image distance measure. An image distance measure compares the similarity of two images in various dimensions such as color, texture, shape, and others. For example a distance of 0 signifies an exact match with the query, with respect to the dimensions that were considered. As one may intuitively gather, a value greater than 0 indicates various degrees of similarities between the images. Search results then can be sorted based on their distance to the queried image. A long list of distance measures can be found in.

5) Color

Computing distance measures based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values (that humans express as colors). Current research is attempting to segment color proportion by region and by spatial relationship among several color regions. Examining

images based on the colors they contain is one of the most widely used techniques because it does not depend on image size or orientation. Color searches will usually involve comparing color histograms, though this is not the only technique in practice.

6) Texture

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image the texture is located. Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated. However, the problem is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as silky, or rough.

7) Shape

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined first applying segmentation or edge detection to an image. Other methods like use shape filters to identify given shapes of an image. In some case accurate shape detection will require human intervention because methods like segmentation are very difficult to completely automate.

8) Applications

Some software producers are trying to push CBIR based applications into the Internet and law enforcement markets for the purpose of identifying and censoring images with skin-tones and shapes that could indicate the presence of nudity, with controversial results.

- We present an algorithm for preprocessing of web log file and a new approach for clustering of user navigation patterns based on the graph partitioning for modeling user navigation patterns.
- For the clustering of user navigation patterns we create an undirected graph based on connectivity between referrer and URI pages and we propose a formula for assigning weights to edges.
- The results represent that our approach can improve the quality of clustering for user navigation pattern in web usage mining systems. These results can be use for predicting user's next request in the huge web sites.

B. Steps of proposed method:

1) Data Cleansing:

Irrelevant records are eliminated during data cleansing. Since target of web usage mining is to get traversal pattern, following two kinds of records are unnecessary and should be removed.

The records of graphics, video and format information.

- The records having filenames suffixes of GIF, JPEG, and CSS and so on, this can be found in cs_uri_stem field of record.

  The records with failed HTTP status codes.
- By examining the status field of every record in the web log, the record with status code over 299 and fewer than 200 are removed. User and Session Identification:
- The task of user and session identification is to find out the different user sessions from the original web access log. A referrer-based method is used for identifying sessions. The different IP addresses distinguish different users.

If the IP addresses are same,

- The different browsers and operation systems indicate different users which can be found by client IP address and user agent who gives information of user's browsers and operating system.

If all of the IP address, browsers and operating systems are same, the referrer information should be taken into account. The Refer URI (cs_referer) is checked, a new user session is identified if the URL in the Data Cleansing User Identification Session Identification

2) Content Retrieval:

- Content Retrieval retrieves content from users query request

3) Path Completion:

- Path Completion should be used acquiring the complete user access path. The incomplete access path of every user session is recognized based on user session identification. If in a start of user session, Referrer as well URI has data value, delete value of Referrer by adding Web log preprocessing helps in removal of unwanted click-streams from the log file and also reduces the size of original file by 40-50%.

Step 3: Generation of Paged:

- Paged is sequence generated numbers like p1, p2, p3….are created for user query/pages/page views.

Step 4: User Navigation Mining:

- Web pages accessed are modeled as undirected graph G= (V, E). The set V of vertices contains the identifiers of the different pages hosted on the Web server.
- Undirected graph created for a single user session using Hash Map
- Hash Map data structure stores the referrer–URI pair and their corresponding weights.
- Weight of edges given by 1or 0. If link between page and referrer exist then weight will be 1 else 0.
- Weight of pages is given as frequency of pages in graph.

- Weight of pages (W) = Frequency(F) of referrer–URI pair(occurrence in user session)

  Applied Depth First Search Algorithm on graph and obtained all possible navigation patterns. Path length is calculated by considering the total weight of the edges in a graph.

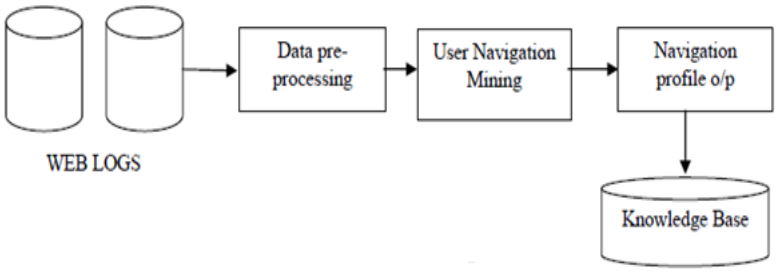


Fig. 1 Steps for navigation pattern mining

Design Consideration and Implementation for Figure 1 show the various steps are involved in identifying clusters of user navigation pattern.

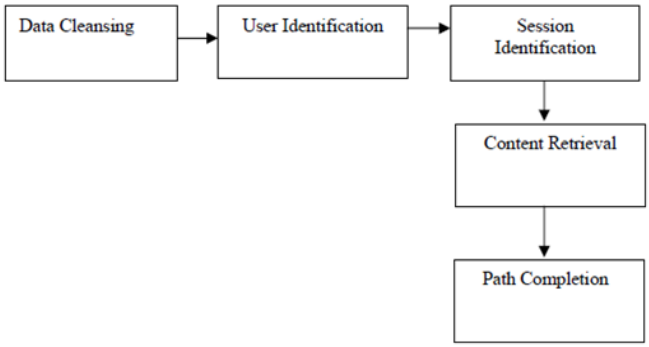
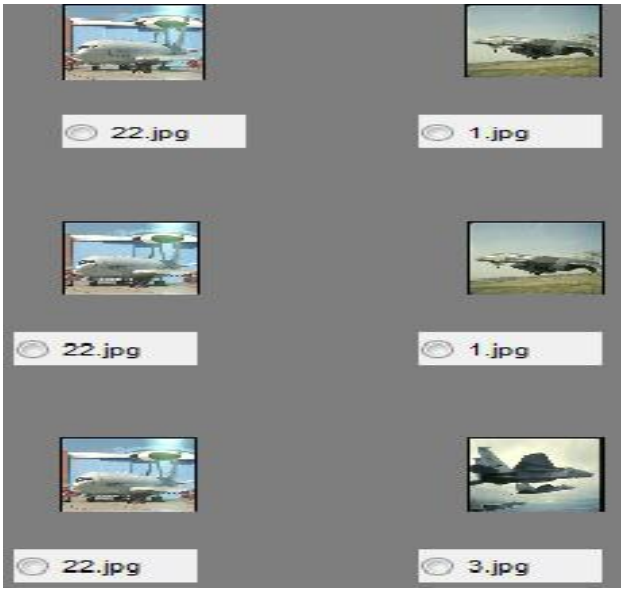


Fig. 2 Block diagram of preprocessing steps

Figure 2 represents the data processing is done. Applications of proposed systems are Web field, Photo shop and medical field.

IV. EXPERIMENTAL RESULTS

The experiments were carried out on a Core i3; 2.4 GHz processor with 4GB RAM using MATLAB.The image database contains 100 kind of different image sets. Fig. 2 shows the image retrieval results using only color Histogram Extraction, and Color and DWT texture and the proposed method with FFE. The image at the centre the query image and the other images are the retrieval results. The performance of a retrieval system can be measured in terms of its recall (or sensitivity) and precision (or specificity)**.** Recall measures the ability of the system to retrieve all models that are relevant, while precision measures the ability of the system to retrieve only models that are relevant.



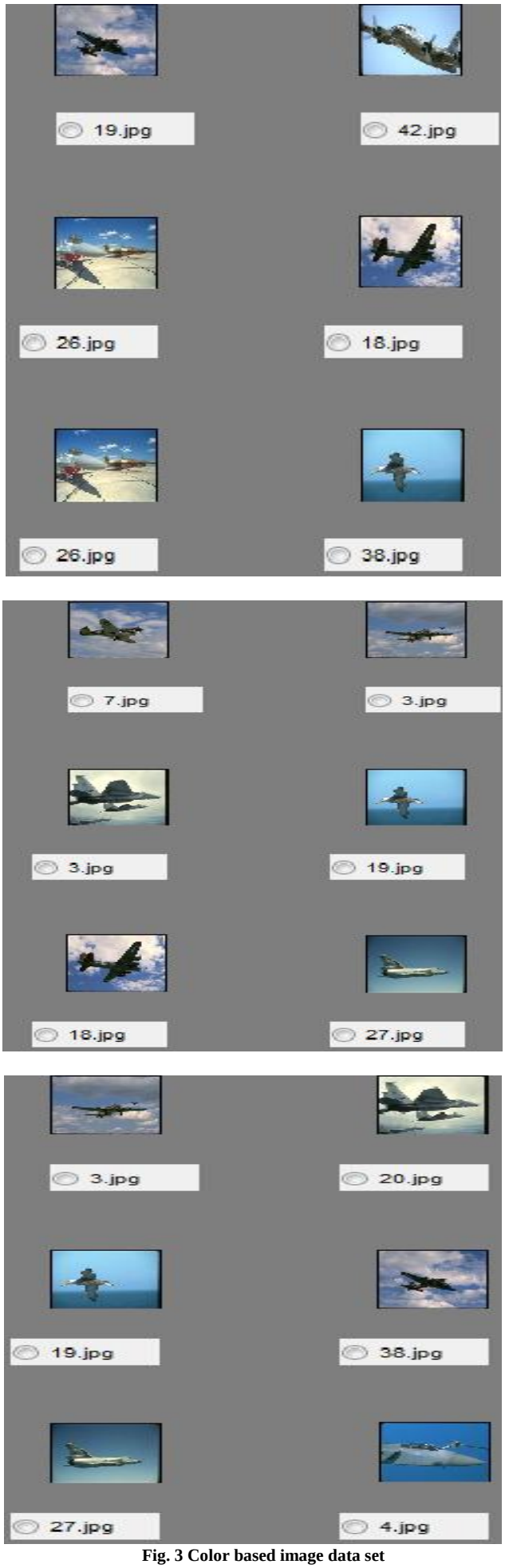


Fig. 3 Color based image data set

Fig. 4 Color Histogram based image retrieval of proposed method

Fig. 4 shows the Color Histogram based image retrieval of proposed method, which is used for to retrieve the image from the huge color image data set in figure 3, based on Content-based image retrieval (CBIR) method.

V. CONCLUSION

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 image retrieval is opposed to concept based approaches (see concept based image indexing).In this paper, a CBIR method has been proposed which uses the combination of dominant color, DWT texture. A total of 39 features covering color, texture proved that the proposed method yielded higher average precision and average recall. In addition, the proposed method almost always showed performance gain of average retrieval time over the other methods. As Further studies, the proposed retrieval method is to be evaluated for more various databases.

References

[1] Xia Yu , , "Image retrieval based on color, texture, shape and SVM relevance feedback ," *IEEE International Conference on CAIDCD*, vol. 1, pp. 779 - 781, Nov. 2010.

[2] W. Niblack et al., "The QBIC Project: Querying images by content using color, texture and shape," *Proc. SPIE* 1908, Storage and Retrieval for Image and Video Databases, pp. 173-187, San Jose, CA, Feb. 1993.

[3] W. Y. Ma and B. S. Manjunath, "Texture features and learning similarity," Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 425-430, San Francisco, CA, June 1996.

[4] J. R. Smith and S. F. Chang, "Local color and texture extraction and spatial query," IEEE Int. Conf. on Image Processing, Vol. 3, pp. 1011-1014, Lausanne, Switzerland, Sep. 1996.

[5] Qian Huang et al., "Foreground/background segmentation of color images by integration of multiple cues," IEEE Int. Conf. on Image Processing, Vol. 1, pp. 246-249, Washington, DC, Oct. 1995.

[6] M. J. Swain and D. H. Ballard, "Color indexing," International Journal of Computer Vision, Vol. 7, No. 1, pp.11-32, 1991.

