International Journal of Engineering Research and General Science Volume 3, Issue 3, May-June, 2015
ISSN 2091-2730

An Image Searching Approach using Image Content

Apurva S. Gomashe[1], Prof. R. R. Keole[2]

[1]Department of Comp.Sci & Engg,

HVPM's College of Engg & Tech, Amravati University, Maharashtra, India

[2]Department of Information Technology,

HVPM's College of Engg & Tech, Amravati University, Maharashtra, India

Abstract: The existing image searching approach is design to analysis the images from query log and its fail to retrieve the new query's which are not present in query log and this process was time consuming and its gives images from log history means if in query log present only two relevant image similar to user giving image but user want new image at that time this system was fail. The text based image retrieval system (TBIR) limitation is annotation problem and in this system it's important to describe query properly and but the content based image retrieval system (CBIR) has become an active research area. CBIR refers to techniques used to index and retrieve images from databases based on their visual content. Visual content is typically defined by a set of low level features extracted from an image that describe the color, texture and/or shape of the entire image. Semantic gap between visual features and human semantics has become a bottleneck in content-based image retrieval. We develop such system using Image content-based image retrieval with different methods to reduced semantic gap. Efficiently proved more desired result that user wants.

In this paper, first introduce basic techniques used for image searching and image retrieving from huge data collection of image, existing systems and then architecture of our proposed system, how it's work.

Keywords: — Image Content Based Image Retrieval, Features Extraction, Re-ranking, Features vector, Color Histogram, Text Based Image Retrieval, Image Searching

1. INTRODUCTION

In the world there are various search engines introduced to search different data, documents, etc, form millions of data collections. Search engines used in PC, laptops, and the program is stored in operating system of PC, laptops internally and personal mobile, basically search engines used in World Wide Web for searching online data which stored in server database. Here we proposed new search engine system for image searching/image retrieving. The interest towards image retrieval is increased due to the rapid growth of the World Wide Web. The need to find a desired image from a collection is shared by many groups, including journalists, engineers, historians, designers, teachers, artists, and advertising agencies. The image needs and usages vary considerably among the users in these groups [1]. The users may require access to the images, based on primitive features, such as color, texture or morphological or associated text. The technology to access these images has also accelerated phenomenally. The current approaches are broad and inter disciplinary, mainly focused on three aspects of image research which are text-based retrieval, content-based retrieval and interactive based image retrieval. Early techniques are based on the textual annotation of images.

Many techniques have been developed for text-based information retrieval [2] and they proved to be highly successful for indexing and querying web sites. Their success may also shed some light on the area of image retrieval, because the relatively mature theories and techniques of text-based information retrieval may be applicable to the image domain. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on standard Boolean queries. Although text-based methods are fast and reliable when images are well annotated, they are incapable of searching in unannotated image collections. The generalization of the information retrieval from the text domain to the image database is, however, non-trivial. The greatest obstacle arises from the intrinsic difference between the text and image in representing and expressing information [3].

This concept emphasizes on use of visual content of image like color, texture, shape etc. for image comparison and retrieval rather than textual query. In common words, visual feature of any image is anything that is seen or felt about that image. It includes any visual variation in the look of that image. These contents are then extracted from images in the database and are described by multidimensional vectors. The feature vectors of the images in database form the feature database. To retrieve images, users provide the retrieval system with example images or sketched figures. The system then converts them into internal representation of feature vectors. The similarities /distances between the feature vectors of the query example or sketch provided and those of the images in the database are calculated and then retrieval is performed. Under this work various factors defining the concerned visual contents are described in details. A content based image retrieval system allows the user to present a query image in order to retrieve images stored in the database according to their similarity to the query image. Content-Based Image Retrieval (CBIR) has become an active research area [4]. However, effective and precise image retrieval still remains an open problem because of the extreme difficulty in fully characterizing images. Successful techniques have been developed for some specific applications, such as face and finger-print recognition [7, 8]. An effective approach for querying and browsing images still remains elusive. In our proposed system we use a novel approach for Content Based Image Retrieval by combining the color and texture features. Our paper is organized as follows. Section 2 reviews existing image searching/image retrieval approaches. In Section 3 Architecture of new proposed system. Section 4 discusses the conclusions and references.

2. EXISTING IMAGE SEARCHING/RETRIEVAL APPROACHES

Existing image retrieval techniques can be classified into three categories: text-based, content-based, and interactive approaches. The text-based approach is a traditional way to search images simply by keyword based search. The images are indexed according to the content, like the caption of the image; filename, title of the web page, and alternate tag, etc. and stored in the database. Processing a user query could involve a stop word removal, stemming and tokenizing. Some of the keyword based image retrieval approaches are bag of words, natural language processing and Boolean model [12]. Image retrieval is then shifted to standard database management capability combined with information retrieval techniques. Some commercial image search engines, such as Google Image Search and Lycos Multimedia Search, are keyword-based image retrieval systems.

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 (see this survey[1] for a recent scientific overview of the CBIR field). Content-based image retrieval is opposed to traditional concept-based approaches (see Concept-based image indexing). "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 searches that rely purely on metadata are dependent on annotation quality and completeness. Having humans manually annotate images by entering keywords or metadata in a large database can be time consuming and May not capture the keywords desired to describe the image. The evaluation of the effectiveness of keyword image search is subjective and has not been well-defined. In the same regard, CBIR systems have similar challenges in defining success. In this approach, the processing of a query image involves extraction of visual features and perform search in the database for similar images [13]. A typical CBIR system views the query image and images in the database (target images) as a collection of features, and ranks the relevance between the query image and any target image in proportion to a similarity measure, calculated from the features. The low level image features can be used to compute similarity between images [14]. Despite the recent progress, content-based image retrieval has its own limitations because of the semantic gap between the low level image features and high level semantic content of images (like sunset, flowers, etc.).

Relevance feedback is a powerful technique, originally used in the traditional text-based information retrieval systems. In CBIR a relevance-feedback-based approach allows the user to interact with the retrieval algorithm by providing information regarding the images which the user believes to be relevant to the query. Based on the user feedback, the model of similarity measure is dynamically updated to give a better approximation of the perception subjectivity. Empirical results demonstrate the effectiveness of a relevance feedback for certain applications. Nonetheless, such a system may add burden to the user especially when more information is required than just a Boolean feedback (relevant or non-relevant). Now currently existing system is basically design to overcome the semantic gap limitation so for that it has only focused on analyzing a particular query appearing in the query logs and extract the images form log history (query log), this existing system work flow is understand by using following framework

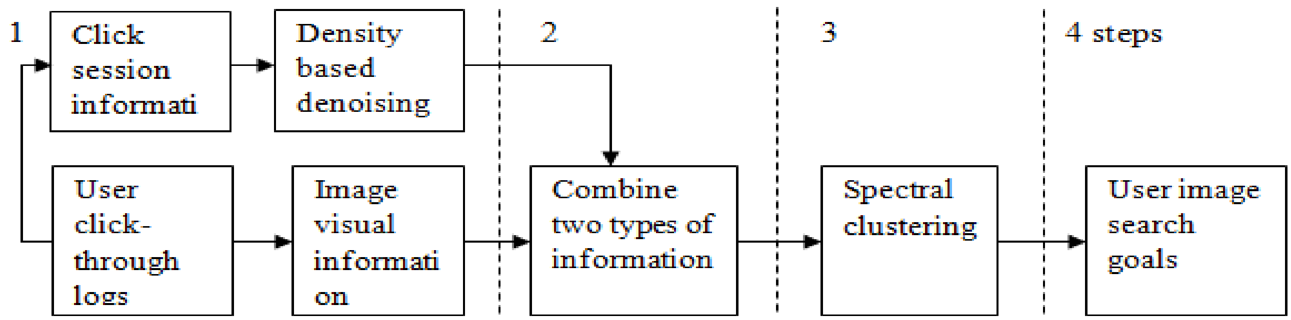

Figure 1: Framework of Existing System

In this framework consist of four steps and each steps perform various operation to further process now we see there working of each steps

Step 1: We first extract the visual information of the clicked images from user click-through logs. Normally, the images clicked by the users with the same search goal should have some common visual patterns, while the images clicked by the users with different search goals should have different visual patterns to be distinguished from each other. For example, for the query apple, there must be some visual patterns to distinguish fruit apples from phones. Therefore, it is intuitive and reasonable to infer user image-search goals by clustering all users' clicked images for a query with image visual information and use each cluster to represent one search goal. In this paper, we extract three types of image visual features (i.e., color, texture, and shape features) containing color moments (CMG) [16], color correlogram (CC) [16], cooccurrence texture (CT) [15], local binary pattern (LBP) , and edge auto-correlogram (EAC) [17]. We concatenate the above five feature channels to get the feature vector for each image. At the same time, we also extract the click session information from user click-through logs. We consider that the clicked images in a session have high correlations, which is under the hypothesis that the user has only one search goal when he submits a query and he just clicks those similar images. However, in the real situation, many users may click some noisy images.

Step 2: Image visual information is combined with click session information for further clustering by one of the two proposed strategies, named edge-reconstruction-based strategy and goal-image-based strategy. It should be noted that these two strategies are alternatives by using different ways to model the clicked images for a query with similarity graph [25].

Step 3: It introduce spectral clustering algorithm to cluster the image graph that contains both image visual information and click session information. Spectral clustering is introduced in this step because clusters representing different user goals may have arbitrary shapes in visual feature space when clustering.

Step 4: It gives the user image search goals, but only from query log not retrieve new query which is not present on that query log so in proposed system our main work is to retrieve the images from query log as well as database at same time.

3. ARCHITECTURE OF PROPOSED SYSTEM

In market there are many search engines are available and they perform in different ways like some of them are used text based image retrieval method, when used this method user gives only text as input and on the basis of text get the output but this process is not show desired output or result because they have their limitations. In existing system used text for content based and this type of method known as hybrid method, but this method is used only to retrieve images form query log, in query log present only those data which access by previous user, so this kind of system not gives new query data or images which is not present on query log, so to overcome this limitation we design a new proposed system. The proposed system is an initiative to solve the above user and developer problem. The system is motivated from the literature survey of the search engine. In our proposed system we used image content based image retrieval method, in which main work is depended upon image features means we used histogram method and match the features of images given by user to features of images in database. Now see the architecture diagram as follows

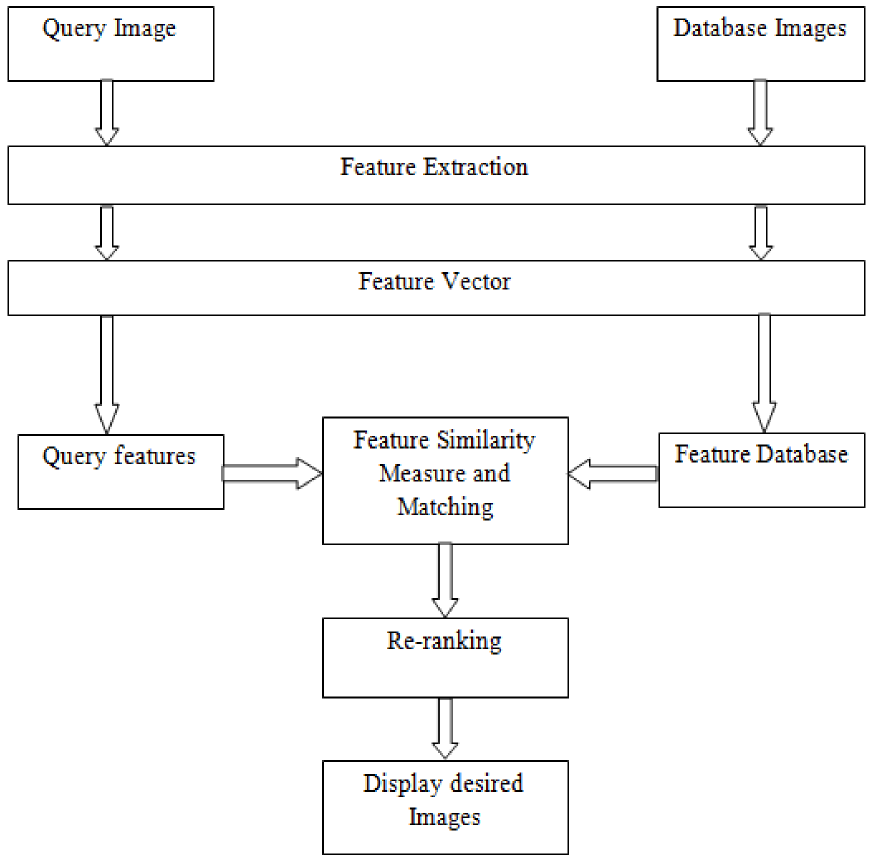

Figure 1: Architecture of Proposed System

- *Query Image:* In our proposed system user gives image as input to get more relevant output in short time period.

- *Database Images:* Malians of images are stored in database and here user can also create its own data collections.

- *Feature Extraction:* The feature is defined as a function of one or more measurements, each of which specifies some quantifiable property of an object, and is computed such that it quantifies some significant characteristics of the object. Features Extraction process is play main role in every type of image search engines because its help at the time of creation of database means adding the entries into the database as well as at the time of searching images its help to find or search matching features from the database.
  We classify the various features currently employed as follows:

  • General features: Application independent features such as color, texture, and shape. According to the abstraction level, they can be further divided into:

  - Pixel-level features: Features calculated at each pixel, e.g. color, location.

  - Local features: Features calculated over the results of subdivision of the image band on image segmentation or edge detection.

  - Global features: Features calculated over the entire image or just regular sub-area of an image.

  • Domain-specific features: Application dependent features such as human faces, fingerprints, and conceptual features. These features are often a synthesis of low-level features for a specific domain.

On the other hand, all features can be coarsely classified into low-level features and high level features. Low-level features can be extracted directed from the original images, whereas high-level feature extraction must be based on low-level features.

- **Color Features:**

The color feature is one of the most widely used visual features in image retrieval. Images characterized by color features have many advantages:

• Robustness. The color histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled. It is also insensitive to changes in image and histogram resolution and occlusion.

• Effectiveness. There is high percentage of relevance between the query image and the extracted matching images.

• Implementation simplicity. The construction of the color histogram is a straightforward process, including scanning the image, assigning color values to the resolution of the histogram, and building the histogram using color components as indices.

• Computational simplicity. The histogram computation has O(X, Y) complexity for images of size X × Y . The complexity for a single image match is linear, O(n), where *n* represents the number of different colors, or resolution of the histogram.

• Low storage requirements. The color histogram size is significantly smaller than the image itself, assuming color quantization.

Typically, the color of an image is represented through some color model. There exist various color models to describe color information. A color model is specified in terms of 3-D coordinate system and a subspace within that system where each color is represented by a single point. The more commonly used color models are *RGB* (red, green, blue), *HSV* (hue, saturation, value) and Y,Cb,Cr (luminance and chrominance). Thus the color content is characterized by 3-channels from some color model. One representation of color content of the image is by using color histogram. Statistically, it denotes the joint probability of the intensities of the three color channels.

Color descriptors of images can be global or local and consist of a number of histogram descriptors and color descriptors represented by color moments, color coherence vectors or color correlogram [9].

Color histogram describes the distribution of colors within a whole or within a interest region of image. The histogram is invariant to rotation, translation and scaling of an object but the histogram does not contain semantic information, and two images with similar color histograms can possess different contents.

- *Feature Vector:* In pattern recognition and machine learning, a feature vector is an n-dimensional vector of numerical features that represent some object. Many algorithms in machine learning require a numerical representation of objects, since such representations facilitate processing and statistical analysis. Feature construction is the application of a set of constructive operators to a set of existing features resulting in construction of new features. Examples of such constructive operators include checking for the equality conditions {=, ≠}, the arithmetic operators {+, −, ×, /}, the array operators {max(S), min(S), average(S)} as well as other more sophisticated operators.

- *Feature Similarity Measure and Matching:* This involves matching these features to yield a result that is visually similar. The commonly used similarity measure method is the Distance method. There are different distances available such as Euclidean distance.

- *Re-ranking*: Re-ranking is used to re-arrange the sequence of resulting images means most relevant images are display first on the list of output.

- *Display desired Images:* Display the relevant images that user's wants and which are appears in database.

4. CONCLUSION

In this paper we discuss the introduction about search engines used for image searching/ retrieval methods and various existing image searching/retrieving approaches like text based image retrieval approach, relevant feedback approaches with their limitations, to overcome this limitations we introduced new proposed system which work on the basis of content based image retrieval techniques means users gives the input in the form of image. A query image can be retrieved efficiently from a large database. CBIR technology has been used in several applications such as fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medicine, historical research. A Database consists of different types of images has implemented on the system. Different Features such as histogram, color mean, Color structure descriptor texture is taken into consideration for extracting similar images from the database. From the experimental result it is seen that combined features can give better performance than the single feature. So selection of feature is one of the important issues in the image retrieval. The system is said to be efficient if semantic gap is minimum. It's understands by using their architecture which describe in this paper.

REFERENCES:

[1] Miller, H., A. Geissbuhler, S. March and, P. Clough. Benchmarking Image Retrieval Applications. – In: Proc. Of the 10th International Conference on Distributed Multimedia Systems, San Francisco, CA, USA, 2004.

[2] Vikhar, P. A. Content Based Image Retrieval (CBIR): State-of-the-Art and Future Scope for Research. – The IUP Journal of Information Technology, Vol. **6**, 2010, No 2, 64-84.

[3] Valova, I., B. Rachev. Retrieval by Color Features in Image Databases. – In: Proc. Of Adbis'04, University of Rousse, Budapest, Hungary, 2004.

[4] Abby, A., Goodrum. Image Information Retrieval: An Overview of Current Research. – Informing Science, Vol. **3**, 2000, No 2, 63-66.

[5] Mezaris, V., H. Doulaverakis. A Test-Bed for Region Based Image Retrieval Using Multiple Segmentation Algorithms and the MPEG-7experiment Model: The Schema Regerence System. Network of Excellence in Content-Based Semantic Scene Analysis and Information Retrieval, IST-2001-32795, 2001.

[6] Borlund, P., P. Ingwersen. The Development of a Method for the Evaluation of Interactive Information Retrieval Systems. – Journal of Documentation, Vol. **53**, 1997, No 3, 225-250.

[7] Kuilenburg, H., M. Wiering, M. Uyl. Model Based Methods for Automatic Analysis of Face Images. – In: Proc. of 16th European Conference on Machine Learning, 2005, 194-205.

[8] Sandeep, K., A. N. Rajagopalan. Human Face Detection in Cluttered Color Images Using Skin Color and Edge Information. – In: Proc. of Indian Conference on Computer Vision, Graphics and Image Processing, 2002.

[9] Deb, Sagarmay, and Zhang, Yanchun. "An Overview of Content-based Image Retrieval Techniques", IEEE, 18th International Conference on Advanced Information Networking and Application (AINA"04), 2004.

[10] Gemert, J. C. Retrieving Images as Text. Master's Thesis, University van Amsterdam, 2003.

[11] Burns, M., K. Leung, A. Rowland, J. Vicke rs, J. V. Hajnal, D. Rueckert, D. L. G. Hill. Information Extraction from Images (IXI): Grid Services for Medical Imaging. – In: Proc. of Distributed Databases and Processing in Medical Image Computing (DiDaMIC), Rennes, France, 2004, 65-73.

[12] Anne, H., H. Ngu, Quan Z. Sheng, Du Q. Huynh, Ron Lei. Combining Multi-Visual Features for Efficient Indexing in a Large Image Database. – VLDB Journel, 2001.

[13] Gemert, J. C. Retrieving Images as Text. Master's Thesis, University van Amsterdam, 2003.

[14] Deselaers, T., T. Wey and, D. Keyers. FIRE in Image CLEF 2005: Combining Content- Based Image Retrieval with Textual Information Retrieval. Working Notes of CLEF Workshop, Austria, 2005.

[15] Z. Zha, L. Yang, T. Mei, M. Wang, Z. Wang, T. Chua, and X. Hua, "Visual query suggestion: Toward capturing user intent in internet image search," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 6, no. 3, p. 13, 2010.

[16] S. Wan, Y. Xue, X. Yu, F. Guan, Y. Liu, and X. Cheng, *ICTNET at Web Track 2011 Diversity Task*. MD, USA: National Instit. Standards Technology, 2011

[17] Z. Lu, X. Yang, W. Lin, X. Chen, and H. Zha, "Inferring users' imagesearch goals with pseudo-images," in *Proc. IEEE Visual Commun. Image Process.*, 2011, pp. 1–4