

A Scalable Approach For Image Extraction Using Content Based Image Retrieval In Peer To Peer Network

Inchara C G
PDA College of engineering
Kalburgi, India
inchara125@gmail.com

Prof Priyadarshini Patil
PDA College of engineering
Kalburgi, India
priyadarshinicpatil@gmail.com

Abstract—The last decade has observed great curiosity in research on content-based image retrieval. This has covered the way for a large number of new techniques and systems, and a growing interest in associated fields to support such systems. Likewise, digital imagery has expanded its horizon in many directions, resulting in an explosion in the volume of image data required to be organized. Content-Based Image Retrieval (CBIR) in Peer to Peer system uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. The distributed nature of these systems, where nodes are typically located across networks, innately hinders the efficient retrieval of information. we consider the searching and retrieval of information that is dispersed on peers of a network. Our approach builds on work in unstructured P2P systems and uses only local knowledge. The reason for using unstructured P2P systems is that ,they impose very small demands on individual nodes and can easily accommodate nodes of varying power Active research in CBIR is geared towards the development of methodologies for analyzing, interpreting cataloging and indexing image database. The quality of response is heavily dependent on the choice of the method used to generate feature vectors and similarity measure for comparison of features we proposed an algorithm which includes the advantages of various other algorithms to improve the accuracy and performance of retrieval. The accuracy of color histogram based matching can be increased by using Color lookup table (CLT) for successive refinement

Keywords—CBIR,P2P,CLT .

I. INTRODUCTION

The global infrastructure of computers and networks creates exciting opportunities for accumulating vast amounts of data and for sharing computers and resources in an original scale. In the last few years, the Peer-to-Peer (P2P) model has become a very powerful for developing Internet scale file systems and sharing resources (that is, CPU cycles, memory, storage space, and network bandwidth) over large-scale geographical areas. We are using the concept of CBIR which defines a technique which uses visual contents to search images from large scale image databases according to user

interests. Image retrieval is a process of retrieving the image from huge dataset. Retrieving images in our peer to peer network is an important challenge. So the overlay network of nodes (peers) is constructed on top of heterogeneous networks. Overlays are flexible and deployable. The first wave of P2P systems implemented unstructured P2P overlays in which no global structure or knowledge is maintained. For searching the data or resources, messages are sent over multiple hops from one peer to another with each peer responding to queries for information it has stored locally. Unstructured P2P systems are very good to share and store documents, because of their decentralized nature allows easy additions, updates, increased storage, and offer fault-tolerant properties through the use of replication and caching. The P2P infrastructure has different processing and communication capabilities, located across different areas. As a result, retrieving images over such Internet-scale environments is subject to greater variations due to unpredictable communication latencies, excessive resource consumption, and changing resource availability.

II. RELATED WORK

A.W. M. Smeulders, M. Worring, S. Santini, and R. Jain [1] describe the “Content-based” such that the search will analyze the actual contents of the image rather than the metadata such as tags, Keywords and/or descriptions associated with the image. The term ‘content’ refers to shapes, textures, color or any other information that can be derived from the image itself. Also humans manually enters the keywords for images in a large database can be inefficient, expensive. Thus a system can filter the images based on their content and provide better indexing and return accurate results. Chun et al. [2] proposed a CBIR method based on an efficient combination of multi resolution color and texture features. The colors and autocorrelograms features of the hue and saturation component images in HSV color space are used. S. Tong and E. Chang [3] describes that the Support vector machines (SVMs, also support vector networks) are supervised learning models associated with learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a

model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. Gulfishan Firdose Ahmed, Raju Barskar [4] describes the basic components of content-based image retrieval system. Image retrieval methods are based on color, texture, shape are analyzed and compared. In many areas of commerce, government and hospitals, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of analogue photographs, diagrams, drawings, paintings, and prints. Usually, the only way of searching these collections was by keyword indexing, or simply by browsing. Digital images databases however, open the way to content-based searching. X. Zhou and T. Huang [5] authors analyze the nature of the relevance feedback problem in a continuous representation space in the context of multimedia information retrieval. An attempt is made to compile a list of critical issues to consider when designing a relevance feedback algorithm. Relevance feedback is a feature of some information retrieval systems. The idea behind relevance feedback is to take the results that are initially returned from a given query and to use information about whether or not those results are relevant to perform a new query.

III. PROPOSED METHOD

In our proposed approach rather than searching the image based on the tag, keywords. Here we are giving image as a input query image. Based on the input query image similar type of images are being retrieved. Image features are extracted based on color, shape. SVM is used as a classifier which results in a higher accuracy. As we are using peer to peer system once the input query image is selected we can get the output images from the other peer. In particular, peers are used to minimize the network distance from their neighboring nodes by establishing connections to nodes that belong to the same network. It is designed to make image search efficient in unstructured P2P networks.

IV. PROPERTIES OF IMAGE

The Properties for extraction the image features is as follows:

A. Color Feature Extraction

A color image can be represented using three primaries of a color space. Since the RGB color space cannot be seen directly by the human eye so the HSV color space has been used in this approach. HSV is an intuitive color space that can be seen directly to visual perception, and it is common for image retrieval systems. Hue is used to represent a pure color, where as saturation represent the amount of white light added to a pure color. Value refers to the light intensity or measures brightness. The very important advantage of HSV color space is good compatibility. The color distribution of pixels in an image contains sufficient information. There are two color properties that is Global color properties and Local color

properties of an image. The following two features to represent the global properties of an image can be used. The mean of pixel colors states the principal color of the image, and the standard deviation of pixel colors represents the variation of pixel colors in an image. The variation degree of pixel colors in an image is called the color complexity of the image. Color correlogram gives the information about the features of colors. It includes spatial color correlations, which describes the global distribution of local spatial correlation of colors and is very easy to compute. Color moment feature is used to distinguish images based on their color features and it is also gives the similarity of color measurement between the images. Then the similarity values are compared with the values of images stored in the database for image retrieval.

B. Texture Feature extraction

Texture is the visual patterns that have properties of homogeneity that do not result from the presence of only a single colour or intensity. The texture descriptors are used to measure the properties such as smoothness, coarseness and regularity. Statistical approaches yield characterization of texture as a smooth, coarse, grainy and so on. Texture provides important information in image classification as it describes the content of many real-world images such as fruit skin, clouds, trees, bricks, and fabric. Hence, texture is an important feature in defining high-level semantics for image retrieval purpose. The texture feature extraction techniques are - Gabor Wavelet Transform, Texture boundary detection, Texture classification, Color texture, Fourier descriptors.

C. Shape Feature Extraction

Shape features of images are extracted by using connected component method. Shape feature provides the semantic information about an image. Shape features are usually described using part or region of an image. The accuracy of shape features depends upon the segmentation scheme used to divide an image into meaningful objects. However, fast and robust segmentation is difficult to achieve. This limits the shape features only to those retrieval applications where objects or region of images are readily available. The shape descriptors are divided into two classes: boundary based descriptor and region based descriptor. Some of the boundary based description techniques are chain codes, polygonal approximations, Fourier descriptor and finite element model. Region based descriptors are statistical moment and area. A good shape feature should be invariant to translation, rotation and scaling.

V. SYSTEM DESIGN

The figure1 describes that the system consists of two phases that is training phase and the testing phase. In the training phase there is a creation of the dataset and in the testing phase we are giving image as a input query image. Here first the user select the image from the database then there is a generation of the dataset once the dataset is generated then the dataset is stored in the database. Next when the user goes to the testing phase here the user gives the input query image then there is a

selection of the dataset once the dataset is selected then there is feature extraction.

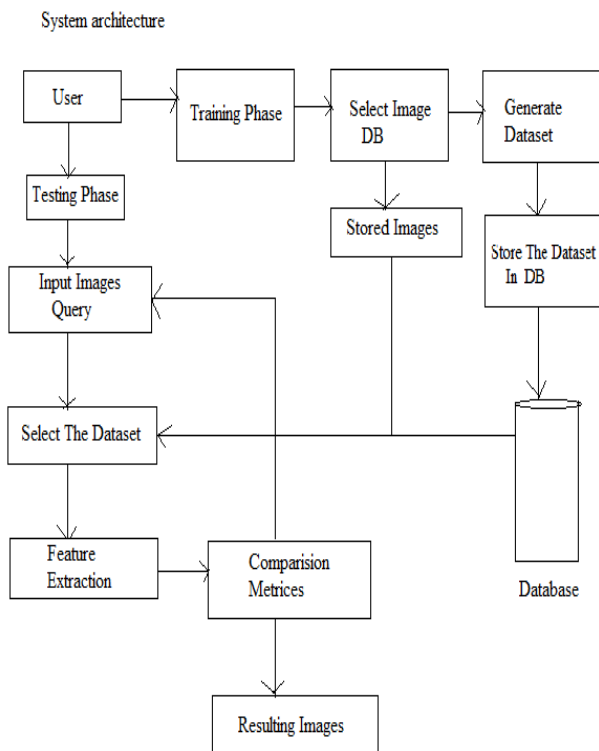


Figure1: System Architecture for retrieving images in Peer to Peer system

The features extracted from the images directly lead to the results. Some of the basic features of the images are color features, shape and texture features. Once the features have been selected then there is a metrics selection here we are using the support vector machine classifier which predicts target value of data instances in the testing set which are given only the attributes. Next step is to retrieve the similar type of images from the given input query image. As we are using peer to peer network the above same steps are followed but the output images retrieved are on the another peer.

A. Classification by SVM

Support vector machine also known as SVM and it is a supervised machine learning method that examines the data and identifies the patterns, used for classification. The advantage of this algorithm is that the classification of the input query object depends on feature vectors and training samples. The features trained and testing images are classified by using the Support Vector Machine classifier to give the result to the user. SVM is a useful technique for data classification. Machine Learning is considered as a subfield of Artificial Intelligence and it is concerned with the development of techniques and methods which enable the computer to learn. Over the period of time many techniques and methodologies were developed for

machine learning tasks. The statistical learning theory provides a framework for studying the problem of gaining knowledge, making predictions, making decisions from a set of data. In simple words, it enables the choosing of the hyper plane space in such a way that closely represents the underlying function in the target space. In statistical modeling we would choose a model from the hypothesis space, which is closest (with respect to some error measure) to the underlying function in the target space. SVM performs better in terms of not over generalization when the neural networks might end up over generalizing easily. SVM is a useful technique for data classification. Even though it is considered that its Neural Networks are easier to use than this, however, sometimes unsatisfactory results are obtained. However the classification task usually involves with training and testing data which consist of some data instances. Each instance in the training set contains one target value and several attributes. The goal of SVM is to produce a model which predicts target value of data instances in the testing set which are given only the attributes.

VI. IMPLEMENTATION

Here the image database consists of 500 images. In which 150 images are used for testing and remaining 350 images are used for training. First of all the images are stored in the database. Then the user has to select an input image from the testing database i.e. the query image. Query images are displayed on screen once the query image is displayed. Next step is to get the similar type of image from the training database. The features from which images are retrieved depend on color, shape and texture. Color features are extracted by using color correlogram, color moment and hsv histogram. SVM method classifies query image to relevant image in image database which results in higher accuracy. As peer to peer system is being used. We are using the two systems and connecting this system by giving the IP address of both. And the system is made online we can give an input query image on system A and load the dataset on same system. Once the system is being loaded we can get the same type of images on the system B. This is how the images are retrieved in peer to peer. The method is found to be effective and robust, in terms of accuracy and variety of images considered. The accuracy of the method is found to be 92.4% which is effective and is comparable with other methods.

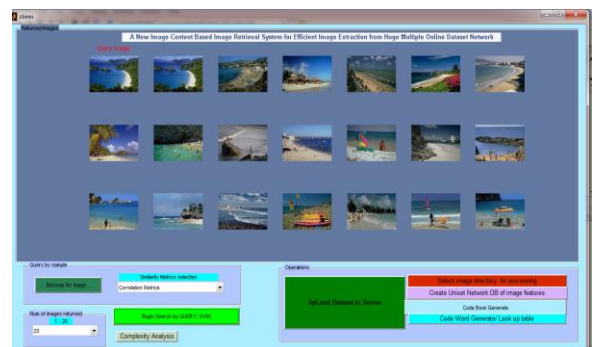


Figure2: Retrieving similar type of images from the given input query image.

The figure2 describes that for example beach image is given as an input query image .the output image obtained is that the similar type of the beach input images.

VII. RESULT ANALYSIS

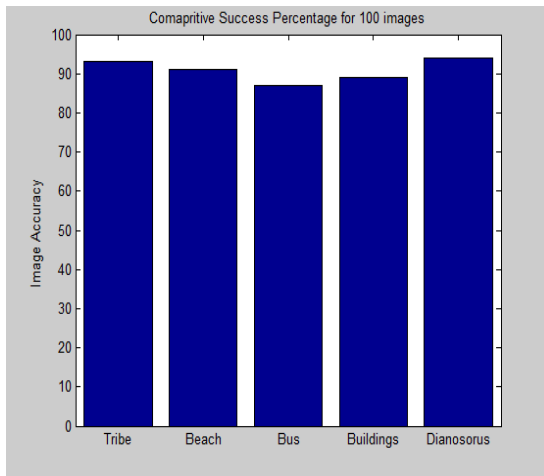


Figure3: Comparative accuracy analysis for 100 images

The figure3 describes the comparative accuracy analysis for 100 images here we are taking five dataset that is the Tribe, Beach, Bus, Buildings, Dinosaurous images. The accuracy analysis for Tribe images is 92%,then the Beach images is 90%,then for the Bus images is 89%,then for the Buildings is the 90%,then last for the Dinosaurous is the 93%.This is how the accuracy analysis for all the five dataset used.

VIII. CONCLUSION

This paper examined the primary parts of the content based image retrieval system traits depiction. Many further studies in this area have been recognized and also usage of this technique is shown in the same. It is concluded that to attain the extraction at the high speed as well as for making it so flexible that it can also adjust with the images of large size, it is required to have an impressive multidimensional approach. The primary aim of this paper is to represent the significance of Support Vector Machine in the efficient retrieval of image. In this SVM is used as the classifier which is performing the task of classifying the image and this process of classification is given to all the traits of the image which are extracted after the feature extraction process. It is mostly used for estimating the highest margin hyper planes within the feature space which is also a high dimensional feature space.

REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans.Pattern Anal.Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380.

- [2] Y. D. Chun, and I. H. Jang, "Content-based image retrieval using multiresolution color and texture features," *IEEE Trans.Multimedia*, vol. 10, no. 6, pp. 1073–1084.
- [3] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. ACM Multimedia*..
- [4] Gulfishan Firdose Ahmed, "A Study on Different Image Retrieval Techniques in Image Processing", in September 2011.
- [5] X. Zhou and T. Huang, "Relevance feedback for image retrieval: A comprehensive review," *Multimedia Syst.*, vol. 8, no. 6, pp. 536–544..
- [6] D. Tao, X. Tang, and X. Li, "Which components are important for interactive image searching?," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 1, pp. 3–11, Jan. 2008.
- [7] Gupta, A., Jain, R.: *Visual Information Retrieval*. Communications of the ACM, vol. 40, 70–79. (1997)
- [8] Faloutsos, C., Flickner, M., Niblack, W., Petkovic, D., Equitz, W., Barber, R.: *Efficient and Effective Querying by Image Content*. (1993)
- [9] Rui, Y., Huang, T.S., Mehrotra, S.: *Relevance feedback techniques in interactive content-based image retrieval*. In *Storage and Retrieval for Image and Video Databases*.
- [10] Smith, J. R., Chang, S.: *Automated binary texture feature sets for image retrieval*. *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Proc*, May. (1996).
- [11] Ono, A., Amano, M., Hakaridani, M.: *A flexible content-based image retrieval system with combined scene description keyword*. *Proc. IEEE Conf. on Multimedia Computing and Systems*. (1996) Ray-I Chang et al. *ComSIS* Vol. 9, No. 4, Special Issue, December 2012..
- [12] K. A. Hua, K. Vu, and J. H. obsammach, "a flexible and efficient sampling based image retrieval technique for large image databases."
- [13] Rajshree S. Dubey, Rajnish Choubey, Jor Bhattacharjee, "Multi feature Content Based Image Retrieval", *International Journal on Computer Science and Engineering (IJCSSE)*, volume: 2, No. 06, 2010, pages: 2145- 2149.
- [14] Jianhua Wu, Zhaorong Wei, Youli Chang, "Color and Texture Feature For Content based image retrieval", *International Journal of Digital Content Technology and its Applications*,2010.
- [15] D. Joshi, R. Datta, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", *ACM Comput. Surv.*, April-2008.
- [16] Tianzhong Zhao, Lu, Yafei Zhang1, "Feature Selection based on Genetic Algorithm for CBIR", *Congress on Image and Signal processing, IEEE*, volume: 02, May-2008, pages: 495-499.
- [17] Jonathan E. Rowe, "Genetic Algorithm theory", *ACM GECCO '07: Proceedings of the 9th annual conference companion on Genetic and evolutionary computation*, July-2007.