A Hybrid Approach for content based image retrieval from large Dataset

Devendra Kurwe
PG Scholar DoCSE
UIT-RGPV, Bhopal (India)

Prof. Anjna Jayant Deen
Asst. Prof. DoCSE
UIT-RGPV, Bhopal (India)

Dr. Rajeev Pandey
Asst. Prof. DoCSE
UIT-RGPV, Bhopal (India)

ABSTRACT

Image processing is one of the methods to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. The volume of digital images generated and uploaded on the internet are very large. The major problem is retrieving the desired images from huge collection of images. To improve the retrieval performance an accurate and efficient system is required. Content based image retrieval technique has been a very useful system. Today content based image retrieval is a required concept, while dealing with multiple activities daily today either with computer or web or mobile we often need to query the database to find efficient required output in short time. In this paper we are proposing a hybrid approach which is the combination of genetic and Bayesian algorithm which giving us the better results in some aspects which overcomes the disadvantages of the existing algorithm and find its suitable in point of efficiency and accuracy. For the valuation of result two standard parameters one is Recall and another is Precision are used which shows better value in comparing to other retrieval algorithms.

Keywords- CBIR, Bayesian algorithm, Genetic algorithm, Feature Extraction.

1. INTRODUCTION

An image retrieval system is employed for browsing, searching and retrieving images from a large number of databases of digital images. More conventional and familiar methods of image retrieval utilize some method of adding meta data such as captioning, keywords or explanations to the images so that retrieval can be performed over the annotation words. Regularly growth of digital images required improved methods for browsing, searching and sorting through ever-growing image databases. These databases are used by various professionals including doctors searching for related clinical cases, editors looking for illustration images and almost everybody wants to organize their personal photos. There is need of image browsing, retrieval and searching tool by the users from various domains including fashion, medicine, publishing, crime prevention, remote sensing, architecture etc. To improve the retrieval performance an accurate and efficient system is required. Content based image retrieval technique has been a very useful, with various techniques developed by various researchers. The idea of content-based image retrieval is not to rely on textual descriptions of image content. in its place, a set of features is used that allow the user to find images that are visually similar to a presented query image. Manual image annotations are consume more time and its expensive to address this, there has been a large amount of research done on automatic image annotation. Image search is a specialized data search employed to find images. Images consist of more information than a normal text contains and if the images are to be retrieved from the source then they got to be done employing all the data that it brings. "Content-based" means that the search analyzes the contents of the image rather than the Meta data such as keywords, tags, or explanations associated with the image. The term "content" in this context might indicates to colors, shapes, textures, or any other data that can be derived from the image itself. CBIR is enticing 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 explain the image.

2. RELATED WORK

To diminish the lack of consistency problem, the image retrieval is carried out according to the image features. Such scheme is called content-based image retrieval (CBIR). The main challenge of the CBIR system is to construct meaningful descriptions of physical attributes from images to expedite efficient and effective retrieval. CBIR has become an dynamic and fast-improving research area in image retrieval in the last few years. the difference between the user's information need and the image representation is called the semantic gap in CBIR systems. In order to reduce the gap, relevance feedback is very helpful into CBIR system. The basic idea behind relevance feedback [1] is to integrate human perception subjectivity into the query and involve user to evaluate the retrieval results. Then depending upon user's integration the similarity measures are automatically refined. There are lots of CBIR algorithms has been proposed and most of them work on the finding effectively specific image or group of relevant image to that query image using similarity computation phase.

Bayesian Algorithm

Named for Thomas Bayes, a mathematician, Bayesian logic is a branch of logic applied to decision making that deals with probability inference. A variant of Bayesian algorithm is called Naïve Bayes. It is one of the most effective and efficient classification algorithms. Figure shows graphically the structure of naïve Bayes, each attribute node has the class node as it parent, but does not have any parent from attribute node. The main advantage of this algorithm is representation of the best relationship among terms. Bayesian network consists of three layers: query layer, feature index layer and relevant image layer. The root node is the query layer representing the query example image given by the user. The intermediate layer is feature index layer which can be further divided into two levels. The first level contains low-level feature representations, such as color, texture and shape. The second level is composed by the components of the feature vectors. The relevant image layer consists of the individual relevant images specified by the user. The network is different with various query examples and different relevant images

Genetic Algorithms

(GAs) are inspired from phenomena found in living nature. The phenomena incorporated so far in GA models include phenomena of natural selection as there are selection and the production of variation by means of recombination and mutation, and rarely inversion, diploid and others. Most Genetic Algorithms work with one large population, i.e. in the recombination step each individual may potentially choose any other individual from the population as a mate. Then GA operators are performed to obtain the new child offspring;

the operators are:

Crossover: - The task of crossover is the creation of a new individual out of two individuals of the current population.

Mutation: - The mutation operator M(chromosome) selects a gene of that chromosome and changes the allele by an amount called the mutation variance (mv), this happens with a mutation frequency (mf).

GAs within the field of evolutionary computation is robust, computational, and stochastic search procedures modelled on the mechanics of natural genetic systems. In general, a GA contains a fixed-size population of potential solutions over the search space. These potential solutions of the search space are encoded as binary, called chromosomes. The initial population can be created randomly or basis on the problem- specific knowledge. GA is a branch of evolutionary computation.

3. PROPOSED METHODLOGY

In past, content based image retrieval is done using one or two low level features such as shape, colour and texture. The Content Based Image Retrieval (CBIR) systems display the large amount of results at the end of the process this will drove the user to spend more time to analyze the output images. In this paper we combine Bayesian and genetic algorithm techniques, advantages and disadvantages of these algorithm, content based image retrieval using genetic algorithm and Bayesian algorithm used for content based image retrieval are discussed in this paper. Proposed work is the hybrid approach for the content based image retrieval which is the combine effort from the genetic and Bayesian algorithm and we have analyzed the result based on the results observed on a sample dataset taken by us. On combining effort of two algorithm which already proven their track of efficiency here we are obtaining best of them and analyzing the result.The procedure is described proposed Bayesian and Genetic algorithm approach is as follows.

1. Add dataset image
2. Create feature database for all image dataset
3. Given query image
4. Classify images using Bayesian classifier
5. Similarity measures using genetic algorithm
6. Shown most similar to query image

Simple categories of images in database

4. EXPERIMENTAL RESULT AND ANALYSIS

In CBIR firstly user adds image data set. Then it's extract and retrieve image through texture and feature algorithm. In texture and feature algorithm firstly browse input image apply this algorithm for retrieving our dataset. All procedure of our system can be illustrate as different matlab screens which shown below as we used in our experiments.

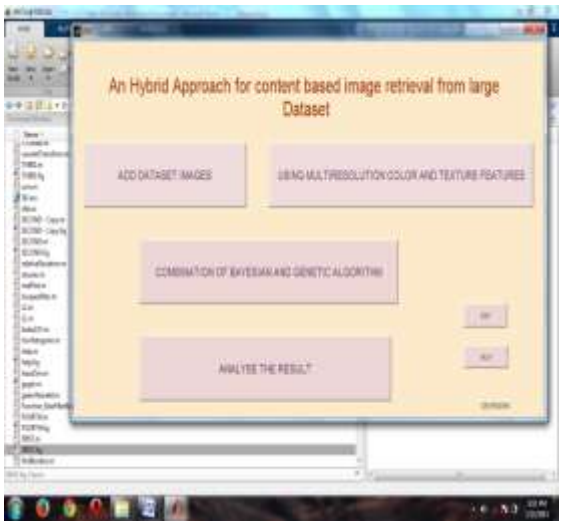

Fig 1: Home screen of ou experiments

In this approach first we need add dataset of images. Which extracts features of all images. After loading of process we can retrieve the query image.

Fig 2: result of texture and multiresolution color feature algorithm

It's extract and retrieve image through multiresolution color and texture features. Now we shown using Bayesian and genetic algorithm. In Bayesian and genetic algorithm firstly browse input image apply this algorithm for retrieving our dataset. Now we get the result which we show through in below figure.

Fig 3: result of Bayesian and genetic algorithm

We compare our result through bit rates and database time. Our algorithm is more efficient and effective as compare to existing once. The proposed CBIR algorithms are implemented using MATLAB software.

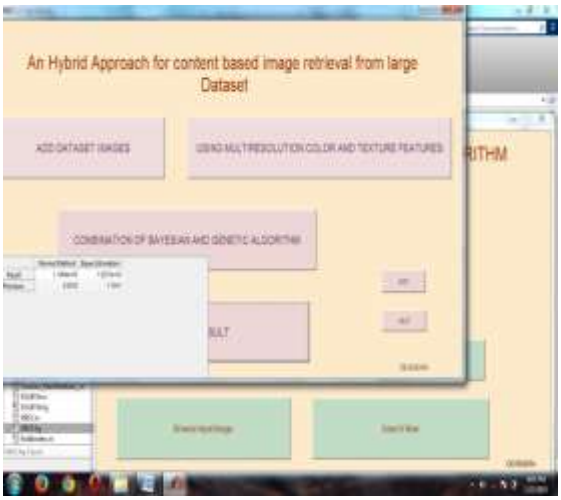

Fig 4: result of texture and feature algorithm

CBIR has done using with two efficient algorithms that are Bayesian and Genetic Algorithm. In CBIR we compare Normal estimation algorithm with Bayesian and GA. In this simulation we analyze that the bit error and database time. Here we are observing our result in two factor or aspects such as Database time in Micro second and Bit error (Error %) with which we can analyze the efficiency of the system we have designed for the CBIR. We compare our result through bit rates and database time. Our algorithm is more efficient and effective as compare to existing once. For the valuation of result two standard parameters one is Recall and another is Precision are used which shows better value in comparing to other retrieval algorithms.

Performance Evaluation Parameters The calculation for the precision and recall has been done according to the following formula and equations. The experimental setup and the calculation of the efficiency for every technique is often calculate using recall and precision and use to observe the results by different image technique while querying the dataset which we retrieved after annotation operation.

For query q,

Precision = A (q) / B (q)

Recall= A (q) / C (q)

A (q) = Similar Images to query in database
B (q) = Relevant Images retrieved by system
C (q) = Total number of images in database

So precision is represents the ratio of the number of images relevant to the query q among retrieved images to the number of retrieved images & Recall is the ratio of the number of images relevant to the query among retrieved images to the number of images relevant to the query in a DB, so all the evaluation process of query the dataset are

always done using these two calculations. After performing the calculation and simulation we have come to the best values which we have got and the values for the precision was calculated and the values for the recall was calculated, we have calculated both the parameter for different type of images and observed the following results, the categories which we have mentioned earlier are also mentioned in table and we observed that the following are the results where the precision value observed by us while compare to the existing technique calculation are best.

Category	Image Count	C & T	IGA	Proposed
African people	0-100	65	72	77
Beach	101-200	61	92	83
Buildings	201-300	68	86	84
Bus	301-400	62	72	78
Dinosaurs	401-500	68	97	83
Average Recall		65	80	81

Table 1 Precision values for three methods

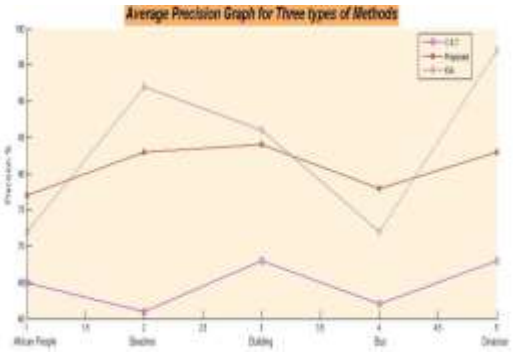

Fig 5: Precision graph for results

Recall Calculation:

Recall also calculated and experimented 5 times in all the 5 categories to observe the best results and presented via graphically to show the experimental results better than the existing one and the observed results are below:

Category	Image Count	C & T	IGA	Proposed
African people	0-100	13	14	16
Beach	101-200	12	18	17
Buildings	201-300	14	17	17
Bus	301-400	12	14	16
Dinosaurs	401-500	14	19	17
Average Recall		13	16	17

Table 2 – Recall values for these methods

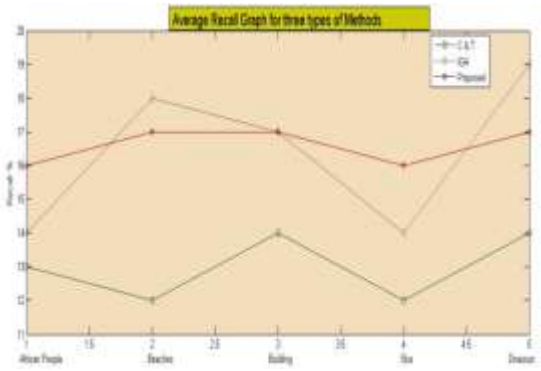

Fig 6: Recall graph for results

Here we have observed the following results and proposed approaches in recall compared with the existing approached the experiments performed by us. On analyzing through line graph drawn for the existing and proposed technique for the recall and precision % and the observed results are showing a better output and better results visually, also in order to show a difference we have taken a range from 1-2.2 so that its showing the clear difference in between our approach and existing approach clearly in recall, also in precision we have taken the small range from 1-2.6 in order to monitor best results to observe.

5. CONCLUSION AND FUTURE WORK

In past, content based image retrieval is done using one or two low level features such as shape, colour and texture. The Content Based Image Retrieval (CBIR) systems display the large amount of results at the end of the process this will drove the user to spend more time to analyze the output images. This

technique provides the optimum solution for the image retrieval as compare to other techniques. In past, content based image retrieval is done using one or two low level features such as shape, colour and texture. In this paper we combine Bayesian and genetic algorithm techniques. Here we are using Bayesian algorithm for classification using genetic algorithm which provide optimal solution for image retrieval. In future work we try to generate an appropriate algorithm for decrease computational time. We can investigate more methods for image retrieval which provides better result from this method.

6. REFERENCES

1. Zhong Su, Hongjiang Zhang, Stan Li, and Shaoping Ma "Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning" IEEE Transactions On Image Processing, August 2003
2. Sumaira Muhammad Hayat Khan, Dr.Ayyaz Hussain and Dr.Imad Fakhri Taha Alshaikhli "Comparative study on Content-Based Image Retrieval (CBIR)" International Conference on Advanced Computer Science Applications and Technologies, 2012.
3. Chih-Chin Lai and Ying-Chuan Chen "A User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm" IEEE Transactions On Instrumentation And Measurement, October 2011
4. Prachi Bhende and Prof. A. N. Cheeran "Content Based Image Retrieval in Medical Imaging" International Journal of Computational Engineering Research.
5. Mohd. Danish, Ritika Rawat and Ratika Sharma "A Survey: Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy" International Journal of Engineering Research and Applications, September 2013.
6. Khlifia Jayech and Mohamed Ali Mahjoub "New approach using Bayesian Network to improve content based image classification systems" IJCSI International Journal of Computer Science Issues, November 2010
7. Nidhi Singhai and Prof. Shishir K. Shandilya "A Survey On: Content Based Image Retrieval Systems" International Journal of Computer Applications, July 2010.
8. Ritika Hirwane "Fundamental of Content Based Image Retrieval" International Journal of Computer Science and Information Technologies, (2012).
9. Neha Jain, Sumit Sharma and Ravi Mohan Sairam "Content Base Image Retrieval using Combination of Color, Shape and Texture Features" International Journal of Advanced Computer Research, March 2013.
10. Amanbir Sandhu and Aarti Kochhar" Content Based Image Retrieval using Texture, Color and Shape for Image Analysis" International Journal of Computers & Technology, AUG 2012.
11. Hatice Cinar Akakin and Metin N. Gurcan "Content-Based Microscopic Image Retrieval System for Multi-Image Queries" IEEE Transactions On Information Technology In Biomedicine, July 2012.
12. Mr. S. Manoharan and Dr. S. Sathappan" a novel approach for content based image retrieval using hybrid filter techniques" International Conference on Computer Science & Education, April 2013.
13. Nishant Singh, ShivRam Dubey, Pushkar Dixit, Jay Prakash Gupta" Semantic Image Retrieval by Combining Color, Texture and Shape Features" International Conference on Computing Sciences, (2012).
14. S.R.Surya, G.Sasikala "An Enhanced Image Retrieval using Contribution-based Clustering Algorithm with Spatial Feature of Texture Primitive and Edge Detection" International Journal of Computer Applications (0975 – 8887) August 2012.
15. Erchan Aptoula and Sebastien Lefevre "Morphological Description of Color Images for Content-Based Image Retrieval" IEEE Transactions on Image Processing, vol. 18, no. 11, November 2009.
16. Mahmoud R. Hejazi, Yo-Sung Ho" An Efficient Approach to Texture-Based Image Retrieval" Gwangju Institute of Science and Technology, 15 October 2007.