CONTENT BASED IMAGE RETRIEVAL WITH HASH CODES

MAITHIILI.K, ELAKKIYA.P, GAYATHRI.G, SANGEETHA.A

***Abstract*: This paper proposes a new CBIR(Content Based Image Retrieval) system using Hash codes performs few major tasks. The first one is feature extraction (FE), where a set of features or feature vector, is generated for accurate representation of content for each image in the Wang dataset comprising of 499 images. A feature vector is much more smaller in size compared to the original image. The second is similarity measurement(SM), where a distance between the query image and each image in the dataset using their features is computed so that the top images can be retrieved. The third is hash functions produce hash values based on the image visual appearance. Hash function calculates similar values for similar images also for dissimilar images dissimilar hash values are calculated. using similarity function to compare two hash values, it can be decided whether two images are different or not. Another task is hamming distance to calculate how many bits are different in hash .**

***Index terms- Hash function, Hamming Distance,* CBIR, FE ,SM**

I. INTRODUCTION

The Content based image retrieval (CBIR) is an automated technique that takes an image as query and returns a set of images similar to the query image. Low–level images features like texture , color are extracted from the images of the database to them in terms of their features .Images of the same classes are expected to have same and similar character. Therefore, when similarity measure is performed on the image features, the resulting set achieves a high level of retrieval performance. CBIR has several advantages over the traditional text based retrieval .Using the visual contents of the query image in CBIR is a more efficient way of finding relevant images than searching based on text annotations .

The Proposed CBIR with hash codes performs the image retrieval with less time also the result images matches the retrieved images maximum upto 98 percentage due to the efficiency of the algorithms used in existing and proposed system. efficient search mechanism, is existing image features are of high dimensions and current image dataset are huge hence comparing a query with every dataset sample is computationally critical and hard.

***MAITHILI.K**, Information Technology , Christ college of engineering and technology, Puducherry, India.*
***ELAKKIYA.P**, Information Technology , Christ college of engineering and technology, Puducherry, India.*
***GAYATHRI.G**, Information Technology , Christ college of engineering and technology, Puducherry, India*
***SANGEETHA.A**, Information Technology , Christ college of engineering and technology, Puducherry, India.*

The Content based image retrieval techniques can capture low –level image features also accept pictorial queries too, Proposed CBIR searches for required images by their features , URL or path queried by the user which are automatically indexing the retrieval system. Proposed CBIR eliminates time required for annotation of keywords in text based image retrieval(TBIR) Also improves proposed system efficiency since it not limited by the number of keywords whereas Image representation and similarity measurements are the main aim of Proposed CBIR. A feature set is saved in a dataset which represents the most important areas of a set of images. Feature vectors are extracted from an image, and their similarity is measured against the Wang dataset.

The, Major problems in the existing image retrieval are searching and retrieving irrelevant images , low matching of input image and output image. To tackle these problems we go for hash function , hamming distance for finding difference of bits in images using algorithms in existing [1] and proposed algorithms the best retrieval results is produced. The rest of the paper is organized as follows; The introduction is followed by the literature survey in section II. Section III describes the existing system. Section IV consists of our proposed system with its explanation , algorithm and advantages followed by conclusion in section V.

II. LITERATURE SURVEY

[1]QAIS is Query Adaptive Image Search introduces an approach which enables query-adaptive ranking of the fetched images with equal Hamming distances value to the query images. Thus the result is achieved by using learning bitwise weights with the hash code algorithm for a different set of predefined semantic concept classes in the dataset. Query-adaptive learning weights are then computed online by testing the proximity between a query image and image in the semantic concept classes in Flickr image dataset . Scalable image search that are based on the matching and similarity measurements have always been a top research topic .**T**he Hamming distances are unique integer values .

ISSN: 2278 – 1323

Large number of images having same Hamming distances which at large scale disturbs the search results with weighed ranking is more important.

[2]In Lost in binarization : Query adaptive ranking for similar image search with compact codes performs During image search, binary code of the input image are used to search against labelled images in the predefined semantic classes of the dataset. It uses a large set of images that are more proximate to the query in the Hamming space to find its semantic content also compute the weights using the pre-computational weights of the relevant semantic classes in dataset. Lastly, images in the dataset are quickly ranked using weighted Hamming space.
[3]This Weakly supervised hashing in Kernel space is very popular in computer research owing to the handling of nonlinearly divided data. The kernel function is a more potential choice for image similarity measurements. It increases the risk for data indexing The two issues motivate weakly supervised hashing in kernel space that proposes a label regularized maximum-margin partition (LAMP) method (3) to improvise hashing performance by using kernel-based similarity approach with additional less number of pair wise constraints .
[4] Complementary Hashing for Approximate Nearest Neighbor Search provides hashing based on Approximate Nearest Neighbor (ANN) algorithm that has been getting lots of attention in computer research. The data-dependent hashing methods, e.g. Locality Sensitive Hashing (LSH) uses a single hash table to retrieve growing number of hash buckets containing the input image. It propose a complementary hashing technique which balances the precision and recall in a more effectively. Thus finds true proximate neighbours.

[5] SPLH can perform hashing technique which uses random projections from the data to derive hashing. Result suffers poor discrimination when compact codes are enabled. Sequential projection learning for hashing with compact codes propose a unique data dependent compact projection learning technique such that every hash function is designed for correcting the errors made by the previous . The proposed technique is easily adaptable to both unsupervised and semi-supervised learning to signify the performance over the two large datasets consisting upto 1 million.
[6]It propose a semi-supervised hash function that is described as minimum empirical error on labelled data with maximum variance and independent hash bits against the labelled and unlabelled data. The proposed methodology handles both metric and semantic similarity measurement. Unsupervised hashing technique has high performance with distances in image , semantic similarity is labelled pair of images .

III EXISTING WORK

A comparison by human on two images based on their own semantics, while a image retrieval system relies on comparison of features or feature vectors corresponding to visually similar image features. Also one of the major tasks of proposed CBIR systems is similarity comparison, extracting feature of every image on its pixel values defining algorithms for comparing images. Images are compared by calculating difference of its feature components to other image descriptors in the database. Based on the similarity value, images are ranked for retrieval and this process is not very efficient in the existing system. Also the user feedback if he is not satisfied with the displayed database images is not provided here.

An image retrieval system like traditional CBIR relies on match checking of feature vectors corresponding to visual image features. One of the major tasks of CBIR systems is similarity comparison, extracting feature of every image on its pixel values defining algorithms for comparing images. Images are compared by calculating difference of its feature components to other image descriptors in the database. Based on the similarity value, images are ranked for retrieval and this process is not very efficient in the existing system.

Existing System Description :

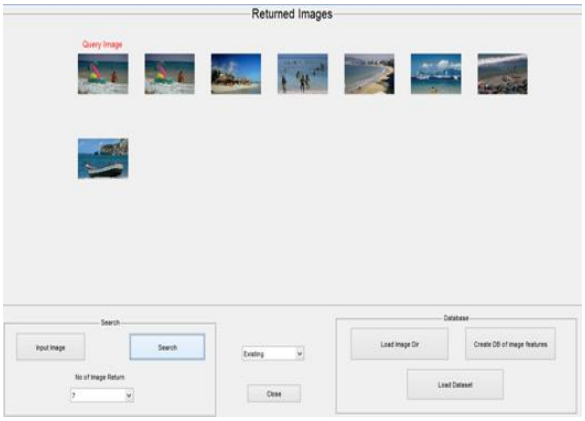

Fig 1: Top 7 Returned Images of existing System

Illustration of Figure 1 is that the existing content based image retrieval system takes the input query as an image and return or fetch the images similar to the query image but it retrieves irrelevant images within top 5 images itself hence the existing system has less effectiveness , efficiency and poor accuracy. It also extracts feature and perform embedding in hash code.

Drawback of the existing system is that it Does not identify the low level features that meet the requirement of a content based image retrieval. Another one is In order to retrieve the desired images from a large database, No effective and efficient algorithm has been used. Also it Does not provide relevant feedback search if the user is not satisfied with retrieved images. Time taken to load the dataset is also high . Thus to tackle these issues a more effective and efficient algorithm was required. Hamming distance has to be calculated to eliminate irrelevant image retrieval .

ISSN: 2278 – 1323

This existing CBIR approach is constructed to use a single set of hash codes for search of image. It is further extends this approach to the case where variety of different sets of hash codes exist. Also there are various different semantic concept classes in dataset,
it gets easy to train dataset and a set of hash codes for every class by using images including the semantic concept.
The class-specific hash codes are always expected to be much more discrete for a specific class of images. class-specific hash codes are used to produce a group of bitwise weights. During online search, The semantic dataset in existing is used both for computing query-adaptive weight and selecting a suitable set of class-specific hash codes for each input image query.

IV Our Proposed Concept

In this project to provide a efficient image retrieval few tasks are performed as Images are being resized to image size of 8 x 8 irrespective of all different image alterations hence we get same hash even if there is little variation in image. This results in number of pixels being reduced to 64.Features such as Red, green and blue component of images are extracted and image is converted to gray scale and RGB mean is being computed. Hamming space value is calculated to find how many bits are different in hashed image.

Many image Retrieval systems take a simple approach by using typically normal-based distances on the extracted feature set as similarity measurement function. The main purpose behind these proposed CBIR systems is that given a good set of features extracted from the images in the dataset (here the images are stored.) then for two images to be similar, their extracted features vectors have to be close to each other.
The techniques or algorithms used in our proposed CBIR are Feature Extraction algorithm in this Active Contour Selection processes the image and compute feature vector. This feature is used for the extraction of information from Wang dataset. Features can be human vision related, low-level features, middle-level features, high-level features and semantic related. Most commonly used features for retrieval are color, shape, texture. Similarity measurement algorithm performs matching which compares between extracted features of images to find whether they are similar or not and up to what level they are matching. also compare two images, distance of zero shows that it is very similar picture , distance of 5 means a few things may be different if 10 means fully different picture. hashing algorithm proposed advocates to generate the hash value for each image in the dataset.The hash value of query image is used to calculate the hamming distance. Finally the retrieval of images from the dataset are of same class and cross class is checked to calculate the ranking of retrieval.

There are 4 modules: They are Administrative Database, Feature Extraction , Similarity Comparison, User Evaluation.

Administrative Database :

This database is maintained for the storage of images. For each image feature vectors such as mean and standard deviation for Red, Green and Blue component of the image. Wang dataset that is being used as a training dataset up to 499 images with different concept classes for feature vector generation in the proposed CBIR . Administrative database performs maintaining the database, update the database regularly according to user request, also classifying images for efficient searching. Query Images are loaded according to user requirement.

Feature Extraction :

The features of the query image and the database images are computed. The features considered are:
(1) Active contour selection (region identification)
(2) Statistical features such as mean and standard deviation values of the color image.

Figure 2 Illustrates system architecture of the proposed system in block diagram here firstly the input image is being queried by the user hence the features of the images are extracted by using the feature extraction algorithm. The features of the image are extracted from the images in the wang dataset.

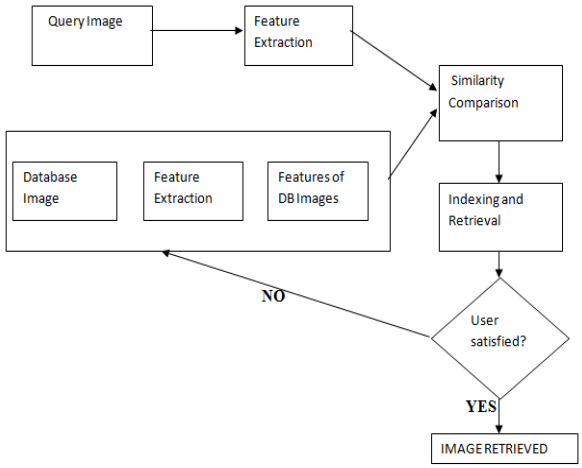

Fig 2: Proposed System

After feature extraction similarity comparison or match checking is performed by using the similarity measurement algorithm to find the relevant image to the query image .Using Hamming distance the bits of image are found to retrieve the similar and dissimilar image . Indexing and Hashing is done to reduce time and for quick image retrieval .The images are returned according to query image and number of retrieval if not satisfied relevance feedback of search with another image is performed.

Similarity Comparison :

Every image present in the database the system identifies region and extract feature vectors. To compare the similarity between query image and images in the database the algorithm proposed here has the ability to improve search performance and they have the advantage that they directly work with a coding of parameter set, search process is carried out.

ISSN: 2278 – 1323

Fig 3: Top 6 returned images of Proposed System

User Evaluation :

If the user is satisfied with the image retrieved, then the process ends. Else, the user is allowed to select and rank the first few retrieved images whose feature vectors are computed again. The iteration continues till the user is satisfied and gets the relevant image to the query image.

Advantages :

Quick retrieval of relevant images using Hash codes. another is Complete User satisfaction with Relevance image as feedback.

V.CONCLUSION AND FUTURE ENHANCEMENT

In this paper, we Propose a Hashing algorithm using Euler's Formula along with the Feature extraction algorithm and existing system algorithm for more better user satisfaction and output is not accepted until the relevance image is retrieved by user providing relevance image as feedback. Thus We leave a full fledged implementation of Content Based Image Retrieval(CBIR) also known as Query By Image retrieval with hash codes having more space and less time consuming. The Quality of results is also high and Implementing this Project Using Visual Basic , MATLAB .

REFEERENCE

[1] Y-G Jiang , J Wang , X Xue and S-F Chang, "Query adaptive Image Search with hash codes ," *IEEE Trans. on Multimedia*, vol. 15, no.2, Feb 2013.

[2] Y.-G. Jiang, J. Wang, and S.-F. Chang, "Lost in binarization: Query- adaptive ranking for similar image search with compact codes," in Proc. ACM Int. Conf. Multimedia ,2011

[3] C. Y. Mu, J. Shen, and S. Yan, "Weakly-supervised hashing in kernel space," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.

[4] Hao Xu , JD Wang , Zhu Li , G Zeng , S Li and Nenghai Yu ," Complementary hashing for approximate Nearest Neighbor Search ,"in Proc IEEE transaction ,2011.

[5] J. Wang, S. Kumar, and S.-F. Chang, "Sequential projection learning for hashing with compact codes" in Proc. Int. Conf. Machine Learning, 2010.

[6] J. Wang , S.Kumar and S-F Chang ," Semi-supervised hashing for scalable image retrieval ," in Proc IEEE Conf. Computer Vision and Pattern Recognition , 2010.

[7] H. Jegou, M. Douze, and C. Schmid, "Improving bag-of-features for largescaleimagesearch,"Int.J.Comput.Vision,vol.87,pp.191–212, 2010

[8] W. Liu, J. Wang, S. Kumar, and S.-F. Chang, "Hashing with graphs," in Proc. Int. Conf. Machine Learning, 2011

[9] R.-S. Lin, D. A. Ross, and J. Yagnik, "Spec hashing: Similarity preserving algorithm for entropy-based coding," in Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2010.

[10] Y. Weiss, A. Torralba, and R. Fergus, "Spectral hashing," in Adv. Neural Inf. Process. Syst., 2008.

Maithili.K, Working as an Senior Assistant professor in Christ College of Engineering and Technology, Puducherry, India

Elakkiya.P, Pursuing B.Tech degree in Christ College of Engineering and Technology, Puducherry, India

Gayathri.G, Pursuing B.Tech degree in Christ College of Engineering and Technology, Puducherry, India.

Sangeetha.A, Pursuing B.Tech degree in Christ College of Engineering and Technology, Puducherry, India

ISSN: 2278 – 1323