International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 05 | May -2017 www.irjet.net p-ISSN: 2395-0072

Retrieving Large Scale Content Based Image Using Privacy Preserving Framework

Komal Satam [1], Sayali Satpute [2], Pranita Umarjikar [3]

[123]*UG DYPIET Ambi, Pune*

Abstract - *Content based information system in which content is used for data retrieval. The main point is that if content is sensitive and user don't want to reveal this query to CBIR system then there is requirement of privacy preservation. In present content based information retrieval system, one method is that both client and server sides are loaded with the same tasks. It supply two layers of protection. First, robust hash value are used as queries to protect original content. Second, in future to decrease the inexactness of the server the client can choose to delete some bits in hash values, robust hash algorithm are used, first one is random projection and second is discrete wavelet transform. This becomes better the performance of retrieval. In the content based image retrieval system, images are classified into different categories based on their characteristics or properties of color. The input is given as image to the system and based on the input image resulting images are made available by seeking its properties and comparing with each image in database. For sensitive image, locking is provided for privacy of the image and it will not be displayed in the destination folder.*

Key Words: Image Retrieval, Multimedia Database, Indexing, Security, Feature Matching, Feature Extraction.

1. INTRODUCTION

Nowadays, multimedia content is produced in large amount and occurred throughout an area. Content based search methods are designed for searching content in large database. These methods are used by Content Based Information/Image Retrieval(CBIR)[1]. In this framework, user send set of queries/Images to system and in return they get proper information or similar images from the database. Some issues like query image or database involves sensitive information. In this environment, the database owner plays important role. User and service provider can work with different parties which they do not believe each other. Privacy problem arises when untrusted party tries to access the secured image. This involves the need of PCBIR(Privacy Content Based Image Retrieval) system[1].Information retrieval handles the representation, storage and retrieval of unstructured data. Classical information retrieval issues mainly with text.The development of multimedia databases and of the web have given new interest to Information Retrieval[1].

In Content Based Image Retrieval, a user has an image knows as query image and he or she is interested to search the similar image from image database. User sends image in order to obtain similar images. Image query refers to the problem of finding object that are relevant to a user query within image database[10].Recently researchers are working on three parameters of CBIR: Recall, Precision and Retrieval time. Precision is ratio of appropriate retrieved images and total retrieved images. Recall is ratio of relevant retrieved images and total relevant images in database. Retrieval time is time taken to retrieve all images[10].

Feature extraction is the process of extracting visual features which includes color, shape, and texture[11]. In this features color is most vital, reliable and widely used feature. The main issue in color feature extraction contain color space, color quantization and the choice of matching function[9].RGB color model is used for extracting color features. RGB color model is an additive color model in which green, red and blue light are added together in various ways to reproduce a broad array of colors. The name of the model comes from the initials of the additive primary colors, green, red and blue. The RGB value of query image is converted into Hue Saturation Value (HSV)[11]. In image database , RGB values of images are also converted into HSV and comparison is done for similarity matching. The HSV of query image is comparing with HSV images in database. Then which images are matches to query image is retrieved by user in particular destination folder[11].

Privacy Content Based Image Retrieval (PCBIR) framework concerns for both public and private database[1]. An effective image retrieval system needs to operate on set of images to seek the relevant images based on input image[8]. According to the instance, cost of privacy protection is effective. Image retrieval is technique of searching digital images from database[6].The Content Based Image Retrieval (CBIR) uses image content to find and retrieve digital images from collection of images. Content based image retrieval is set of techniques for retrieving semantically relevant images from an image database[7].

The rest of work is organized as follows: Section II is brief literature review. In Section III there are existing system information. Then in section IV Proposed work is done. In section V and VI system architecture and algorithm done respectively. In VII and VIII Mathematical model and screen

shots is given respectively and finally IX concluding the work.

2. LITERATURE SURVEY

Content-based multimedia information retrieval: State of the art and challenges[2]. This paper proposed algorithms for searching nearest neighbor matches to high dimensional data, these two useful algorithms are: the randomized k-d forest and the priority search k-means tree. To scale large data sets that would not fit in a single machine memory, they proposed a distributed nearest neighbour matching framework. All this research has been reveal as an open source library such as fast library for approximate nearest neighbours (FLANN). It gives accuracy up to 99 percent as a result. It has advantages like extensible for large High Dimensional data and high accuracy.

Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends [3]. This paper presents Discrete Cosine Transform (DCT) hashing method for creating index structures for face descriptors. A hash index is created, and further queried to search the images more similar to the query image. DCT hashing algorithm has best retrieval accuracy and more efficient compared to other popular state of- the-art hash algorithms. It provides 88 percent retrieval precision. It has advantages like speedy and computationally cheap and outperforms than LSH, E2LSH and KLSH for nearest neighbour recall. The disadvantage is cost of computing the hash.

On the use of LSH for privacy preserving personalization [4]. This paper presents a new Nearest Neighbour (NN) framework: Robust Sparse Hashing (RSH). This paper appeal to inspire by the concept of dictionary learning for sparse coding. For precise and fast Nearest Neighbour retrieval, the main idea is to sparse the code data by using learned dictionary, and then produce hash codes out of these sparse codes. Results tell that RSH provides different accuracy with various dataset that is 92 percent - MNIST dataset, 100 percent - SIFT dataset. It has advantages like fast Hash creation and best precise performance on SIFT and MNIST.

One-way private media search on public databases: The role of signal processing [5]. This paper derived the Projected Residual Vector Quantization (PRVQ) algorithm. The effectiveness of PRVQ algorithm is consolidated on two kinds of high dimensional vectors: GIST and vector of locally aggregated descriptors (VLAD). PRVQ outperforms existing methods, for example product quantization (PQ), transform coding (TC), and Residual Vector Quantization (RVQ). It gives 30 ms per vector as a result of search time/ speed up parameter. It's advantages are high Precision and disadvantage is no cooperative framework.

3. EXISTING SYSTEM

The existing Content Based Information Retrieval system without privacy policy in which, client can generate query based on the content. This query is generated without any privacy policy. The whole query is known to the server. When this query is send as an input query to the server. The role of server is to check the information related to query in database. Server checks nearest images or document in database using indexing algorithm. The generated nearest images or document based on input query are grouped into list which is generated by server knows as candidate list. This candidate list is send to the client and client can choose the relevant output for the query based on the candidate list. Following diagram shows the existing system.

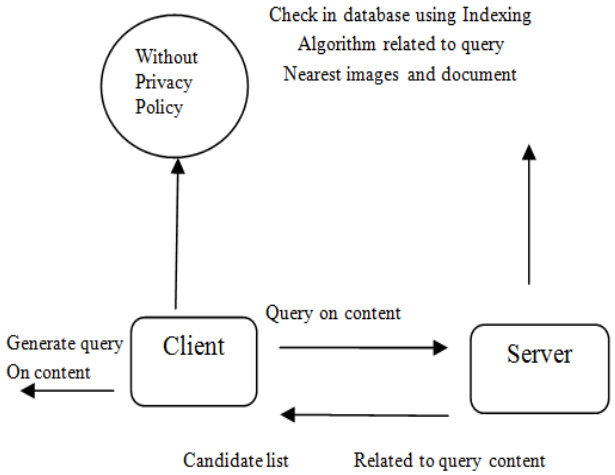

Fig 3.1:- Existing System

4. PROPOSED SYSTEM

In proposed system, user provides the image itself as input and on the basis of input image user can get the similar images as result. All the resulting images are retrieved from database and stored in output folder or destination folder. For the classification of images in database this system uses clustering algorithm that is k-means. All the images in database are clustered according to the input image. In a particular cluster nearest images of input image is stored and then these images user obtained in destination folder and choose relevant image as output.

Images get classified into different categories according to given input image. Users have following option for

classification of image- Bus, Horse, Flower, Profile Photos, Group Photos and Abstract etc. Users can select source folder of dataset and destination folder for classification of images by clicking browse button. Using abstract option user can give one photo as input and user can see related images in panel. Security to particular images will be given prior to retrieving. Those secured images will not be displayed in destination folder. Content Based Image Retrieval is used to retrieve images based on given input image.

5. SYSTEM ARCHITECTURE

Content Based Image Retrieval system is used to retrieve relevant images based on the input image from database. Fig 5.1 shows the system architecture of proposed system. There are few steps for retrieving images, firstly user provides input as an image to searching similar images from image database. Then the features of image are extracted by feature extraction. In feature extraction visual features of image is extracted such as color. When choosing the image retrieval based on color, system traverses each image in image database. The image database is collection of different images in which features of each image are extracted. Then at the time of feature matching, system calculates characteristics of images and measures the similarity with key image. Then system calculates similarity distance of images present in database. When the distance is less, then there are more chances to get similar feature image based on key image. All these relevant images are retrieved by user and stored in user defined destination folder.

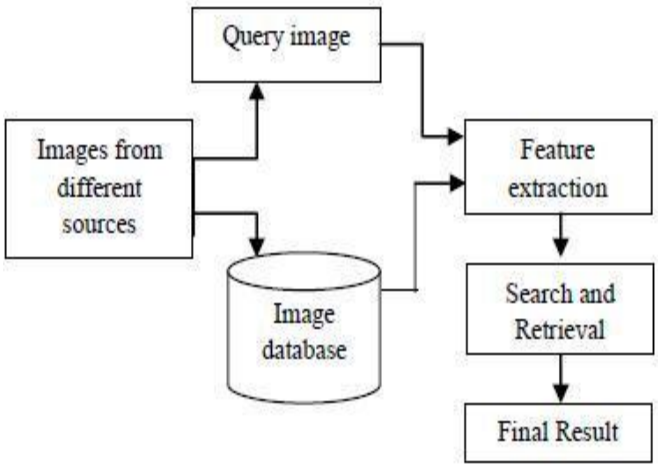

Fig 5.1:- System Architecture

6. ALGORITHM

6.1 K-means Algorithm

In proposed system, K-means clustering algorithm is used.K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Step 1: Input as an image.
Step 2: Initialize cluster centroid.
Step 3: Calculate Euclidean Distance.
Step 4: Move on next observation and calculate Euclidean Distance.
Step 5: Calculate Euclidean distance for the next observation, assign next observation based on minimum Euclidean distance and update the cluster centroid.
Step 6: Continue the steps until all observations are assigned.

Euclidean Distance: It measures the similarity between the two different feature vectors. Q and D are feature vectors of the query image and database image.

7 . MATHEMATICAL MODEL

A mathematical model is a description of a system using mathematical language and concept. The process of developing a mathematical model is termed mathematical modeling. Mathematical models are used in the natural sciences and engineering disciplines as well as in the social sciences . A model may help to explain a system and to study the effects of different components, and to make predictions about behavior.Following is the mathematical model for system.

Identify the processes as P.
S= {I, O, P....}
P= {IQ, IE}

Where,
IQ is Image Query
IE is Information Extraction.
P is processes.
IQ= {U, MAX, EF}

Where,
U=input Query
MAX = {1, 2, 3, ... , n}

IE= {EF, FM Techniques, Info}

Where,

EF is input which is extracted features of queried image given to IE
FM is used for matching features of queried image with all database image.
Output is similar images matching with queried image.

8. Evaluation

Step 1:- Selection of source and destination folder with the help of browse button and specify the path. The retrieved images will be stored in the destination folder.

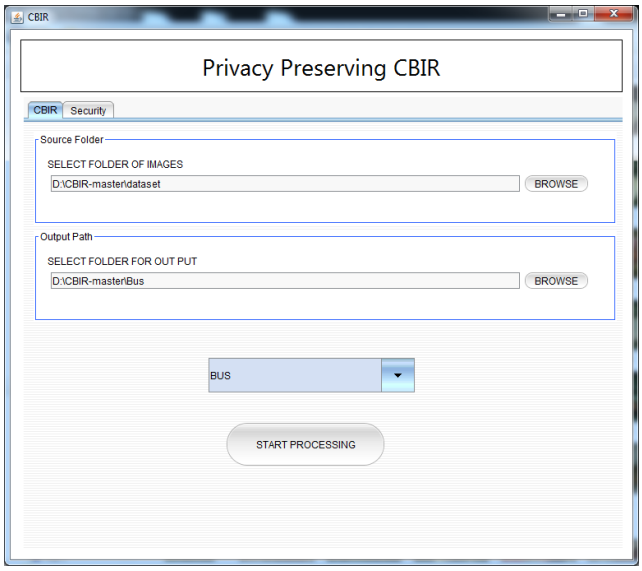

Fig 8.1: Privacy Preserving CBIR Framework.

Step 2:-.Using drop down button user can select category of images such as bus, flower, horse, etc.On clicking the start processing button, processing will be done.

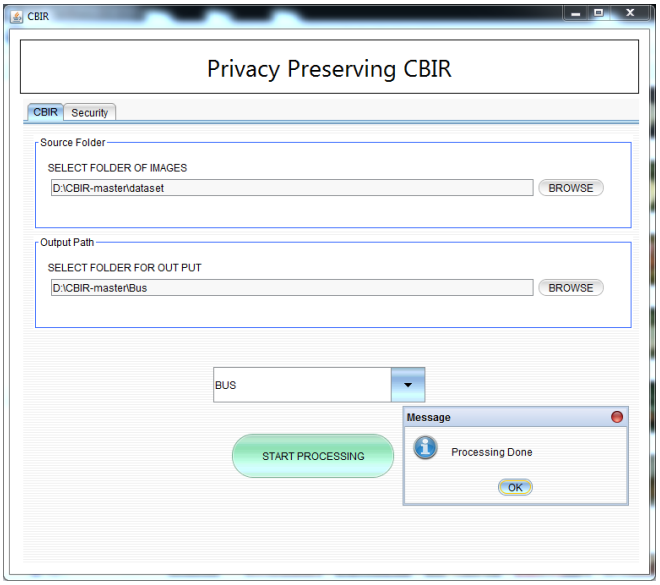

Fig 8.2: Classification and processing

Step 3:-After processing, system uses clustering for grouping of similar images. Based on feature matching of images, it displays related images in the destination folder.

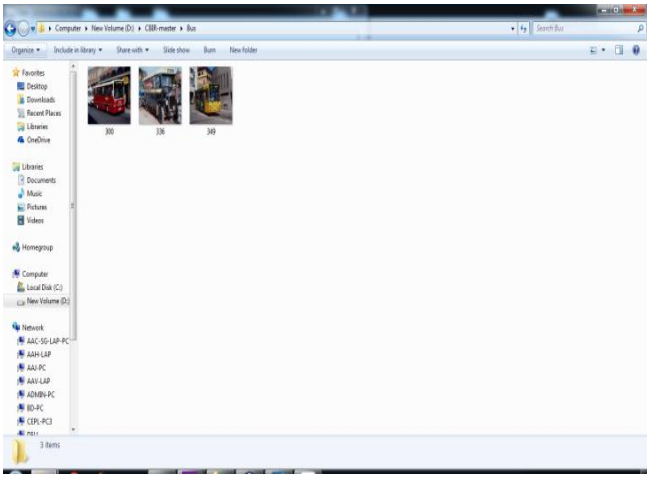

Fig 8.3: Retrieved Resulting Images.

Step 4:-User can give particular image as an input using browse button. By applying CBIR system related images will be displayed in the panel itself.

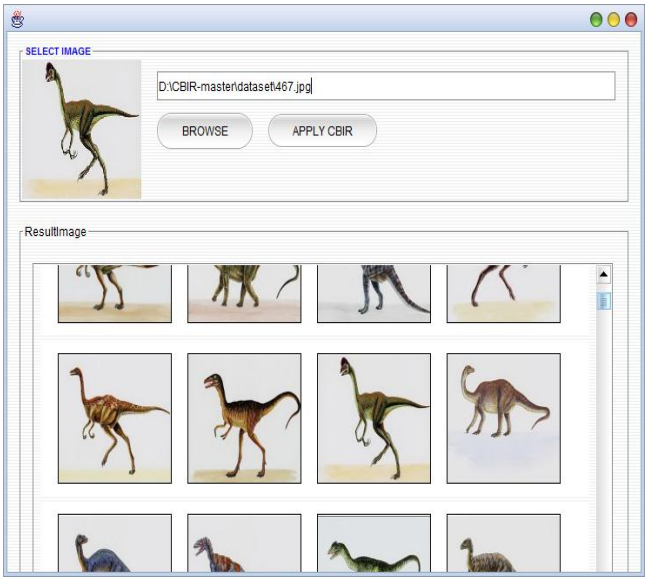

Fig 8.4: Specific Image as an input.

Step 5:-To provide security, user can browse the path of particular image. Multiple images can be added to the security list one by one. By clicking the secure images button, user can check out the previously secured images.

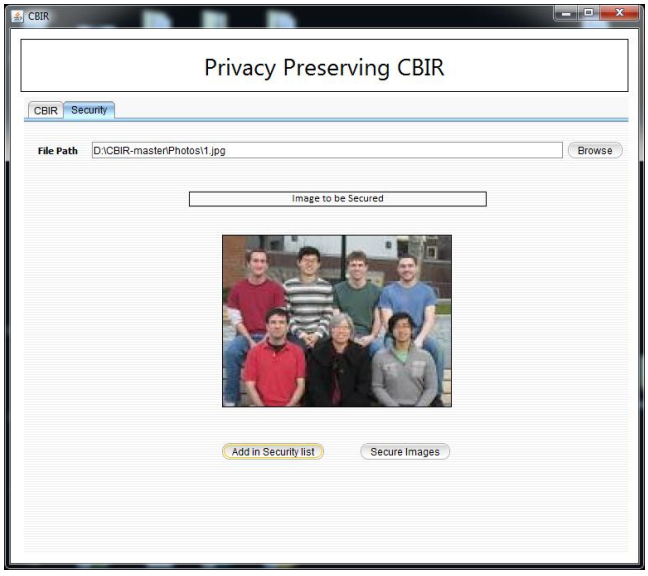

Fig 8.5: Security to particular image

9. CONCLUSIONS

A new protected framework is particularly designed for the search, and retrieval of content based images.It works as a redefine search system in which input is given as image and based on the input image the resulting images are made available. For classification purpose partial clustering is used. K-means clustering algorithm is applied depending on which the particular image is stored in the specific cluster. This framework will give security by giving lockingfor particular images. Those secured images will not be displayed in output folder and cannot be retrieved by users.

The proposed work has used multiple clusters for storing images and increases the size of the dataset. Future work will provide features like file extraction and includes more categories to retrieve images. It will also be used in face recognition system.

ACKNOWLEDGEMENT

Authors would like to thank Li Weng, Laurent Amsaleg, April Morton.

REFERENCES

[1] Li Weng, Laurent Amsaleg, and April Morton, A Privacy-PreservingFramework for Large-Scale Content-Based Information Retrieval ieee trans-actions on information forensics and security, vol. 10,no. 1, january 2015.

[2] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," ACM Trans.Multimedia Comput., Commun., Appl., vol. 2, no. 1, pp. 1–19, Feb. 2006.

[3] J. Bringer, H. Chabanne, and A. Patey, "Privacy-preserving biometric identification using secure multiparty computation: An overview and recent trends," IEEE Signal Process. Mag., vol. 30, no. 2, pp. 42–52, Mar. 2013.

[4] A. Aghasaryan, M. Bouzid, D. Kostadinov, M. Kothari, and A. Nandi, "On the use of LSH for privacy preserving personalization," in Proc. 12th IEEE Int. Conf. Trust, Secur., Privacy Comput.Commun. (TrustCom), Jul. 2013, pp. 362–371.

[5] G. Fanti, M. Finiasz, and K. Ramchandran, "One-way private media search on public databases: The role of signal processing," IEEE SignalProcess. Mag., vol. 30, no. 2, pp. 53–61, Mar. 2013.

[6] Shereena V.B. and Julie M. David, "content basedimage retrieval :a review," csit 2014, pp.65-77.

[7] Neha Jain, Sumit Sharma, Ravi Mohan Sairam,"Result Analysison content based image retrieval using combination of color,shape and texture features" Volume-2 Number-4 Issue-7December-2012, pp.35-42.

[8] Hui Hui Wang, Dzulkifli Mohamad, N.A. Ismail, "Approaches,Challenges and Future Direction of Image Retrieval,"journal of computing, volume 2, issue 6, june2010, pp.93-199.

[9] Amandeep Khokher, and Rajneesh Talwar, "Content-basedImage Retrieval: Feature Extraction Techniques andApplications," Proceedings published in International Journal of Computer Applications, pp.9-14.

[10] Jahnavi Shukla, Jignesh Vania, "A Survey on CBIR Features Extraction Techniques", International Journal Of Engineering And Computer Science ISSN:2319-7242, Volume 3 Issue 12 December 2014, Page No. 9555-9559.

[11] http://www.google.com