ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

A Privacy Preserving and Content Based Image Retrieval from Cloud using Gabor Features and HOG

Raju B N

Dept. of Computer Science, Akkamahadevi Women's University, Vijayapura, India

ABSTRACT: In Cloud Computing, clients are commonly reduce their management cost by outsource their image to the cloud storage. To protect an image on the cloud, Encryption on the outsourced information is a promising way, but it also introduce much difficulty to performing efficient searches over encrypted information, which can makes the search technology on plaintext unusable. Imposing confidentiality and scalability on cloud information increases the complexity of the cloud computing. In this paper we implementing the system for efficient method for data analysis by extracting the features using Gabor features and Histogram of oriented gradients (HOG). We present the new method, CBIR over encrypted images. SVM method is helps us to find the similarity of the image which adapting the machine learning approach. SVM applies a hierarchical micro-clustering algorithm that accounts the entire data set only once to provide an SVM with high quality samples. SVM tries to generate the best SVM boundary for encrypted sets given limited amount of resources.

KEYWORDS: Gabor Features, Histogram of Oriented Gradients, SVM, Encryption, Cloud Computing.

I. INTRODUCTION

Cloud computing is a rapidly growing technology where resources such as storage devices, platform and applications are shared over the internet and is widely used by multiple users in small and medium business. Cloud services can be provided and delivered remotely by vendors such as Amazon or Microsoft as "public clouds", or the resources are designed, installed, monitored and controlled internally as "private clouds".

Cloud data retrieval is an important service to be considered as certain specific data files the users are interested during a given session must be retrieved in an efficient way and quickly. Cloud computing innovation is generally developed and most associations need to utilize this innovation in their business forms. Be that as it may, then again, the utilization of this innovation is difficult and numerous associations are worried about putting away their touchy information in their server farms as opposed to putting away them in the distributed storage focuses. In the distributed computing environment, trust, as an answer for improve the security, has pulled in the consideration of specialists. Trust is a standout amongst the most imperative approaches to enhance the dependability of distributed computing assets gave in the cloud environment and has a critical part in the business situations.Some people exclusively associate big data with semi-structured and unstructured data of that sort, but consulting firms like Gartner Inc. and Forrester Research Inc. also consider transactions and other structured data to be valid components of big data analytics applications. Data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics, data mining, text analytics and statistical analysis.

Cloud storage is a modeled scenario of enterprise storage systems where information is stored in virtualized segmented pools of storage which are generally maintained by anonymous third party service providers. There are many benefits which cloud can provide customers with ranging from scalability of the services, cost availability, convenience etc.

Cloud security is the major disadvantage for its adoption. The different security issues include Data loss, multenancy issues, DDOS attacks, availability etc. Users are more concerned about their stored information into cloud and retrieval of data from the cloud. Thus, effective measures need to be taken to secure the users data. For this purpose, the paper proposes an Effective Secure Storage and Retrieve system which assures the secure managing of the user's stored data

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

and its retrieval. The proposed system adopts the encryption algorithms such as AES and FHE to secure users data. The locking mechanism is employed to allow only the authorized users to access the data.

The most common form of CBIR is an image search depends on visual Content- based image retrieval. This technique, use visual contents to search images from large scale image databases according to users interests Content-based image retrieval, such as color, shape, texture, and spatial layout to represent and index the image. Content Based Image Retrieval (CBIR) is the method that helps to organize digital image archives by visual content basis. By this definition, anything ranging from an image similarity function to a robust image annotation engine falls under the purview of CBIR.

II. RELATED WORK

Jianqing Fan et.al in his paper [1] discusses statistical methods for estimating complex correlation structure from large pharmacogenomic datasets. We selectively overview several prominent statistical methods for estimating large covariance matrix for understanding correlation structure, inverse covariance matrix for network modeling, large-scale simultaneous tests for selecting significantly differently expressed genes and proteins and genetic markers for complex diseases, and high dimensional variable selection for identify important molecules for understanding molecule mechanisms in pharma cogenomics. Their applications to gene network estimation and biomarker selection are used to illustrate the methodological power. Several new challenges of Big data analysis, including complex data distribution, missing data, measurement error, spurious correlation, endogeneity, and the need for robust statistical methods, are also discussed.

Cloud computing is obviously one of today's most tempting innovation ranges due, in any event to a limited extent, to its cost-effectiveness and adaptability. In any case, notwithstanding the surge in action and enthusiasm, there are noteworthy, industrious worries about distributed computing that are obstructing energy and will in the long run trade off the vision of distributed computing as another IT acquisition model.

IR can likewise cover different sorts of information and data issues past that predetermined in the center definition above. The expression "unstructured information" refers to information which does not have clear, semantically unmistakable, simple for-a-PC structure. It is the inverse of organized information, the sanctioned case of which is a social database, of the sort organizations generally use to keep up item inventories and staff records. Truly, no information is genuinely "unstructured".

This is certainly valid for all content information on the off chance that you tally the inactive etymological structure of human dialects. In any case, not withstanding tolerating that the planned idea of structure is unmistakable structure, most content has structure, for example, headings and sections and commentaries, which is regularly spoken to in archives by express markup, IR is also used to facilitate “semi structured” search such as finding a document where the title contains Java and the body contains threading

As of late, a few truths demonstrate that individuals give careful consideration to the protection of their pictures, and the expanding stress over security might be the key point later on cloud-based picture administrations. For instance, a cell phone application KeepSafe [9] conceals pictures and recordings from unapproved clients by basically scrambling picture organizers with a PIN code, and it possesses 13 million clients and picked up a speculation of 2 millions dollars. Besides, the face look usefulness of Facebook has been relinquished for a long time because of the protection worry from clients and governments [10].

Subsequently, huge picture need effective yet private substance based hunt works earnestly. Existing largescale picture indexing and pursuit frameworks as a rule disregard the security issue and don't bolster security insurance systems. A few frameworks label pictures with watchwords and metadata, and hunt on scrambled catchphrases, e.g., [11]–[14]. In any case, catchphrase based quest strategies can't be embraced for substance based picture seek, since they just inquiry the events of encoded labels in a precise match way and can't gauge the separation between high dimension highlight vectors. In addition, they for the most part uncover the list items to the cloud. To accomplish content-based picture look, we have to quantify the separation between sets of highlight vectors and in addition to ensure the vectors and list items.

Figure 1 shows the block diagram of proposed system. The system used to retrieve similar images based on the query image selected while retrieving the images will be in the form of encrypted manner and while displaying the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

images will be in the form of De-crypted manner and this process is known as IES-CBIR (Image Encryption Scheme-Content Based Image Retrieval).

Figure 2 shows the CBIR process. The methodology consists of two major phases. One is the training phase and the other is the testing phase. Training phase consists of reading the data, pre-processing, extraction of useful features and train the feature vector. The first step is reading and pre-processing of data from the database consists of data that are to be trained. Pre-processing is the step that makes data to accessible form. This step includes resizing, rearranging, etc. After the data is pre-processed we extract the useful and unique features that help for differentiating the data. Using CB-support vector machine these feature vectors are trained. Trained data will be stored in a knowledge base.

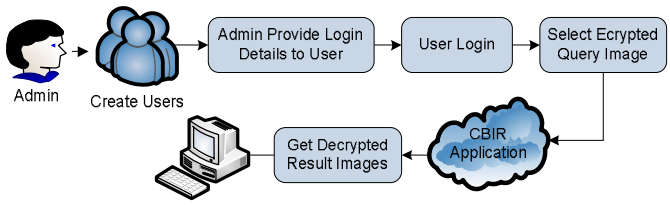

Figure1: Proposed CBIR Cloud Architecture

Testing phase is the evaluation phase. Here input data is given for test. Input data involved for pre-processing and feature extraction. After feature vector is extracted from the test data the feature vector is recognized using CB-SVM classifier by using trained knowledge. Finally the test data is categorized based on the classification result.

A. Preprocessing

Input image is applied with image pre-processing steps to enhance the visual appearance of images. Color conversion to gray image and resizing is applied. Pre-processing is followed by feature extraction.

B. Gabor Feature

The Gabor filter is generally utilized as a part of the image features. The Gabor filter wavelet is the type of sine wave adjusted by the Gaussian coefficient. The Gabor filter is helpful for extracting local and global data. The Gabor filter is tunable band pass channel, multistage and multi resolution filter.

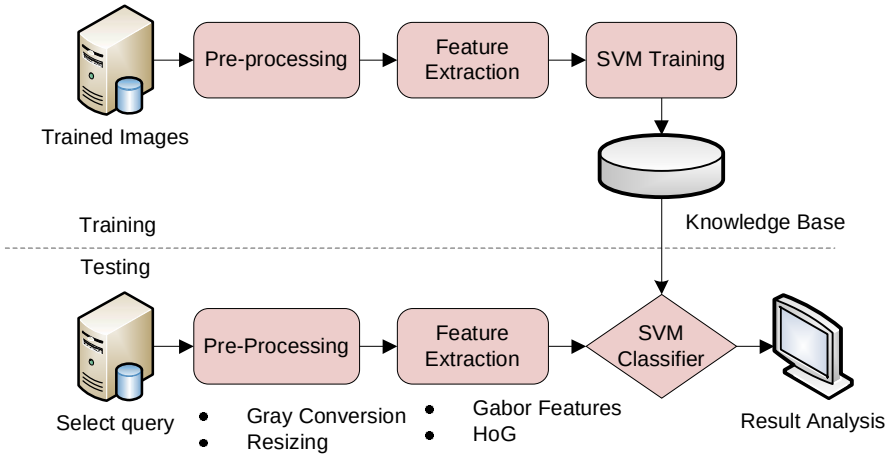

Figure 2: Architecture of Proposed System

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

The Gabor filter eq. (1) is utilized as a part of texture segmentation, image representation. It offers ideal resolution in space and time domain. It gives better visual representation in the involved composition pictures. Be that as it may, the current Gabor parameter requires additional time utilization for feature extraction. The Gabor filter works on the frequency, orientation and Gaussian kernel.

$$\text{Gabor}(x, y, \theta, \varphi) = X.Y \quad (1)$$

$$X = \exp(-(x^2 + y^2) \div \sigma^2) \quad (2)$$

$$Y = \exp(2\pi\theta(x\cos\theta + y\sin\theta)) \quad (3)$$

The terms x and y (eq. 2 and eq. 3) is the position of the filter relative to the input signal [5]. The angular representation of the filter is represented as 'θ'. The angular orientation of the filter is represented as 'Ø'.

C. Histogram of Oriented Gradients:

A standard feature extraction algorithm is i.e. histogram oriented gradient is considered for the extraction of the features form the image. Histogram of the gradient is nothing but it is a feature descriptor used in computer vision and image processing. Patches of multiple scales of images is analyzed by the gradient histogram algorithm at many image locations. The patches will be cropped and further resized. The horizontal and vertical gradients are calculated by the algorithm. The absolute value of x-gradient, absolute value of y-gradient and finally the magnitude of gradient is calculated in order to extract the features.

The image sub-divided into small regions that are interconnected, they are termed as cells. The image will be filtered by sobel kernel to achieve image gradient along both x and y directions.

$$\begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Horizontal mask Vertical mask

$G_x and\ G_y$ referred to the image gradient along both the directions. Magnitude and angle of the image gradient is calculated to each pixel of the image. In these $|G|$ is the magnitude of gradient, θ_g is the angle of gradient, i and j indicated rows and columns of the images respectively. To determine the gradient of histogram, angle is subdivided into n equal distances. Where, n denotes the gradient directions. In order to calculate the histogram, each pixel inside the cell votes to one of the histogram bins based on its gradient angle. These votes are weighted in the pixel based on gradient size based on Eq. (01) and Eq. (02).

$$|G(i,j)| = \sqrt{|G_x(i,j)|^2 + |G_y(i,j)|^2} \quad (01)$$

$$\theta_g(i,j) = \tan^{-1}\frac{G_y(i,j)}{G_x(i,j)} \quad (02)$$

After calculating the gradient histogram in each cell, the histogram is allocated to the cell central pixel. Therefore for each pixel and n dimensional vector, that indicates its surrounding histogram is calculated.

D. Encryption Process:

Encryption is performed by partitioning the images into several blocks and applying encryption process to each blocks using AES encryption algorithm. For each block of image following steps will be achieved.

Step 1: Setup and create the account on the server for sharing of data. This account is generated by data owner.

Step 2: KeyGen algorithm is used for the generation of public key. The data owner generates a public secrete key to encrypt the data over cloud. It also creates an aggregate key to access the block of ciphers of limited size.

Step 3: Encrypts the data provided by the data owner by using the secrete key. This encrypted data is then share among the cloud.

Step 4: The aggregate key is used for extracting the particular block of the ciphers from the cipher file. But other encrypted data remains secure.

Step 5: Decrypt: The encrypted data is then decrypted by using the same secrete key which is use for encryption.

E. SVM Classifier

This paper presents a new approach for scalable and reliable SVM classification. The method, called Clustering-Based SVM, is specifically designed for handling very large data sets. When the size of the data set is large, SVMs tend

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

to perform worse with training from the entire data than training from a fine quality of samples of the data set. Selective sampling (or active learning) techniques with SVMs try to sample the training data intelligently to maximize the performance of SVMs, but they normally require many scans of the entire data set.

Our SVM using the similar idea applies a hierarchical micro-clustering algorithm that scans the entire data set only once to provide an SVM with high quality samples that carry the statistical summaries of the data such that the summaries maximize the benefit of learning the SVM. SVM is scalable in terms of the training efficiency while maximizing the performance of SVMs.

The key idea of SVM is to use a hierarchical micro-clustering technique to get finer description closer to the boundary and coarser description farther from the boundary, which can be efficiently processed as follows: SVM first constructs two micro-cluster trees from positive and negative training data respectively. In each tree, a node in a higher level is a summarized representation of its children nodes.

After constructing the two trees, SVM start training an SVM only from the root nodes. Once it generates the "rough" boundary from the root nodes, it selectively decluster only the data summary near to the boundary into lower (or finer) levels using the tree structure. The hierarchical representation of the data summaries is a perfect base structure for SVM to perform the selective de-clustering effectively. SVM repeats this selective de-clustering to the leaf level.

F. Cloud Storage:

- Cloud storage stores the encrypted data of data owner. We can categorize the storage parts into several groups.
- User module is to retrieve the images from the cloud by giving encrypted query image. The retrieved images will be decrypted successfully only for the valid users. If user wants the decrypted file means user must have the secret key. If user is valid secrete key automatically retrieved from the user and the image files are decrypted
- Admin module is used to add the client to search the file using the file content.

G. User Regestration and User Login Module

When the cloud user wants to get query results from the cloud he must a member in cloud. Therefore, he must be register first. Registration module consists of the fields like username, password, email id and mobile number. User registration process is done by the admin. Without registration can't perform the other operations so initially user register then go for the login. After entering all fields, user details will be stored in the Cloud.

In user module, the fields are username and password. And before login he should be registered as a user then only he can login and use the secured system in cloud computing. Then all the CBIR process will be achieved.

III. RESULTS AND DISCUSSION

Below Figure 2 shows the input image of testing phase. Input image is applied with pre-processing steps like color conversion as show in figure 3.

Figure 3: Input Color Image

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

Image preprocessing is followed by feature extraction, in our proposed system Gabor features of the image is extracted and trained using SVM method.

Figure 4: Color Converted Input Image

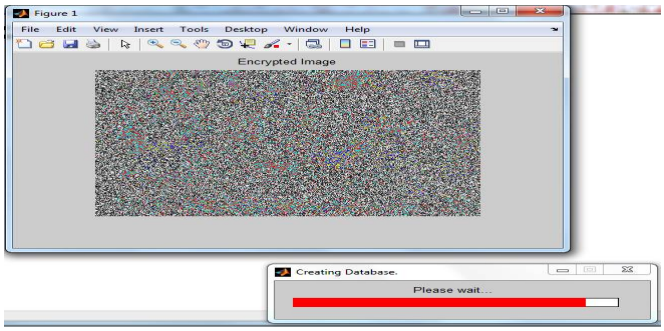

Figure 5: Encrypted Image

Using CB-SVM classifier image is retrieved, which is shown in the Figure 6.

Figure 6: Retrieved Images

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 2, February 2018

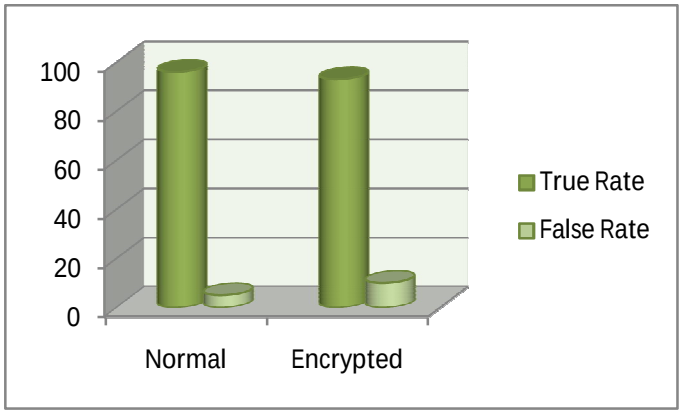

Figure 7: Performance Analysis of CBIR Process

Our proposed system tested more than 100 times. Error rate and true rate is estimated by analyzing accuracy of result in each execution. Comparison is done on both types of inputs such normal images as well as encrypted images. As shown in Figure 7.

IV. CONCLUSION

The proposed system is effectively provides secure information uploading and retrieval by using AES. AES provides more security to the system. The proposed system also increases the efficiency and availability of user data in the cloud. Proposed mechanism provides such a practical mechanism design which fulfills input/output privacy, cheating resilience, and efficiency in the cloud. Also our image encryption method has given proficient result on maintaining security and preserving privacy in cloud environment. The proposed system integrates scalable encryption method with an SVM classifier which can suitable for image database in cloud server. The primitive SVM is not feasible for handling encrypted data set due to complexity in features. SVM applies a hierarchical micro-clustering algorithm that scans the entire data set only once to provide an SVM with high quality micro-clusters.

REFERENCES

[1] Jianqing Fan and Han Liu, "Statistical Analysis of Big Data on Pharmacogenomics", Advanced Drug Delivery Reviews, 2013.
[2] Liran Einav, Jonathan Levin, "The Data Revolution and Economic Analysis", National Bureau of Economic Research, 2014.
[3] Yuri Demchenko, Paola Grosso, Cees de Laat and Peter Membrey, "Addressing Big Data Issues in Scientific Data Infrastructure", Collaboration Technologies and Systems, 2013.
[4] Puneet Singh Duggal and Sanchita Paul, "Big Data Analysis: Challenges and Solutions", International Conference on Cloud, 2013.
[5] Stephanie E Hampton, Carly A Strasser, Joshua J Tewksbury, Wendy K Gram, Amber E Budden, Archer L Batcheller, Clifford S Duke, and John H Porter, "Big data and the future of ecology", The Ecological Society of America, 2013.
[6] Hwanjo yu, jiong yang, jiawei han, xiaolei li. "making svms scalable to large data sets using hierarchical cluster indexing". submission to data mining and knowledge discovery: an international journal.
[7] Jia Deng,Alexander C. Berg Stony Brook University Li Fei-Fei, Hierarchical Semantic Indexing for Large Scale Image Retrieval.2014
[8] Ritendra Datta Jia Li James Z. Wang. Content-Based Image Retrieval Approaches and Trends of the New Age.2013.
[9] "Keepsafe application," https://www.getkeepsafe.com/.
[10] D. M. Reporter, "Facebook to switch off controversial facial recognition feature following data protection concerns," http://www.dailymail.co.uk/news/article-2207098/Facebookswitch- controversial-facial-recognition-feature-following-dataprotection- concerns.html, Sep. 2012.
[11] D. X. Song, D.Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy. IEEE, 2000, pp. 44–55.
[12] P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.
[13] Y. Lu and G. Tsudik, "Privacy-preserving cloud database querying," Journal of Internet Services and Information Security, vol. 1, no. 4, pp. 5–25, 2011.
[14] Ni H, Guo Z. and Huang B, "Binary Patent Image Retrieval Using the Hierarchical Oriented Gradient Histogram", IEEE, pp. 23-27, 2015.