

# Secure Privacy-Preserving Content-Based Retrieval in Cloud Image Repositories

**DR.M.A.THALOR,**

Associate Professor, Computer Engineering, AISSM'S IOIT, Pune.

**AISHWARYA BIRANGAL, PRIYANKA JADHAV, SUJAY METHA, KIRTI NIMBALKAR.**

Student, Computer Engineering, AISSM'S IOIT, Pune

**Abstract**— Recently, digital content has become a significant and inevitable asset for any enterprise and the need for visual content management is on the rise as well. There has been an increase in attention towards the automated management and retrieval of digital images owing to the drastic development in the number and size of image databases. This was a crucial feature for the acquisition of data outsourcing solutions based on cloud. Although the outsourcing of data storage in the cloud causes new security issues that must be correctly resolved. We present a secure scheme for the storage and recovery of the privacy protection in huge collections of shared images. Our approach is built on IES-CBIR, a novel Encryption technique of the image that presents image restoration features based on its content. The scheme permits encrypted storage as well as image search using content-based image retrieval queries while maintaining privacy against honest but bizarre cloud administrators. We have developed a model of the proposed scheme, formally examined and tested its safety properties, and experimentally assessed its performance and accuracy of recovery. Our proposal shows that IES-CBIR adds more security and it enables more impressive features than the existing schemes, in terms of complexity of space and time, and introduces a new technique of practical solution.

**Keywords** – Data and Storage Outsourcing; Encrypted Data Processing; Searchable Encryption; Content-Based Image Retrieval

## I. INTRODUCTION

Moving towards the goal of supporting multimedia data, a new solution is proposed for multiple users sharing storing and searching images in the cloud with security guarantees. This is basically based on a new encryption concept designed for images called IES-CBIR. It is image encryption scheme with content based image retrieval, one key driving factor for cloud services has been growth in production and sharing of

multimedia data, both in corporate and personal use scenario. Image content retrieval and Content-based visual information retrieval is also known as Content-based image retrieval (CBIR) and CBIR is a query by image content and content based visual information retrieval is the use of computer perception to the image retrieval problem of searching for digital images in large dataset. 'content-based' means that the searching will analyse the actual or similar contents of the image. the term 'content' in this circumstances might refer colours, shapes, texture, pixel information can be derived from the image itself. Without the capacity to inspect the image content, searches should be based on metadata, to give an example such as titles or keywords. These metadata must be generated by a human itself and stored every image in the databases. The image retrieval system provides an effective way to introduced, explore and recover a series of related images in real-time applications. it is now probable to create huge and extensive digital image databases. Basically these collections can contain millions of images and terabytes of data. it is essential to design good and effective research methods. Previously the automatic indexing methods, the image databases were indexed depend on the keywords that a human classifier had decided and inserted. Unluckily, this usage has two serious short comings. First of all, because a database becomes bigger and bigger day by day, the work requisite to index each image becomes less practical. Secondly, two different people, or even the same person on two different days, can index similar images in an incompatible way. The result of these inefficiencies is a search result that is not optimal for the end user of the system. The fact that a computer performs indexing based on a CBIR scheme tries to solve the human-based indexing deficiencies. Because a computer is able to process images at a much higher rate, without ever getting tired. For example, each CBIR system must be adjusted for its particular use in order to achieve optimal results.

Various way has been developed to catching the image content information by directly analysing the characteristics of an image. The characteristics of the image are constructed directly from the typical concentrate data sequence depends on

block or semitone truncation encoding without executing the decoding procedure. These image retrieval schemes include two phases, indexing building and searching, to come up with a set of similar images from the databases. The index building phase extracts the characteristics of all images in the cloud image repositories, which is then stored in the cloud database as a feature vector. In the search phase, user encrypt the query image with the mentioned encryption method as a trapdoor, retrieval system derives the characteristics of the image of an image sent by a user (as a query image).

#### A. Motivation

1. Image search on encrypted data for security.
2. Improve accuracy in image search.

#### B. Problem Statement

Image search is an important method to find images contributed by social users in such websites. However, how to make the result relevant and with diversity is challenging.

## II. RELATED WORK

The important step of research is to Literature survey first. Before where we actually start developing our project we need to study the previous papers which should be of our domain which we are working on. The basis of study we can generate the drawback or predict. Then start working with the reference of previous papers.

Here, we review the related work briefly on Image search and their different techniques.

**Reference paper [1]** proposed show importance of privacy-preserving SIFT (PPSIFT) and this paper also shows scale-invariant feature transform (SIFT) which has been widely used in different-different fields, and it shows the problems of secure SIFT feature extraction and representation of the image retrieval and also encrypted domain. Limitation of this paper is that it just checks similar search and that's why there is low accuracy and more time complexity and space complexity taken in this method.

**From Reference paper [2]** shows that IES-CBIR is more secure, operations are done more efficiently than existing, in terms of space complexity as well as time complexity. It also enables more reliable applications in practical. This was just proposed system rather than actual implementation only.

**Reference paper [3]** Results out offers protection of two layers. First layer which contains of robust hash values which are used to prevent revealing original features or content which are as queries. Second layer has client omit certain bits in a hash value to further increase the ambiguity for the server and client can choose it. It does not explore more efficient

optimization algorithms to improve the security and image retrieval.

**From Reference paper [4]** Another paper stating Aggregating local image descriptors into compact codes, addresses the problem of large-scale image search. Implementation bag-of-visual words is done over here. It also Required Expensive hardware and this gives problem for large scale image for searching of it. There are Three constraints are considering efficiency, memory usage, accuracy. Firstly we present it and then evaluate image descriptors in different ways of aggregating local into a vector which shows that the Fisher kernel achieves better performance than the reference bag of visual words approach for any given vector dimension.

**From Reference paper [5]** Is a paper in which texts of surroundings are too noisy often or they can be few to accurately describe the content image. Annotations of human are expensive resourcefully therefore scaled up is not possible. This lately aims a latent subspace to create with the ability in the original incomparable for comparing information. This paper only consider color feature.

**Reference paper [6]** Execution of this paper multimedia hashing for global retrieval from heterogeneous data sources can be classified into text and content based retrieval. using a query image, we may obtain textual results. In this paper, we present a new inter media hashing from heterogeneous data sources. Cross-media retrieval methods are not able to deal with global inter media databases. For that purpose many more hashing standard procedure have been proposed. they are divide into two important categories random projection and machine learning based hashing methods

**From Reference paper [7]** In paper of CMFH plays fundamental role in many more important applications like data mining, information retrieval, and computer vision. the main important features are to be hashing based nearest neighbor search. CMFH integrate information from different databases sources into final integrated hash code. MVSH integrate information from multiple view source into binary code and it referred to CMFH. It also increase search accuracy by merging MVSH.

**Reference paper [8]** paper can be used for both unsupervised and supervised data embedding, unsupervised embedding for PCA and supervised embedding for canonical correlation analysis (CCA). in that paper binary coding method is used to perform PCA to ensure good generalization ability for binary code to reduce search complexity and data structure have been proposed. the first step is computing binary code after that next step is binarize the data for unsupervised embedding In that paper we used CCA to prove the effective tools for extracting a common latent space from robust to noise .CCA gives both clean and noisy information ,for clean scenario setup analogy used and for noisy we learn CCA from tiny image

**From Reference paper [9]. Propose** based on short-term and long-term cross-media retrieval. This method basically focuses on two typical types of media data, which is image data and audio data. firstly, set mapping function from high dimensional to low dimensional latent space. Secondly the multidimensional index like efficient retrieval, classification, event detection space is built, regarding many issues. We proposed cross-media retrieval and optimization methods for image data and audia data.

**Reference paper [10]** in paper of deep visual-semantic alignment for generating image descriptor from deep neural network model developed between segment of sentences and the region of image .and multi model recurrent neural network architecture introduced. Important feature is to be generating descriptor.

### III. EXISTING SYSTEM

Together with the increasing requirements, in recent times research activities through the media have received significant attention. Since each technique has various ways of representation and correlational structures, we have studied a variety of methods of the problem from the learning aspect correlations between different techniques. The current methods proposed for the use of canonical correlation analysis (CCA), double-wing harmonics, and deep automatic encoder, multiple learning and Boltzmann deep machine to solve the problem. The Efficiency of this hash-based techniques, there is also an extensive line of work done to address the issues of mapping high-quality multimode data to low-level hash codes, such as latent hashing (LSSH), discriminant coupling dictionary of hashing (DCDH), hashing of cross-sight (CVH), etc.

As we studied the various papers on image search techniques, the issues are related to searchable image search and the challenge is to address image search problem using content based features such as using global color algorithm.

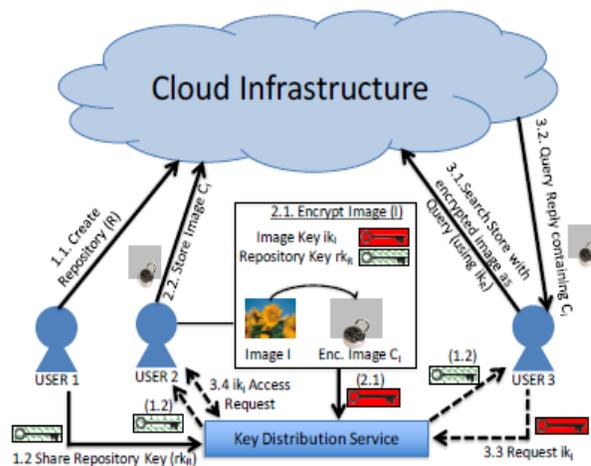
### IV. PROPOSED APPROACH:-

We propose a secure scheme for the storage and recovery of the subcontracted privacy protection in large archives of shared images. Our proposal is based on CBIR, a novel Encryption scheme of the image that presents image recovery properties based on content. The scheme which permits encrypted storage and search by using content-based features extraction queries image retrieval and maintain privacy against bizarre cloud administrators but honest. Here we built the proposed scheme model , examined formally and experimentally assessed its performance ,tested its security parameters and accuracy of retrieval. This results shows that CBIR is safe, permitting more efficient operations than the

existing methodologies, both in terms of complexity of time and space, and leads the way to new perspective of practical application.

### V. CONCLUSION

This Paper which we proposed is a new secure framework for the external storage of privacy protection, research and recovery of image archives on large scale, where the general expenses reduction of the customer. Our framework we has proposed new cryptography scheme, specifically designed for images, called CBIR. The key to its design is the observation that in the images, color information can be separated from the plot information, allowing the use of different cryptographic techniques with different properties for each and allowing to preserve privacy Image recovery based on the content that will be created from unreliable third-party cloud servers.



**Fig. System Architecture**

We formally analyze the safety of our proposals and further experiments the evaluation of the implemented prototypes revealed that our approach reaches an interesting exchange between precision and I remember in the CBIR, while exhibiting high performances and scalability compared to alternative solutions.

### REFERENCES

- [1] C.-Y. Hsu, C.-S. Lu, and S.-c. Pei, Image Feature Extraction in Encrypted Domain With Privacy-Preserving SIFT," IEEE Trans. Image Process., vol. 21,no. 11, pp. 4593-4607, 2012.
- [2] B. Ferreira, J. Rodrigues, J. Leit-ao, and H. Domingos, Privacy-Preserving Content-Based Image Retrieval in the Cloud," in SRDS'15. IEEE, 2015.

- [3] L. Weng, L. Amsaleg, A. Morton, and S. Marchand-maillet, "A Privacy-Preserving Framework for Large-Scale Content-Based Information Retrieval," *TIFS*, vol. 10, no. 1, pp. 152-167, 2015.
- [4] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704-1716, Sep. 2011.
- [5] Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui, "Clickthrough-based cross-view learning for image search," in *Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval*, pp. 717-726, 2014.
- [6] J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen, "Inter-media hashing for large-scale retrieval from heterogeneous data sources," in *Proc. Int. Conf. Manage. Data*, pp. 785-796, 2013.
- [7] G. Ding, Y. Guo, and J. Zhou, "Collective matrix factorization hashing for multimodal data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2083-2090, 2014.
- [8] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916-2929, Dec. 2013.
- [9] H. Zhang, J. Yuan, X. Gao, and Z. Chen, "Boosting cross-media retrieval via visual-auditory feature analysis and relevance feedback," in *Proc. ACM Int. Conf. Multimedia*, pp. 953-956, 2014.
- [10] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in *Proc. IEEE Conf. USA*, pp. 3128-3137, Jun. 2015.