International Journal of Current Engineering and Technology
©2021 INPRESSCO®, All Rights Reserved

E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

Proposed approach for content-based image retrieval in cloud repositories

Miss.Sweta Kadam and Prof. Dr.Archana C Lomte

Department of Computer Engineering Bhivarabai Sawant Institute of Technology and Research Savitribai Phule Pune University Pune, India

Received 10 Nov 2020, Accepted 10 Dec 2020, Available online 01 Feb 2021, **Special Issue-8 (Feb 2021)**

Abstract

Now days, Cloud computing it is been playing a crucial role in terms of data storing and reducing the overall cost to entrepreneurs. Storage requirements for visual data have increased in recent years, following the appearance of many interactive multimedia services and applications for mobile devices in personal and business scenarios. This was a key a determining factor for the adoption of cloud-based data outsourcing solutions. However, even the outsourcing of data storage in the cloud leads to new security challenges that must be carefully addressed. We propose a secure framework for the storage and recovery of the subcontracted privacy protection in large archives of shared images. Our proposal is based on IESCBIR, a novel Encryption scheme of the image that presents image recovery properties based on content. The framework allows both encrypted storage and search using content-based image retrieval queries while preserving privacy against honest but curious cloud administrators. We have built a prototype of the proposed framework, formally analyzed and tested its safety properties, and experimentally assessed its performance and accuracy of recovery. Our results show that IES-CBIR is probably safe, allowing more efficient operations that the existing proposals, both in terms of complexity of time and space, and opens the way to new scenarios of practical application.

Keywords: *Encrypted Data Processing; Searchable Encryption; Content-Based Image Retrieval, Storage*

Introduction

Content-based image retrieval is also known as image content retrieval and content-based visual information retrieval is the use of artificial vision for the problem of image retrieval of large digital image search database size. ”Content-based” means that the research will analyze the actual content of the image. The term ”content” in this context may refer to colors, shapes, textures or any other information that may be derived from the image itself. Without the ability to examine the image content, searches should be based on metadata such as titles or keywords. These metadata must be generated by a human being and stored exactly every image in the database. An image retrieval system returns a series of images from a collection of images in the database to meet the demand of users with a similarity rating like the similarity of the image content, the similarity of the border motif, the similarity of the color, etc. The image recovery system provides an effective way to access, explore and recover a series of similar images in real-time applications. As a result of recent advances in digital storage technology, it is now possible to create large and extensive digital image databases. These collections can contain millions of images and terabytes of data. In order for users to take full advantage of these databases, it is necessary to design effective research methods. Before the automatic indexing methods, the image databases were indexed based on the keywords that a human classifier had decided and inserted. Unfortunately, this practice has two serious shortcomings. First of all, because a database becomes bigger and bigger, the work required 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 inconsistent 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. A recovery system designed to consult medical X-ray images will probably be a poor system for recovering satellite images of tropical forests in South America. Furthermore, the algorithms currently used cannot consistently extract the abstract features from the images, such as the emotional response, which would be relatively easy to observe for a human being.

Various approaches have been developed to capture the image content information by directly calculating the image characteristics of an image. The characteristics of the image are constructed directly from the typical compressed data sequence based on block or semitone truncation encoding without executing the decoding procedure. These image recovery schemes include two phases, indexing and searching, to retrieve a set of similar images from the database. The indexing phase extracts the image characteristics of all images in the database, which is then stored in the database as a feature vector. In the search phase, the recovery system derives the characteristics of the image of an image sent by a
user (as a query image).

A. Motivation

- Content-based image retrieval (CBIR) applications have been rapidly developed along with the increase in the quantity, availability and importance of images in our daily life.
- CBIR scheme has been limited by it's the severe computation and storage requirement for visual data have been increasing in recent years, following the emergence of many highly interactive multimedia services and applications for mobile devices in both personal and corporate scenarios.
- This has been a key driving factor for the adoption of cloud-based data outsourcing solutions. However, outsourcing data storage to the Cloud also leads to new security challenges that must be carefully addressed, especially regarding privacy

Review of Literature

This System provides privacy protection for photo sharing and searching without leakage of query contents and result. Personalized private content can be defined using checkbox configuration. This system either automatically or manually determines rectangular ROP (Region of Privacy), and then ROP is separated into public and secrete part. To prevent sensitive information, secrete part is encrypted. Only legitimate users can access secrete part and retrieve ROP with key. For ROP separation the technique reviewed is Mask, P3, and Blur [1].

This system aims to minimize redundant data for enhancing query proficiency and minimizing operation cost. The main function is to fast identify similar images from massive image dataset in cloud [2].

In this system, without exposing owner's data privacy the feature descriptors which are based on secrete data are acquired. First each image set is encrypted and then cipher text is distributed to two independent servers. Then server returns encrypted feature descriptors to owner of data who is able to retrieve actual feature descriptors [3].

The proposed system supports CBIR over encrypted images without leaking the sensitive information to the cloud server. Firstly, for representing images their feature vectors are extracted. Then, by using locality- sensitive hashing the pre-filter tables are constructed to enhance search proficiency. The water marking technology used to prevent illegal distribution of images. Watermark is directly implanted into images that are encrypted. It also allows searching over encrypted images [4].

In this work problem multi-keyword ranked search over encrypted data with privacy-protection in cloud computing (MRSE) is defined and solved, performed multi keyword ranked search over encrypted data. Encrypted index that is searchable from data documents is used. For measuring similarity coordinate matching and inner product similarity is used [5].

The proposed framework provides storage and retrieval of images in large image repositories with privacy protection. It is based on Image Encryption Scheme called IES CBIR that displays properties dependent on Content-Based Image Retrieval. All data sent to cloud is encrypted for ensuring users privacy. Image texture is encrypted using probabilistic encryption for protection purpose. Color information is encrypted using deterministic encryption. Color information is used for image retrieval and content based image indexing. The solution enables encrypted storage as well as searching using CBIR queries with privacy protection [6].

This proposed work, allow to deploy the CBIR service and image database to the cloud with privacy protection without displaying the real content of the database to the cloud server. Uses the local features for retrieving image based on its content. Uses EMD –Earth Movers Distance for calculating similarity of images. For improving search efficiency similar images are grouped together. The owner is responsible for producing searchable index before forwarding data to cloud. This scheme allows searching and CBIR on encrypted data. Authorized user uses encrypted query for searching image on cloud [7].

This framework is designed to store search and retrieve images that are dynamically updated with privacy protection on cloud. The main aspect is to reduce overhead of client. Image color information and texture information is separated that allows to use different encryption techniques. Global Color features are encrypted using deterministic encryption and used for indexing and searching of images based on similarity. Encrypted images stored on cloud and Search query is encrypted [8].

In this paper SIFT (Scale Invariant Feature Transform) and homomorphic encryption is used for preserving privacy of images Difference of Gaussian transform is executed for extracting the feature points. The images are twisted together with Gaussian Filters. Dissimilarity is calculated between two adjoining Gaussian blurred images. Using homomorphic encryption image is encrypted to maintain user's privacy. Focus on homomorphic comparison of encrypted data. Two encrypted data are compared based on their locations. Pixels, locations are not encrypted and thus SIFT feature location is public and will not break privacy as feature vectors related to them are in encrypted form. Demanding issue of homomorphic comparison is resolved in this paper [9]. This system is designed to perform social discovery based on images to increase the friends list of user depending on their common interest securely and efficiently using encryption. The social interest of user determined based on BOW (Bag Of Word) representation. Then compact and secure similarity index is designed which enables fast and scalable similarity based search on millions of encrypted images of user's profile vectors and is done by using BOW model by extracting visual content with similarity [10].

In this work, the problem of verifiable privacy preserving multiparty computation is focused. They presented ranging protocol based on two party thresholds which is justifiable for both input and output and provides privacy protection. They also proposed testable ranking protocol for participant and aggregator model [11].

In this work, without using secure communication channel or trusted key issuers the privacy protected sum and product calculation protocols are accomplished. They proposed some protocols that give assurance of data privacy under semi honest cloud model. Then also proposed some advanced protocols which sustain up to k passive adversaries who do not interfere with computation [12].

Proposed Methodology

We propose a secure framework 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 framework allows both encrypted storage and search using content-based image retrieval queries while preserving privacy against honest but curious cloud administrators. We have built a prototype of the proposed framework, formally analysed and tested its safety properties, and experimentally assessed its performance and accuracy of recovery. Our results show that CBIR is probably safe, allowing more efficient operations that the existing proposals, both in terms of complexity of time and space, and opens the way to new scenarios of practical application.

A. Advantages of proposed system

- This proposed approach support image data privacy protection
- Using Shape-Based Invariant Texture Index (SITI) descriptor, proposed work will get more accurate result.
- In proposed approach, data stored in encrypted format so more security will be provided.
- Security will be maintained using access control.
- Also security will be maintained using users image secret key.

B. Architecture

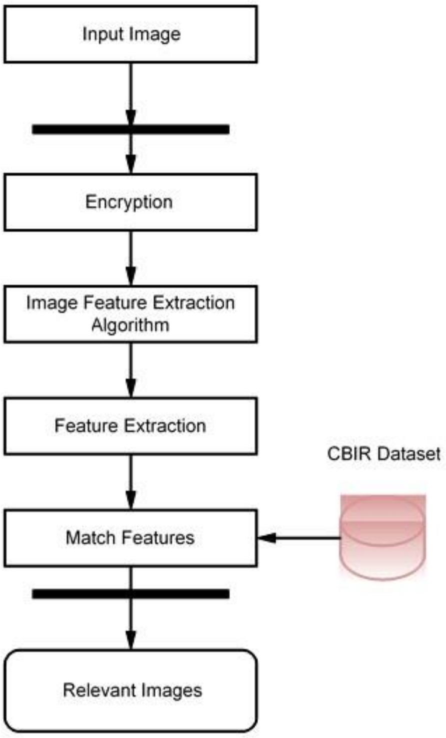

Fig. 1. Proposed System Architecture

C. Algorithm

1. Blowfish Algorithm

It is symmetric algorithm. It used to convert plain text into cipher text .The need for coming with this algo is weakness in DES/RSA. The 56 bit key of des is no longer safe against attacks based on exhaustive key searches and 64-bit block also consider asweak.Blowfish was to be used 128-bit block with 128-bit keys.

Input:

128 bit /192 bit/256 bit input (0, 1) Secret key (128 bit) +plain text (128 bit).

Process:

10/12/14-rounds for-128 bit /192 bit/256 bit input

Xor state block (i/p)

Final round:10,12,14

Each round consists: sub byte, shift byte, mix columns, add round key. Output: cipher text(128 bit)

2. Shape-Based Invariant Texture Index (SITI) Feature Extraction image content identification.

Steps:

1. Color feature is one of the most widely used visual featuresin image retrieval, for its invariance with respect to image scaling, rotation, translation. In this work, an image is divided into four equal sized blocks and a centralized image with equal-size. For each

block, a 9-D color moment is computed, thus the dimension of color comment for each image is 45. The 9-D color moment of an image segment is utilized, which contains values of mean, standard deviation and skewness of each channel in HSV color space.

2. Edge Detection: Most of the shape information of animage is enclosed in edges. So first we detect these edges in an image and by using these filters and then by enhancing those areas of image which contains edges, sharpness of the image will increase and image will become clearer.

Canny Edge Detection:

Canny edge detection is a technique to extract useful structural information from different vision objects and dramatically reduce the amount of data to be processed. It has been widely applied in various computer vision systems. Canny has found that the requirements for the application of edge detection on diverse vision systems are relatively similar. Thus, an edge detection solution to address these requirements can be implemented in a wide range of situations. The general criteria for edge detection include:

1. Detection of edge with low error rate, which means thatthe detection should accurately catch as many edges shown in the image as possible
2. The edge point detected from the operator should accuratelylocalize on the center of the edge.
3. A given edge in the image should only be marked once,and where possible, image noise should not create false edges.

Results and Discussion

The section shows overall accuracy of Existing Algorithm(Global Color) and Proposed System algorithm(SITI Algorithm) . So this works gives better content based image retrieval results compare to existing method.

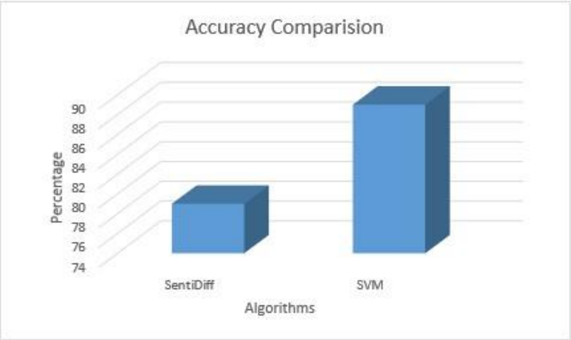

Fig. 2. Accuracy Graph

	Existing System	Proposed System
Accuracy	79.32	89.77

In these experiments we used the one image dataset to analyze the performance of our system, with and without image compression. Following graph shows the experimental results for the SEARCH OPERATION, comparing existing approaches. The results showed here represent the performance for searching in the image dataset with a random image chosen from the collection as query (the results represent the average of 100 random runs each). The Encrypted image and Encrypted image feature represent local processing done by the querying user, while the Cloud column represents not only the network time for transmitting the query and receiving its results, but also the time elapsed by the server in processing and calculating those results.

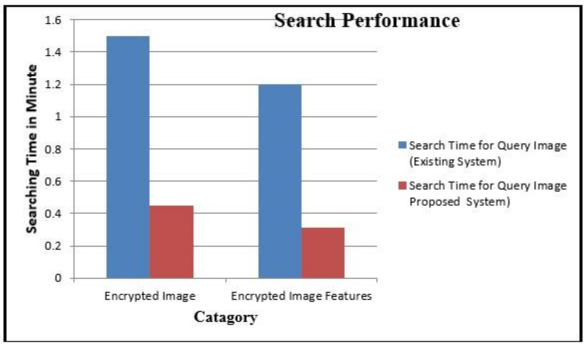

Fig. 3. Search Performance Graph

	Existing System	Proposed System
Encrypted Image	1.5min	0.45min
Encrypted Image Features	1.2min	0.43min

Conclusion

In this Paper, we have proposed a new framework for the external storage of privacy protection, research and recovery of large-scale dynamic image archives, where the reduction of the general expenses of the customer is central appearance. At the base of our framework there is a new cryptography scheme, specifically designed for images, called content based image retrieval. 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. 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]. L. Zhang, T. Jung, C. Liu, X. Ding, X.-Y. Li, and Y. Liu, “Pop: Privacypreserving outsourced photo sharing and searching for mobile devices,” in ICDCS. IEEE, 2015.

[2]. Y. Hua, H. Jiang, and D. Feng, "Real-time semantic search using approximate methodology for large-scale storage systems," IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 4, pp. 1212–1225, 2016.

[3]. [S. Hu, Q. Wang, J. Wang, Z. Qin, and K. Ren, "Securing sift: Privacypreserving outsourcing computation of feature extractions over encrypted image data," IEEE Transactions on Image Processing, vol. 25, no. 7, pp. 3411–3425, 2016.

[4]. Z. Xia, X. Wang, L. Zhang, Z. Qin, X. Sun, and K. Ren, "A privacypreserving and copy-deterrence content-based image retrieval scheme in cloud computing," IEEE Transactions on Information Forensics and Security, vol. 11, no. 11, pp. 2594– 2608, Nov 2016.

[5]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacypreserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, Jan 2014.

[6]. Bernardo Ferreira, Jo˜ao Rodrigues, Jo˜ao Leit˜ao, Henrique Domingos, "Privacy Preserving Content- Based Image Retrieval in the Cloud". 2015 IEEE 34th Symposium on Reliable Distributed Systems

[7]. Zhihua Xia, Yi Zhu, Xingming Sun, Zhan Qin, Kui Ren, "Towards Privacy preserving Content-based Image Retrieval in Cloud Computing". IEEE transactions on computer computing, vol. , no. , September 2015

[8]. Bernardo Ferreira, Joao Rodrigues, Joao Leitao, Henrique Domingos, "Practical Privacy-Preserving Content-Based Retrieval in Cloud ImageRepositories". IEEE Transactions on Cloud Computing, Year: 2017, Volume: PP, Issue: 99

[9]. 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.

[10]. X. Yuan, X.Wang, C.Wang, A. Squicciarini, and K. Ren, "Enabling Privacy-preserving Image-centric Social Discovery," in ICDCS'14. IEEE, 2014.

[11]. L. Zhang, X.-Y. Li, Y. Liu, and T. Jung, "Verifiable private multiparty computation: Ranging and ranking," in IEEE INFOCOM, 2013.

[12]. T. Jung, X.-Y. Li, and M. Wan, "Collusion-tolerable privacy preserving sum and product calculation without secure channel". In IEEE TDSC, 2014.