SIMILARITY SEARCH FOR ENCRYPTED IMAGES IN SECURE CLOUD COMPUTING

P.Vijay Bhaskar Reddy, Associate Professor, Department of Master of Computer Applications, Narayana Engineering College(Autonomous), Gudur.SPSR Nellore, AP
K.Nandini, PG Scholar, Department of Master of Computer Applications, Narayana Engineering College(Autonomous), Gudur.SPSR Nellore, AP

Abstract

With the growing popularity of cloud computing, more and more data owners are willing to outsource their data to the cloud. However, private data should be encrypted before outsourcing for security requirements, which obsoletes data utilization like content-based image retrieval. In this paper, we propose a secure similarity image search scheme, which allows data owners to outsource their encrypted image database to the cloud server without revealing the real content of images. The proposed scheme supports both global and local feature based image retrieval under various distance metrics, such as earth mover's distance. Firstly, the data owner extracts either global features or local features from images to represent the images. Then, these features are used to generate a searchable index. Finally, both image database and searchable index are encrypted before outsourcing to the cloud server. When a query image coming, the data user extracts feature from the query image and generates the search trapdoor. The trapdoor is then sent to the cloud server and used to compare the similarity with the searchable index. Extensive experiments are conducted to show the efficiency and applicability of our proposed similarity image search system.

Keywords: similarity image retrieval, linear programming, earth mover's distance, cloud computing.

Introduction

Due to the low-cost storage, the number of images is growing rapidly and images are starting to play an important role in our daily life. The storage and retrieval of large-scale image databases has become a big problem. Fortunately, content-based image retrieval is helpful to real-world image retrieval applications. For example, doctors may retrieve the similar cases of their patients in the medical image database to help them make the right decision. However, millions of images are usually included in a large-scale image database and every image is high-dimensional. So, this kind of image retrieval service usually has high computational complexity and intensive storage requirement. Thanks to the strong data storage and management ability of cloud computing, data owners are willing to store their data on the cloud server without maintaining the image database locally. Authorized data users can retrieve similar images from the cloud server without interacting with the data owner. Despite of various advantages of cloud services, the privacy of image database becomes the main concern of image database outsourcing. Patients do not hope to reveal any information of their medical images to unauthorized user. We study the whole problem thoroughly and propose a practical similar search solution over encrypted images which protect the privacy of image database. In this paper, we study the secure similarity image search problem and propose a practical solution. We exploit techniques from security, image processing and information retrieval domains to achieve secure and efficient searching over encrypted images. The solution supports both global and local features under different distance metrics, such as Euclidean distance and earth mover's distance based methods. In particular, a secure linear programming (LP) transformation is designed such that the cloud server is able to determine the earth mover's distance with input and output privacy. By leveraging the computation power of the cloud, the proposed scheme costs low local computation while achieving high retrieval accuracy. The extensive experiments are conducted to validate the efficiency and applicability of the proposed solution. The reminder of this paper is organized as follows. Section introduces the system architecture and preliminaries. In Section, we give the design of similarity image search, which supports both global and local feature based solution. In Section, we formally analyze the security of the proposed schemes. In Section, we implement our proposed scheme and study its efficiency, applicability and local computational savings. Finally, Section summarizes related work, and a conclusion is given in Section.

Statement of the Problem

A similarity search problem involves a collection of objects (documents, images, etc.) that are characterized by a collection of relevant features and represented as points in a high-dimensional attribute space. Given queries in the form of points in this space, we are required to find the nearest (most similar) object to the query. Our scheme is designed to not only support similarity search, but also prevent the leaking of information about the image database. In this study, we consider a cloud data system involving three different entities as Image owner, image user and cloud server.

With the emergence of intelligent terminals, the Content-Based Image Retrieval (CBIR) technique has attracted much attention from many areas (i.e., cloud computing, social networking services, etc.).

Although existing privacy-preserving CBIR schemes can guarantee image privacy while supporting image retrieval, these schemes still have inherent defects (i.e., low search accuracy, low search efficiency, key leakage, etc.).

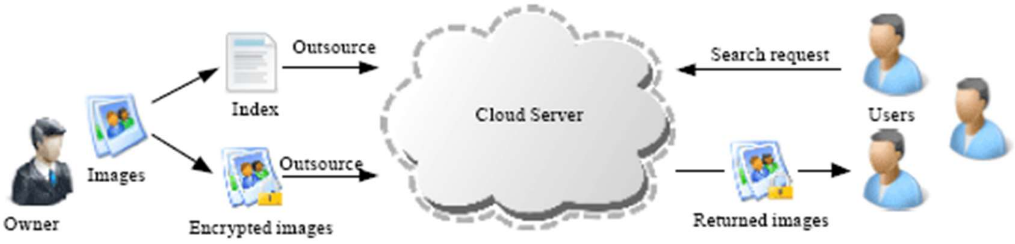
Objectives of the study

To enable secure and efficient similarity search over encrypted images in cloud under the aforementioned model, our scheme should achieve the following design goals:

- Accuracy: The proposed scheme in this study should have high retrieval accuracy, achieving the same results as standard CBIR system.
- Efficiency: Our scheme should reduce the computational complexity and communication overhead as far as possible.
- Security: In the procedure of search, we should ensure the security of sensitive data without leaking information such as image databases and search index

Review of Literature

The proposed similarity image search system involves three types of entities: data owner, data user and cloud server. The data owner has a large-scale of image database to be outsourced, where n is the number of images in the database. The data owner extracts features from images and generates a searchable index. To protect the sensitive information of images, the image database and searchable index are encrypted before outsourcing to the cloud server. To support the similarity image search over encrypted images, the data owner has to construct a searchable encryption scheme. The authorized data user extracts feature vector from the query image and generates an encrypted query. Then, the encrypted query is submitted to the cloud server. The cloud server receives the encrypted query and compares the similarities with the searchable index. Then, the top-ranked similar encrypted images are returned to the data user. Finally, the data user decrypts the received images. The data owner and data user are always trusted, but the cloud server is considered to be “honest but curious”. The cloud server tries to derive sensitive information by analyzing the communication history. Our proposed similarity image search solution is designed to prevent the cloud server from knowing either the image database or users' queries. we propose a practical content-based image retrieval solution over encrypted images in cloud computing. Besides, we exploit p-stable LSH in content-based image retrieval. Our scheme is based on the global feature of images. By leveraging the computation power of the cloud, the proposed scheme costs low local computation while achieving high retrieval accuracy. In this work, we give rigorous security analysis and conclude that under certain security model our similarity image search scheme is secure. At last, extensive experiments are conducted and acquire satisfying experiment results and search efficiency.



Research Methodology

In this section, we describe our encrypted image search scheme in two phase. To search similar images outsourced to the cloud, data owner should construct a secure index and outsource it along with the encrypted images to the cloud server. Then, server performs search on the index according to the requests submitted by the data users. The cloud server learns nothing about the query request or the image databases itself.

Setup phase: In the setup phase, image owner needs to build a secure index and encrypt the images. Then, the index and the encrypted images are uploaded to the cloud.

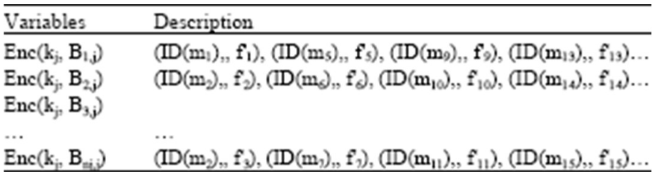
Key generation: Firstly, the image owner generates private keys kj(j = 0, 1, 2,…, L) and km. In particular, kj is the secret key for the encryption of keywords Bi,j and km is the secret key for the encryption of image database. Let kj, km∈{0,1}ψ be the secret keys of size ψ. Besides, a (l+1)x(l+1) invertible matrix R is generated to encrypt the feature vectors.

Feature extraction: For each image fi in the image database, we extract a global feature mi from it. So, every image mi in the image database has one corresponding global feature fi.

Secure index construction: Features of the images are high-dimensional vectors and p-stable LSH scheme can work directly on these vectors in the Euclidean space without embedding them. Learning from the p-stable LSH algorithm, we have a locality sensitive function:

$$h_{a,b}(v) = \left\lfloor \frac{a \cdot v + b}{r} \right\rfloor$$

The data owner constructs the hash function g(fi) = (h1(fi), h2(fi),…, hλ(fi)) where hi is randomly chosen from hash family H. Then, the data owner applies g(fi) = (h1(fi), h2(fi),…, hλ(fi)) to all global features Image for - Enabling Similarity Search over Encrypted Images in Cloud so as to build the search index. Let Image for - Enabling Similarity Search over Encrypted Images in Cloud denote the derived set of LSH hash values, where, N refers to the total number of cluster. In fact, the g(fi) function maps a d-dimensional feature vector to a λ-dimensional vector. So, each Bi,j is a λ-dimensional vector. However, in general, such LSH function is not necessary to have one-way property.



Variables	Description
$Enc(k_j, B_{1,j})$	$(ID(m_1),, f_1), (ID(m_5),, f_5), (ID(m_9),, f_9), (ID(m_{13}),, f_{13})$…
$Enc(k_j, B_{2,j})$	$(ID(m_2),, f_2), (ID(m_6),, f_6), (ID(m_{10}),, f_{10}), (ID(m_{14}),, f_{14})$…
$Enc(k_j, B_{3,j})$	
…	…
$Enc(k_j, B_{ni,j})$	$(ID(m_2),, f_3), (ID(m_7),, f_7), (ID(m_{11}),, f_{11}), (ID(m_{15}),, f_{15})$…

Therefore, it is necessary to encrypt the keywords Bi,j in the hash table for the security of them before outsourcing to the cloud. Here, we utilize kj as the secret key, let Image for - Enabling Similarity Search over Encrypted Images in Cloud be a pseudorandom permutation.

Global features are used to compare the distance between images in the search phase. So, we must encrypt global features in such way that the cloud server is able to compute the distance of them.

Then, the data owner randomly picks a(R+R)x(R+R) invertible matrix R to encrypt the modified feature vector as Image for - Enabling Similarity Search over Encrypted Images in Cloud.

We assume every image mi has an unique identifier ID(mi), (ID(mi), f'i), (ID(mj), f'j)…, (ID(mk), f'k) be a list of image identifiers and their corresponding modified feature vector. Finally, one search index I is as shown in Table 1. Furthermore, to increase the clustering accuracy, we repeat the index construction process L times by generating L hash tables as Table 1.

Upload: After constructing the index, image owner encrypts the image database with the key km and sends all encrypted images, encrypted features along with the search index (Table 1) to the cloud server for search purpose. Though the encrypted images outsourced, authenticated users should be able to retrieval images from the cloud server. To reach this goal, image owner should share some information with authenticated users:

Upload: After constructing the index, image owner encrypts the image database with the key km and sends all encrypted images, encrypted features along with the search index (Table 1) to the cloud server for search purpose. Though the encrypted images outsourced, authenticated users should be able to retrieval images from the cloud server. To reach this goal, image owner should share some information with authenticated users:

kj(j = 0, 1, 2,…, L): Secret key for keyword Bi,j encryption
km: Secret key for image database encryption
R: Secret matrix
g: p-stable LSH function used in the index construction

Search phase: In search phase, image user wants to retrieve images that are similar to one query image from the cloud server. In order to avoid the information leaking, image user generates a secure trapdoor with the query image. Then, the trapdoor is submitted to the cloud server. With the trapdoor, the cloud server returns k most similar images by searching on the Index.

Trapdoor generation: The authorized user first needs to extract a global feature fq from the query image when he wants to search images similar to the query image. Then, we apply L p-stable LSH functions g(fq) = (hq(fi), h2(fq),…, hλ(fq)) on the feature fq and therefore, generate L hash values for the feature fq. Finally, to protect the security of the trapdoor, the data user applies pseudorandom permutation Enc on each hash value such that T(fq) = Enc(k1, g1(fq)), Enc(k2, g2(fq)),…, Enc(kL, gL(fq))). Besides, for a query feature vector fq = (f1,q,…, fR,q), the data user first constructs a modified vector as:

$$\hat{f}_q = (-2f_{1,q}, -2f_{2,q}, \cdots, -2f_{\ell,q}, 1)^T$$

He next chooses a random positive value r∈ú and uses it with the secret matrix R to encrypt the modified query vector as Image for - Enabling Similarity Search over Encrypted Images in Cloud. At last, the data user sends T(fq) along with the f'q to the cloud server.

Search: Once receiving a search request, the cloud server performs search on the index for each component of the T(fq). Then, the cloud server gets an identifier list of images corresponding to every component of T(fq). The global feature extracted from the image is mapped to L hash tables during the index period. So, during search period, the global feature extracted from the query image is also mapped to L hash tables with L hash functions. If the query image and one specific image in the collection satisfies the condition that g(fq) = g(fi), then they are considered to be similar. Finally, cloud server gathers all these image identifiers that have the same hash value with the query image. However, in order to get more accurate results, the cloud server need to compute the distance between the query image and all these similar images. The cloud server conducts the scalar product:

$$f_q^{'T} \cdot f_i^{'} = (rR^{-1} \cdot \hat{f}_q)^T \cdot R^T \hat{f}_i = r(\hat{f}_q)^T \cdot \hat{f}_i = r(\|f_i\|_2^2 - 2\sum_{j=1}^{\ell} f_{j,i} f_{j,q}) = r(\|f_q - f_i\|_2^2 - \|f_q\|_2^2)$$

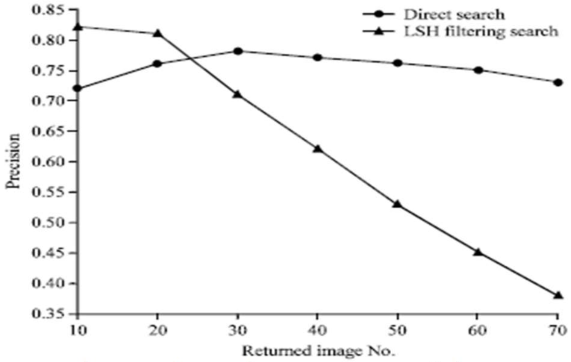
The distance ||fq-fi||22 is hidden by the secret scalar r and the unknown ||fq||22. When x≥0, the function f(x) = x3 is order preserving, i.e., f(x1)>f(x2) implies x1>x2. Also, because for each query fq the values of ||fq||22 and r is fixed, the cloud server can directly find closest feature vectors by simply sorting out the set of scalar products r(||fq-fi||22-||fq||22), without knowing the sensitive information from the feature vectors.

After computing the distances, the cloud server ranks these images according to their distance with the query image. Finally, the cloud server sends the top k most similar encrypted images back to the data user as search results.

Once receiving the encrypted images returned by the cloud server, data user decrypts these images with the secret key km shared by the data owner and obtains the plain images similar to the query image. Now, the round of search is completed.

Results and Discussion

In this section, we describe the details of our experiments and analysis the experimental results of our scheme. And an image database consisting of 130000 images is used to test the performance of the proposed scheme.



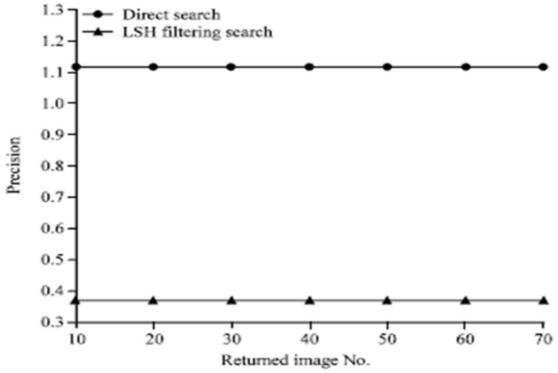
Result accuracy: This criterion ıs used to evaluate the correctıon of the returned results, is evaluated by precision defined as follows:

$$\text{Precision} = \frac{\text{num}_{\text{hit}}}{\text{num}_{\text{return}}}$$

Where, $\text{num}_{\text{return}}$ is the number of the returned images, num_{hit} is the number of correct similar images. The accuracy of the scheme is mainly decided by the feature extraction method in common image retrieval systems. The accuracy of a query in our scheme is shown in the Fig. In our experiment, we compare the retrieval performance of our proposed scheme with the direct distance computation scheme. The direct distance computation scheme compares the distance between different feature vectors as shown in Eq.The direct search scheme linearly scans all images in the database and finds the similar images with the query image.

Time complexity: In theory, the LSH filtering search scheme should search faster than the direct search scheme since the direct search time is linear to the total number of image in the database. The experiment also shows the prospective results as the theoretical analysis as shown in Fig. 3. The experiment results illustrate that the LSH filtering scheme efficiently improves search efficiency and saves search time. We find that search time of different returned image number is same. That is because we always compute the distance of all remaining images with the query image and finally return specific number of images to the user.

Though our LSH scheme does not improve the precision, it reduces the computational costs of similarity image search by filtering images using search index.



The direct search scheme finds the similar images by linearly scanning all images in the database. So the time complexity of the direct search scheme is O(n) which n is the total number of image in the database. However, our scheme saves search time by exploiting LSH algorithm. The experiment result shows that LSH algorithm makes sense for similarity image search. Our work aims to prove that LSH is suit for similarity search over encrypted image and eventually gets the prospective results. To further improve the retrieval accuracy, we will exploit local image features in our future work.

Conclusion

In this study, we proposed an efficient similarity search scheme over encrypted images in the cloud. We exploit the widely used p-stable locality sensitive hashing for the construction of index in our scheme. P-stable LSH algorithm is often used in high dimension spaces and is used to construct secure search index in this study. Our scheme enables efficient similarity image search and less time consuming. In the cloud computing, it is critical to prove the security of outsourced images. This study gives a rigorous security definition and proved the security of the proposed scheme under the provided security model. To verify the proposed scheme, we conduct experiments on a image library with 130000 images. Experiment results show that our scheme is suit for similarity search over encrypted images. In short, our works have a great significance to the further development of similarity CBIR in the cloud.

References

1. Boneh, D. and B. Waters, 2007. Conjunctive, subset and range queries on encrypted data. Proceedings of the 4th Theory of Cryptography Conference, February 21-24, 2007, Amsterdam, The Netherlands, pp: 535-554.
2. D. X. Song, "Practical techniques for searches on encrypted data", Security and Privacy, S&P 2000.Proceedings, 2000 IEEE Symposium on, ed: IEEE, (2000), pp. 44-55.
3. Cao, N., C. Wang, M. Li, K. Ren and W. Lou, 2011. Privacy-preserving multi-keyword ranked search over encrypted cloud data. Proceedings of the 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, April 10-15, 2011, Shanghai, China, pp: 829-837.
4. C. Wang, "Achieving usable and privacy-assured similarity search over outsourced cloud data", INFOCOM,2012 Proceedings IEEE, (2012), pp. 451-459.
5. J. Li, "Fuzzy keyword search over encrypted data in cloud computing", INFOCOM, 2010 Proceedings IEEE,(2010), pp. 1-5.
6. M. Chuah and W. Hu, "Privacy-aware bed tree based solution for fuzzy multi-keyword search over encrypted data", Distributed Computing Systems Workshops (ICDCSW), 2011 31st International Conference, (2011),pp. 273-281.
7. D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Theory of cryptography, ed: Springer, 2007, pp. 535-554.
8. X. Zhiyong, Efficient Multi-Keyword Ranked Query on Encrypted Data in the Cloud", Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference, (2012), pp. 244-251.
9. J. Katz, "Predicate encryption supporting disjunctions, polynomial equations, and inner products", Advances in Cryptology–EUROCRYPT 2008, ed: Springer, (2008), pp. 146-162.
10. C. Ning, "Privacy-preserving multi-keyword ranked search over encrypted cloud data", INFOCOM, 2011 Proceedings IEEE, (2011), pp. 829-837.