IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 2, 2019, accepted October 24, 2019, date of publication November 4, 2019, date of current version November 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951175

A Large-Scale Secure Image Retrieval Method in Cloud Environment

YANYAN XU[1], XIAO ZHAO[1], AND JIAYING GONG[2]
[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[2]Ping An Technology (Shenzhen) Company, Ltd., Shenzhen 518000, China

Corresponding author: Yanyan Xu (xuyy@whu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41571426, in part by the National Key Research and Development Program of China under Grant 2017YFB0504202, in part by the Wuhan Applied Basic Research Programs under Grant 2017010201010114, and in part by the Fundamental Research Funds for the Central Universities.

ABSTRACT With the rapid development of cloud computing technology, more and more users choose to outsource image data to clouds. To protect data's confidentiality, images need to be encrypted before being outsourced to clouds, but this brings difficulties to some basic yet important data services, such as content-based image retrieval. Existing secure image retrieval methods generally have some problems such as low retrieval accuracy and low retrieval efficiency, which cannot meet requirements for large-scale image retrieval in cloud environment. In this paper, we propose a large-scale secure image retrieval method in cloud environment. The Hamming embedding algorithm is utilized to generate binary signatures of image descriptors. A frequency histogram combined with binary signatures is generated to provide a more precise representation of image features in an image and thus the retrieval accuracy is improved. Visual words are selected from the histogram by the random sampling method before the min-Hash algorithm is performed on binary signatures of selected visual words to generate a secure index. The random sampling method and min-Hash algorithm can not only ensure the security of the search index, but also greatly improve the image retrieval efficiency. This method achieves the balance among security, accuracy and efficiency of large-scale secure image retrieval in public clouds. The security analysis and experimental results show the effectiveness of the proposed method.

INDEX TERMS Large-scale image retrieval, secure image retrieval, min-Hash, hamming embedding, secure inverted index.

I. INTRODUCTION

With the popularization of digital cameras and smart phones, multimedia data such as images and videos become much easier to capture and have shown an explosive growth. Cloud computing platforms integrating grid computing, parallel computing and distributed computing can provide powerful support for massive data services and application processing with its low cost, powerful computing capacity and unlimited resource pool. Therefore, users tend to upload image data to the cloud for storage and processing. Although this brings great convenience, it also causes serious security threats. Data outsourced to the cloud is completely out of its owner's direct physical control thus can be attacked by outside hackers or "honest-but-curious" cloud service providers (CSP), and may be at risk of being leaked or abused.

In order to protect users' privacy and enhance data confidentiality, sensitive images need to be encrypted before being uploaded to clouds, but the encryption process can make some common operations in the cloud environment difficult, such as image retrieval. Content based image retrieval (CBIR) is a very promising method in the image retrieval field. It is characterized by extracting various image features and then comparing the distance between features automatically. However, after the image is encrypted, the distance between image features becomes difficult to maintain due to the randomness brought by encryption process, which makes it challenging to use the CBIR method.

Many methods were proposed to solve this problem. Information retrieval on encrypted domain originated from retrieval on text documents. Song *et al.* [1] proposed

The associate editor coordinating the review of this manuscript and approving it for publication was Zhitao Guan.

a ciphertext scanning method based on streaming cipher to make sure the search term exists in the ciphertext. Boneh *et al.* [2] proposed a keyword search method based on public-key encryption, where the server can identify whether messages encrypted by users' public key contain specific keywords, but learn nothing else. Swaminathan *et al.* [3] explored techniques to securely rank documents and extract the most relevant documents from an encrypted collection based on the encrypted search queries. Wang *et al.* [4] utilized an order-preserving symmetric encryption (OPSE) method to achieve both security and privacy preserving, but its security is compromised. Cao *et al.* [5] proposed a privacy-preserving multi-keyword ranked search over encrypted data. Although secure text search techniques can be extended to image retrieval based on user-assigned tags, the extension to CBIR is not straightforward. CBIR typically relies on comparing the distance of image features, but comparing similarity among high dimensional vectors using cryptographic primitives is challenging [6].

In recent years, several methods have been proposed to solve the problem of secure CBIR. Lu *et al.* [6] proposed three privacy preserving image search schemes over encrypted multimedia data set, where bit plane randomization, random projection and random unary encoding are applied to low-level features such as color histograms. Lu *et al.* [7] proposed a secure image retrieval scheme over encrypted domain where the secure index for matching visual strings in the encrypted domain is constructed through order preserving encryption and randomized hash functions. Xia *et al.* [8] proposed a Bag-of-Encrypted-Words (BOEW) model. The image is first encrypted by substitution and permutation. The local histograms are clustered together to generate a BOEW model. As a result each image can be represented as a normalized histogram of the encrypted visual words. Xia *et al.* [9] proposed a secure Local Binary Pattern (LBP) feature extraction method, where block and pixel permutation are used together to provide a privacy-protected LBP extraction scheme in the ciphertext domain. These schemes are efficient but the security is compromised. Ferreira et al. proposed a novel scheme in [10] where color information is encrypted by deterministic encryption techniques to enable privacy-preserving image retrieval while texture information is encrypted by probabilistic encryption algorithms for better security. Xu et al. proposed a privacy-preserving content-based image retrieval method [28]. The images are divided into two different components based on orthogonal decomposition, for which encryption and feature extraction are executed separately. CSP can extract global histograms as features from Discrete Cosine Transform (DCT) coefficients of encrypted images and compare it to the features of the queried images. Erkin et al. proposed to use secure multiparty computation to retrieve private information in [11], [12], but it requires many interactive rounds and the communication complexity is too high. Hsu *et al.* [13] proposed a secure retrieval scheme that applies the homomorphic encryption algorithm on SIFT. Similar methods base on homomorphic encryption scheme are proposed in [11], [14], [15]. Although it is sufficiently secure, homomorphic encryption algorithm will cause serious expansion of ciphertext data and the computational cost is very high. Xia et al. [16] designed a secure retrieval framework based on local features (SIFT), where Earth Mover's Distance (EMD) is transformed in a way that the CSP can evaluate the similarity between images without learning sensitive information. This method is secure, but two-rounds communication are needed between CSP and users before the CSP obtains top-k ranked images, which is time-consuming and its communication cost is high. Qin et al [17] proposed a secure image retrieval method based on Harris Corner optimization and local sensitive hash (LSH), but the retrieval efficiency is not ideal, especially for large-scale image datasets.

There exists two problems in the aforementioned works. Firstly, there is a trade-off between security and efficiency. Methods based on lightweight encryption such as permutation or substitution are efficient but not secure; methods based on multiparty computation or homomorphic encryption algorithm are secure, but the computation cost is too high to be practical. Secondly, there is a trade-off between retrieval accuracy and efficiency. many methods utilize low-level features in image retrieval, such as color, texture, shape, etc., the retrieval accuracy can hardly meet requirements of practical applications because of the ''semantic gap'' between the visual features and the richness of human semantics. Some methods are based on local features such as SIFT. Although the retrieval precision is higher, the efficiency is unsatisfactory. Therefore, most of these methods cannot satisfy requirements of large-scale image retrieval in cloud environment.

To solve these problems, a large-scale secure image retrieval method in cloud environment is proposed in this paper. The Hamming embedding (HE) algorithm is utilized to generate binary signatures of image descriptors. A frequency histogram combined with binary signature is constructed to provide a more precise representation for image features, thereby improving the retrieval accuracy. Visual words are selected from the histogram by the random sampling method before the min-Hash algorithm is performed on binary signatures of selected visual words to generate a secure index. The random sampling method and min-Hash algorithm can not only ensure the security of the search index, but also greatly improve the image retrieval efficiency. This method achieves the balance among security, accuracy and efficiency of large-scale secure image retrieval in public clouds. The security analysis and experimental results prove the effectiveness of the proposed method.

The rest of this paper is organized as follows. In Section II, we give a brief introduction of the system model and the threat model. We introduce preliminary knowledge in Section III. We present our scheme in Section IV. In Section V, we provide security analysis. Experimental results and performance analysis are given in Section VI. Conclusions and future work are given in Section VII.

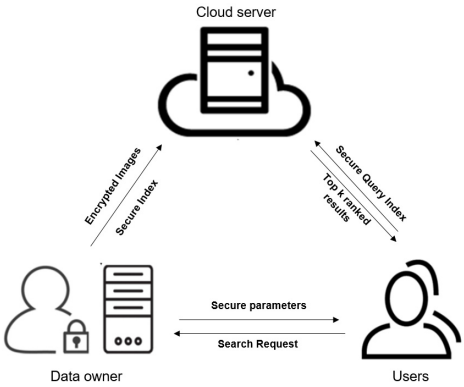


FIGURE 1. System model.

II. PROBLEM FORMULATION

A. SYSTEM MODEL

Figure. 1 illustrates a secure image retrieval model in the Cloud. There are three entities involved in this model: data owner, CSP, and users. Data owner extracts features from images and constructs secure search index outsourced together with encrypted images to the cloud. CSP stores cipher-images and secure index, performs secure image retrieval and returns top k ranked results when receiving users' query. Users generate trapdoor and send it to CSP, decrypt cipher-images returned by CSP and get requested images.

B. THREAT MODEL

We consider data owner and users are always trusted. The security threats are mainly from the "honest-but-curious" CSP and external attackers. CSP faithfully follows the designated operations yet intends to infer data from data owner and users to obtain private information. External attackers try to get image content. Thus the security of images and index outsourced by data owner, the query information sent by users, and the security of the query process executed between users and the CSP should be considered in the proposed method.

III. PRELIMINARIES

A. BAG-OF-VISUAL-WORDS MODEL

BoW (Bag-of-Words) model is first being used in natural language processing and information retrieval. Sivic extended this idea to computer version and proposed Bag-of-Visual-Words (BoVW) model [18], which has been successfully adopted to enable fast indexing and retrieval of large image collections [19], [20]. In this model, local features are extracted from all images in the database and then jointly clustered. The cluster centers are used as 'visual words' to form a vocabulary. Image features are mapped to one or more visual words and the image can be represented by a frequency histogram of visual words. Images with the closest histogram distance are returned as retrieval results. This method can reduce the impact of "semantic gap" on the accuracy of image retrieval and has shown good performance in image retrieval task, but it still suffers from some problems, such as insufficient discriminative power of visual words, quantization error caused by assigning descriptors to visual words, low efficiency caused by comparing distance between high dimensions of vectors, etc. In order to solve these problems, Jégou etc. improved the BoVW model and proposed the HE algorithm in [21]. This algorithm constructs binary signature vectors for the features assigned to the same clustering center, and then a threshold function are used to filter out features that are in the same cluster but have large differences from other features, so that the retrieval accuracy can be improved.

B. MIN-HASH ALGORITHM

The min-Hash algorithm is originally used in detecting the similarity of two documents [22], [23], and now it has been extended to detect the similarity of images expressed by visual words [24], [25]. It is based on the theory of Jaccard similarity, for the given sets A and B, the Jaccard similarity is defined as:

$$\mathrm{J}(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{A + B - |A \cap B|} \tag{1}$$

For a random array with uniform distribution, the probability that two sets have the same value of min-Hash equals the Jaccard similarity of the two sets.

Suppose a number of random hash functions are given: $f_j{:}F \rightarrow R$, assigning a real number to each feature vector. Let X_a and X_b be different feature vectors from the vector set F. The random hash functions have to satisfy two conditions [22]: $f_j(X_a) \neq f_j(X_b)$, $\mathrm{P}(f_j(X_a) < f_j(X_b)) = 0.5$. The functions also have to be independent. Note that each function f_j infers an ordering on the set of feature vectors $X_a < X_b$ if $f_j(X_a) < f_j(X_b)$. We define a min-Hash as an element of a set ∂ under ordering induced by f_j in (2):

$$\mathrm{m}\left(\partial, f_j\right) = \operatorname{argmin} f_j(X) \quad X \in \partial \tag{2}$$

For each set ∂_i and each hash function f_j the min-Hash $\mathrm{m}\left(\partial_i, f_j\right)$ are recorded. The probability of $\mathrm{m}\left(\partial_1, f_j\right) = \mathrm{m}\left(\partial_2, f_j\right)$ is:

$$\mathrm{P}\left(\mathrm{m}\left(\partial_1, f_j\right) = \mathrm{m}\left(\partial_2, f_j\right)\right) = \mathrm{J}\left(\partial_1, \partial_2\right) = \mathrm{sim}\left(\partial_1, \partial_2\right) \tag{3}$$

IV. SECURE IMAGE RETRIEVAL SCHEME

The proposed method consists of the following process: on the data owner side, the HE algorithm is utilized to generate the binary signatures of image descriptors, and thus a frequency histogram combined with binary signatures of features is constructed. Visual words are permuted and selected from the histogram with the random sampling method; min-Hash algorithm is performed on binary signatures of selected visual words to generate a secure index. On the user side, it generates a secure query index of requested image and initiates a query request to the CSP. The CSP performs secure image retrieval, which compares the distance between query index and secure image index stored in the CSP and returns

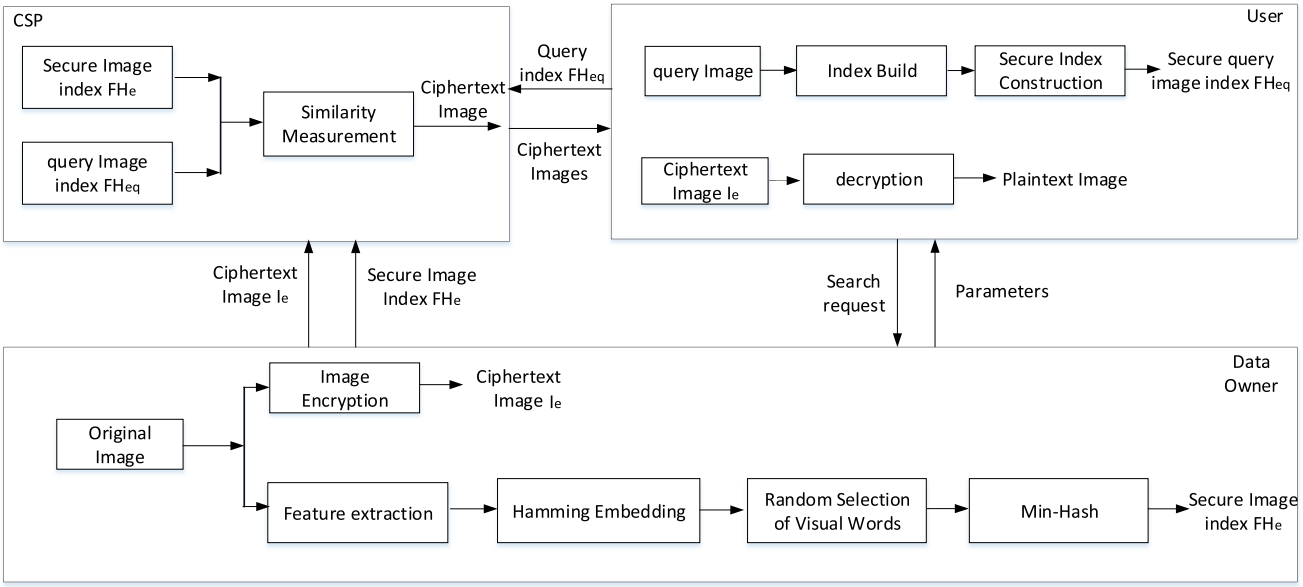


FIGURE 2. Flowchart of the proposed method.

the most similar ciphertext images. Finally, the user decrypts to obtain the plaintext image. The flow chart of the proposed scheme is shown in Figure. 2.

A. CONFIGURATION STAGE

The main task of this stage is to build a visual dictionary, generate a median matrix and encryption/decryption keys at the data owner. The processing flow is as follows:

1) FEATURE DESCRIPTOR QUANTIZATION

SIFT descriptors x_i, $i \in [1, n_w]$, are extracted from a training image dataset TB, n_w is the number of descriptors in TB. The quantizer q is defined in (4), which is a function that maps a descriptor $x_i \in R^d$ to an integer index $w_l = q(x_i)$, w_l is the visual word assigned to x_i, $l \in [1, k]$, k is the number of integer index, and q is obtained by performing k-means clustering algorithm on descriptors. Visual dictionary $VD = \{w_{1,\ldots\ldots,}w_k\}$ is constructed by combination of visual words.

$$\begin{aligned} q &: R^d \rightarrow [1, k] \\ x_i \rightarrow w_l &= q(x_i) \quad i \in [1, n_w], \quad l \in [1, k] \end{aligned} \tag{4}$$

2) MEDIAN MATRIX GENERATION

The median matrix $\tau = \{\tau_{w_1}, \ldots\ldots, \tau_{w_k}$ of all SIFT descriptors is generated, which consists of median vector τ_{w_l} of descriptors assigned to the same visual word w_l.

3) ENCRYPTION/DECRYPTION KEY GENERATION

$KeyGen(\cdot)$ is used to generate random scrambling key K_1, encryption key K_2 of the min-Hash algorithm, and AES algorithm encryption/decryption key K_3, where α is a security parameter used to generate key, as shown in (5).

$$KeyGen\left(1^{\alpha}\right) \rightarrow (K_1, K_2, K_3) \tag{5}$$

B. DATA OWNER SIDE

Data owner is mainly responsible for generating cipher-images and secure image index that will be uploaded to CSP. Images are encrypted by AES algorithm to obtain cipher-images. Secure image index is constructed by three steps: binary signature generation; random selection of visual words; min-Hash secure image index construction.

1) BINARY SIGNATURE GENERATION

SIFT descriptors are extracted from an image and denoted as $x_i = \{x_{i1}, x_{i2}, \ldots, x_{i128}\}$, $i \in [1, n_w]$; x_i is mapped to the visual word w_l according to VD; then it is compared with median vector τ_{w_l} of w_l and a 128-bits binary vector $h(x_i) = \{h_1(x_i), h_2(x_i), \ldots, h_{128}(x_i)\}$ is constructed according to (6):

$$h_j(x_i) = \begin{cases} 1 & if \ \ x_{ij} > \tau_{w_{l,j}} \\ 0 & otherwise \end{cases}, \mathrm{j} \in [1, 128] \tag{6}$$

Different feature descriptors have different discriminative capabilities for the image. We assign weight information to x_i based on term of frequency (TF) and inverse document frequency (IDF), where TF shows the number of times that a visual words occurs in an image; IDF reflects the informativeness of visual words, that is, visual words that appear in many different images are less informative than those that appear rarely [25]. The weight information $wgt(x_i)$ is calculated by (7):

$$\begin{aligned} wgt(x_i) &= \frac{idf^2(x_i)}{\sqrt{tf(x_i)}} \\ tf(x_i) &= \frac{n_{l,j}}{\sum_k n_{k,j}}, \quad idf(x_i) = log\frac{|I|}{1 + \left|\{j{:}w_l \in I_j\}\right|} \end{aligned} \tag{7}$$

where $n_{l,j}$ means the number of times that visual word w_l occurs in the image I_j; $\sum_k n_{k,j}$ represents the total number of

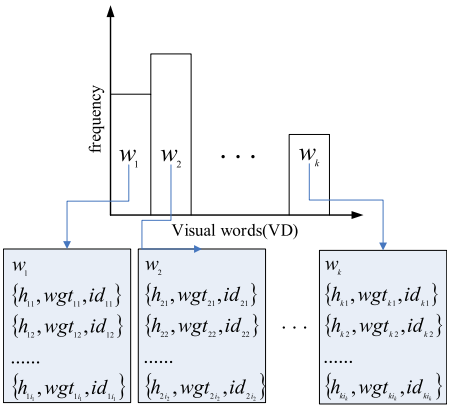


FIGURE 3. Frequency histogram combined with binary signatures.

times that all visual words occur in the image I_j; $|I|$ represents the total number of images in the dataset I; $\left|\left\{j{:}w_l \in I_j\right\}\right|$ represents the number of images containing the visual word w_l.

After each descriptor x_i is processed by the method mentioned above, it can be expressed as $s(x_i)$. Then a frequency histogram FH combined with binary signatures of descriptors is constructed, as shown in (8):

$$\begin{aligned} s\left(x_i\right) &= \left\{h\left(x_i\right), wgt\left(x_i\right), id\left(x_i\right)\right\} \\ w_l &= \left\{s\left(x_1\right), \ldots, s\left(x_{\mathrm{n}_l}\right)\right\}, \quad FH = \left\{w_1, w_2, \ldots, w_k\right\} \end{aligned} \tag{8}$$

where n_l is the number of descriptors that belongs to w_l, $id(x_i)$ is the image ID that x_i belongs to. The frequency historgram is shown in Figure 3.

2) RANDOM SELECTION OF VISUAL WORDS

Although binary signatures of descriptors can refine visual words and significantly improve retrieval precision, it also introduces additional information and increases the memory usage of the index. In a large dataset, the distribution of features in each image in a k-dimensional visual dictionary is relatively sparse, we consider to reduce the redundant information and improve retrieval performance to suit for large-scale image retrieval. We first permute image frequency histogram $\boldsymbol{FH}$ with $\boldsymbol{K_1}$, then use random selection function to select $\boldsymbol{v}$ ciphertext visual words randomly from $\boldsymbol{FH}$ according to (9), where $\boldsymbol{Perm}()$ represents a permutation operation, $\boldsymbol{RanSel}()$ represents a random selection operation. We obtain a permuted and condensed image frequency histogram $\boldsymbol{FH'}$ that includes $\boldsymbol{v}$ visual words.

$$RanSel\left(Perm\left(FH, K_1\right), v\right) \rightarrow FH' \quad v < k \tag{9}$$

3) SECURE INDEX CONSTRUCTION

Min-hash algorithm is utilized in $s(x_i) \subset \mathrm{w}_l$ to generate secure index $s_e(x_i) \subset w_l^e$, $l \in [1, v]$. The detailed process is given in (10-12): using the encryption key K_2 to generate m ($m < 128$) independent random hash functions $f_j, j \in [1, m]$; performing f_j on binary vectors $h(x_i)$ of $s(x_i)$, then 128-bits binary vector $h(x_i)$ is transformed to m-bits binary vector $h_e(x_i)$, m < 128.

At last the secure inverted index table FH_e is constructed as (12).

$$HashFunGen(K_2, \mathrm{m}) \rightarrow F\_\mathrm{Set}(f_j) \quad \mathrm{j} \in [1, \mathrm{m}] \tag{10}$$

$$\begin{aligned} h_e\left(x_i\right) &= \left\{h_{e1}\left(x_i\right), h_{e2}\left(x_i\right), \ldots, h_{em}\left(x_i\right)\right\} \\ &= argmin(f_1\left(h(x_i)\right), f_2(h(x_i)) \ldots f_m((x_i)) \end{aligned} \tag{11}$$

$$\begin{aligned} s_e\left(x_i\right) &= \left\{h_e\left(x_i\right), wgt\left(x_i\right), id\left(x_i\right)\right\} \\ \mathrm{w}_l^e &= \left\{s_e\left(x_1\right), \ldots, s_e\left(x_{\mathrm{n}_l}\right)\right\}, \\ FH_e &= \left\{\mathrm{w}_1^e, \mathrm{w}_2^e, \ldots, \mathrm{w}_v^e\right\} \end{aligned} \tag{12}$$

C. USER SIDE

Users send query requests to the data owner. After identity authentication, the data owner sends parameters VD, τ, K_1, K_2, K_3 to users securely.

1) INDEX CONSTRUCTION

User extracts SIFT descriptors from a query image Img_q and constructs a frequency histogram FH_q according to (6-8).

$$IndexBuild\left(Img_q, \mathrm{VD}, \tau\right) \rightarrow FH_q \tag{13}$$

2) SECURE INDEX CONSTRUCTION

The secure query index FH_{eq} is generated according to (9-12). Users send the secure index FH_{eq} to the CSP for retrieving requested images.

$$ConstructSecureindex\left(FH_q, K_1, K_2\right) \rightarrow FH_{eq} \tag{14}$$

3) CIPHER IMAGE DECRYPTION

After receiving the requested cipher image Img_{eq} from the CSP, users obtain the search result Img by decrypting Img_{eq} via K_3.

$$Dec\left(Img_{eq}, K_3\right) \rightarrow Img \tag{15}$$

D. CLOUD SERVER

After receiving the query request, the CSP compares the distance between the secure query index FH_{eq} and the secure image index FH_e stored in the CSP and returns the most similar ciphertext images.

Assume $\mathrm{s}_{eq}(x_i) \subset FH_{eq}$, where $x_i = (x_1, x_2, \ldots, x_t)$; $\mathrm{s}_e(y_j) \subset FH_e$, where $y_j = (y_1, y_2, \ldots, y_p)$. Then the similarity of x_i and y_j can be evaluated as following:

$$\mathrm{f}\left(x_i, y_j\right) = \begin{cases} wgt\left(x_i\right) * \left(\exp\left(\dfrac{-L\left(x_i, y_j\right)^2}{\sigma^2}\right)\right) \\ \quad if \;\; w'\left(x_i\right) = w'\left(y_j\right) \;\; and \;\; L\left(x_i, y_j\right) \leq h_d \\ 0 \quad otherwise \end{cases} \tag{16}$$

$$L\left(x_i, y_j\right) = \frac{\sum_{v=1}^{m} x_i(v) \oplus y_j(v)}{m} \tag{17}$$

$$\mathrm{sim}\left(A, B\right) = \sum\nolimits_{i=1\ldots t} \sum\nolimits_{j=1\ldots p} \mathrm{f}\left(x_i, y_j\right) \tag{18}$$

where σ is the weight parameter, h_d is a calculation threshold, $L(a, b)$ is the quantized Hamming distance between the computed vectors a, b, and m is the dimension of vector generated by min-Hash, sim (A, B) is the matching score for image B to image A.

Finally, the retrieved ciphertext images are arranged in descending order of matching scores to return to users.

V. SECURITY ANALYSIS

The AES algorithm is used to encrypt images and its security is widely accepted. Therefore we will focus on analyzing the security of the secure index under different attack models: Ciphertext Only Attack (COA) and Known Plaintext Attack (KPA).

A. COA

In COA model, the server can only access to the secure index. The secure inverted index table contains v visual-word IDs $w_l^e, (l \in [1, v])$ that are randomly selected and permuted from visual dictionary. For CSP, the difficulty of getting the visual words from the visual dictionary to build a secure index is $O(k!)$ in which k is the total number of visual words in the database. In a large scale dataset, k is usually very large (in our experiment, $k = 20,000$), so, the cloud server cannot infer the distribution of the original visual words of each image based on the visual word numbers in the secure index. In addition, min-Hash algorithm is utilized to construct the secure index, and the maximum probability of attacking binary signature to get the image feature vector is:

$$P_r = \frac{1}{\left(C_d^{\lceil \frac{d}{2} \rceil}\right)^n * \left(A_d^m\right)^n} \tag{19}$$

where d is the dimension of feature vector and n is the total number of images features. For massive image feature points, it is difficult to obtain image features by attacking binary signatures.

B. KPA

In this attack mode, the CSP knows some plaintext image pairs and the corresponding secure index information. Therefore, it can generate visual word representations for known images and compare them to the secure index to obtain part of visual words being used to construct the secure index. With this information, the entire permutation order of visual words may be revealed so that the CSP can easily obtain more information about the ciphertext image. However, in a large-scale dataset, the visual dictionary dimension is usually very large and the visual words are randomly selected to build a secure index. The difficulty for CSP to obtain the original visual word distribution of each image is $O(k^v \cdot v!)$ where v is the number of the visual-word IDs contained in the secure inverted index table. Thus it is difficult for the CSP to obtain visual words and infer image content. Further, min-Hash algorithm is employed to construct the secure index. Even if the CSP knows the distribution of the visual words of the image, it cannot obtain useful information about the random hash function F_set without the key. As a result, it is difficult for attackers to build a secure index for any other images.

Therefore, in the KPA model, the server can only know which encrypted images in the ciphertext image dataset may be similar to the known images through the index information of the known images, but cannot obtain detailed content information of the ciphertext images.

VI. EXPERIMENT RESULTS AND ANALYSIS

In this chapter, we present the experimental results of the proposed method. We perform experiments on INRIA Holidays dataset (1491 images, 4.455M descriptors) [21] and Oxford dataset (5000 images, 4.977M descriptors) [26]. To evaluate large-scale secure image search we also introduce two distractor image collections randomly selected from Flickr 1M [27]: Flickr10k (10,000 images, 10.28M descriptors), Flickr100k (100,000 images, 103.72M descriptors). The proposed method is evaluated in terms of the search precision, search efficiency and security. All experiments are implemented by C++ language on Windows 7 (64-bit) operating system, Intel (R) Core (TM) i3 CPU @ 3.07 GHz, 6.00GB.

A. PARAMETER ANALYSIS

In the experiments, the parameters value are set as follows: the dimension of visual dictionary $k = 20,000$; binary signature dimension $l = 128$; similarity comparison threshold $h_d = 56$, $\sigma = 28$.

In the process of constructing the secure inverted index, a random sampling feature selection method is proposed to select visual words randomly from the visual dictionary. In order to verify the effect of different size visual dictionary on retrieval performance in the random sampling method, we conduct multiple sets of experiments with the random sampling method on a visual dictionary of size $k = 20,000$ on the holiday image database, and obtain mean average precision (mAP) and time cost for different size of visual dictionary. The results are shown in Figure. 4.

From the Figure. 4(a), we can see when $k < 12,000$ in the random sampling, the retrieval accuracy decreases rapidly. Considering the time cost spent in searching process shown in Figure. 4(b), we select 12,000 visual words from a visual dictionary of $k = 20,000$ with the random sampling method. Table 1. shows the experimental results of multiple groups which verify the effect of the random sampling on the accuracy and efficiency of the secure retrieval. It can be seen that the average retrieval accuracy of multiple experiments is around 74.0%, and the average retrieval time is around 0.2s. Compared with the average retrieval accuracy of 76.56% and the retrieval time of 0.47s without the random sampling, the retrieval time is significantly shortened and the retrieval efficiency is highly improved although the retrieval accuracy of random sampling is reduced by about 2%.

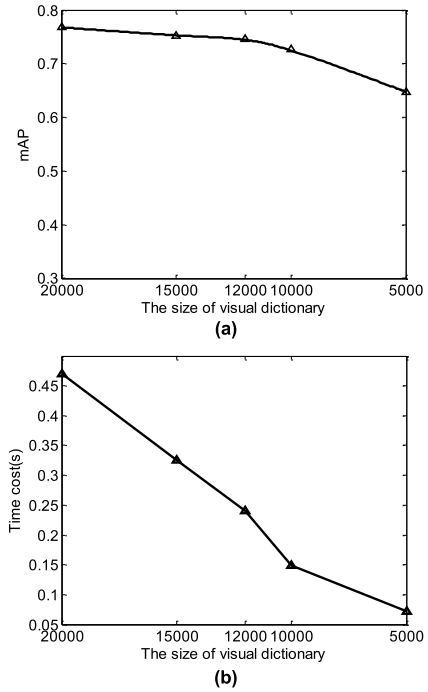


FIGURE 4. The effect of different size of visual dictionary on retrieval performance in the random sampling method: (a) mAP (b) time cost.

TABLE 1. The effect of the random sampling method on retrieval accuracy and efficiency.

Group	The number of features	mAP	Retrieval time(s)	Original Retrieval Time(s)
1	2673119	73.8%	0.2108	
2	2661046	73.9%	0.1885	
3	2678594	74.6%	0.2272	0.47
4	2671388	74.1%	0.2253	
5	2657129	73.5%	0.1912	
Average	2668255	74.0%	0.2086	

Besides the random sampling method, the number of hash functions m in min-Hash algorithm impacts retrieval performance as well. Experiments are performed on the holiday database in terms of $m = 128, 96, 64, 32$ respectively and experimental results are given in Figure 5. We can see that the retrieval accuracy decreases as m decreases and when m is smaller than 64, the retrieval accuracy decreases rapidly. Considering the search accuracy and efficiency, we set $m = 64$ in our experiments.

B. TIME CONSUMPTION

In this section, we test the time consumption of the proposed method, which primarily consists of secure index construction time in data owner side and the secure image retrieval time in CSP side.

1) TIME CONSUMPTION OF SECURE IMAGE INDEX CONSTRUCTION

In the process of configuration, generating a median matrix τ is a one-time process. Its time complexity is related to the

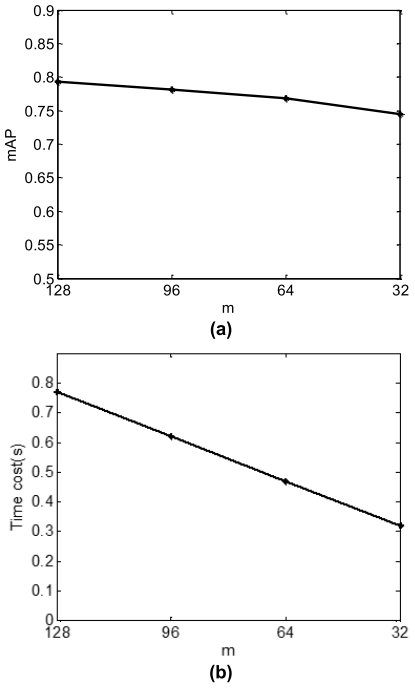


FIGURE 5. The effect of the number of hash functions on retrieval performance in min-Hash algorithm: (a) mAP (b) time cost.

TABLE 2. Comparison Results of time complexity and actual running time.

Phases	Time complexity	Running time(s) for holiday	Running time(s) for oxford
Binary signature	$O(n)$	923	1899
Random sampling	$O(k)$	79	78
Min-hash	$O(m)$	524	532

number of the feature descriptors n in training dataset, that is, $O(n)$. We use 2M descriptors of holiday dataset to train the median matrix, and the time cost is 1048s.

The generation of secure image index includes three phases: binary signature generation, random selection of visual words and min-Hash generation. The time complexity and actual running time are shown in Table 2. The time consumption of binary signature generation is proportional to the number of feature descriptors n; the time consumption of the random sampling is related to the visual dictionary dimension k; the time consumption of min-Hash index generation is related to the number of hash functions m. In these three phases, the time cost of generating a binary signature is the most time consuming part, but this part is necessary to improve the retrieval accuracy, and it is only a one-time process, so the time cost can be accepted.

2) TIME CONSUMPTION OF SECURE IMAGE RETRIEVAL

In the proposed method, the random selection and min-Hash algorithm are utilized together to establish a secure index, which play an important role in reducing search time in the CSP side. In order to verify its effect, we conducted

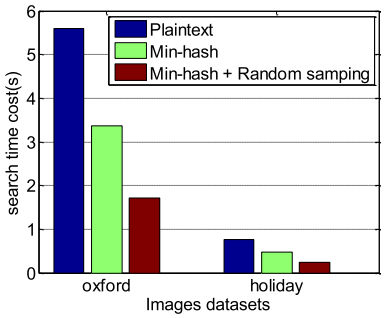


FIGURE 6. Comparison results of search time cost in the CSP side.

TABLE 3. Comparison of storage consumption on different datasets.

Datasets	Holiday(MB)	Oxford(MB)
HE Plaintext	216.03	851.1
Min-hash	106.86	426.5
Min-hash+ random sampling	63.17	237.8

TABLE 4. mAP comparison.

Dataset	BoVW[20]	BoVW+ minHash[7]	HE +min-Hash	The proposed method
Holiday	55.91%	52.86%	76.56%	74.0%
oxford	25.59 %	22.13%	48.77%	45.7%

experiments on different datasets and obtained comparison results of search time consumption shown in Figure. 6. It shows that the proposed method is good at reducing the dimension of the image feature vector and improving retrieval efficiency.

C. STORAGE CONSUMPTION

Table 3 shows the comparison results of storage consumption from different datasets. From the table we can see our method can greatly reduce the storage pressure of data in clouds compared to HE plaintext, which makes the retrieval method more suitable for large-scale image datasets.

D. SEARCH PRECISION

We conducted experiments on INRIA holiday and Oxford datasets respectively, and obtained the mAP results shown in Table 4. We can see that the retrieval result of our scheme is much better than the original BoVW algorithm proposed in [19] and Lu's method proposed in [7]. This is because we utilize HE method to generate binary signatures that refine visual words, which provides more precise representation for image features and similarity measure for descriptors assigned to the same visual word, and thus significantly improve the retrieval precision.

Although the retrieval accuracy is reduced by about 2% compared the method without any feature selection, the retrieval time was reduced significantly because of the min-Hash and random sampling. Therefore, the proposed method greatly improves the retrieval efficiency at the expense of slightly reduced retrieval precision.

TABLE 5. mAP comparison results on large-scale image search.

Dataset	BoVW[20]	BoVW+ minHash[7]	HE +min-Hash	The proposed method
Holiday+ Flickr 10k	47.6%	45.27%	66.31%	62.72%
Holiday+ Flickr 100k	38.91%	35.23%	55.93%	51.56%

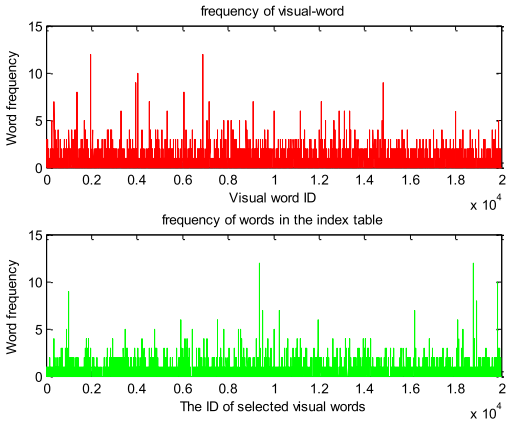


FIGURE 7. Comparison of frequency histogram distribution.

To evaluate large-scale secure image search, we randomly select images from Flickr 1M, which is a distractor image dataset for large-scale image search, and construct two different size distractor datasets: Flickr10k and Flickr 100k, and then combine the Holiday dataset with these two datasets. The mAP results are shown in Table 5. We can see that the proposed method has higher retrieval precision than other methods, and it can fulfill the needs of large-scale secure image retrieval in the cloud environment.

E. SECURITY

In order to prevent the data attacker from illegally acquiring the distribution characteristics of the original image features and the mapping relationship between the image features and visual words, the visual dictionary is permuted and visual words are randomly selected from the encrypted visual dictionary. Figure. 7 shows the word-frequency histogram comparison results between original word frequency histogram and randomly selected visual-words histogram. From the figure we can see that the proposed method completely disturbs the original word frequency distribution of the image so that the CSP can no longer calculate the correct word frequency distribution according to the uploaded index table, thereby can protect the word frequency information of the image effectively.

VII. CONCLUSION

In this paper, a large-scale secure image retrieval method in cloud environment is proposed. The HE algorithm is utilized to generate binary signatures for image features and filter the

mismatched feature points assigned to the same visual words, which greatly increase the retrieval accuracy. The random sampling selection and the min-Hash algorithm are combined to generate secure image index, which can not only decrease the redundancy information and improve retrieval efficiency, but also can guarantee the security of the index. The proposed method achieves the balance among security, accuracy and retrieval efficiency of large-scale secure image retrieval in public clouds. The security analysis and experimental results prove the effectiveness of the proposed method. Future research will focus on exploring retrieval efficiency further.

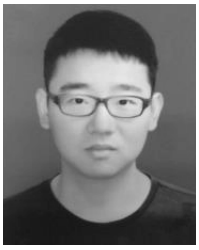
ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their helpful and constructive suggestions and comments.

REFERENCES

[1] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. IEEE Symp. Secur. Privacy*, Berkeley, CA, USA, May 2000, pp. 44–55.
[2] D. Boneh, G. D. Crescenzo, and R. Ostrovsky, "Public key encryption with keyword search," in *Proc. EUROCRYPT*, Interlaken, Switzerland, 2004, pp. 506–522.
[3] A. Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, M. Wu, and D. W. Oard, "Confidentiality-preserving rank-ordered search," in *Proc. ACM Workshop Storage Secur. Survivability*, Alexandria, VA, USA, Oct. 2007, pp. 7–12.
[4] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in *Proc. IEEE ICDCS*, Genova, Italy, Jun. 2010, pp. 253–262.
[5] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 222–233, Jan. 2014.
[6] W. Lu, A. L. Varna, A. Swaminathan, and M. Wu, "Secure image retrieval through feature protection," in *Proc. IEEE Conf. Acoust., Speech Signal Process.*, Taipei, Taiwan, Apr. 2009, pp. 1533–1536.
[7] W. Lu, A. Swaminathan, A. L. Varna, and M. Wu, "Enabling search over encrypted multimedia databases," *Proc. SPIE*, vol. 7254, Feb. 2009, Art. no. 725418.
[8] Z. Xia, L. Jiang, D. Liu, L. Liu, and B. Jeon, "BOEW: A content-based image retrieval scheme using bag-of-encrypted-words in cloud computing," *IEEE Trans. Services Comput.*, to be published, doi: 10.1109/TSC.2019.2927215.
[9] Z. Xia, X. Ma, Z. Shen, X. Sun, N. N. Xiong, and B. Jeon, "Secure image LBP feature extraction in cloud-based smart campus," *IEEE Access*, vol. 6, pp. 30392–30401, 2018.
[10] B. Ferreira, J. Rodrigues, J. Leitão, and H. Domingos, "Practical privacy-preserving content-based retrieval in cloud image repositories," *IEEE Trans. Cloud Comput.*, vol. 7, no. 3, pp. 784–798, Sep. 2019.
[11] R. L. Lagendijk, Z. Erkin, and M. Barni, "Encrypted signal processing for privacy protection: Conveying the utility of homomorphic encryption and multiparty computation," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 82–105, Jan. 2013.
[12] Z. Erkin, J. Li, A. P. O. S. Vermeeren, and H. de Ridder, "Privacy-preserving emotion detection for crowd management," in *Proc. AMT*, Warsaw, Poland, 2014, pp. 359–370.
[13] 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, Nov. 2012.
[14] W.-T. Chu and F.-C. Chang, "A privacy-preserving bipartite graph matching framework for multimedia analysis and retrieval," in *Proc. ICMR*, Shanghai, China, Jun. 2015, pp. 243–250.
[15] W. Lu, A. L. Varna, and M. Wu, "Confidentiality-preserving image search: A comparative study between homomorphic encryption and distance-preserving randomization," *IEEE Access*, vol. 2, pp. 125–141, 2014.
[16] Z. Xia, Y. Zhu, X. Sun, Z. Qin, and K. Ren, "Towards privacy-preserving content-based image retrieval in cloud computing," *IEEE Trans. Cloud Comput.*, vol. 6, no. 1, pp. 276–286, Mar. 2018.
[17] J. Qin, H. Li, X. Xiang, Y. Tan, W. Pan, W. Pan, W. Ma, and N. N. Xiong, "An encrypted image retrieval method based on Harris corner optimization and LSH in cloud computing," *IEEE Access*, vol. 7, pp. 24626–24633, 2019.
[18] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. ICCV*, Nice, France, Oct. 2003, pp. 1470–1477.
[19] F.-F. Li and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. CVPR*, San Diego, CA, USA, Jun. 2005, pp. 524–531.
[20] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. CVPR*, New York, NY, USA, Jun. 2006, pp. 2161–2168.
[21] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Proc. ECCV*, Marseille, France, 2008, pp. 304–317.
[22] A. Broder, M. Charikar, A. M. Frieze, and M. Mitzenmacher, "Min-wise independent permutations," in *Proc. STOC*, Dallas, TX, USA, 1998, pp. 327–336.
[23] A. Z. Broder, "On the resemblance and containment of documents," in *Proc. SEQUENCES*, Salerno, Italy, Jun. 1997, pp. 21–29.
[24] O. Chum, J. Philbin, M. Isard, and A. Zisserman, "Scalable near identical image and shot detection," in *Proc. CIVR*, Amsterdam, The Netherlands, Jul. 2007, pp. 549–556.
[25] O. Chum, J. Philbin, and A. Zisserman, "Near duplicate image detection: Min-hash and TF-IDF weighting," in *Proc. BMVC*, Leeds, U.K., Sep. 2008, pp. 812–815.
[26] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. CVPR*, Minneapolis, MN, USA, Jun. 2007, pp. 1–8.
[27] *Flickr 1M*. Accessed: Aug. 30, 2019. [Online]. Available: https://press.liacs.nl/mirflickr/
[28] Y. Xu, J. Gong, L. Xiong, Z. Xu, J. Wang, and Y.-Q. Shi, "A privacy-preserving content-based image retrieval method in cloud environment," *J. Vis. Commun. Image Represent.*, vol. 43, pp. 164–172, Feb. 2017.

YANYAN XU received the B.E. degree from the Xi'nan Institute of Technology, China, in 1995, the M.E. degree in electrical engineering from the Hubei University of Technology, China, in 2000, and the Ph.D. degree in communication and information system from Wuhan University, China, in 2007. She is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. Her research interests include multimedia information security and multimedia communication systems.



XIAO ZHAO received the B.E. degree in computer science and technology from Zhengzhou University, China, in 2017. He is currently pursuing the M.E. degree in communication and information system with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. His research interests include multimedia information security and multimedia network communication.

JIAYING GONG received the M.E. degree in communication and information system from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China, in 2018. She is currently a Software Engineer with Ping An Technology Company, Ltd., Shenzhen, China. Her research interest includes multimedia information security.