A Privacy-preserving Image Retrieval Method Based on Improved BoVW Model in Cloud Environment

Jiaying Gong, Yanyan Xu and Xiao Zhao
State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, People’s Republic of China

ABSTRACT
With the rapid development of cloud computing technology, more and more users choose to outsource image data to clouds. To protect users’ privacy and guarantee data’s confidentiality, images need to be encrypted before being outsourced to CSP, but this brings new difficulties to some basic yet important data services, such as content-based image retrieval (CBIR). In this paper, a privacy-preserving image retrieval method based on an improved BoVW model is proposed. An improved BoVW method based on Hamming embedding can provide binary signatures that refine the matching based on visual words; therefore, retrieval precision is improved significantly; orthogonal transformation is utilized to implement privacy-preserving image retrieval, where image features are divided into two different fields with orthogonal decomposition, for which encryption and distance comparison are executed separately, and two kinds of operation results are fused in the final vector with orthogonal composition. As a result, cloud server can extract components from encrypted features directly and compare distance with those of query image, without violating the privacy of images and features. Any algorithm can be used to encrypt features, which enhance the practicability of the proposed method. The security analysis and experimental results prove its security and retrieval performance.

KEYWORDS
Image retrieval; Secure retrieval; BoVW; Cloud environment; Orthogonal transformation

1. INTRODUCTION
Along with the development of cloud computing technology, public cloud service can provide unlimited storage space and computing ability for massive multimedia. At the same time, it brings serious security problems because data owners lose physical control of data while CSP (cloud service provider) is considered “honest but curious”, which means that it may lead to unauthorized access to data and cause leakage of personal privacy [1–3]. Given such circumstance, data must be encrypted before being outsourced to the CSP to protect its confidentiality. Although data security can be guaranteed by encryption, it also brings many new challenges to data management and data sharing service, such as image retrieval. Content-based image retrieval (CBIR) is a very promising method in image retrieval field which is characterized by extracting image features and comparing the distance between features automatically, and it has grown rapidly and made progresses in both the derivation of new features and the construction of signatures based on these features [4–6]. However, conventional CBIR methods cannot be applied under cloud environment directly because encrypted data fail to preserve the distance between feature vectors. If users want to retrieve images from CSP, encrypted image need to be decrypted first, then retrieval can be operated on plaintext, which makes the sensitive information being exposed to attackers, breaks privacy, and hence is not desired. Therefore, it is important to develop technologies of CBIR over encrypted domain, which is also called private/privacy-preserving CBIR (PCBIR) [7–9].

Many advances have been achieved in image retrieval over cryptographic domain. Early technology mainly originated from retrieval on text document for example, Song et al. [10] proposed a ciphertext scanning method based on streaming cipher to make sure whether the search term exists in the ciphertext; Boneh et al. [11] proposed a keyword search method based on public-key encryption; Swaminathan et al. [12] ranked the order the documents securely and extracted the most relevant documents from an encrypted collection; Cao et al. [13] proposed privacy-preserving multi-keyword ranked search over encrypted data. Although these methods can be extended to image retrieval based on user-assigned tags, 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.
Recently, several methods have been proposed to solve the problem of PCBIR. Lu et al. [14] proposed three distance-preserving methods: bit plane randomization, random projection, and random unary encoding, which are applied on low-level features such as color histogram. Karthik et al. [15] proposed a transparent privacy-preserving hash method, which keeps the statistical rules of encrypted AC coefficients but ignores the spatial information distribution of the image. Xu et al. [16] proposed a secure retrieval method for JPEG images, which preserves the distribution of the AC coefficients and the statistical rule of color after the decoding of encrypted images. Ferreira et al. proposed a method in [17], where color information is encrypted by deterministic encryption techniques to support color-feature-based CBIR, and texture information is encrypted by probabilistic encryption algorithms for better security. Zhang et al. [18] use Pallier encryption algorithm to protect some lower level features such as color, texture, and shape, and achieve secure retrieval effects. Some local-feature-based PCBIR methods are also proposed. Hsu et al. [19] put forward a homomorphic encryption-based secure SIFT method which has good security at the cost of serious ciphertext extensions. Xia et al. [20] proposed to use SIFT features and transform earth mover’s distance (EMD) in a way that the cloud server can evaluate the similarity between images without learning sensitive information. Huang et al. [21] convert the high-dimensional VLAD descriptors to compact binary codes, and then adapt the asymmetric scalar-product-preserving encryption to design PCBIR to achieve the privacy requirements in the cloud environment. In contrast with global feature approaches, local-feature-based PCBIR methods achieve higher retrieval accuracy, but it requires quite complex methods to implement distance preserving, which is not suitable for large-scale image retrieval under cloud environment.

It should be noted that the above research results are all relying on specific encryption methods, such as shuffling and homomorphic encryption, which limit its universality. For example, homomorphic encryptions are too computation- and communication-intensive to be used in low-profile devices and large-scale systems, while shuffling is not suitable for some situations that have high requirements to security. Besides, these methods are all relying on global features or local features to evaluate the similarity of images, whose retrieval accuracy can hardly meet requirements of practical applications because of “semantic gap” between the visual features and the richness of human semantics [22]. Aiming to solve these problems, we propose a privacy-preserving image retrieval method under cloud environment in this paper, where an improved BoVW model is employed to improve retrieval precision, orthogonal transformation is combined together to implement privacy-preserving retrieval under cloud environment. An improved BoVW method based on Hamming embedding can provide binary signatures that refine the matching based on visual words; therefore, retrieval precision is improved significantly; orthogonal transformation are used to divide features into two different domains, for which encryption and distance comparison are executed separately. As a result, cloud server can extract components from encrypted features directly and compare distance with those of query image, without violating the privacy of images and features. Because the encryption operation and distance comparison operation are independent, any algorithm can be used to encrypt features, which enhance the practicability of the proposed method. The experimental results prove its effectiveness and security.

The organization of this paper is as follows: Chapter 2 introduces system architecture and preliminaries; Chapter 3 proposes our scheme. Chapter 4 provides experimental results and performance analysis, and Chapter 5 presents conclusions.

2. SYSTEM ARCHITECTURE AND PRELIMINARIES

2.1 System Architecture

The system model used in this paper is given in Figure 1.

There are three entities involved in this model: content owner, CSP, and user. Content owner trains visual dictionaries, extracts features from images, and constructs search index, which are then encrypted together with images and outsourced to the CSP. CSP stores cipher images and secure index and performs image retrieval...
when receiving user request. Users send retrieval request to content owner and CSP, decrypt cipher-images returned by CSP, and get requested images.

2.2 Preliminaries

2.2.1 BoVW Algorithm

Bag of visual words (BoVW) model is extended from natural language processing and information retrieval field to computer vision in [23]. Descriptors are quantized into visual words with the k-means algorithm, and then image can be represented by the frequency histogram of visual words obtained by assigning each descriptor of the image to the closest visual word. Fast access to the frequency vectors is achieved by an inverted file system. BoVW has been successfully adopted to enable fast indexing and retrieval of large image collections [24]. The retrieval process consists of four steps: image content is described by means of a set of visual descriptors such as SIFT and SURF; descriptors are clustered into visual words by means of a set of visual descriptors such as SIFT and SURF.; descriptors are clustered into visual words with the k-means algorithm, and then indexing and retrieval of large image collections [24].

Although BoVW has shown good performance in image retrieval task, it still suffers from some problems, such as insufficient discriminative power of visual words, quantization error caused by assigning descriptors to visual words, and low efficiency caused by comparing distance between high dimensions of vectors. Many methods are proposed to improve the performance of BoVW, such as [25,26]. Among these methods, [27] is the one that achieves higher accuracy for large-scale image search based on Hamming embedding (HE). The main idea of HE is to select low number of centroids k in k-means clustering to form a rough visual dictionary and refine the quantized index q (x_i) (the index of the centroid closest to the descriptor x_i) with those of a d_b-dimensional binary signature, thus the Euclidean distance between two descriptors can be mapped into the Hamming distance, and mismatch points can be removed by setting proper threshold; thus, retrieval performance is improved. This method includes four steps:

(1) Assign descriptors x to their closest centroid, resulting in q(x); project x by means of a d_b × d orthogonal projection matrix P, producing a vector z = P^T x = [z_1, . . . , z_d_b]^T; compute the median value τ_i of all vectors that belong to the same centroid;

(2) Compute the signature b(x) = (b_1(x), . . . , b_d_b(x)) as follows:

\[ b_i(x) = \begin{cases} 1 & \text{if } z_i > \tau_q(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (1) \]

(3) Use HE matching function to match the distance of x, y:

\[ f_{HE}(x, y) = \begin{cases} (t_f - idf(q(x)))^2 | f_q(x) = q(y) \\ \text{and } h(b(x), b(y)) \leq h_t \\ 0 & \text{otherwise} \end{cases} \quad (2) \]

where \( t_f - idf() \) weights the visual words according to their frequency, and \( h() \) is the Hamming distance defined as follows:

\[ h(b(x), b(y)) = \sum_{i=1}^{d_b} b_i(x) \oplus b_i(y) \quad (3) \]

\( h_t \) is a fixed Hamming threshold such that \( 0 \leq h_t \leq d_b \).

(4) Given a query image represented by its local descriptor \( y_f \), and a set of database image \( j = 1, . . . , n \) represented by their local descriptor \( x_{i,j} \), then a voting system is used to evaluate the similarity between the query image and each database image. The image score \( s^*_j \) is given as follows:

\[ s^*_j = g_j \left( \sum_{i=1}^{m'} \sum_{j=1}^{m_j} f(x_{i,j}, y_f) \right) \quad (4) \]

where \( g_j() \) is a post-processing function, \( m_j \) is the number of local descriptors of the \( j \)-th database image, \( m' \) is the number of local descriptors of the query image. The score reflects the number of matches between the query and each database image.

2.2.2 Distance-Preserving Method Based on Orthogonal Transformation

Orthogonal transformation is a kind of vector representation method that consists of two processes: orthogonal decomposition and orthogonal composition. Any vector can be expressed as a sum of a set of component coefficients by orthogonal decomposition, and component coefficients are fused into one composite vector by orthogonal composition.
Suppose vector $X = (x_1, x_2, \ldots, x_n)$, orthogonal matrix $B = (b_1, b_2, \ldots, b_n)$, then $X$ can be expressed as follows:

$$X = B \cdot Y$$  \hspace{1cm} (5)

Vector $Y$ can be expressed as follows:

$$Y = B^{-1} \cdot X$$  \hspace{1cm} (6)

If we divide $B$ into two sub-matrix $R = (b_1, b_2, \ldots, b_m)$ and $S = (b_{m+1}, b_{m+2}, \ldots, b_n)$, then $Y$ is divided into $Y = (Y_1, Y_2)^T$ correspondingly, and we can obtain the following:

$$X = R \cdot Y_1 + S \cdot Y_2$$  \hspace{1cm} (7)

If encryption is operated on $Y_1$ and feature is extracted from $Y_2$, respectively, then we can obtain the following:

$$X_{ef} = R \cdot Enc(Y_1, K_e) + S \cdot FeaExt(Y_2, K_f) = R \cdot Y_{1e} + S \cdot Y_{2f} = B \cdot Y_{ef}$$  \hspace{1cm} (8)

Two different operations will be independent and will not interfere with each other because of the independence of the orthogonal decomposition, while two kinds of operation results are fused in final vector $X_{ef}$ because of the fusion of the orthogonal composition. We can also obtain the following:

$$D(X_{ef}, K_d, R) = R \cdot Dec(Y_{1e}, K_d) + Y_{2f}$$  \hspace{1cm} (9)

$$FeaExt(X_{ef}, K_f, S) = S \cdot FeaExt(Y_{2f}, K_f)$$  \hspace{1cm} (10)

From these equations, we can see that features can be extracted from encrypted vectors directly, and the distance between the component vectors is similar to the one between the original vectors because of the distance-preserving characteristic of the orthogonal transformation.

### 3. THE PROPOSED METHOD

Based on algorithms mentioned in Section 2, we propose a privacy-preserving CBIR method combined with an improved BoVW model and orthogonal transformation.

The specific retrieval process is given as follows:

#### 3.1 Configuration Stage

In configuration stage, visual dictionary, encryption/decryption key, and Gaussian orthonormal matrix are constructed for later usage.

- $(VD) \leftarrow VDGen(TB)$: Content owner generates visual dictionary $VD$ from training image dataset $TB$, where two steps are included: SIFT descriptors are extracted from $TB$, which have been widely used due to its distinctiveness and robust matching across a substantial range of affine distortion, addition of noise, and change in illumination; Descriptors are quantized into visual words to form a visual dictionary with k-means clustering method.
- $(K_I, K_J) \leftarrow KeyGen(\alpha)$: Encryption key $K_I$, $K_J$ are generated with $KeyGen()$ function, where $\alpha$ is a security parameter used to generate key.
- $(P, B) \leftarrow MatrixGen()$: A random Gaussian orthonormal matrix is generated as $P$, $B$, $R$, and $S$ are selected from $B$ in terms of users practical needs on security or retrieval performance.
- $(\tau_q(x_j)) \leftarrow MedianMatrixGen(P, q(x_j), x_j)$: The descriptors $x_j$ from $VD$ are assigned to their closest centroid $q(x_j)$ and projected by $P$ to $z = P x_j = [z_1, \ldots, z_m]^T$, then median value $\tau_q(x_j)$ of all vectors that belong to the same centroid $q(x_j)$ is generated.

#### 3.2 Content Owner Side

- $(S_{x_j}) \leftarrow SigBuild(Img, VD, P, \tau_q(x_j))$: Sift descriptors $x_j$ are extracted from images $Img$ that will be outsourced to CSP, assigned to particular cluster centers $q(x_j)(j = 1, \ldots, m_j)$ in $VD$, and $x_j$ is mapped to vectors with an orthogonal projection matrix $P$, then binary signature of $x_j$ is generated by comparing with median matrix $\tau_q$ according to Eq.(1). Therefore a descriptor $x_j$ can be represented as a signature $S_{x_j} = (q(x_j), b(x_j))$, where $b(x_j) = (b_1(x_j), \ldots, b_{m_j}(x_j))$.
- $(q_{ef}(x_j), b_{ef}(x_j)) \leftarrow SecureIndexBuild(S_{x_j}, K_I, R)$: Orthogonal Decomposition is applied to $S_{x_j}$ to obtain encryption field $S_{1x_j} = (q_1(x_j), b_1(x_j))$, encryption algorithm AES are executed on it to obtaining $q_{ef}(x_j), b_{ef}(x_j)$ according to Equation (8), finally encrypted features $(q_{ef}(x_j), b_{ef}(x_j))$ are obtained by means of orthogonal composition. TF-IDF is used to construct secure image index.
- $(Img_e) \leftarrow SecureImgBuild(Img, K_J)$: Images that will be outsourced to CSP are encrypted by $K_J$ and AES encryption algorithm, cipher-images $Img_e$ are obtained.

Data owner uploads cipher-images $Img_e$ and corresponding secure image index to CSP.

#### 3.3 User Side

- User sends query request to the data owner. After identity authentication, owner sends secure parameters $(R, S, P, K_I)$ to users securely.
(S_q) ← SigBuild(Img_q, VD, P): Sift descriptors x_q are extracted from Img_q, and binary signature S_q = (q(x_q), b(x_q)) is generated with the method we mentioned in 3.2.

(q_{ef}(x_q), b_{ef}(x_q)) ← TrapdoorBuild (s_q, R, K_f): Orthogonal Decomposition is applied on signature S_q to obtain encryption field S_q{\text{ef}}, encryption algorithm AES and encryption key K_f are executed on S_q{\text{ef}}, then orthogonal composition is applied to obtain encrypted features (q_{ef}(x_q), b_{ef}(x_q)) as trapdoor. User sends trapdoor to the CSP as query.

Img ← Dec(Img_{eq}, K_f): After receiving retrieval results Img_{eq} from CSP, user decrypts cipher images with K_f and obtains requested images.

### 3.4 CSP Side

- CSP obtains sub-matrix S content data owner securely, operates orthogonal decomposition on encrypted feature (q_{ef}(x_q), b_{ef}(x_q)) and search trapdoor (q_{ef}(x_q), b_{ef}(x_q)) sent by users according to Equation (10), then distance comparison field is obtained to compare the distance between (q_{ef}(x_q), b_{ef}(x_q)) and search trapdoor (q_{ef}(x_q), b_{ef}(x_q)).
- (Img_{eq}) ← DistComp((q_{ef}(x_q), b_{ef}(x_q)), (q_{ef}(x_q), b_{ef}(x_q))): The matching scores are available by the similarity comparison between (q_{ef}(x_q), b_{ef}(x_q)), (q_{ef}(x_q), b_{ef}(x_q)) according to Equations (11–13). The higher the matching scores, the more similar the corresponding image is to the query image, and the encrypted images Img_{eq} with the highest scores are sent to the requested user.

\[
f(x_{i,j}, x_q) = \begin{cases} 
(idf(q_{ef}(x_q))^2 \times \exp\left(-\frac{L(x_{i,j}, x_q)^2}{\sigma^2}\right))^\gamma & \text{if } q_{ef}(x_{i,j})q_{ef}(x_q) \text{ and } L(x_{i,j}, x_q) \leq L_d \\
0 & \text{otherwise}
\end{cases}
\]

\[
L(x_{i,j}, x_q) = \sqrt{(b_{ef}(x_{i,j}) - b_{ef}(x_q))^2}
\]

\[
s = \sum_{q=1}^{m'} \sum_{j=1}^{m_j} f(x_{i,j}, x_q)
\]

where idf () is a measure of the general importance of a visual word and indicates the ratio of the number of images containing a given word to the total number of images; tf describes the number of occurrences of a given visual word in an image; L_d and \(\sigma\) are empirical values obtained from experiment; \(m_j\) is the number of local descriptors of the j-th database image, \(m'\) is the number of local descriptors of the j-th database image, \(m'\) is the number of local descriptors of the j-th database image, \(m'\) is the number of

### 4. EXPERIMENTAL RESULTS

In this chapter, we present an experimental evaluation of the proposed method and compare retrieval results with several classical privacy-preserved image retrieval methods, including random projection, bit-plane randomization, and randomized unary encoding proposed in [14], Pallier algorithm put forward in [19]. We perform experiment on INRIA Holidays dataset [28] containing 1491 images, which are divided into 500 queries and 991 corresponding relevant images. The performance of our scheme is evaluated in terms of security and retrieval performance.

#### 4.1 Security

The orthogonal decomposition is implemented on image features instead of images, and images can be encrypted by any encryption method. We use AES-128 encryption algorithm to encrypt images. Even with the most powerful biclique attacks, the computational complexity is \(2^{126.1}\), which means the security of AES will not be broken. For the encrypted features, since S is known by CSP, and as a sub-matrix of B, R, and S are not completely independent, so there is a potential secure problem. The complexity of attack B and R in the case of known the sub-matrix S is as follows:

\[
d_B \sim 2^{n^2k-k_m+2^{(m+1)(k-1)+k_0\frac{1}{2}}} \]  

\[
d_R \sim 2^{3klog_2(m(n-k)+m-k_k+2^{(m+1)k+k_0\frac{1}{2})}} \]

where the size of B is n, element of B is k-bit fixed word-length integer, m is the number of columns of the sub-matrix S, and \(m_k\) is a factor that includes the efficiency improvement achieved over an exhaustive search by using a different algorithm. Therefore, the complexity of attacking B and attacking R tend to O(n^2k) and O(m^2k), which means even if orthogonal sub-matrix S is known, the computational complexity of attacking the orthogonal matrix B and the orthogonal sub-matrix R is exponentially increasing; therefore, the security of features can also be guaranteed.

Three different metrics are used to evaluate the security of features: autocorrelation function, information entropy, and histogram distribution.

(1) Autocorrelation function
The autocorrelation function of a feature vector measures how correlated the neighboring feature elements are [29]. Suppose n is the size of feature vectors, T is the delay of signals, then the autocorrelation function \( R(T) \) is calculated as follows:

\[
R(T) = \sum_{i=1}^{n} x(i)x(i-T)
\]  

The autocorrelation function for the raw color histogram, randomized features of three different retrieval methods of Lu, random uniform vector, and the encrypted features using algorithm proposed in this paper are shown in Figure 2. We can see that the raw color feature has non-negligible correlation, which indicates there exists strong association between adjacent vectors, while our method has similar autocorrelation as random features encrypted by Pallier’s algorithm and is much lower than Lu’s three methods, which shows encrypted features using our method has low autocorrelation thus has good confidentiality.

(2) Information entropy

Information entropy describes the uncertainty of random variables. A higher entropy of encrypted features means that it has a distribution closer to uniform, thus higher randomness and higher security. The entropy is defined as follows:

\[
H = - \sum_{i=0}^{n} p(x(i)) \times \log_2 p(x(i))
\]  

where \( x(i) \) is the feature, \( n \) is the length of \( x(i) \), \( p(x(i)) \) is the probability of \( x(i) \). A higher entropy of encrypted features means it has a distribution closer to uniform, thus higher randomness and higher security. The entropy of encrypted features for different algorithm is given in Table 1.

From this table, we can see that both the color histogram and visual words have low entropy, which means that there is strong inherent correlation among feature values. The features encrypted by the proposed method achieve higher entropy than Lu’s three methods but lower than uniform random features encrypted by Pallier’s algorithm. Although its security is not as good as features encrypted by complex homomorphic algorithm, its time complexity consumed by encryption is much lower than Pallier’s algorithm; therefore, it is a more practical method.

(3) Histogram distribution

Histogram distribution is usually employed to evaluate security of encrypted data. In the proposed method, secure image index of each image is outsourced to CSP, which consists of two parts: \( (q_{ef}(x_j), b_{ef}(x_j)) \), \( q_{ef}(x_j) \) means the encrypted cluster centers that feature \( x_j \) belonging to; \( b_{ef}(x_j) \) means the encrypted binary signature of \( x_j \). The CSP can obtain only the frequency histogram distribution of encrypted visual words, but it can hardly acquire image features from its encrypted binary signature. Therefore it is difficult for the CSP to only use frequency histogram distribution to recover the image features and infer the image content without correct decryption key.

Figure 3 shows the comparative results of frequency histogram distribution between the encrypted visual words and the original visual words of an image. From Figure 3, we can see that the histogram of encrypted visual words is different from that of original visual words, and it is difficult for attackers to infer the image content information only through frequency histogram distribution of the encrypted visual words.

### Table 1: Entropy for different methods

| Feature type                     | Entropy |
|----------------------------------|---------|
| Color histogram                  | 1.18    |
| Visual words                     | 2.59    |
| Bit-plane randomization          | 6.12    |
| Random projection                | 5.94    |
| Randomized unary encoding        | 6.69    |
| Proposed scheme                  | 6.83    |
| Uniform random features          | 7.96    |

4.2 Analysis of Time Complexity

Assume the size of the selected operation data \( X \) is \( n \), the dimension of sub-matrix \( S \) is \( k(k < n) \), \( g \) is the number
of bit-planes to randomize, $m$ is the dimension of projected features, $M$ is the eigenvector of the feature vector, then the time complexity comparison results with other methods are given in Table 2.

From Table 2, we can see that the time complexity of the Pallier homomorphism method is the highest; Randomized unary encoding is higher; the proposed method is lower than these methods and similar to random projection proposed by Lu.

### 4.3 Retrieval Precision

Retrieval precision is evaluated by mean average precision (mAP) that is widely used to measure image retrieval performance for a group of queries [30]. It is defined as follows:

$$mAP = \frac{\sum_{q=1}^{Q} aveP(q)}{Q}$$  \hspace{1cm} (18)

where $Q$ is the number of queries, $avePq$ represents the average value of all the precisions measured each time a new relevant image is retrieved. In this experiment, a group of 500 query images are retrieved in a database containing 991 images, and the MAP of query images for different encryption methods are computed as Table 3 shows.

From the table, we can see that the retrieval result of our scheme is much better than those of other methods. Meanwhile, we also compute the mAP of the unencrypted HE method, and its MAP is 80.59 which is roughly equal to the result of our encrypted scheme, namely, our method can hardly reduce retrieval accuracy of original HE.

Through the above analysis of retrieval performance, it is proved that the proposed method has a good trade-off between security and retrieval performance, it not only guarantees the security of retrieval in cloud environment but also ensures the high precision of retrieval for practical applications.

### 5. CONCLUSION

A privacy-preserving image retrieval method based on improved BoVW model in cloud computing is proposed in the paper. An improved BOVW method based on Hamming embedding which provides binary signatures refining visual words are combined with orthogonal transformation to achieve privacy-preserving image retrieval under cloud environment. Experiments show that our scheme has obvious advantages in security and retrieval precision compared with other methods. Future research will focus on exploring retrieval efficiency further.

### DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

### FUNDING

This research is supported by the National Natural Science Foundation of China [grant number 41571426], National Key Research and Development Program of China under [grant number 2017YFB0504202], and Wuhan Applied Basic Research Programs [grant number 2017010201010114].
REFERENCES

1. S. Subashini, and V. Kavitha, “A survey on security issues in service delivery models of cloud computing,” J. Netw. Comp. Appl., Vol. 34, no. 1, pp. 1–11, 2011.

2. L. Ferretti, F. Pierazzi, M. Colajanni, and M. Marchetti, “Security and confidentiality solutions for public cloud database services,” in Proceedings of the Seventh International Conference on Emerging Security Information, Systems and Technologies, 2013.

3. R. Huang, X. Gui, S. Yu, and Z. Wei, “Study of privacy-preserving framework for cloud storage,” Comp. Sci. Inf. Syst. Vol. 8, no. 3, pp. 801–19, 2011.

4. R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: ideas, influences, and trends of the new age,” ACM Comput. Surv. Vol. 40, no. 2, pp. 5–60, 2008.

5. Y. Liu, D. Zhang, L. Lu, and W. Ma, “A survey of content-based image retrieval with high-level semantics,” Pattern Recogn. Vol.40, pp. 262–282, 2007.

6. J. Wang, T. Li, Y-Q. Shi, S. Lian, and J. Ye, Forensics feature analysis in quaternion wavelet domain for distinguishing photographic images and computer graphics. Multimed Tools Appl., 76, no. 22, pp. 23721–37, 2017.

7. J. Shashank, P. Kowshik, K. Srinathan, and C. V. Jawahar, "Private content based image retrieval," in Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, 2008, pp. 1–8.

8. 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. 125–41, 2014.

9. L. Wang, L. Amsaleg, A. Morton, and S. Marchand-Maillet, “A privacy-preserving framework for large-scale content-based information retrieval,” IEEE Trans. Inf. Forensics Secur., Vol. 10, no.1, pp. 152–67, 2015.

10. D. X. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proceedings of the 2000 IEEE Symposium on Security and Privacy, 2000, pp. 44–55.

11. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in Proceedings of the International Conference On the Theory and Applications of Cryptographic Techniques, Springer, Berlin, 2004, pp. 506–22.

12. A. Swaminathan, et al., “Confidentiality-preserving rank-ordered search,” in Proceedings of the 2007 ACM Workshop on Storage security and Survivability, 2007, pp. 7–12.

13. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” IEEE Trans. Parallel and Distributed Syst., Vol. 25, no.1, pp. 222–33, 2014.

14. W. Lu, A. L. Varna, A. Swaminathan, and M. Wu, “Secure image retrieval through feature protection,” in Proceedings of the IEEE Conference on Acoustics, Speech Signal Processing, 2009, pp. 1533–6.

15. K. Karthik, and S. Kashyap, “Transparent hashing in the encrypted domain for privacy preserving image retrieval,” Signal Image Video Process, Vol. 7, no.4, pp. 647–64, 2013.

16. 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. 164–72, 2017.

17. B. Ferreira, J. Rodrigues, J. Leitao, and H. Domingos, “Privacy-preserving content-based image retrieval in the cloud,” in Proceedings of the IEEE 34th Symposium on Reliable Distributed Systems (SRDS), 2015, pp. 11–20.

18. L. Zhang, T. Jung, C. LiuX. Ding, X-Y. Li, and Y. Liu, “Pop: privacy-preserving outsourced photo sharing and searching for mobile devices,” in Proceedings of the IEEE 35th International Conference on Distributed Computing Systems (ICDCS), 2015, pp. 308–17.

19. 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, pp. 276–86, 2018.

20. K. Huang, M. Xu, S. Fu, and D. Wang, “Efficient privacy-preserving content-based image retrieval in the cloud,” in Proceedings of the International Conference on Web-Age Information Management. Springer International Publishing, 2016, pp. 28–39.

21. J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.

22. C. Wang, L. Zhang, and H. J. Zhang, “Learning to reduce the semantic gap in web image retrieval and annotation,” in Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2008, pp. 355–62.

23. 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. 4593–607, 2012.

24. S. Avila, N. Thome, M. Cord, E. Valle, and A. de A. Araújo, “Bossa: extended bow formalism for image classification,” in Proceedings of the 18th IEEE International Conference on Image Processing (ICIP), 2011, pp. 2909–12.
25. C. Wengert, M. Douze, and H. Jégou. “Bag-of-colors for improved image search,” in Proceedings of the 19th ACM International Conference on Multimedia, 2011, pp. 1437–40.

26. J. Yang, J. Liu, and Q. Dai. “An improved bag-of-words framework for remote sensing image retrieval in large-scale image databases,” International Journal of Digital Earth, Vol. 8, no. 4, pp. 273–92, 2015.

27. H. Jegou, M. Douze, and C. Schmid. “Hamming embedding and weak geometric consistency for large scale image search,” Computer Vision–ECCV 2008, Vol. 5302, pp. 304–17, 2008.

28. http://lear.inrialpes.fr/people/jegou/data.php

29. S. Haykin. “Nonlinear methods of spectral analysis,” Optica Acta International Journal of Optics, Vol. 27, no. 2, 142, 1979.

30. H. Müller, and W. Müller, D M G. Squire, S. Marchand-Maillet, and T. Pun, “Performance evaluation in content-based image retrieval: overview and proposals,” Pattern Recognition Letters, Vol. 22, no. 5, pp. 593–601, 2001.

Authors

Jiaying Gong received her master’s degree in communication and information system from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. Her research interest is multimedia information security.  
E-mail: 2015206190036@whu.edu.cn

Yanyan Xu is a professor of the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, China. Her main research interests include cloud computing security and big data privacy protection, multimedia network communications.  
Corresponding author. E-mail: xuyy@whu.edu.cn

Xiao Zhao is a master’s degree candidate of communication and information system in 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.  
E-mail: 281713836@qq.com